STUDIA INFORMATICA

Formerly: *Zeszyty Naukowe Politechniki Śląskiej, seria INFORMATYKA*
Quarterly

Volume 32, Number 4D (103)

Bartłomiej ZIELIŃSKI

PROTOKOŁY WARSTWY LINIOWEJ W BEZPRZEWODOWYCH SIECIACH KOMPUTEROWYCH

Silesian University of Technology Press
Gliwice 2011

STUDIA INFORMATICA **Volume 32, Number 4D (103)**

Formerly: *Zeszyty Naukowe Politechniki Śląskiej, seria INFORMATYKA* Nr kol. 1859

Editor in Chief

Editorial Board

STUDIA INFORMATICA is indexed in INSPEC/IEE (London, United Kingdom)

PL ISSN 0208-7286, QUARTERLY
Printed in Poland
The paper version is the original version

SPIS TREŚCI

CONTENTS

WPROWADZENIE

Bezprzewodowe sieci komputerowe są od kilkunastu lat przedmiotem intensywnych badań. Z jednej strony bowiem perspektywa korzystania z sieci bez konieczności podłączania przewodów wydaje się niezmiernie atrakcyjna, z drugiej strony – efektywne wykorzystanie bezprzewodowych mediów transmisyjnych pociąga za sobą konieczność zaprojektowania odpowiednich protokołów komunikacyjnych. Okazuje się bowiem, iż protokoły sieci przewodowych nie nadają się do stosowania w sieciach bezprzewodowych bądź też mogą spowodować obniżenie – nierzadko znaczne – wydajności sieci [127]. Dotyczy to szczególnie protokołów warstwy liniowej sieci.

Warstwa liniowa powinna zapewnić bezbłędną komunikację między stacjami [15, 113]. Zagadnienie to obejmuje detekcję, a nierzadko także korekcję błędów. Aby było to możliwe, informacje na poziomie warstwy liniowej są wymieniane w ramkach o określonej strukturze, podczas gdy warstwa fizyczna przesyła pojedyncze bity bez wnikania w ich znaczenie. Z tego względu można przyjąć, iż warstwa liniowa jest najniższą warstwą odpowiedzialną za realizację określonych mechanizmów, np. gwarancji jakości usług (QoS).

W wielu przypadkach, szczególnie w sieciach lokalnych, przyjmuje się, iż warstwa liniowa składa się z dwóch podwarstw, odpowiedzialnych za realizację dostępu stacji do łącza (podwarstwa dostępu) oraz prawidłową wymianę informacji między stacjami (podwarstwa łącza logicznego).

Podwarstwa dostępu do łącza jest jednym z najistotniejszych zagadnień projektowania każdej sieci komputerowej. Jest to bowiem najniższa warstwa, umożliwiająca wprowadzenie określonych mechanizmów, np. gwarantujących zachowanie determinizmu czasowego czy też jakości usług (QoS). Z drugiej strony, podwarstwa ta musi ściśle współpracować z warstwą fizyczną, m. in. prawidłowo oceniać stan łącza na podstawie tak dostarczonej informacji. W sieciach bezprzewodowych, a szczególnie w sieciach ad hoc – pozbawionych centralnej stacji sterującej, jak i posiadających nieregularną i często zmieniającą się strukturę – ocena stanu łącza na podstawie „klasycznego" mechanizmu wykrywania nośnej może prowadzić do występowania kolizji, zmniejszających stopień wykorzystania łącza. Z tego powo-

du dla sieci bezprzewodowych są projektowane nowe protokoły, korzystające z dodatkowych mechanizmów unikania kolizji, które uzupełniają, a nawet zastępują wykrywanie nośnej.

Podwarstwa łącza logicznego jest odpowiedzialna za dostarczenie warstwom wyższym usług transmisji informacji. W tym celu są definiowane zasady wymiany informacji, a także formaty używanych ramek. W przeciwieństwie do sieci przewodowych, sieci bezprzewodowe realizują w tej warstwie także dodatkowe mechanizmy. Przykładowo, sieć Packet Radio wykorzystuje protokół AX.25, który, prócz elementów typowych dla warstwy liniowej, zawiera także pewne rozwiązania charakterystyczne dla warstwy sieciowej i transportowej. Z kolei protokół standardu IEEE 802.11, stosowany powszechnie we współczesnych bezprzewodowych sieciach lokalnych, wprowadza całą grupę ramek służących do zarządzania pracą sieci, ze szczególnym uwzględnieniem bezpieczeństwa sieci. Protokół ten wspiera także różne warianty dostępu do łącza dzięki dodatkowym ramkom sterującym. Należy mieć jednak świadomość, iż w każdym przypadku formaty ramek, jak i zasady ich wymiany, nie pozostają bez wpływu na wydajność protokołu, a co za tym idzie, na uzyskiwaną w praktyce efektywną prędkość transmisji. W przypadku sieci bezprzewodowych ulega ona dalszej degradacji ze względu na określone właściwości sprzętu transmisyjnego oraz wpływ otoczenia.

Celem pracy jest określenie zależności między parametrami protokołu a jego osiągami. Z tego powodu przeprowadzone badania dotyczą głównie przypadku transmisji w warunkach idealnych bądź też jak najbardziej do nich zbliżonych.

Dalsza część monografii jest podzielona na trzy rozdziały, co w pewnym sensie odpowiada wskazanym powyżej właściwościom warstwy liniowej.

Pierwszy rozdział jest poświęcony zagadnieniu protokołów dostępu do łącza w sieciach bezprzewodowych, ze szczególnym uwzględnieniem sieci ad hoc. W części tej opisano zjawiska, występujące w sieciach bezprzewodowych, a mające istotny wpływ na działanie protokołu dostępu do łącza. Następnie przedstawiono metody unikania i wykrywania kolizji, możliwe do zrealizowania w sieciach bezprzewodowych. Na podstawie tych opisów określono warunki, w których wykrywanie kolizji – pomimo występowania efektu przechwytywania – jest możliwe w sieciach wykorzystujących promieniowanie podczerwone z wiązką rozproszoną. Porównano także zachowanie dwóch metod unikania kolizji w sieci ad hoc, zawierającej stacje ruchome, i określono kryterium skuteczności unikania kolizji metodą wymiany ramek sterujących. Kolejny fragment rozdziału przedstawia rywalizacyjne protokoły dostępu do łącza, zaprojektowane dla sieci bezprzewodowych. W protokołach tych są stosowane opisane wcześniej metody unikania i wykrywania kolizji. Dla wybranych protokołów przeprowadzono analizę wydajności w różnych warunkach pracy sieci, włączając warunki typowe dla kilku przypadków istniejących sieci bezprzewodowych. Wybrane protokoły zostały także zaimplementowane w małej, doświadczalnej sieci bezprzewodowej, w której doko-

nano pomiaru ich wydajności dla kilku wybranych konfiguracji. Uzyskane wyniki doświadczalne odbiegają nieco od wyników analitycznych, co może świadczyć o występowaniu w sieci zjawisk, które nie zostały uwzględnione w modelu, np. efektu przechwytywania.

Drugi rozdział poświęcono sieci Packet Radio oraz stosowanemu w niej protokołowi AX.25 i kontrolerom TNC, które są używane jako adaptery tej sieci. W tej części opisano zasady działania protokołu AX.25 z uwzględnieniem najnowszej wersji (2.2) oraz ważniejszych różnic w stosunku do wersji wcześniejszych (szczególnie najczęściej stosowanej wersji 2.0). Opisano także ogólną zasadę pracy kontrolerów TNC. Jako że są to układy mikroprocesorowe, zbudowane z seryjnych i stosunkowo łatwo dostępnych elementów, dokonano przeglądu konstrukcji obecnie dostępnych kontrolerów i porównano ich parametry konstrukcyjne i użytkowe. Opisano także główne funkcje oprogramowania sterującego pracą kontrolera.

Na podstawie opisu protokołu AX.25 stworzono jego model analityczny, umożliwiający określenie jego wydajności, efektywnej prędkości transmisji oraz opóźnień występujących podczas przesyłu danych. W modelu uwzględniono najważniejsze parametry protokołu. Model pozwala określić zachowanie protokołu w warunkach idealnych, może zatem stanowić punkt odniesienia dla wyników osiągniętych w warunkach rzeczywistych. Przez porównanie takich wyników można wówczas ocenić wpływ implementacji protokołu oraz sprzętu i oprogramowania transmisyjnego na rzeczywiste osiągi sieci. Wykorzystując stworzony model, przeanalizowano wpływ poszczególnych parametrów protokołu na jego wydajność dla obu wariantów łącza radiowego, pracującego z różnymi prędkościami transmisji. Wykorzystując model protokołu AX.25, stworzono także model analityczny kontrolera TNC. Model ten pozwala oszacować teoretyczny wpływ kontrolera na efektywną prędkość oraz opóźnienia transmisji, pozwala także oszacować pojemność bufora w kontrolerze TNC, która gwarantuje ciągłość transmisji po stronie nadawczej.

Wykonano także liczne testy w doświadczalnej sieci Packet Radio. Sieć ta składała się z komputera osobistego klasy IBM PC, do którego podłączano dwa kontrolery TNC. Transmisja między kontrolerami odbywała się przewodowo, a to w celu uniknięcia wpływu zakłóceń na przebieg transmisji. Dzięki temu warunki pracy kontrolerów były możliwie jak najbardziej zbliżone do idealnych. Wyniki badań wykazały silną zależność parametrów użytkowych sieci zarówno od mocy obliczeniowej kontrolera TNC, jak i od oprogramowania sterującego jego pracą, a zwłaszcza od szczegółów implementacji protokołu AX.25.

Trzeci rozdział monografii poświęcono bezprzewodowym sieciom lokalnym zgodnym ze standardem IEEE 802.11. Standard ten można uznać obecnie za najistotniejsze rozwiązanie w zakresie bezprzewodowych sieci lokalnych. W monografii opisano topologie sieci, określone w standardzie, a także wybrane elementy warstwy fizycznej i liniowej. Opisano także formaty ramek stosowane na poziomie warstwy fizycznej, a dokładniej podwarstwy PLCP,

ponieważ mają one kluczowe znaczenie dla oceny wydajności sieci. Kolejny fragment poświęcono warstwie liniowej protokołu, wyróżniając aspekty takie, jak formaty i typy ramek, zasady ich wymiany w różnych wariantach dostępu do łącza (DCF, PFC, EDCA, HCCA) oraz wybrane elementy zarządzania siecią. Opisano także mechanizm potwierdzenia blokowego. Osobny fragment poświęcono najnowszemu rozwiązaniu – dodatkowi IEEE 802.11n, który został ukończony w trakcie pisania monografii (listopad 2009). W tym fragmencie opisano wybrane aspekty warstwy fizycznej – szczególnie formaty ramek na poziomie podwarstwy PLCP – oraz rozszerzenia warstwy liniowej – agregację ramek A-MSDU i A-MPDU.

Opisane powyżej mechanizmy poddano analizie pod kątem ich wpływu na wydajność protokołu w różnych warunkach, a co za tym idzie, możliwej do uzyskania maksymalnej teoretycznej prędkości transmisji. Na podstawie opisu zasad wymiany ramek na poziomie warstwy liniowej oraz formatów stosowanych na poziomie podwarstwy PLCP wyznaczono zależności analityczne, które pozwalają na oszacowanie wydajności protokołu dla różnych wariantów warstwy fizycznej, prędkości transmisji, pojemności pola danych ramki oraz przyjętej zasady wymiany ramek. Wyprowadzone zależności pozwalają oszacować wydajność protokołu oraz efektywną prędkość transmisji na poziomie warstwy liniowej. Uzyskane wyniki mogą więc być nieco lepsze niż uzyskane w rzeczywistej sieci, w której występuje bardziej rozbudowany stos protokołów, a w szczególności protokół TCP/IP obecny nad warstwą liniową. Dla celów obliczeń przyjęto, że transmisja odbywa się w cyklicznie powtarzających się, identycznych fragmentach, które nazwano cyklami transmisyjnymi. Liczba tych cykli zależy oczywiście od całkowitej objętości przesyłanej informacji, ale w ramach jednego cyklu przesyła się ściśle określoną ilość informacji pochodzącej z wyższych warstw sieci, zależną tylko od zasad wymiany ramek oraz parametrów protokołu. Dzięki temu, analizując transmisję na poziomie pojedynczego cyklu transmisyjnego, można uzyskać wyniki niezależne od całkowitej objętości przesyłanej informacji.

Wykorzystując wyprowadzone zależności, oszacowano efektywność protokołu na poziomie warstwy liniowej dla wszystkich opisanych warstw fizycznych i metod wymiany ramek. Uzyskane wyniki są w przybliżeniu zgodne z wynikami, jakie można uzyskać w rzeczywistej sieci bezprzewodowej. Drobne odstępstwa są spowodowane tym, że w obliczeniach uwzględniono osiągi warstwy liniowej, podczas gdy w sieci rzeczywistej są stosowane także wyższe warstwy sieci. Może to świadczyć o wystarczającej dokładności przyjętego modelu sieci. Uzyskane wyniki pokazują, że już przy obecnie stosowanych warstwach fizycznych wydajność protokołu jest niewystarczająca, a to przez zbyt duży narzut, wprowadzany głównie przez warstwę fizyczną. Natomiast przy zastosowaniu potwierdzenia blokowego czy agregacji ramek wydajność sieci jest znacznie wyższa i pozwala na efektywne wykorzystanie prędkości transmisji określonych w obecnych warstwach fizycznych.

Każdy z opisanych powyżej rozdziałów zawiera krótkie podsumowanie, w którym podkreślono oryginalny wkład Autora. Z tego względu – aby uniknąć powtórzeń – podsumowanie całości monografii ma charakter bardziej ogólny. Wskazuje ono m. in. na problemy, z jakimi można się spotkać podczas badań nad efektywnością protokołów w sieciach bezprzewodowych. W zakończeniu wymieniono też wiele interesujących zagadnień związanych z budową warstwy liniowej bezprzewodowych sieci komputerowych, które – ze względu na charakter pracy – nie zostały w niej omówione. Pokrótce omówiono także wpływ zastosowania transmisji bezprzewodowej na wyższe warstwy sieci.

1. PROTOKOŁY DOSTĘPU DO ŁĄCZA

Protokoły dostępu do łącza są jednym z najważniejszych aspektów projektowania sieci komputerowych [103, 104], szczególnie niższych warstw. Jeżeli bowiem protokół jest źle dobrany do aplikacji, sieć może nie osiągnąć oczekiwanych wartości parametrów, jak np. przepustowość czy opóźnienie. Oczywiście zależą one także od innych warstw sieci, jednak podwarstwa dostępu do łącza jest odpowiedzialna za realizację wsparcia dla określonych cech sieci na najniższym poziomie. Z drugiej strony, bliskość warstwy fizycznej powoduje, że projektując protokół dostępu do łącza, należy wziąć pod uwagę cechy medium transmisyjnego.

W sieciach bezprzewodowych występuje wiele problemów, nieznanych z sieci przewodowych, a mających istotny wpływ na pracę protokołu dostępu do łącza [11]. Są one szczególnie widoczne w sieciach ad hoc [76], nieposiadających ani ustalonej struktury, ani centralnej stacji nadzorującej pracę pozostałych węzłów sieci. Wymienione właściwości powodują, że wiele typów protokołów dostępu do łącza w ogóle nie nadaje się do zastosowania w takiej sieci. Przykładowo, brak stacji centralnej uniemożliwia stosowanie protokołów wykorzystujących odpytywanie (ang. *polling*). Podobnie, nieregularna i często zmieniająca się struktura sieci, będąca skutkiem m. in. poruszania się węzłów sieci, praktycznie uniemożliwia stosowanie protokołów opartych na przekazywaniu żetonu (ang. *token passing*). W sieci ad hoc można spodziewać się, iż protokół taki większość czasu łącza zużyłby na ciągłą rekonfigurację obiegu żetonu, a nierzadko także na odtwarzanie jego obiegu wskutek jego zgubienia lub zdublowania. W rezultacie, w sieciach ad hoc używa się niemal wyłącznie protokołów rywalizacyjnych. Znane są, co prawda, propozycje protokołów rezerwacyjnych dla sieci ad hoc [78, 118], jednak – jak dotychczas – nie mają one większego znaczenia praktycznego.

Protokoły rywalizacyjne są często stosowane w lokalnych sieciach komputerowych, zarówno przewodowych (Ethernet), jak i bezprzewodowych (802.11). Z pewnością wpływ na to ma ich względna prostota oraz wystarczająca wydajność, a także stosunkowo niewielkie opóźnienia przy niskim obciążeniu sieci. Z drugiej strony, protokoły takie tracą stabilność przy wysokich obciążeniach sieci, nie gwarantują także uzyskania dostępu do łącza w możliwym do określenia czasie. Protokoły rywalizacyjne, prawdopodobnie dzięki swej prostocie, znajdują także zastosowanie jak protokoły pomocnicze w rozwiązaniach o większym stopniu

złożoności. Przykładowo, w protokołach rezerwacyjnych kanał zgłaszania żądań działa zazwyczaj zgodnie z regułą dostępu rywalizacyjnego, najczęściej Aloha lub CSMA.

Skuteczność protokołu rywalizacyjnego opiera się na umiejętności unikania kolizji przed wysłaniem ramki bądź też na wykrywaniu kolizji już w trakcie transmisji. O ile oba mechanizmy można dość łatwo zrealizować w sieci przewodowej (np. Ethernet), w sieciach bezprzewodowych wykrywanie kolizji najczęściej nie jest możliwe, zaś unikanie kolizji metodami znanymi z sieci przewodowych, np. przez wykrywanie nośnej, nie zawsze prowadzi do prawidłowej oceny stanu łącza, a co za tym idzie, do podjęcia słusznej decyzji o rozpoczęciu bądź dalszym powstrzymaniu nadawania. Stanowi to zresztą swoisty paradoks [84], ponieważ pierwsza w historii sieć oparta na rywalizacji (Aloha) była właśnie siecią bezprzewodową. Jednakże trudności m. in. w zapewnieniu odpowiednich – tj. zarówno skutecznych, jak i łatwych w implementacji – metod unikania kolizji dla sieci bezprzewodowych spowodowały, że rozwój sieci bezprzewodowych znacznie się opóźnił.

1.1. Właściwości bezprzewodowych mediów transmisyjnych

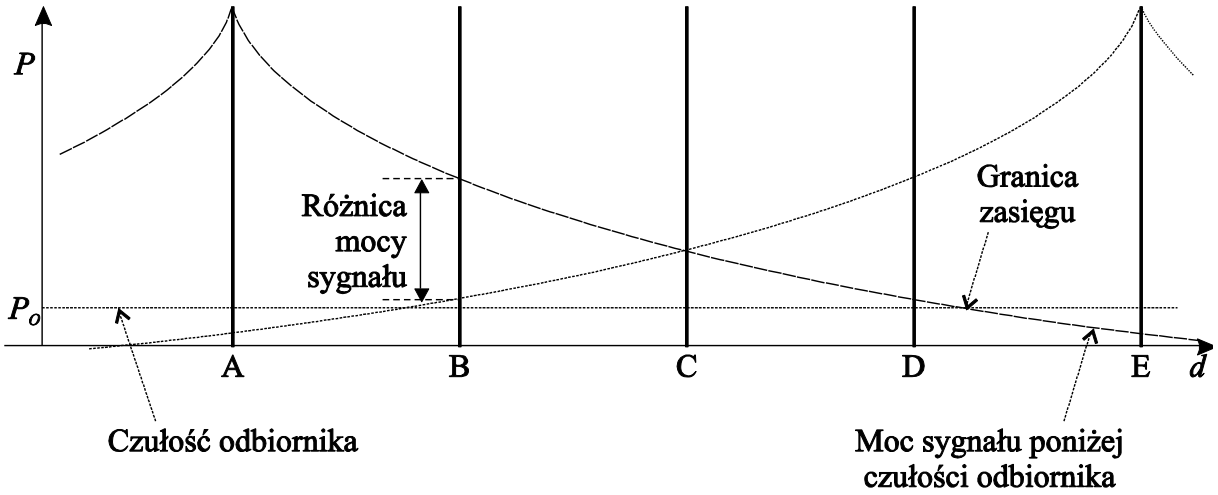
Opisując właściwości bezprzewodowych mediów transmisyjnych, należałoby rozpocząć od fizycznych właściwości fal elektromagnetycznych, propagacji fal radiowych i optycznych z poszczególnych zakresów, sposobów modulacji i kodowania sygnału. Jednakże zagadnienia te dotyczą raczej budowy warstwy fizycznej łącza bezprzewodowego. Jako że niniejsza praca dotyczy protokołów warstwy liniowej, w dalszej części zostaną opisane tylko wybrane właściwości, mające wpływ na budowę tej właśnie warstwy [43].

Można przyjąć, iż większość problemów charakterystycznych dla sieci bezprzewodowych wynika z faktu stopniowego zmniejszania się mocy sygnału wraz z odległością. Pomimo że stwierdzenie to opisuje sieci bezprzewodowe jedynie w przybliżeniu, może stanowić punkt wyjścia do dalszych rozważań (rys. 1.1). Należy jednak mieć na uwadze, że w sieciach bezprzewodowych – zarówno radiowych, jak i optycznych – występuje zjawisko propagacji wielodrogowej. Jest ono szczególnie dotkliwe w sieciach radiowych, gdyż powoduje powstawanie zaników, będących skutkiem nakładania się na siebie fal docierających do odbiornika różnymi drogami, a więc różniących się co do fazy sygnału. Jeśli wśród odebranych sygnałów znajdują się wyłącznie fale odbite, mówi się o zanikach Rice'a, gdy natomiast występuje także fala bezpośrednia – o zanikach Raileigha [110].

1.1.1. Transmisja dwukierunkowa

W łączach bezprzewodowych oczywiście można prowadzić transmisję dwukierunkową, gdyż możliwe jest uzyskanie dwukierunkowości metodą podziału częstotliwości lub czasu. Jednakże nie jest możliwa realizacja systemu transmisji radiowej, w którym można jednocze-

śnie nadawać i odbierać na tej samej częstotliwości. Zazwyczaj jest konieczne wyłączenie odbiornika podczas nadawania, ponieważ moc sygnału pochodzącego z własnego nadajnika jest za duża dla odbiornika. Gdyby nawet jednoczesne nadawanie i odbiór na tej samej częstotliwości były możliwe, i tak własny nadajnik zagłuszyłby sygnały pochodzące z pozostałych stacji (por. rozdz. 1.2.4). Nie jest zatem możliwe wykrywanie kolizji w sposób znany z sieci przewodowych, jak np. Ethernet, a więc przez nasłuch podczas nadawania. Wobec tego, aby zachować możliwie wysoką wydajność protokołu, należy stosować efektywne metody unikania kolizji.



Rys. 1.1. Uproszczona zależność między odległością (d) a mocą sygnału (P)
Fig. 1.1. Simplified relation between distance (d) and signal power (P)

1.1.2. Błędna ocena stanu łącza

W prawidłowo działającej sieci przewodowej – szczególnie sieci lokalnej – każda stacja posiada bezpośrednią łączność z wszystkimi pozostałymi stacjami, a więc może wykryć prowadzone przez nie transmisje. Właściwość ta nie jest jednak wymagana w sieciach bezprzewodowych, szczególnie sieciach ad hoc. Ponieważ moc sygnału zmniejsza się wraz z odległością, niektóre stacje, pomimo przynależności do jednej sieci, mogą znajdować się poza swoim zasięgiem. Zależność ta może nawet być asymetryczna, tj. stacja A może być w zasięgu stacji B, ale B poza zasięgiem A. Nietrudno zauważyć, że brak bezpośredniej łączności znacznie utrudnia – a nierzadko wręcz uniemożliwia – prawidłową ocenę stanu łącza za pomocą wykrywania nośnej.

1.1.3. Zmienność łącza bezprzewodowego w czasie

Właściwości łącza bezprzewodowego są zmienne w czasie, a to z powodu mobilności stacji oraz innych obiektów znajdujących się w otoczeniu sieci [50, 80, 110]. Ze względu na propagację wielodrogową odebrany sygnał jest złożeniem sygnału bezpośredniego, odbitego, ugiętego i rozproszonego. Ponieważ składniki te mogą różnić się co do fazy sygnału, moc

sygnału odbieranego jest różna w różnych miejscach i w różnym czasie. W rezultacie, można wyróżnić miejsca, w których moc sygnału jest za niska, aby możliwy był prawidłowy odbiór. Stacje sieci mogą więc tymczasowo „znikać" i „pojawiać się" w sieci. Warto zauważyć, że stacja „znikająca" może utracić część przesyłanej informacji, podobnie stacja „pojawiająca się" może nie posiadać informacji wystarczającej do prawidłowej oceny stanu łącza. Skala tego zjawiska może zależeć od parametrów łącza, jak np. prędkość transmisji, częstotliwość czy format ramki, a także od prędkości poruszania się stacji [119]. Dokładniejsza analiza tego przypadku znajduje się w dalszej części pracy (por. rozdz. 1.3.7).

Zmienności charakterystyki łącza bezprzewodowego nie można zniwelować za pomocą anten o rozmiarach porównywalnych z długością fali, jest to więc zjawisko charakterystyczne dla sieci radiowych. Natomiast w sieciach opartych na promieniowaniu podczerwonym długość fali jest znacznie mniejsza od rozmiarów czujnika podczerwieni, zatem sieci takie będą wolne od opisywanego zjawiska [50].

1.1.4. Błędy transmisji

Ze względu na opisaną powyżej zmienność łącza bezprzewodowego w czasie, błędy transmisji występują tu o wiele częściej niż w sieciach przewodowych [43]. Ponadto, są one spowodowane nie tylko szumem i zakłóceniami, lecz także zanikami. Jeśli bowiem moc sygnału odbieranego tylko nieznacznie przekracza czułość odbiornika, nawet niewielkie wahanie poziomu sygnału może spowodować błąd transmisji. Skutki opisanego zjawiska można w pewnym stopniu wyeliminować przez wprowadzenie korekcji błędów i zmniejszenie rozmiaru ramki, ale są to techniki wykraczające poza architekturę protokołu dostępu do łącza.

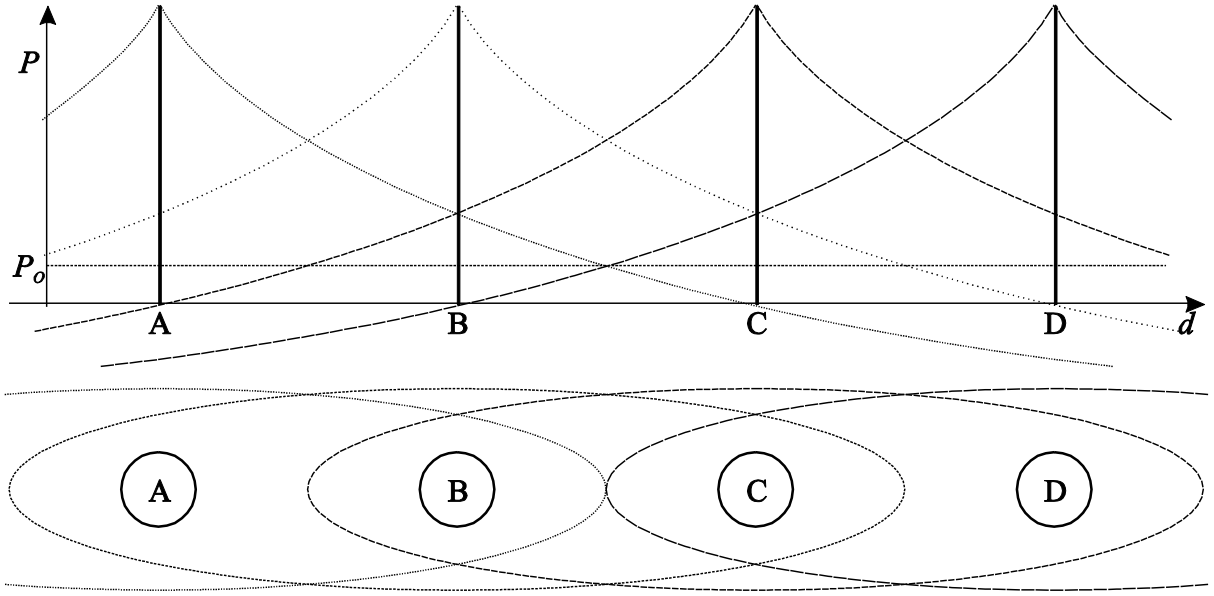
1.2. Zjawiska występujące w sieciach bezprzewodowych

Opisane powyżej właściwości fizyczne łącza bezprzewodowego są przyczyną występowania zjawisk, istotnych z punktu widzenia projektowania protokołu dostępu do łącza. Rozważmy sieć bezprzewodową, w której zasięg każdej stacji obejmuje tylko najbliższe stacje sąsiednie w każdym kierunku, jak pokazano na rys. 1.2. Przyjmijmy także, iż wykrywanie nośnej jest możliwe tylko w obszarze zasięgu użytecznego danej stacji[1].

W rozważanej sieci można zaobserwować:

- zjawisko stacji ukrytej,
- zjawisko stacji odkrytej,
- efekt przechwytywania (zdobywania),
- zakłócenia (interferencje).

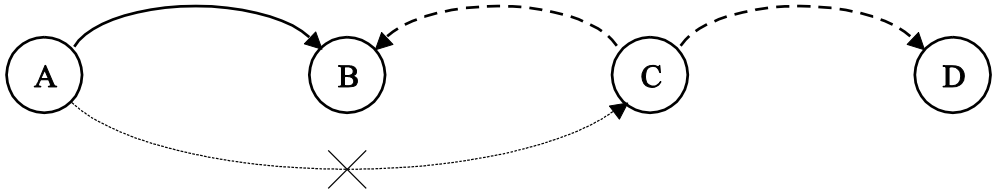
[1] Większość prac dotyczących rozważanego tematu także przyjmuje takie założenie. Jednak np. praca [19] podaje, że wykrywanie nośnej jest możliwe także poza obszarem zasięgu użytecznego.



Rys. 1.2. Przykładowa sieć bezprzewodowa i zasięgi transmisji poszczególnych stacji
Fig. 1.2. Example of wireless network and transmission ranges of individual stations

1.2.1. Zjawisko stacji ukrytej

Zjawisko stacji ukrytej (ang. *hidden station, hidden terminal*) [11, 107] może wystąpić w sieciach, w których nie wszystkie stacje posiadają bezpośrednią, wzajemną łączność. Stację nazywamy ukrytą, jeśli znajduje się w zasięgu odbiorcy pewnej informacji, ale poza zasięgiem nadawcy. Sytuacja taka jest przedstawiona na rys. 1.3.



Rys. 1.3. Ilustracja zjawiska stacji ukrytej
Fig. 1.3. An illustration of hidden station phenomenon

W rozważanym przypadku stacja A nadaje do stacji B. Ponieważ stacje A i C znajdują się poza swoim zasięgiem, stacja C nie wykrywa nośnej pochodzącej ze stacji A, przyjmuje zatem, iż łącze jest wolne. Stacja C może więc rozpocząć swoją transmisję do stacji B lub D. Transmisja ta spowoduje w stacji B kolizję z danymi przesyłanymi ze stacji A. Kolizja ta jest źródłem błędów transmisji, co z kolei powoduje zmniejszenie wydajności łącza wskutek konieczności retransmisji, lecz także z powodu utraty części pakietów i spowodowanych tym opóźnień, także w protokołach wyższych warstw sieci.

Warto zauważyć, iż pojęcie stacji ukrytej jest silnie uzależnione od wzajemnego rozmieszczenia stacji. Stacja może być ukryta wobec pewnej transmisji, a odkryta wobec innej. Jeśli protokół warstwy liniowej stosuje mechanizm potwierdzeń, stacja, która była ukryta

podczas transmisji danych, staje się stacją odkrytą (por. rozdz. 1.2.2) podczas transmisji potwierdzenia.

Wśród stacji ukrytych można rozróżnić ukryte nadajniki i odbiorniki [10, 11]. Rozróżnienie takie jest możliwe w sieciach, w których transmisji danych towarzyszy przesłanie dodatkowej informacji sterującej, na podstawie której stacja ocenia stan łącza; pozwala to stacjom na stwierdzenie, czy są ukryte lub odkryte wobec bieżącej transmisji danych.

1.2.1.1. Ukryty nadajnik

Ukryty nadajnik to stacja ukryta, która ma informacje do wysłania. Aby nie zakłócała ona przesyłu danych między nadawcą a odbiorcą, musi zostać powiadomiona o tej transmisji. Ponieważ stacja ukryta znajduje się poza zasięgiem nadawcy, odpowiedzialność za powiadomienie stacji ukrytej spoczywa na odbiorcy informacji.

1.2.1.2. Ukryty odbiornik

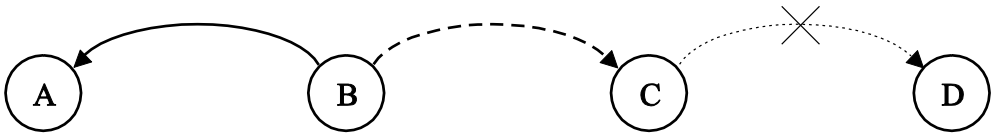
Ukryty odbiornik natomiast to stacja ukryta, która jest adresatem informacji przesłanej od innej stacji, znajdującej się poza zasięgiem zarówno nadawcy, jak i odbiorcy, aktualnie przebiegającej transmisji. Ukryty odbiornik może wprawdzie odebrać wywołanie, jednak nie może na nie odpowiedzieć tak długo, jak długo pozostaje stacją ukrytą. Odpowiedź taka spowodowałaby bowiem kolizję z już przebiegającą transmisją. Brak odpowiedzi od ukrytego odbiornika można zinterpretować na wiele sposobów. Do najbardziej typowych można zaliczyć następujące sytuacje:

1) stacja ukryta powstrzymuje się z wysłaniem odpowiedzi,
2) wywołanie zostało zniszczone w wyniku kolizji lub błędów transmisji,
3) odpowiedź na wywołanie została zniszczona w wyniku kolizji lub błędów transmisji,
4) ukryta stacja tymczasowo zniknęła z zasięgu stacji próbującej się z nią skomunikować,
5) ukryta stacja jest wyłączona.

W każdym z opisanych przypadków zachowanie stacji wywołującej stację ukrytą powinno być inne. Wprawdzie w większości sytuacji wystarczy dokonać retransmisji, jednak opóźnienie, jakie musi upłynąć przed ponowieniem próby skomunikowania się ze stacją ukrytą, jest w każdej sytuacji inne. Zakres tego opóźnienia we współczesnych sieciach lokalnych może wahać się od około kilkuset mikrosekund (przypadki 2 i 3), poprzez kilka milisekund (przypadek 1) do kilku lub kilkunastu sekund w przypadku 4. Jeśli ukryty odbiornik jest wyłączony, retransmisja jest praktycznie bezcelowa. Dla prawidłowego działania sieci jest zatem niezbędne, aby stacja wywołująca miała możliwość rozpoznania stanu stacji ukrytej. Jest to jednak niemożliwe tak długo, jak długo stacja wywoływana jest ukryta, gdyż, informując o swoim stanie, mogłaby zakłócić inną transmisję. Wysłanie odpowiedzi będzie natomiast możliwe po zakończeniu transmisji, wobec której rozważana stacja jest ukryta.

1.2.2. Zjawisko stacji odkrytej

Zjawisko stacji odkrytej (ang. *exposed station, exposed terminal*) [11, 107] również może wystąpić w sięciach, w których nie wszystkie stacje posiadają bezpośrednią, wzajemną łączność. Stację nazywamy odkrytą, jeśli znajduje się w zasięgu nadawcy pewnej informacji, ale poza zasięgiem odbiorcy. Sytuacja taka jest przedstawiona na rys. 1.4.



Rys. 1.4. Ilustracja zjawiska stacji odkrytej
Fig. 1.4. An illustration of exposed station phenomenon

W rozważanym przypadku stacja B nadaje do stacji A. Ponieważ stacje B i C znajdują się w swoim zasięgu, transmisja ta zostaje wykryta przez stację C. Stacja ta zatem uważa łącze za zajęte i powstrzymuje transmisję do stacji D. Transmisja ta nie zakłóciłaby jednak przesyłu z B do A, ponieważ A i C są poza swoim zasięgiem. Można więc stwierdzić, że stacja odkryta niepotrzebnie wstrzymuje transmisję, powodując tym samym spadek wydajności łącza. Tym niemniej, jeśli nadawca (stacja B) oczekuje jakiejkolwiek informacji zwrotnej (np. potwierdzenia) od odbiorcy, stacja odkryta powinna wstrzymać transmisję w czasie, gdy owa informacja jest oczekiwana. W najprostszym przypadku stacja odkryta powinna powstrzymać się od nadawania podczas całej wymiany informacji między nadawcą a odbiorcą.

Pojęcie stacji odkrytej – podobnie jak i ukrytej – jest silnie uzależnione od wzajemnej lokalizacji stacji. Przykładowo, stacja będąca odkrytą wobec pewnego przesyłu danych, może stać się stacją ukrytą podczas transmisji potwierdzenia przez odbiorcę.

Podobnie jak w przypadku stacji ukrytej, także i tutaj można wyróżnić odkryte nadajniki i odbiorniki [10, 11].

1.2.2.1. Odkryty nadajnik

Odkryty nadajnik to stacja odkryta, mająca dane do wysłania. Wprawdzie znajduje się ona poza zasięgiem odbiorcy przesyłanej informacji, jednak w wielu przypadkach wskazane jest, by powstrzymywała się od nadawania przez cały czas trwania tej transmisji. Wynika to z możliwości przesyłania informacji zwrotnej od odbiorcy (A) do nadawcy (B). Warto zauważyć, że dla zapewnienia prawidłowego działania opisanego mechanizmu nie wystarczy stosowanie wykrywania nośnej. W tym bowiem przypadku stacja odkryta mogłaby rozpocząć transmisję w chwili zwolnienia łącza przez nadawcę informacji, uniemożliwiając mu odebranie informacji zwrotnej od odbiorcy.

1.2.2.2. Odkryty odbiornik

Odkryty odbiornik to stacja odkryta, wywoływana przez stację znajdującą się poza zasięgiem i nadawcy i odbiorcy aktualnie przesyłanej informacji. Jest to sytuacja zbliżona do przy-

padku ukrytego odbiornika. Odbiornik ukryty jednak może odebrać wywołanie, podczas gdy odbiornik odkryty nie ma takiej możliwości, gdyż znajduje się w zasięgu nadawcy. W związku z tym, w chwili zwolnienia łącza przez nadawcę, odkryty odbiornik nie jest nawet świadom, że była do niego kierowana jakakolwiek transmisja. Przypadek ten jest zatem trudniejszy do rozwiązania niż przypadek odkrytego nadajnika.

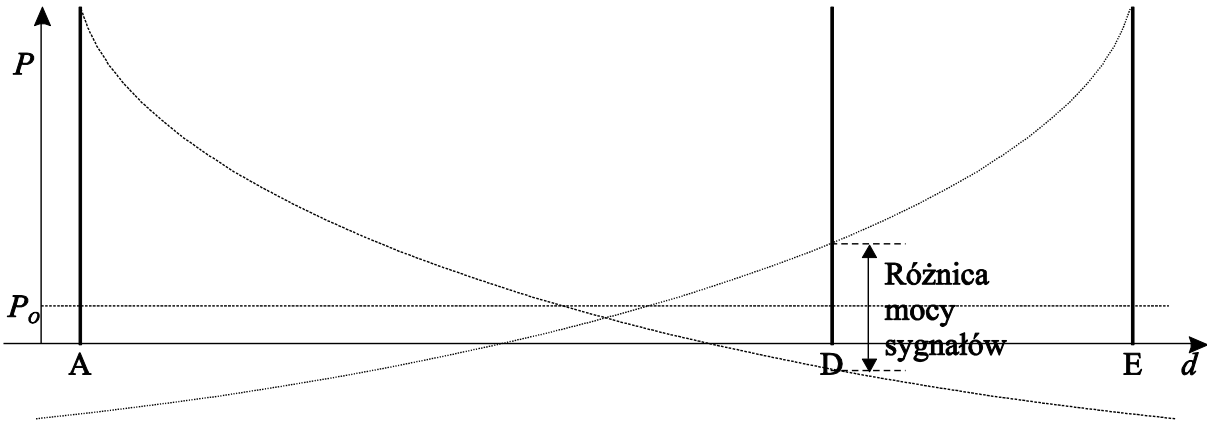
1.2.3. Interferencje

W sieciach bezprzewodowych można wyróżnić wiele źródeł zakłóceń. Wiele z nich jest nierozerwalnie związanych z techniką transmisji radiowej [7, 99, 110]. Zakłócenia te nie będą tu rozważane. W sieciach bezprzewodowych istnieją jednak także zakłócenia pochodzące od innych stacji nadawczych, prowadzących transmisje na tej samej lub zbliżonej częstotliwości.

Można przyjąć, że prawidłowy odbiór informacji jest możliwy, gdy:

- moc odbieranego sygnału przekracza próg czułości odbiornika,
- moc sygnału użytecznego jest wystarczająco większa od mocy sygnałów zakłócających i szumów.

Zależności te pokazano na rys. 1.5.



Rys. 1.5. Warunki prawidłowego odbioru sygnału
Fig. 1.5. Conditions of proper signal reception

Stację zakłócającą można zdefiniować jako znajdującą się poza zasięgiem zarówno nadawcy, jak i odbiorcy pewnej informacji. Jednocześnie stacja ta jest wystarczająco blisko, aby zakłócać prowadzoną transmisję. W przeciwieństwie do zjawiska stacji ukrytej i odkrytej ani nadawca, ani odbiorca nie jest w stanie poinformować stacji zakłócającej o przebiegającej transmisji. Na rys. 1.5 stacją zakłócającą jest A, o ile jest prowadzona transmisja ze stacji E do D.

Zjawisko stacji zakłócającej wynika z faktu istnienia zasięgu zakłóceniowego, którym charakteryzuje się każdy nadajnik radiowy. Można przyjąć, iż zasięg ten jest około 1,5÷2 razy większy od zasięgu użytkowego danego nadajnika [112]. Aby zatem wyeliminować lub przynajmniej zmniejszyć negatywny wpływ stacji zakłócających na pracę sieci, konieczne jest

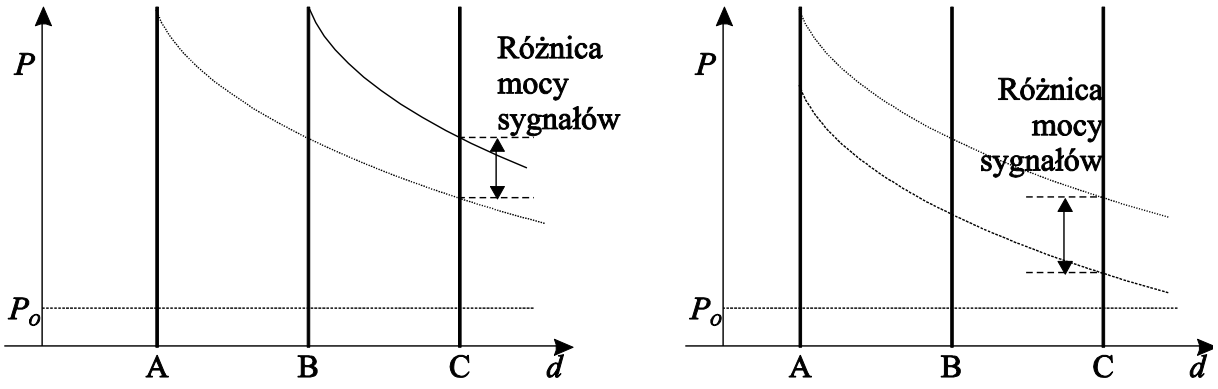
przesłanie informacji o stanie łącza w taki sposób, aby była ona dostępna na odpowiednio większym obszarze.

1.2.4. Efekt przechwytywania

Efekt przechwytywania (ang. *capture effect*) [72, 86, 104] występuje w sieciach bezprzewodowych wówczas, gdy moc sygnałów odbieranych z różnych nadajników jest różna w miejscu odbioru. Zjawisko to jest zresztą charakterystyczne nie tylko dla sieci komputerowych, lecz także dla innych bezprzewodowych – szczególnie radiowych – systemów komunikacyjnych. Można się z nim spotkać m. in., gdy:

- nadajniki pracujące z tą samą mocą są umieszczone w różnych odległościach od odbiornika (przypadek typowy dla sieci komputerowych),
- nadajniki pracujące z różnymi mocami są umieszczone w tej samej odległości od odbiornika (przypadek typowy dla sieci radiowo-telewizyjnych).

Wymienione sytuacje wyjaśniono na rys. 1.6. Oczywiście efekt przechwytywania może wystąpić także wówczas, gdy zarówno moce nadajników, jak i ich odległości od odbiornika są różne.



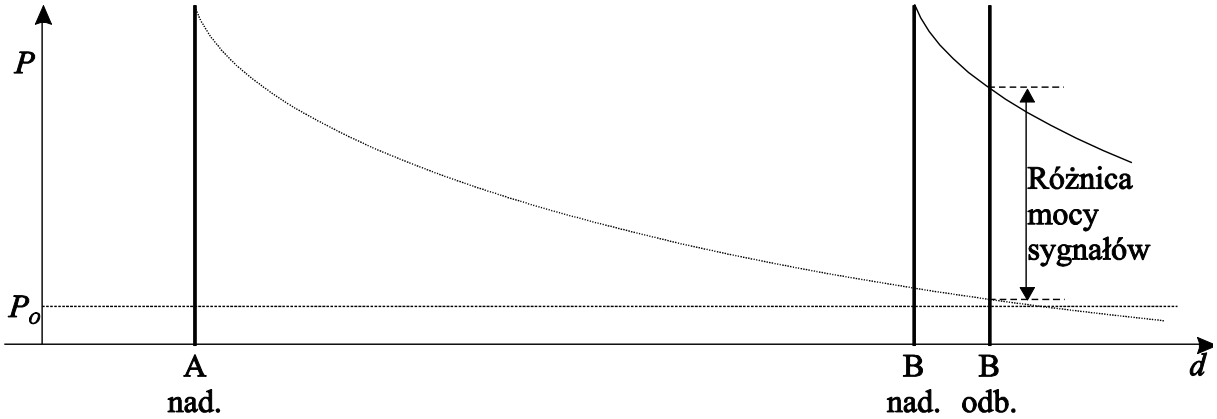
Rys. 1.6. Ilustracja efektu przechwytywania
Fig. 1.6. An illustration of capture effect

W wyniku efektu przechwytywania najsilniejszy sygnał zostaje odebrany poprawnie, zagłuszając sygnały pozostałe. Spośród kilku kolidujących ramek jedna – a mianowicie ta odebrana z najwyższą mocą – może wówczas zostać odebrana poprawnie, dzięki czemu efektywna przepustowość kanału transmisyjnego nie ulega zmniejszeniu. Można przyjąć, że w sieciach radiowych efekt przechwytywania występuje, gdy moc sygnału najmocniejszego przekracza moc kolejnego sygnału o około 1,5÷3 dB [86].

Efekt przechwytywania ma jednak również i negatywne skutki. Jednym z nich jest brak możliwości uzyskania równomiernego podziału łącza. Pewne stacje, znajdujące się bliżej swych odbiorców bądź też – ogólniej rzecz ujmując – umieszczone w miejscach zapewniają-

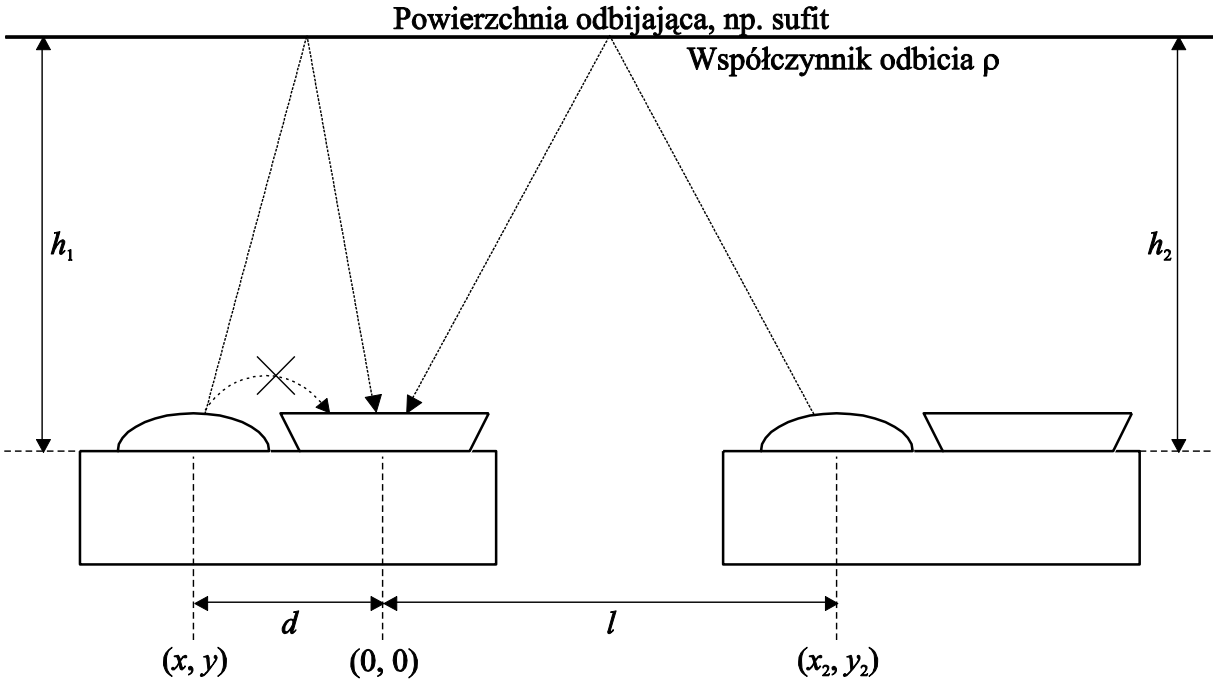
cych lepszą propagację sygnałów, mogą zawłaszczać łącze kosztem stacji rozmieszczonych mniej korzystnie.

Innym negatywnym skutkiem efektu przechwytywania jest niemożność prowadzenia nasłuchu w czasie nadawania, a co za tym idzie, nie jest możliwe wykrywanie kolizji w sposób znany np. z przewodowych sieci Ethernet (niektóre źródła, np. [25, 83], błędnie podają, że wykrywanie kolizji w sieciach radiowych jest jednak możliwe). Nawet gdyby układy nadajników-odbiorników radiowych mogły prowadzić jednoczesne nadawanie i odbiór, sygnał pochodzący z własnego nadajnika zagłuszyłby wszelkie inne sygnały pochodzące z innych stacji. Jest to szczególny przypadek sytuacji, gdy nadajniki pracują z jednakową mocą, a ich odległość od odbiornika jest różna. Jeden z nadajników znajduje się wówczas bardzo blisko odbiornika, co praktycznie uniemożliwia odbiór sygnałów z innych nadajników. Sytuację taką przedstawiono na rys. 1.7.



Rys. 1.7. Ilustracja efektu przechwytywania z nadajnikiem lokalnym i odległym
Fig. 1.7. An illustration of capture effect with local and remote transmitter

Wydaje się, że wykrywanie kolizji przez nasłuch podczas nadawania jest możliwe w sięciach opartych na promieniowaniu podczerwonym z wiązką rozproszoną (ang. *diffuse infrared*) [50]. Warunkiem koniecznym dla zapewnienia działania takiej sieci jest optyczna separacja toru nadawczego i odbiorczego w taki sposób, że sygnał z własnego nadajnika może dotrzeć do odbiornika tylko dzięki odbiciom. W tej sytuacji moc sygnału odebranego z nadajnika własnego i oddalonego będzie – w pewnych warunkach – porównywalna. Ideę takiego rozwiązania ilustruje rys. 1.8.



Rys. 1.8. Wykrywanie kolizji w sieci z podczerwienią rozproszoną
Fig. 1.8. Collision detection in a diffuse infrared network

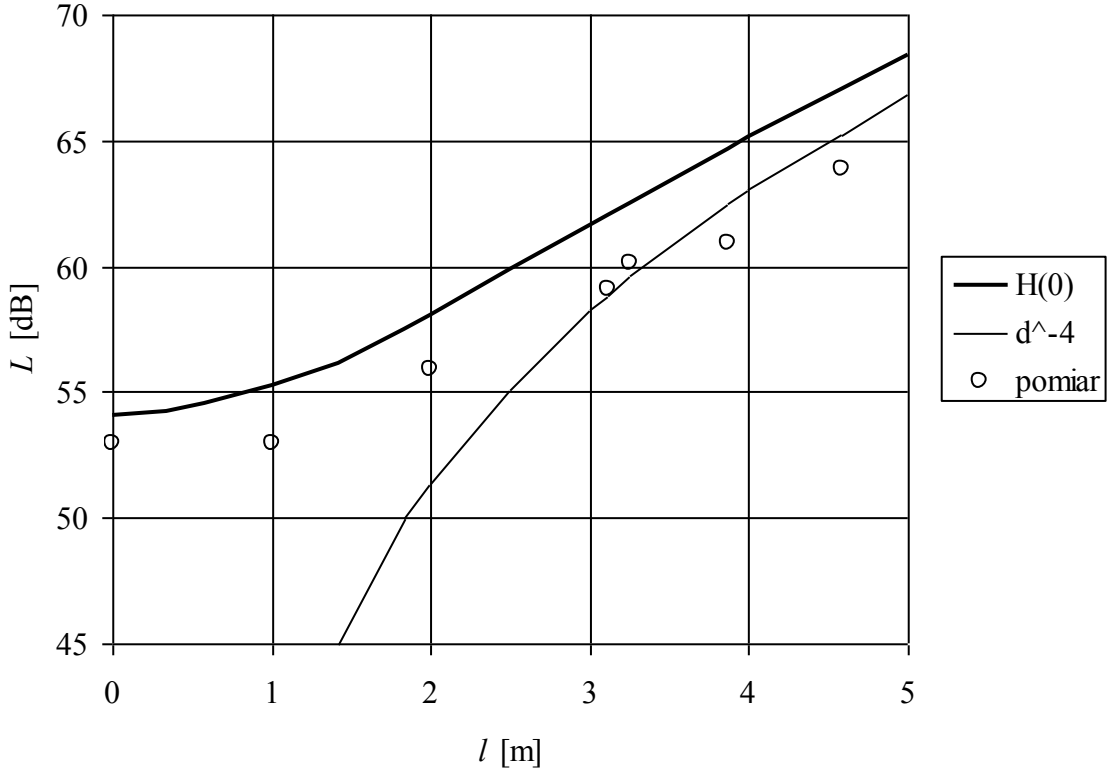
Rozważmy bezprzewodową sieć optyczną z wiązką rozproszoną, jak pokazano na rys. 1.8. W takiej sieci zysk energetyczny trasy można w przybliżeniu – zakładając tylko jedno odbicie sygnału – określić następującą zależnością [60]:

$$H(0) = \frac{\rho T_s g A h_1^2 h_2^2}{\pi^2} \cdot \iint_{sufit} \frac{dx\,dy}{\left(h_1^2 + x^2 + y^2\right)^2 \left[h_2^2 + \left(x - x_2\right)^2 + \left(y - y_2\right)^2\right]^2}, \tag{1.1}$$

gdzie: g – zysk koncentratora optycznego, ρ – współczynnik odbicia, A – powierzchnia detektora, zaś T_s opisuje transmisję sygnałów przez filtr kierunkowy. Wartości x, x_2, y, y_2, h_1 oraz h_2 pokazano na rys. 1.8. Zależność (1.1) ilustruje rys. 1.9. Na rysunku tym zamieszczono także krzywą, reprezentującą zależność wynikającą z prawa czwartej potęgi, a także, dla porównania, pokazano wyniki pomiarów rzeczywistej sieci [60].

Jak widać na rys. 1.9, przy odległości między nadajnikiem a odbiornikiem nieprzekraczającej 1 m tłumienność trasy jest praktycznie niezmienna [60], zatem moc sygnału pochodzącego z własnego odbiornika jest w przybliżeniu równa mocy sygnału pochodzącego z innej stacji. Odpowiednia konstrukcja nadajników-odbiorników podczerwieni oraz ich rozmieszczenie (rys. 1.8) może wówczas pozwolić na wykrywanie kolizji w sposób znany z przewodowych sieci Ethernet – podczas nadawania. Wraz ze wzrostem odległości dzielącej stacje skuteczność tej metody będzie coraz mniejsza w związku z różnicą mocy – przy odległości 2 m sygnał jest około dwukrotnie słabszy niż przy 1 m. Może to sprzyjać występowaniu efektu przechwytywania, podobnie jak w sieciach radiowych. Przy odległości pomiędzy stacjami przekraczającej 3 m tłumienność trasy rośnie w przybliżeniu zgodnie z prawem czwartej po-

tęgi. Można zatem przypuszczać, że przy takiej odległości i odpowiadającemu jej stosunkowi mocy sygnałów efekt przechwytywania nie wystąpi, a sygnał najsilniejszy całkowicie zagłuszy wszelkie pozostałe.



Rys. 1.9. Tłumienność trasy w funkcji odległości dla sieci z podczerwienią rozproszoną [60]
Fig. 1.9. Path loss as a function of distance for diffuse infrared network

1.3. Metody unikania i wykrywania kolizji

W sieciach przewodowych wykrywanie nośnej jest wystarczająco skuteczną metodą unikania kolizji, ponieważ wszystkie stacje mają możliwość wzajemnej komunikacji. Po uzupełnieniu tej metody wykrywaniem kolizji można uzyskać efektywność i stabilność wystarczającą dla większości zastosowań niewymagających deterministycznego dostępu do sieci. Natomiast w sieciach bezprzewodowych wykrywanie nośnej nie jest wystarczające ze względu na opisane powyżej zjawiska stacji ukrytej i odkrytej. W dodatku w większości sieci bezprzewodowych nie można stosować wykrywania kolizji znanego z sieci przewodowych, jak np. Ethernet. Zapewnienie wysokiej wydajności sieci bezprzewodowych wymaga zatem stosowania nowych, wydajniejszych metod unikania i wykrywania kolizji, zaprojektowanych specjalnie dla tych sieci [131, 132, 144, 145, 149].

We współczesnych sieciach bezprzewodowych, prócz wykrywania nośnej, można stosować następujące metody unikania kolizji:

- wykrywanie tonu zajętości,

- wymianę ramek sterujących,
- przesył informacji sterującej w osobnym kanale transmisyjnym.

Wymienione metody mogą zastąpić lub uzupełnić wykrywanie nośnej. Niektóre z nich można także łączyć ze sobą w celu uzyskania wyższej skuteczności.

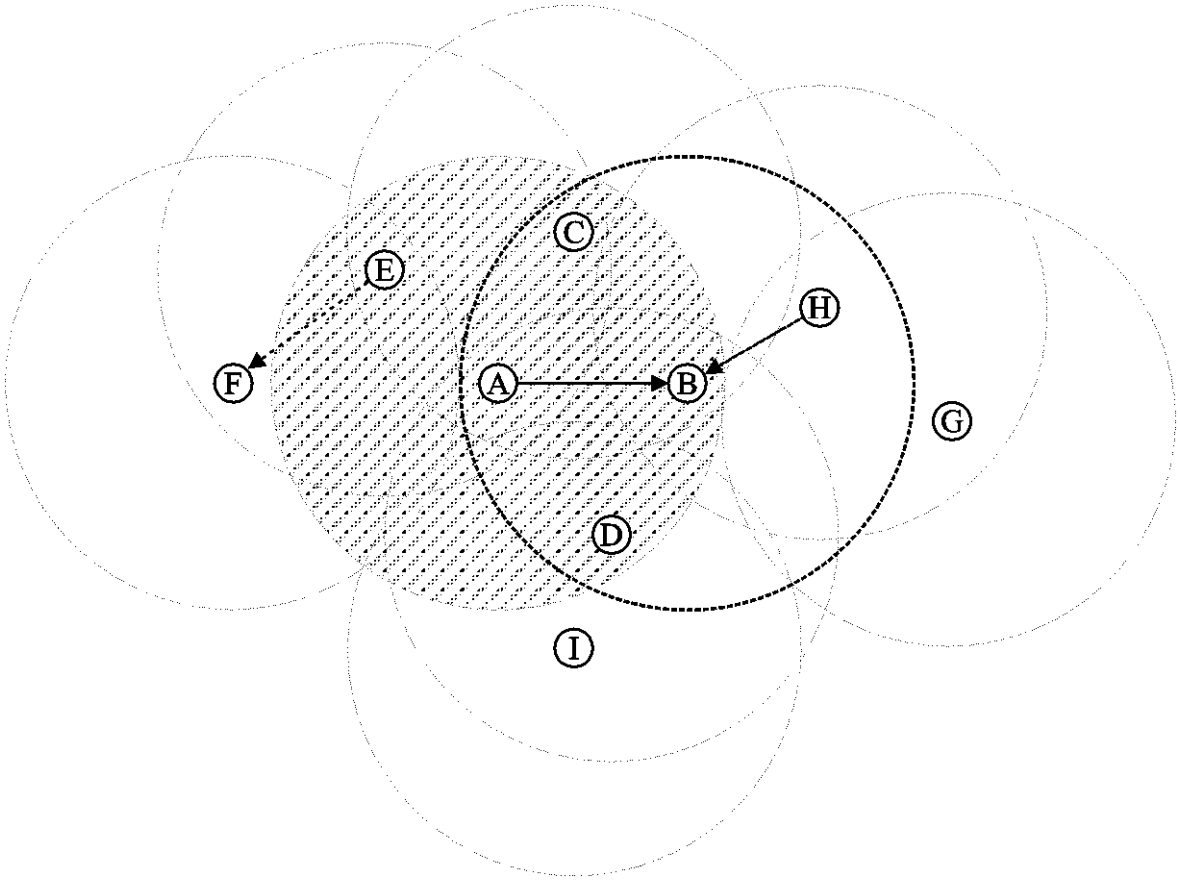
1.3.1. Wykrywanie nośnej

Wykrywanie nośnej jest jedną z podstawowych metod unikania kolizji zarówno w sieciach przewodowych, jak i bezprzewodowych. O ile w tych pierwszych charakteryzuje się wystarczająco dobrą skutecznością, o tyle w drugich może prowadzić do błędnej oceny stanu łącza. Problem ten opisano w rozdziale 1.2. W celu zilustrowania tego zagadnienia oraz porównania wykrywania nośnej z innymi metodami unikania kolizji, zaproponowanymi specjalnie dla sieci bezprzewodowych, można wykorzystać przykładową sieć, pokazaną na rys. 1.10. Sieć ta składa się z kilku stacji z zaznaczonymi zasięgami transmisji. Stacje te w dalszych rozważaniach pełnią następujące role:

- A – nadawca,
- B – odbiorca,
- C i D – znajdują się w zasięgu zarówno nadawcy, jak i odbiorcy i mogą prawidłowo ocenić stan łącza, korzystając z mechanizmu wykrywania nośnej (ang. *carrier detection*),
- E – stacja odkryta (ang. *exposed*),
- H – stacja ukryta (ang. *hidden*),
- I – stacja zakłócająca (ang. *interfering*).

Wszystkie stacje znajdujące się dokoła nadawcy (C, D i E) rozpoznają transmisję do odbiorcy, korzystając z mechanizmu wykrywania nośnej. Stacja ukryta (H), zgodnie z definicją, jest umieszczona w zasięgu stacji B, lecz poza zasięgiem A. Stwarza ona zatem niebezpieczeństwo interferencji z przebiegającą transmisją z A do B. Z drugiej strony, stacja odkryta (E) powstrzymuje się od nadawania do stacji F. Sygnał tej stacji nie interferowałby jednak z sygnałem przesyłanym z A do B, gdyż stacje B i E, podobnie jak A i F, znajdują się poza swoim zasięgiem. Transmisje te można zatem prowadzić jednocześnie.

Jak widać na rys. 1.10, wykrywanie nośnej zapewnia ochronę w całym otoczeniu nadawcy, pozostawiając jednak bez ochrony część otoczenia odbiorcy. W tej części znajdują się właśnie stacje ukryte. Można zatem uznać, że przedstawiony rysunek tłumaczy niewystarczającą skuteczność wykrywania nośnej w sieciach bezprzewodowych i konieczność poszukiwania innych metod unikania kolizji. Pomimo to wykrywanie nośnej jest chętnie stosowane także w sieciach bezprzewodowych, ze względu na relatywną – w porównaniu z innymi metodami – łatwość implementacji. Wiele układów radiowych, w tym układów przeznaczonych do sieci o małym zasięgu, udostępnia informację o obecności nośnej w łączu.



Rys. 1.10. Obszar zajęty przez wykrywanie nośnej dla transmisji A→B
Fig. 1.10. Area occupied by carrier sense for A→B transmission

1.3.2. Wykrywanie tonu zajętości

Wykrywanie tonu zajętości może zastąpić wykrywanie nośnej. Pasmo częstotliwości jest podzielone na dwa kanały [107]:

- kanał komunikatów, używany dla transmisji danych i zajmujący większość pasma,
- kanał zajętości, używany dla sygnalizacji stanu łącza, relatywnie wąski.

Jeżeli stacja ma ramkę gotową do wysłania, przed rozpoczęciem transmisji musi sprawdzić stan kanału. Obecność tonu zajętości świadczy o zajętości łącza, transmisję należy zatem wstrzymać. W przeciwnym przypadku ramkę można natychmiast wysłać.

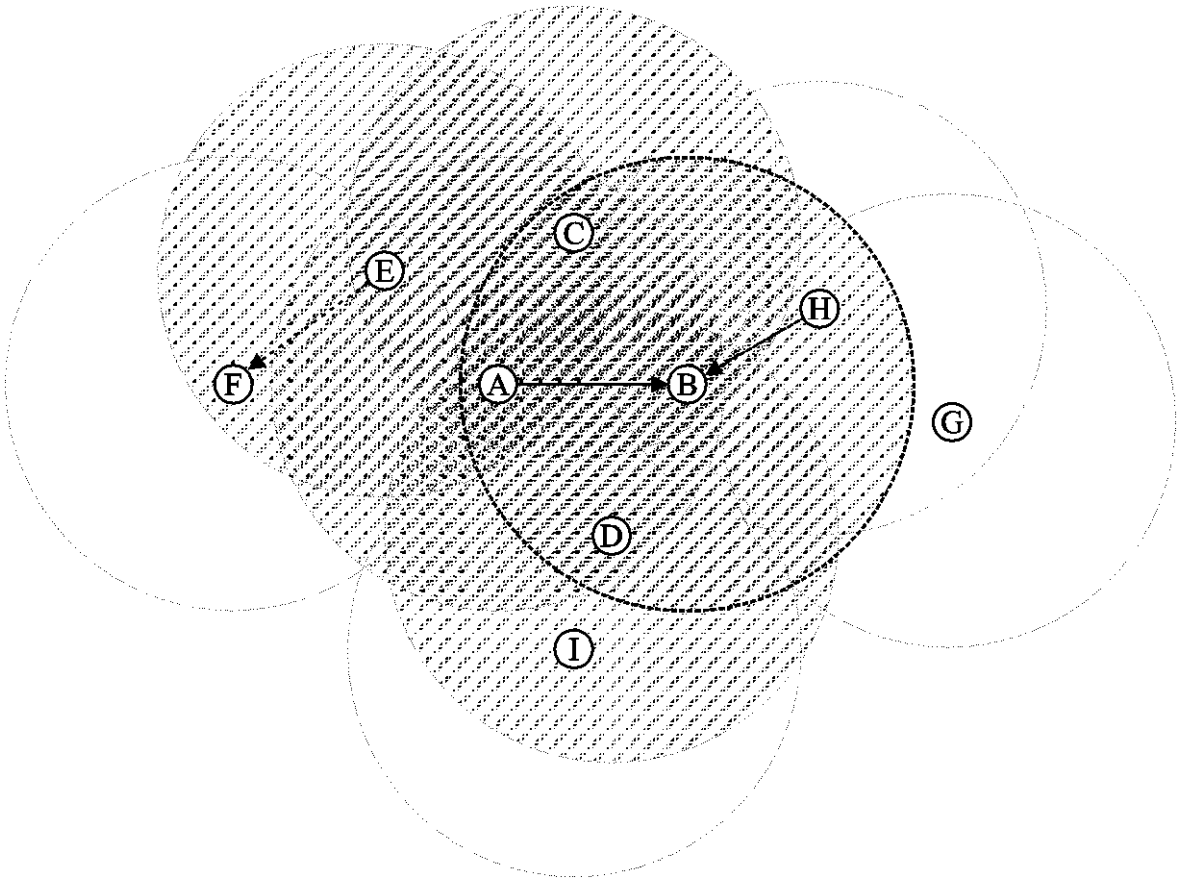
Sygnał zajętości jest zazwyczaj falą sinusoidalną, którą może wytworzyć:

1) każda stacja, odbierająca dane z kanału komunikatów,
2) wyłącznie adresat informacji,
3) początkowo każda stacja wykrywająca transmisję, później natomiast – wyłącznie adresat.

1.3.2.1. Wytwarzanie tonu zajętości przez wiele stacji

Pierwsza z metod [107] jest najprostsza i bardzo efektywna w zakresie zmniejszania liczby stacji ukrytych. Niestety, słabością jej jest znaczny i niepotrzebny wzrost liczby stacji od-

krytych. Można powiedzieć, że obszar zajęty przez określoną transmisję jest znacznie większy niż rzeczywiście potrzeba. Zjawisko to ilustruje rys. 1.11.



Rys. 1.11. Obszar zajęty przez wykrywanie tonu zajętości dla transmisji A→B (metoda 1)
Fig. 1.11. Area occupied by busy-tone sense for A→B transmission (1st method)

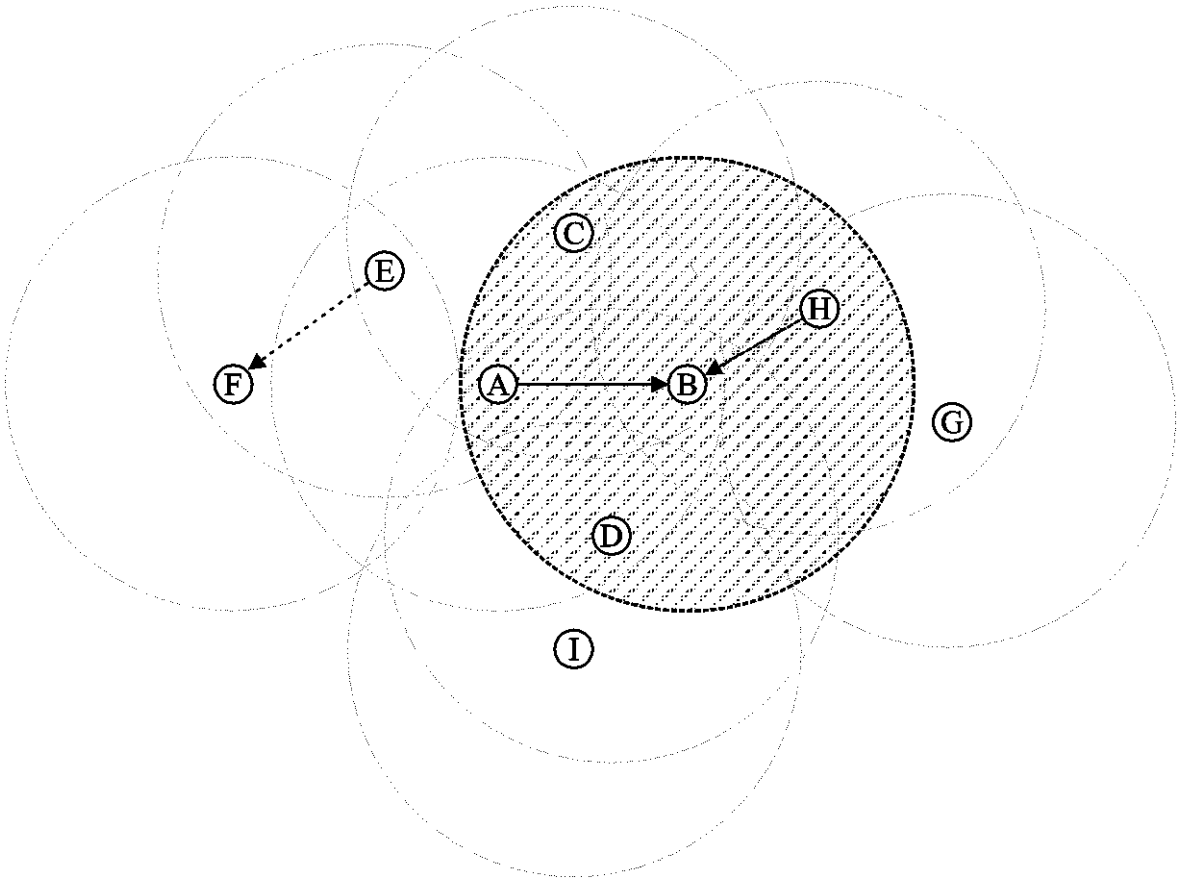
W przedstawionej sieci wszystkie stacje znajdujące się w zasięgu nadawcy (A), włączając odbiorcę (B), wytwarzają ton zajętości. Stacja ukryta (H) jest zatem poinformowana o przebiegającej transmisji. Tym niemniej, stacja odkryta (E) musi powstrzymać się od nadawania, pomimo że znajduje się poza zasięgiem stacji B i nie przeszkadzałaby rozważanemu przesyłowi z A do B. Warto zaznaczyć, że – w przeciwieństwie do CSMA – stacja E byłaby zmuszona do powstrzymania transmisji także, gdyby była umieszczona poza zasięgiem nadawcy. Dla przykładu, w rozważanej sieci ewentualna transmisja ze stacji F również musi zostać opóźniona. Można zatem powiedzieć, że powstrzymać się od nadawania powinna każda stacja, znajdująca się w zasięgu stacji, będącej w zasięgu nadawcy. Jak widać na rys. 1.11, nie jest to jednak konieczne dla zachowania prawidłowej ochrony sygnału wokół odbiorcy.

Pewnego rodzaju efektem ubocznym przyjętego rozwiązania jest możliwość poinformowania stacji zakłócającej (I) o prowadzonej transmisji. W rozpatrywanym przykładzie stacja ta znajduje się bowiem w zasięgu stacji D, wytwarzającej ton zajętości zgodnie z przyjętymi regułami. Należy jednak wyraźnie zaznaczyć, iż jest to możliwe wyłącznie dzięki korzystnej

lokalizacji stacji D – gdyby znajdowała się ona np. w pobliżu stacji C, stacja zakłócająca byłaby już poza jej zasięgiem.

1.3.2.2. Wytwarzanie tonu zajętości wyłącznie przez adresata

Aby uniknąć zwiększonej liczby stacji odkrytych, wytwarzanie tonu zajętości można ograniczyć tylko do adresata informacji [101]. Metoda ta chroni ramkę przed kolizją wokół odbiorcy – a więc w jedynym miejscu, w których ochrona taka jest rzeczywiście potrzebna. Nie ma bowiem potrzeby ochrony ramki wokół nadawcy lub stacji, do których ramka nie jest kierowana. Metoda ta także jest stosunkowo prosta, ale nie chroni ramki do chwili rozpoznania adresu nadawcy. Na początku transmisji ramka jest zatem czuła na kolizje. Obszar, zajęty przez transmisję, pokazano na rys. 1.12. Można zauważyć, że – mimo iż stacja ukryta (H) nadal jest informowana o transmisji – stacja E, poprzednio odkryta, nie musi już powstrzymywać transmisji do stacji F. Co więcej, stacja F także nie jest już odkryta i, w razie potrzeby, może nadawać równolegle z transmisją pomiędzy stacjami A i B. Można się jednak spodziewać, iż – zależnie od odległości od stacji B – będzie ona w pewnym stopniu zakłócać tę transmisję, podobnie jak stacja I.

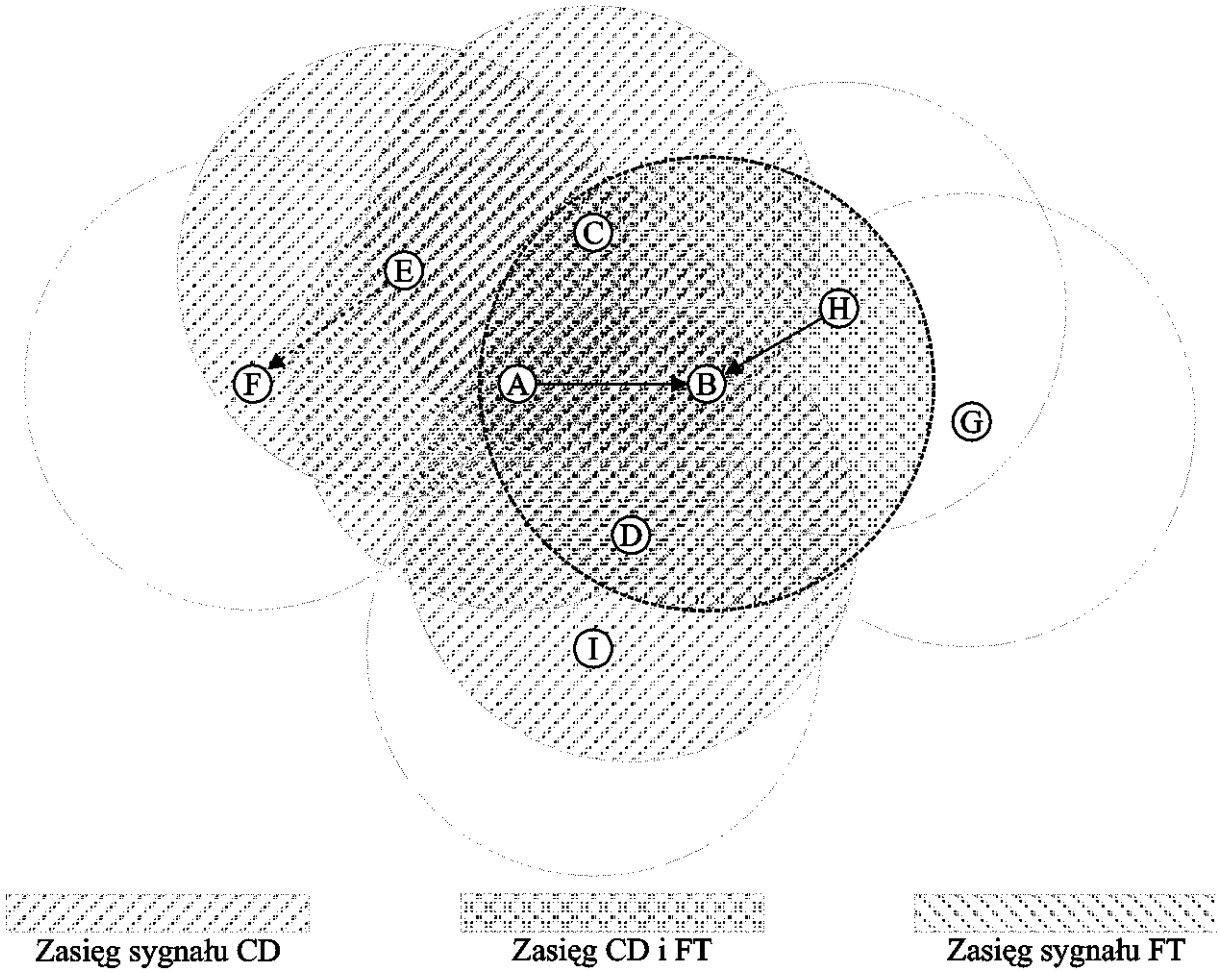


Rys. 1.12. Obszar zajęty przez wykrywanie tonu zajętości dla transmisji A→B (metoda 2)
Fig. 1.12. Area occupied by busy-tone sense for A→B transmission (2nd method)

Pewnym mankamentem tej metody jest konieczność reorganizacji formatu ramek tak, aby adres docelowy przesłany był możliwie jak najwcześniej, a to w celu jak najszybszego wytworzenia tonu zajętości przez adresata. Warto także zadbać o to, by adres ten był chroniony sumą kontrolną, aby zapobiec błędom identyfikacji adresata wskutek błędów transmisji.

1.3.2.3. Hybrydowe wytwarzanie tonu zajętości

Trzecia metoda jest kombinacją metod opisanych powyżej. Charakteryzuje się ona występowaniem dwóch różnych tonów zajętości [42]. Pierwszy z nich (CD, ang. *carrier detect*) jest wytwarzany przez wszystkie stacje znajdujące się w zasięgu nadajnika, gdy tylko zostanie wykryta transmisja w kanale danych. Stan ten trwa do chwili rozpoznania adresu przeznaczenia. Wówczas adresat informacji rozpoczyna wysyłanie drugiego tonu zajętości (FT, ang. *feedback tone*), którego obecność świadczy o prawidłowym rozpoznaniu adresu przez adresata. Pozostałe stacje wyłączają generowanie tonu zajętości CD, gdy tylko okaże się, że nie są one adresatami przesyłanej informacji. Sygnały CD i FT są odbierane na obszarze pokazanym na rys. 1.13.



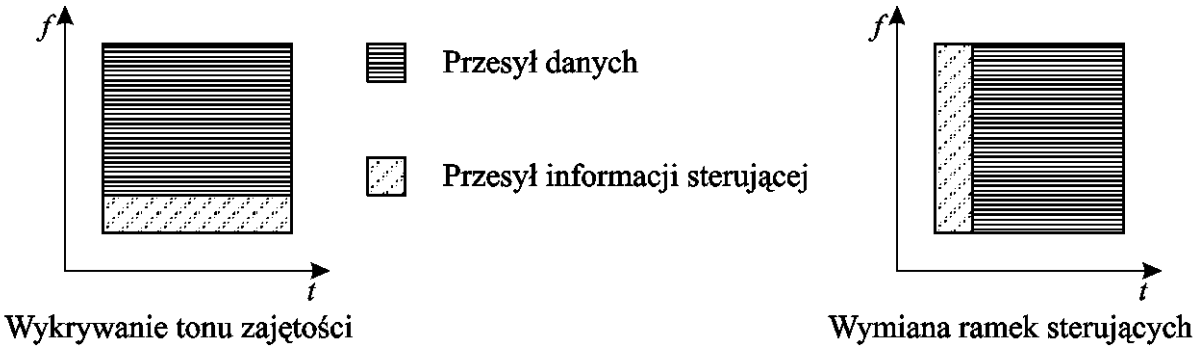
Rys. 1.13. Obszar zajęty przez wykrywanie tonu zajętości dla transmisji A→B (metoda 3)
Fig. 1.13. Area occupied by busy-tone sense for A→B transmission (3rd method)

Warto zauważyć, że ostatnia metoda ma pewne cechy wykrywania kolizji [42]. Jeżeli ramkę podzieli się na preambułę (zawierającą adres docelowy) i rdzeń (zawierający dane), można wykryć kolizję przed rozpoczęciem wysyłania danych. Po wysłaniu preambuły nadajnik czeka na otrzymanie tonu zajętości od adresata. Brak takiego tonu może oznaczać kolizję lub jakikolwiek inny błąd podczas transmisji preambuły, a nawet nieobecność adresata. W każdym z wymienionych przypadków nie ma sensu wysyłania pozostałej części ramki. Z tego punktu widzenia ten sposób jest podobny do wymiany ramek sterujących.

Pewną wadą każdej odmiany wykrywania tonu zajętości jest możliwość łatwego zablokowania sieci przez zagłuszanie kanału zajętości. Ponadto, metody te nie są łatwe do zaimplementowania w praktyce, ponieważ wymagają dwu- lub trójkanałowych układów nadawczo-odbiorczych. Z tego powodu, pomimo interesujących właściwości, nie mają zbyt wielu praktycznych zastosowań.

1.3.3. Wymiana ramek sterujących

Jako że wykrywanie nośnej nie jest wydajną metodą unikania kolizji w obecności stacji ukrytych lub odkrytych, a wykrywanie tonu zajętości jest kosztowne w implementacji, można poprzedzić transmisję danych wymianą informacji sterującej [62]. Można uznać takie podejście za pewną formę jednokanałowej implementacji wykrywania tonu zajętości, przy czym ton zajętości jest tu zastąpiony przez ramki sterujące, przesyłane w tym samym kanale co dane. O ile jednak ton zajętości chroni ramkę w sposób ciągły, o tyle ramki sterujące jedynie poprzedzają transmisję danych. Ilustruje to rys. 1.14. Opisana różnica może mieć znaczenie np. w sieciach zawierających stacje ruchome.



Rys. 1.14. Sposoby przesyłania informacji sterującej w dwóch metodach unikania kolizji
Fig. 1.14. Control information transmission in two collision avoidance methods

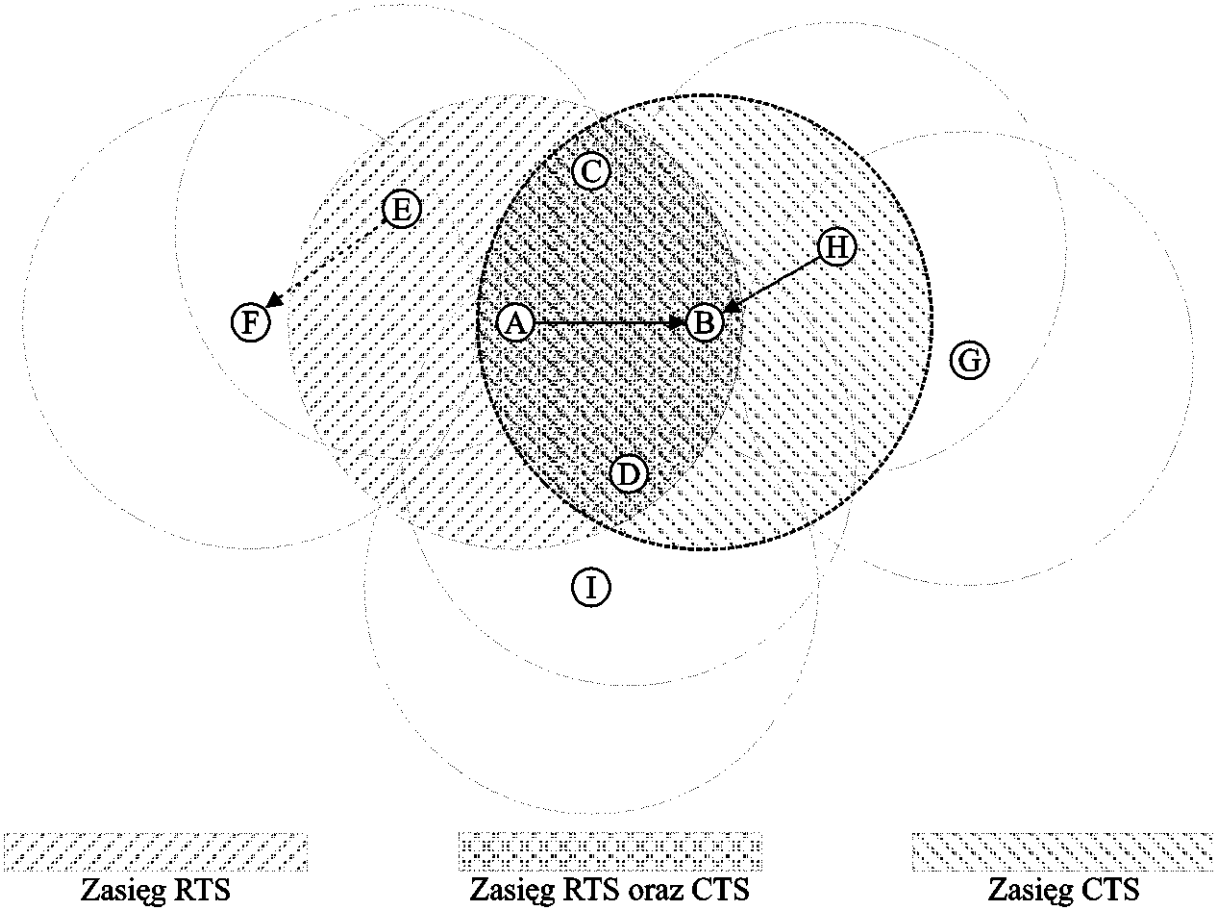
Gdy stacja ma ramkę do wysłania, wysyła najpierw ramkę sterującą, zwaną RTS (ang. *Request To Send*), adresowaną do odbiornika informacji. Ramka powinna zawierać informację dotyczącą długości ramki danych lub przewidywanego czasu transmiji. Jeżeli adresat odbierze tę ramkę poprawnie, odpowiada nadawcy ramką CTS (ang. *Clear To Send*). Ramka ta również powinna zawierać informację o długości ramki danych, która może jednak być inna niż w ramce RTS. Gdy nadajnik poprawnie odbierze ramkę CTS, można przyjąć, że łącze jest

zarezerwowane na potrzeby transmisji między tymi dwiema stacjami przez czas wynikający z zawartości ramek RTS i CTS.

Ramki RTS i CTS mogą oczywiście być odbierane także przez inne stacje znajdujące się w otoczeniu nadajnika i odbiornika. Stacje te można podzielić na następujące grupy:

- stacje odbierające tylko ramkę RTS,
- stacje odbierające tylko ramkę CTS,
- stacje odbierające obie ramki.

Jeżeli stacja odbiera tylko ramkę RTS, znajduje się w zasięgu nadajnika, ale poza zasięgiem odbiornika; można więc przyjąć, iż jest to stacja odkryta. Z drugiej strony, jeśli stacja odbiera tylko ramkę CTS, znajduje się w zasięgu nadajnika, ale poza zasięgiem nadajnika, jest więc stacją ukrytą. Stacja, odbierająca zarówno RTS, jak i CTS, nie jest ani odkryta, ani ukryta i mogłaby poprawnie rozpoznać stan kanału używając wyłącznie wykrywania nośnej. Zasięg poszczególnych typów ramek pokazano na rys. 1.15.



Rys. 1.15. Obszar zajęty przez wymianę ramek sterujących dla transmisji A→B
Fig. 1.15. Area occupied by control frames exchange for A→B transmission

Można wykazać [32], że aby poprawnie chronić ramki danych przed kolizjami spowodowanymi przez stacje ukryte lub odkryte, długość ramki CTS powinna być większa niż RTS. Warto przy tym zauważyć, że w standardzie IEEE 802.11 – najpopularniejszym obecnie stan-

dardzie sieci bezprzewodowych wykorzystującym wymianę ramek sterujących – warunek ten nie jest spełniony.

W niektórych przypadkach celowe jest użycie dodatkowych ramek sterujących. Na przykład, po poprawnej negocjacji RTS-CTS nadajnik może wysłać ramkę DS (ang. *Data Sending*) [9, 11]. Jest to szczególnie użyteczne, gdy z pewnych powodów odbiorca informacji nie może przyjąć ramki o długości zaproponowanej przez nadawcę. W takim przypadku odsyła w ramce CTS inną długość, jednak informacja ta może nie dotrzeć do wszystkich stacji sąsiadujących z nadawcą (np. stacji odkrytych). Ramka DS pozwala zatem przekazać im nowe ustalenia dotyczące czasu rezerwacji łącza. Jest to o tyle ważne, że stacja odkryta, która odebrała ramkę RTS – lecz nie CTS – mogłaby prowadzić transmisję przez czas przesłany w ramce RTS, zakładając, że nie będzie to zakłócać wymiany informacji między nadawcą a odbiorcą. Jeżeli jednak protokół zakłada przesłanie potwierdzenia przez odbiorcę, potwierdzenie takie mogłoby wówczas ulec kolizji z danymi przesyłanymi przez stację odkrytą, uniemożliwiając nadawcy odebranie potwierdzenia. Przesłanie przez nadawcę ramki DS przed wysłaniem danych pozwala uniknąć tej niedogodności. Warto zauważyć przy tym, że stacja odkryta wobec transmisji danych staje się stacją ukrytą wobec przesyłu potwierdzenia.

Inną techniką godną uwagi jest wytwarzanie potwierdzeń na poziomie podwarstwy dostępu do łącza [11]. Oczywiście, wyższe warstwy również mogą wykryć nieprawidłowości transmisji, jednak warstwa dostępu może uczynić to szybciej, zmniejszając czas potrzebny na wykrycie kolizji lub innego błędu oraz zażądanie retransmisji.

Wymiana ramek sterujących jest względnie prostą i wydajną metodą unikania kolizji w obecności stacji ukrytych lub odkrytych. Tym niemniej zastosowana w pojedynkę nie zapewnia wystarczającej skuteczności ze względu na ryzyko kolizji między ramkami sterującymi [34]. Kolizje mogą także wystąpić w środowiskach mobilnych, np. gdy nowa stacja zbliża się do stacji zaangażowanych w transmisję. Stacja taka mogła być poza zasięgiem sieci podczas negocjacji RTS-CTS, ale podczas przesyłu danych znajduje się już w zasięgu. Nie posiadając zatem wiedzy o rezerwacji łącza, może zakłócić przebiegającą transmisję danych. Zjawisko takie można wyeliminować, wspierając wymianę ramek sterujących albo wykrywaniem nośnej [34], albo wykrywaniem tonu zajętości [27, 28].

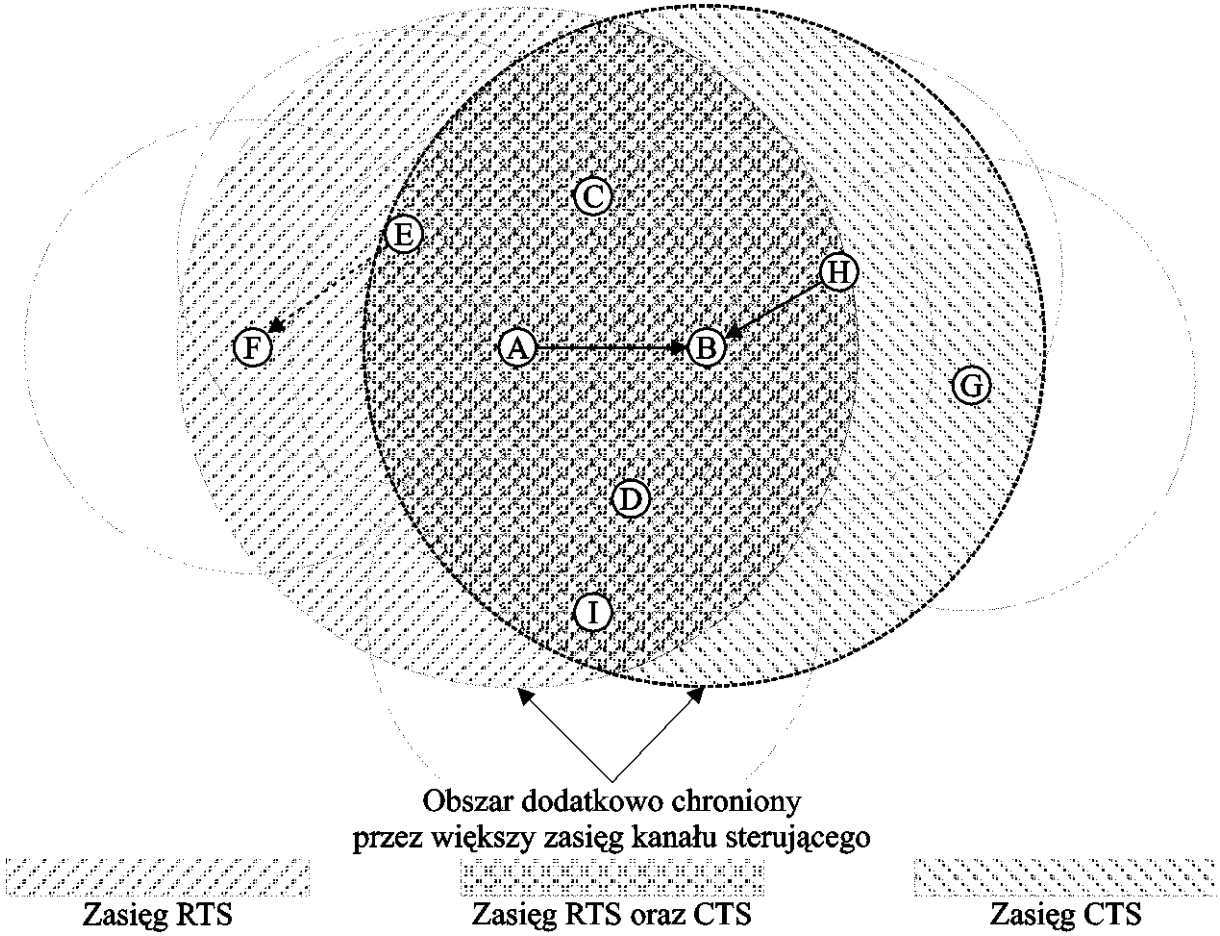
1.3.4. Wydzielony kanał sterujący

Zarówno wykrywanie tonu zajętości, jak i wymiana ramek sterujących rozwiązują jedynie problem ukrytego nadajnika i odkrytego nadajnika [9]. Aby rozwiązać także problem ukrytego lub odkrytego odbiornika, należy zapewnić, by:

- ukryty odbiornik mógł przesłać informację o wstrzymywaniu transmisji,
- odkryty odbiornik mógł odebrać informację sterującą, nawet gdy odbiera dane.

W obu przypadkach niezbędne jest wprowadzenie dodatkowego kanału sterującego. Pozwala to na uniknięcie kolizji między ramkami sterującymi i danych. Ukryty odbiornik może zatem odpowiedzieć na wywołanie, nie zakłócając transmisji – przykładowo, może poinformować stację wywołującą o swojej sytuacji. Podobnie, odkryty odbiornik może odebrać dowolną informację sterującą i odpowiedzieć na nią.

Warto zauważyć, że zwiększenie zasięgu transmisji kanału sterującego polepsza warunki przesyłu informacji. Stacje, które w kanale danych znajdują się poza zasięgiem zarówno odbiornika, jak i nadajnika informacji, mogą znajdować się dostatecznie blisko, aby zakłócać transmisję. Zwiększenie zasięgu kanału sterującego zmienia te stacje w odkryte lub ukryte, tak więc można już poinformować je o stanie łącza i przygotowaniach do transmisji. Zjawisko to ilustruje rys. 1.16.



Rys. 1.16. Obszar zajęty przez wymianę ramek sterujących w osobnym kanale o zwiększonym zasięgu
Fig. 1.16. Area occupied by control frames exchange in a separate channel of extended range

Podobnie do wykrywania tonu zajętości, osobny kanał sterujący – szczególnie gdy ma większy zasięg – jest względnie trudny w implementacji i wymaga bardziej złożonego (i droższego) układu transmisyjnego. Tak więc, mimo swoich zalet, metoda ta nie jest praktycznie stosowana.

1.3.5. Wykrywanie kolizji

W bezprzewodowych sieciach ad hoc najczęściej nie jest możliwe wykrywanie kolizji w sposób znany np. z sieci Ethernet. Dzieje się tak dlatego, że stosowane urządzenia nadawczo-odbiorcze uniemożliwiają nasłuch łącza w czasie nadawania. Gdyby zresztą nawet było to możliwe, ze względu na efekt przechwytywania, sygnał z własnego nadajnika zagłuszałby wszelkie inne sygnały docierające do odbiornika. Tak więc jedynym sposobem na wykrycie kolizji jest potwierdzanie poprawnego odbioru ramki. Brak takiego potwierdzenia może świadczyć jednak nie tylko o kolizji, lecz także o innym błędzie transmisji.

W przypadku gdy konieczne jest wykrywanie kolizji już w czasie transmisji ramki, można zastosować metodę polegającą na chwilowej przerwie w nadawaniu [68, 69, 87]. Metoda ta może działać w każdym łączu dwukierunkowym naprzemiennym, wliczając w to kanały radiowe lub optyczne. Ocena stanu kanału odbywa się przy tym na zasadzie wykrywania nośnej. Po stwierdzeniu, że kanał nie jest zajęty, stacja może rozpocząć transmisję ramki, ale po losowo wybranym czasie przerywa na chwilę nadawanie i ponownie bada stan kanału. Brak nośnej w tym czasie oznacza brak kolizji i stacja może dokończyć przesyłanie ramki. Natomiast obecność nośnej świadczy o kolizji ramek pochodzących z różnych stacji. Stacja, która wykryła kolizję, kontynuuje transmisję jeszcze przez pewien czas, aby inne stacje także mogły wykryć kolizję. Mechanizm ten nie gwarantuje wykrycia wszystkich kolizji, ponieważ dwie stacje mogą przerwać transmisję i badać stan kanału w tej samej chwili [33].

W niektórych przypadkach, szczególnie gdy czasy przełączania między nadawaniem a odbiorem są większe niż czas propagacji sygnału w kanale, zagłuszanie kanału przez nadajnik może nie wystarczyć, by poinformować wszystkie stacje o wystąpieniu kolizji [33]. Wówczas zagłuszanie powinno być także wykonane przez stacje niezaangażowane bezpośrednio w transmisję.

Wykrywanie kolizji w sposób znany z sieci Ethernet jest możliwe w niektórych sieciach optycznych, opartych na promieniowaniu rozproszonym [50] (por. rozdz. 1.2.4). Warunkiem koniecznym jest brak bezpośredniej drogi sygnału z nadajnika do odbiornika w ramach stacji. Do odbiornika docierają wówczas jedynie sygnały odbite, dzięki czemu moc sygnałów pochodzących z różnych stacji jest porównywalna i efekt przechwytywania nie występuje.

1.3.6. Możliwości łączenia metod unikania kolizji

Opisane powyżej metody unikania kolizji, a w szczególności wykrywanie tonu zajętości i wymiana ramek sterujących, można stosować jako jedyne w projektowanym protokole dostępu do łącza. Metody te projektowano z myślą o zastąpieniu wykrywania nośnej, nieefektywnego w niektórych sieciach bezprzewodowych. Możliwe jest jednak połączenie obu metod z wykrywaniem nośnej.

W przypadku wymiany ramek sterujących wprowadzenie dodatkowego wykrywania nośnej pozwala na podniesienie wydajności łącza przez zmniejszenie liczby kolizji między ramkami sterującymi. Skutkuje to zwiększeniem prawdopodobieństwa udanej rezerwacji łącza. Jeżeli ramki sterujące nie są chronione wykrywaniem nośnej, informacja sterująca przekazywana tymi ramkami jest dostępna dopiero po odebraniu całej ramki. Natomiast w przypadku ochrony metodą wykrywania nośnej pewna forma informacji sterującej jest przekazywana w postaci nośnej. Dokładną analizę działania protokołów z wymianą ramek sterujących, opcjonalnie wspomaganą wykrywaniem nośnej, można znaleźć w [34].

W przypadku wykrywania tonu zajętości korzyści wypływające z wykrywania nośnej nie są tak oczywiste; co więcej, połączenie takie nie było, jak dotąd, proponowane w literaturze. Wydaje się jednak, że rozwiązanie takie może być skuteczniejsze od samego wykrywania tonu zajętości – zależy to jednak od sposobu jego wytwarzania.

Jeśli ton zajętości jest wytwarzany przez wszystkie stacje wykrywające transmisję w kanale danych, wykrywanie nośnej mogłoby przynieść korzyści niektórym stacjom.

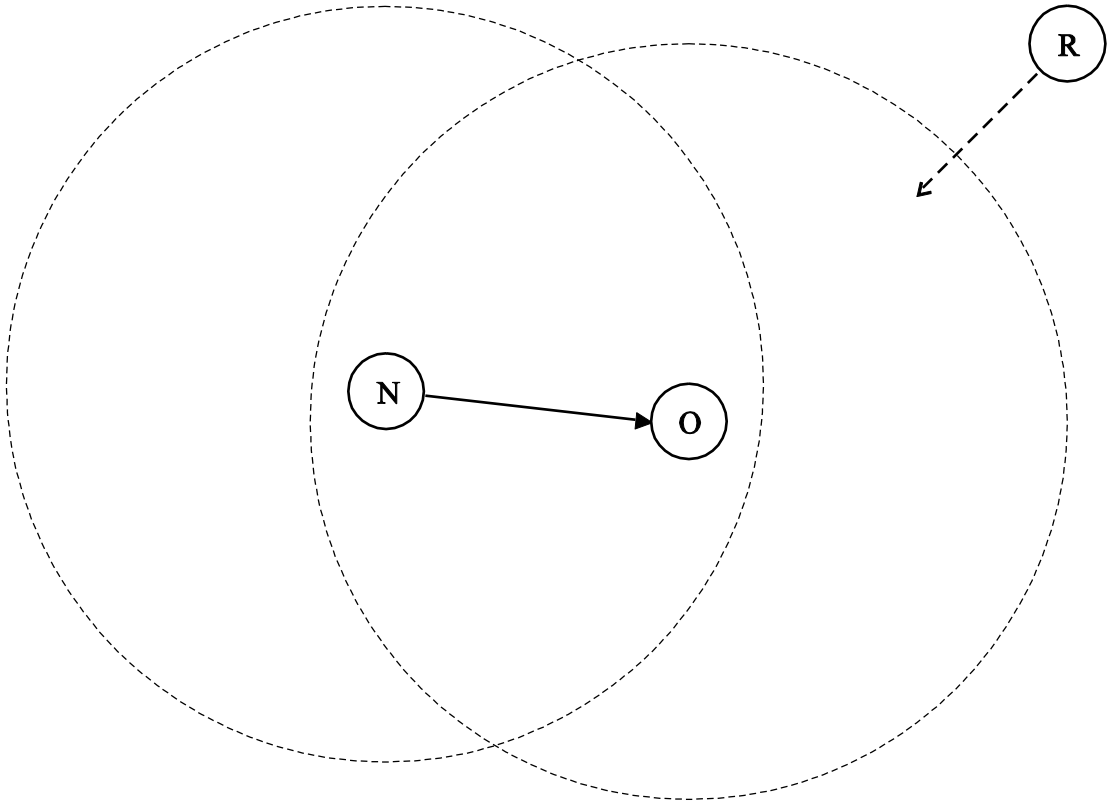
1.3.7. Porównanie wybranych metod unikania kolizji

Przedstawione powyżej metody unikania kolizji, jakkolwiek wydają się całkowicie odmienne, są w istocie zbliżone do siebie. To „ukryte" podobieństwo dotyczy szczególnie wykrywania tonu zajętości i wymiany ramek sterujących. W obu tych metodach część pasma transmisyjnego jest bowiem poświęcona na przesłanie dodatkowej informacji sterującej, określającej stan łącza. Różnica tkwi jednak w sposobie przesyłania tej informacji. Metoda polegająca na wykrywaniu tonu zajętości informuje o stanie łącza w sposób ciągły przez cały czas transmisji danych, podczas gdy wymiana ramek sterujących, na podstawie których można określić stan łącza, jedynie poprzedza przesył informacji (rys. 1.14). Jakkolwiek nie stanowi to zagrożenia w przypadku sieci stacjonarnych lub zawierających niewielką liczbę stacji o ograniczonej mobilności, w przypadku złożonej sieci ad hoc z dużą liczbą ruchliwych stacji może spowodować spadek wydajności sieci [119, 150].

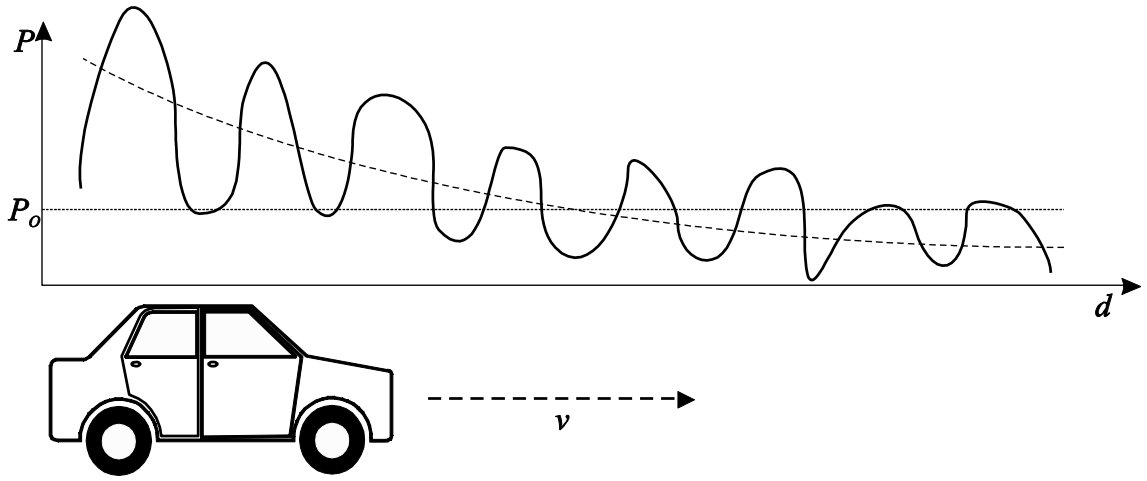
1.3.7.1. Opis problemu

Rozważmy bezprzewodową sieć ad hoc pokazaną na rys. 1.17. Stacja N (nadawca) prowadzi transmisję danych do stacji O (odbiorca). W kierunku tych stacji, od strony odbiorcy, zbliża się stacja ruchoma R. Jest ona zatem poza zasięgiem nadawcy. Jeżeli odbiorca chroni transmisję danych tonem zajętości, to stacja R może uzyskać prawidłową informację o stanie łącza niezależnie od chwili, w której znajdzie się ona w zasięgu odbiorcy. Gdy natomiast nadawca i odbiorca stosują wymianę ramek sterujących, ochrona nie jest już tak skuteczna. Mianowicie, stacja R powinna znajdować się w zasięgu odbiorcy już w chwili, gdy rozpoczyna on transmisję ramki CTS. Jakiekolwiek opóźnienie spowoduje brak (całkowitego) odbioru tej ramki, co pociąga za sobą utratę skuteczności tej metody w opisywanym przypadku.

Dla rzeczywistej sieci ad hoc opisany przypadek jest zbyt uproszczony. W sieciach bezprzewodowych, szczególnie radiowych, występują bowiem zaniki Rayleigha lub Rice'a [110]. Powodują one, że spadek mocy sygnału wraz z odległością jest nie monotoniczny, lecz w dużym stopniu nieregularny. Przykładową zależność mocy sygnału od odległości pokazano na rys. 1.18.

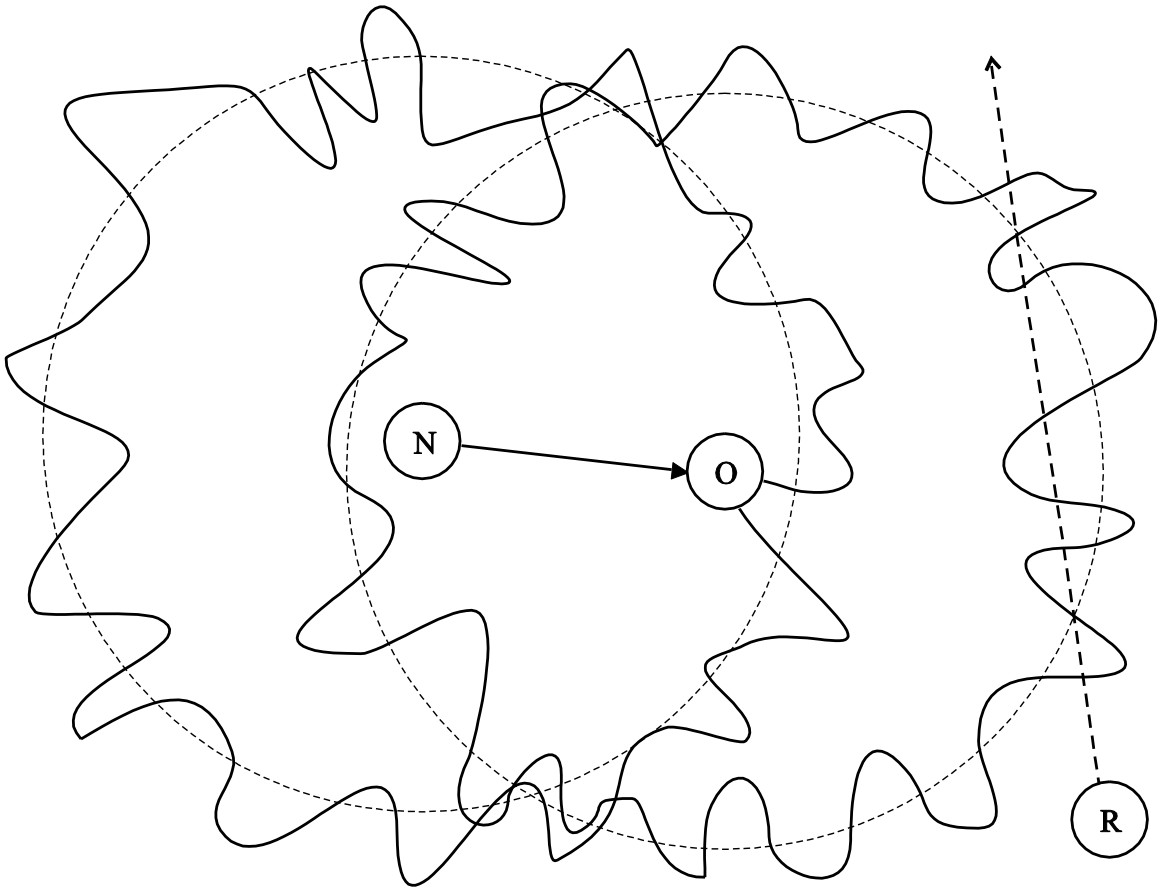


Rys. 1.17. Przykładowa sieć ad hoc z wyidealizowanym zasięgiem stacji
Fig. 1.17. An example of ad-hoc network with idealized station range



Rys. 1.18. Zależność między prędkością stacji ruchomej (v), odległością (d) i mocą sygnału (P)
Fig. 1.18. A relation between mobile station's velocity (v), distance (d) and signal power (P)

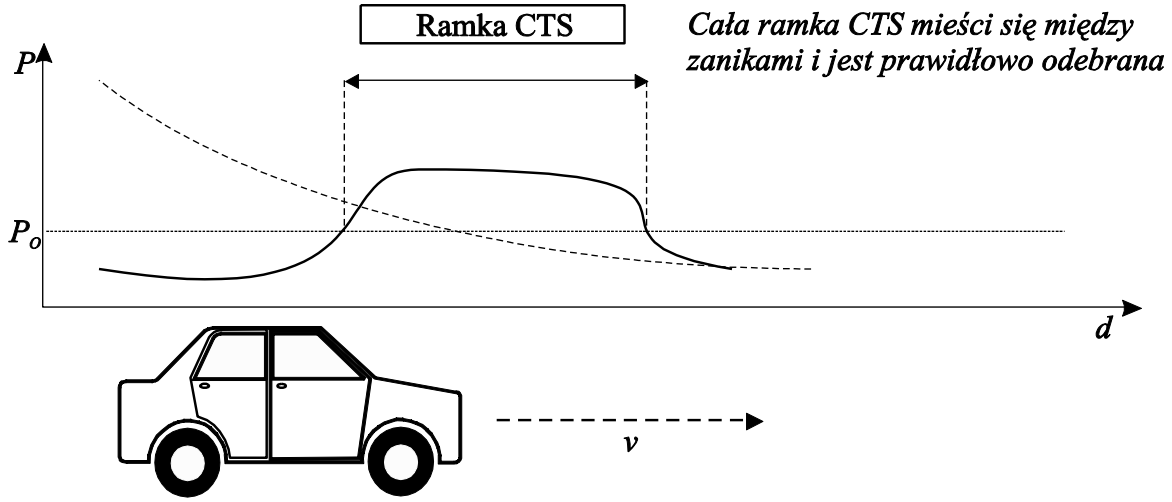
Lokalne minima mocy sygnału odległe są o około połowę długości fali ($\lambda/2$) [14, 50, 110]. Jeśli takie minimum wypada poniżej czułości odbiornika, występuje zanik. W takiej sytuacji zasięgi stacji powinny być pokazane nie jako okręgi, lecz w sposób bardziej nieregularny, np. jak na rys. 1.19. Dodatkowo, im szybciej stacja się porusza, tym bardziej odczuwa negatywne skutki zaników. Efekt ten występuje nie tylko w sieciach radiowych przesyłających informację zakodowaną cyfrowo, lecz także w tradycyjnej radiofonii i radiokomunikacji. Przykładowo, słuchając radia w poruszającym się samochodzie, w pewnych warunkach można usłyszeć, że jakość sygnału radiowego jest niestabilna i zmienia się nawet kilka razy na sekundę.



Rys. 1.19. Przykładowa sieć ad hoc z bardziej realistycznym zasięgiem stacji
Fig. 1.19. An example of ad-hoc network with more realistic station range

Podobnie jak poprzednio, stacja N przesyła dane do stacji O. W pobliżu granicy zasięgu drugiej z wymienionych porusza się stacja R. Ze względu na nieregularny przebieg tej granicy stacja ta na przemian pojawia się i znika z zasięgu odbiorcy. Wraz ze zmianą położenia zmienia się także słyszalność tonu zajętości, można więc założyć, że metoda ta zapewnia wystarczającą skuteczność unikania kolizji. Inaczej jest jednak w przypadku wymiany ramek sterujących. Aby była ona skuteczna, stacja ruchoma musi mieć możliwość prawidłowego odbioru ramki sterującej, szczególnie CTS. W tym celu cała taka ramka musi „zmieścić się”

między dwoma sąsiednimi zanikami. Oznacza to, że czas transmisji tej ramki musi być krótszy niż czas potrzebny stacji ruchomej na przebycie odległości między dwoma najbliższymi zanikami. Wyjaśnienie podanych warunków prawidłowego odbioru ramki sterującej CTS przez stację ruchomą pokazano na rys. 1.20.



Rys. 1.20. Wyjaśnienie warunków prawidłowego odbioru ramki przez stację ruchomą
Fig. 1.20. Explanation of conditions of proper frame reception by a mobile station

1.3.7.2. Kryterium skuteczności metod opartych na wymianie ramek sterujących

Przyjmijmy następujące założenia [119, 150]:

- zasięgi kanału tonu zajętości i kanału danych są identyczne,
- czas wykrycia nośnej i tonu zajętości jest pomijalny,
- długość fali radiowej wynosi λ [m], a częstotliwość – f [Hz],
- stacja ruchoma R porusza się z prędkością v [m/s].

Czas, jaki stacja R potrzebuje na przebycie odległości $\lambda/2$, wynosi:

$$t_{fad} = \frac{\lambda}{2v}. \tag{1.2}$$

Aby ramka CTS została prawidłowo odebrana, czas jej transmisji (wraz z elementami warstwy fizycznej, jak np. preambuła czy czas przełączania odbiór-nadawanie) nie może być większy niż czas pomiędzy dwoma kolejnymi zanikami, tj. średnio[2]

$$t_{CTS} \le \frac{t_{fad}}{2}. \tag{1.3}$$

W praktyce może okazać się, że czas ten powinien być jeszcze krótszy. Wpływ na to mogą mieć warunki pracy sieci lub szczegóły implementacji protokołu.

[2]Jeśli odległość między zanikami wynosi $\lambda/2$, to, jak widać na rys. 1.18, poziom sygnału odbieranego przekracza czułość odbiornika średnio w dwukrotnie mniejszych odległościach.

Biorąc powyższe pod uwagę, aby zachować skuteczność unikania kolizji metodą wymiany ramek sterujących, prędkość poruszania się stacji R jest ograniczona i nie może przekraczać

$$v \le \frac{\lambda}{4 \cdot t_{CTS}} = \frac{c}{4f \cdot t_{CTS}}, \tag{1.4}$$

gdzie c – prędkość światła [m/s]. Przykładowo, w standardzie IEEE 802.11 [52], przy prędkości transmisji 1 Mb/s i rozpraszaniu widma metodą kluczowania bezpośredniego, czas transmisji ramki CTS wynosi około 320 μs, z czego 192 μs stanowi preambuła i nagłówek warstwy fizycznej. Przy częstotliwości 2,4 GHz otrzymujemy

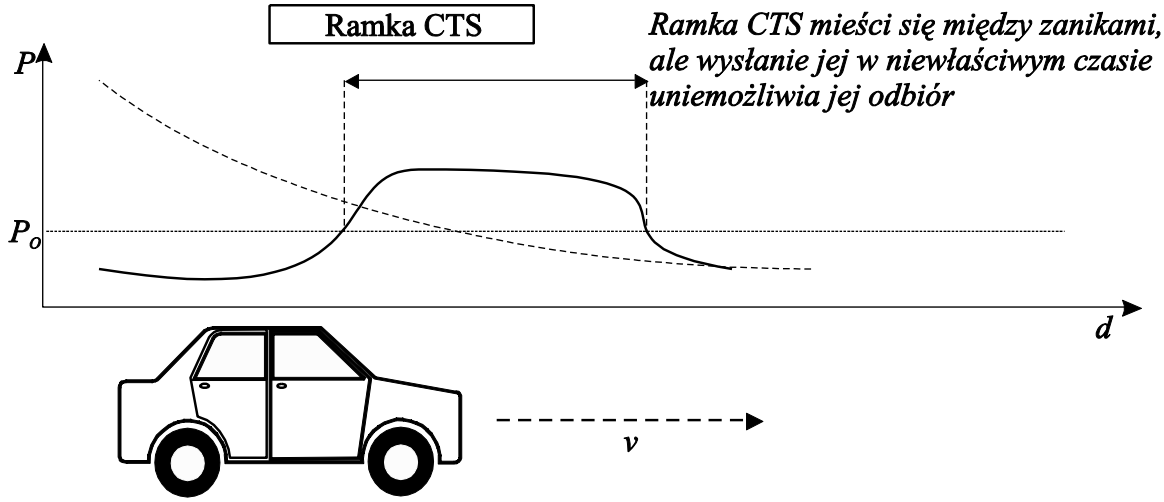
$$v \le \frac{c}{4f \cdot t_{CTS}} = \frac{3 \cdot 10^8 [\text{m/s}]}{4 \cdot 2{,}4 \cdot 10^9 [\text{Hz}] \cdot 320 \cdot 10^{-6} [\text{s}]} = 97{,}7 [\text{m/s}]. \tag{1.5}$$

Warto zauważyć, że inne wersje standardu (802.11b, 802.11g) charakteryzują się krótszymi preambułami oraz większymi prędkościami transmisji, co pozwala na jeszcze większą mobilność stacji. Wynika z tego, że w wielu zastosowaniach ograniczenie prędkości transmisji nie jest istotne. Należy jednak mieć na uwadze, że obsługa krótkiej preambuły jest opcjonalna, zaś ramki sterujące mogą nie być przesyłane z maksymalną prędkością (w standardzie 802.11 dla przesyłu ramek sterujących używa się zwykle niższych prędkości niż dla ramek danych). W takich warunkach mobilność stacji będzie ograniczona.

Przedstawione kryterium ma jedynie charakter orientacyjny. Na rys. 1.18 widać bowiem, że wraz ze wzrostem poziomu sygnału wydłuża się odległość między zanikami, co umożliwia poprawną pracę sieci przy większych prędkościach poruszania się stacji. Analogicznie, przy słabszym sygnale dopuszczalna prędkość jest znacznie mniejsza. Pomimo to spełnienie podanego kryterium można uważać za warunek konieczny skutecznego działania unikania kolizji metodą wymiany ramek sterujących w obecności ruchomych stacji ukrytych.

Spełnienie podanego powyżej kryterium nie gwarantuje, iż ochrona ramką CTS będzie zawsze skuteczna. Może się bowiem zdarzyć – nawet gdy stacja ruchoma porusza się z wystarczająco niską prędkością – że ramka CTS nie zostanie całkowicie wysłana między zanikami, a więc nie będzie prawidłowo odebrana. Przypadek ten pokazano na rys. 1.21. Aby ramka CTS była zawsze wysłana skutecznie – tj. między zanikami – przekraczanie granic zasięgu przez stację ruchomą powinno być zsynchronizowane z transmisjami ramek CTS. Niestety, osiągnięcie takiego stanu nie wydaje się technicznie możliwe, a gdyby nawet, wymagałoby bardzo złożonego sterowania, którego koszt mógłby być wyższy niż korzyści wynikające ze zwiększonej skuteczności unikania kolizji. O ile zatem można oszacować możliwość spełnienia przez daną sieć warunku koniecznego, o tyle nie można mieć gwarancji spełnienia warunku wystarczającego skutecznego działania unikania kolizji metodą wymiany ramek sterujących w obecności ruchomych stacji ukrytych.

Warto zauważyć, że – w zakresie przyjętych założeń – powyższe ograniczenie nie dotyczy metody z wykrywaniem tonu zajętości. W rzeczywistości, ciągła ochrona ramki danych jest całkowicie odporna na poruszanie się stacji, o ile tylko ton zajętości będzie słyszalny nie później niż w chwili wejścia stacji ruchomej w zasięg odbiorcy.



Rys. 1.21. Wyjaśnienie warunków nieprawidłowego odbioru ramki przez stację ruchomą
Fig. 1.21. Explanation of conditions of improper frame reception by a mobile station

1.4. Rywalizacyjne protokoły dostępu do łącza

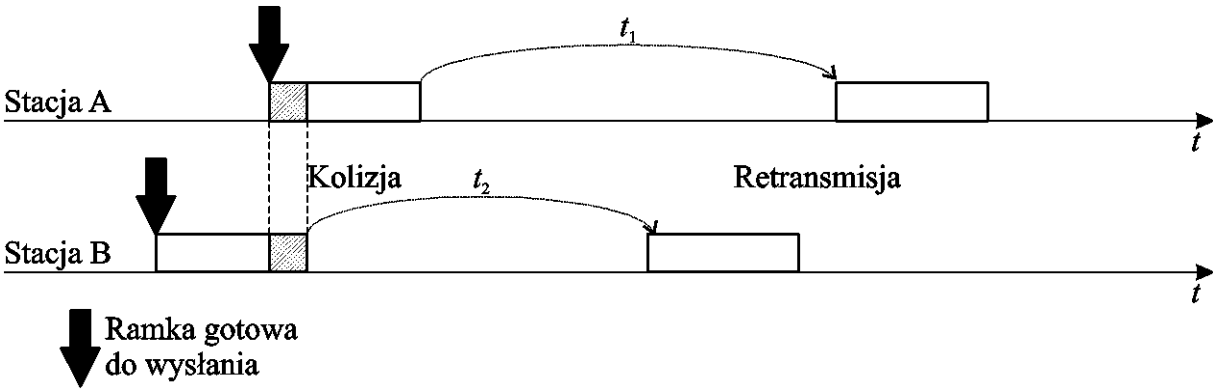
Rywalizacyjne protokoły dostępu do łącza są nierozerwalnie związane z historią sieci bezprzewodowych. Wiąże się to z użyciem takiego właśnie protokołu w sieci Aloha. Była to bodaj pierwsza w historii sieć bezprzewodowa i pierwszy w historii protokół rywalizacyjny. Obecnie protokoły tej klasy są daleko bardziej złożone, ale mimo to są znacznie prostsze w realizacji od protokołów innych klas, np. rezerwacyjnych. Rywalizacja jest też często stosowana jako mechanizm pomocniczy protokołów tej grupy.

1.4.1. Protokoły rodziny Aloha

Najstarszym i najszerzej znanym protokołem dostępu do łącza dla sieci bezprzewodowych jest protokół stosowany w sieci Aloha [2, 3]. Jest on uznawany za pierwszy stworzony protokół rywalizacyjny i może być rozpatrywany jako poprzednik mechanizmów CSMA i CSMA/CD, szeroko stosowanych w sieciach przewodowych [84, 104].

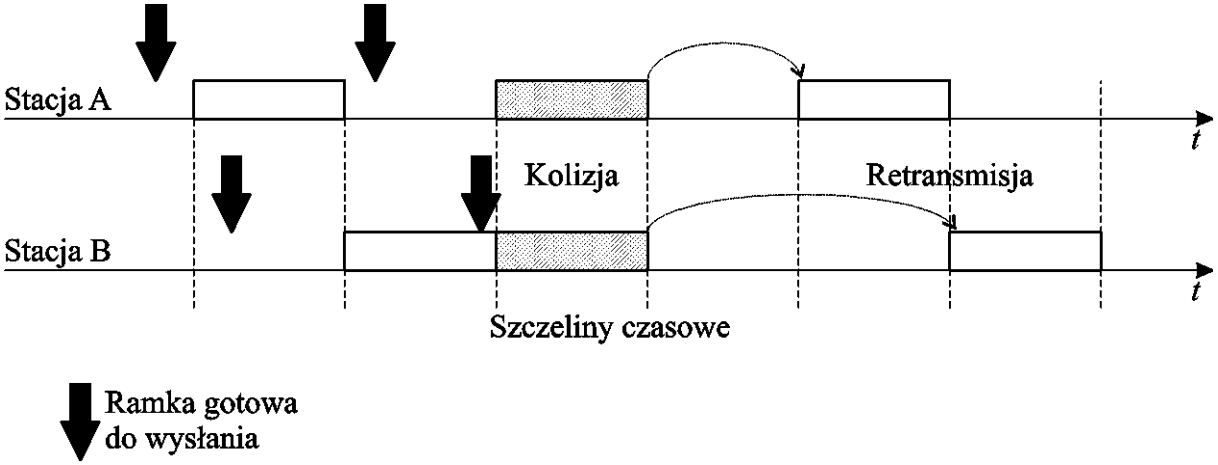
W protokole Aloha transmisja może rozpocząć się w dowolnym momencie, niezależnie od stanu łącza, jeśli jakakolwiek stacja ma ramkę przygotowaną do wysłania. Mogą zatem występować liczne kolizje ramek wysłanych przez różne stacje. Jako że odbiorca informacji potwierdza każdą prawidłowo odebraną ramkę, brak takiego potwierdzenia może oznaczać kolizję ramek (lub inny błąd transmisji). Dla uniknięcia kolizji potwierdzenia zazwyczaj przesyła się osobnym kanałem częstotliwościowym, nie jest to jednak warunek konieczny

działania protokołu Aloha. Niepotwierdzona ramka jest następnie wysyłana ponownie, po pseudolosowo wyznaczonym czasie t_i, stosownie do tych samych reguł, a więc także z ryzykiem wystąpienia kolejnych kolizji. Ideę protokołu przedstawia rys. 1.22.



Rys. 1.22. Zasada działania protokołu Aloha
Fig. 1.22. Aloha protocol operation rules

Innym wariantem protokołu jest tzw. szczelinowy protokół Aloha (ang. *slotted Aloha*) [65]. W tym protokole czas jest podzielony na szczeliny o czasie trwania równym czasowi transmisji ramki. Każda stacja, mająca ramkę gotową do wysłania, musi wstrzymać rozpoczęcie transmisji do najbliższego początku szczeliny. Jeśli nadawanie rozpocznie więcej niż jedna stacja, wystąpi oczywiście kolizja. Jeśli jednak szczelina będzie zajęta przez tylko jedną stację, ramka nie zostanie zniszczona (zakładając, że jedynym powodem zniszczenia ramki mogą być kolizje). Jeśli zatem nie wystąpi kolizja już na samym początku szczeliny, ramka powinna dotrzeć do adresata bez przeszkód. Mechanizm szczelin pozwala zatem uniknąć zniszczenia bezkolizyjnie przesyłanej ramki tuż przed zakończeniem jej transmisji. Ideę protokołu wyjaśnia rys. 1.23.



Rys. 1.23. Zasada działania szczelinowej odmiany protokołu Aloha
Fig. 1.23. Slotted Aloha protocol operating rules

Pomimo że protokoły Aloha były stworzone z myślą o sieciach scentralizowanych (tj. zawierających stację bazową odpowiedzialną za koordynację pracy sieci), można je stosować

także w sieciach ad hoc. Ich zaletą jest prostota, a co za tym idzie – łatwość implementacji. Wydajność tych protokołów – w przeciwieństwie do opisanych dalej protokołów rodziny CSMA – nie zależy także od opóźnienia propagacyjnego w sieci [110]. Zalety te są jednak okupione niskim stopniem wykorzystania kanału, nieprzekraczającym 18,5% dla protokołu Aloha i 37% dla wersji szczelinowej. Ponadto, obie wersje protokołu tracą stabilność już przy niewielkich obciążeniach.

Warto zauważyć, że szczelinowy protokół Aloha jest często używany w protokołach rezerwacyjnych dla sieci bezprzewodowych, np. w protokołach dla bezprzewodowych sieci ATM [124, 125, 142, 143] lub w sieciach cyfrowej telefonii komórkowej, np. GSM [48, 49, 110]. Przy użyciu tego protokołu stacje ruchome przesyłają do stacji bazowej żądania rezerwacji. Wydaje się, iż wykorzystanie prostego protokołu jest w tym przypadku uzasadnione – żądania rezerwacji są stosunkowo krótkimi ramkami, a więc w razie kolizji utrata przepustowości łącza jest niewielka. Ponadto, w sieci zawierającej wiele stacji ruchomych, jak np. komórka sieci GSM, zjawiska stacji ukrytej i odkrytej i tak mogłyby uniemożliwić skuteczne stosowanie bardziej złożonych protokołów, jak np. CSMA.

1.4.2. Protokoły CSMA

Protokoły rodziny CSMA (ang. *Carrier Sense Multiple Access*) [81, 103, 104] są stosowane m. in. w amatorskiej sieci Packet Radio [6] oraz w bezprzewodowych sieciach lokalnych pracujących zgodnie ze standardem IEEE 802.11 [52].

Po przygotowaniu ramki do wysłania stacja musi sprawdzić stan łącza. Jeśli jest ono wolne, transmisja może rozpocząć się natychmiast; w przeciwnym przypadku należy wstrzymać nadawanie i ponowić ocenę stanu łącza po pewnym czasie. Podobnie jak w protokole Aloha, odbiorca musi przesłać potwierdzenie prawidłowego odbioru każdej ramki. Mechanizm ten umożliwia retransmisję ramek utraconych w wyniku kolizji lub innych błędów transmisji.

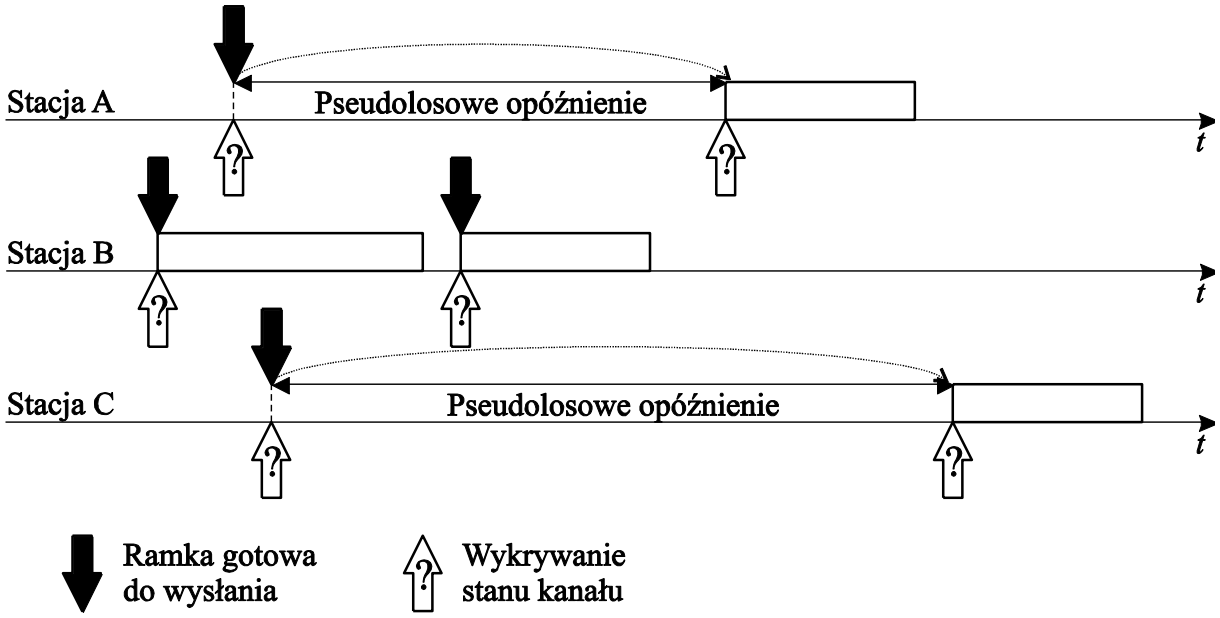
Kolizja może wystąpić, gdy więcej niż jedna stacja rozpocznie transmisję w tym samym czasie. Ze względu na opóźnienie propagacyjne, występujące szczególnie w sieciach rozległych terytorialnie, a także ze względu na niezerowy czas wykrywania nośnej, od chwili rozpoczęcia nadawania przez jedną stację mija pewien czas, gdy mechanizm unikania kolizji nie jest jeszcze skuteczny, a zatem ramka nie jest chroniona przed kolizją.

Ze względu na sposób wyznaczenia momentu kolejnej próby dostępu po stwierdzeniu zajętości łącza można wyróżnić kilka odmian protokołu CSMA [65].

1.4.2.1. Protokół CSMA w wersji nietrwałej

W nietrwałej (ang. *non-persistent*) odmianie protokołu CSMA stacja, po stwierdzeniu zajętości łącza, pseudolosowo dobiera czas, po upływie którego dokonuje kolejnej próby dostępu. Mechanizm ten dobrze się sprawdza przy wysokich obciążeniach łącza. W przypadku bowiem, gdy wiele stacji próbuje uzyskać dostęp do zajętego łącza, próby transmisji mogą

być rozłożone w czasie, co zmniejsza ryzyko wystąpienia kolizji, a zatem zwiększa efektywną przepustowość protokołu. Z drugiej jednak strony, przy niedużych obciążeniach łącza nie wykorzystuje się pełnej jego przepustowości – po jego zwolnieniu stacje oczekujące nie rozpoczynają transmisji natychmiast, lecz dopiero po upływie wyznaczonego czasu opóźnienia. Może to także prowadzić do niesprawiedliwego podziału łącza, ponieważ stacje, które dokonały pierwszej próby dostępu w korzystnym momencie, mogą przesłać pewną liczbę ramek, podczas gdy pozostałe stacje czekają na zakończenie czasu opóźnienia. Sytuacja taka jest przedstawiona na rys. 1.24. Stacja B uzyskała dostęp do łącza w pierwszej próbie. W czasie prowadzonej przez nią transmisji stacje A i C próbują uzyskać dostęp do łącza. Ponieważ jest ono zajęte, stacje losowo dobierają czas opóźnienia, po którym mogą ponowić próbę dostępu. Przez ten czas nie rozpoczną one transmisji, nawet jeśli łącze będzie wolne. Jeśli jednak wyznaczony przez nie czas opóźnienia jest odpowiednio długi, inna stacja (B na rys. 1.24) może rozpocząć transmisję stwierdziwszy, iż łącze jest wolne. Tak więc stacje mogą prowadzić transmisje w zupełnie innej kolejności, niż przystępowały do prób dostępu do łącza.



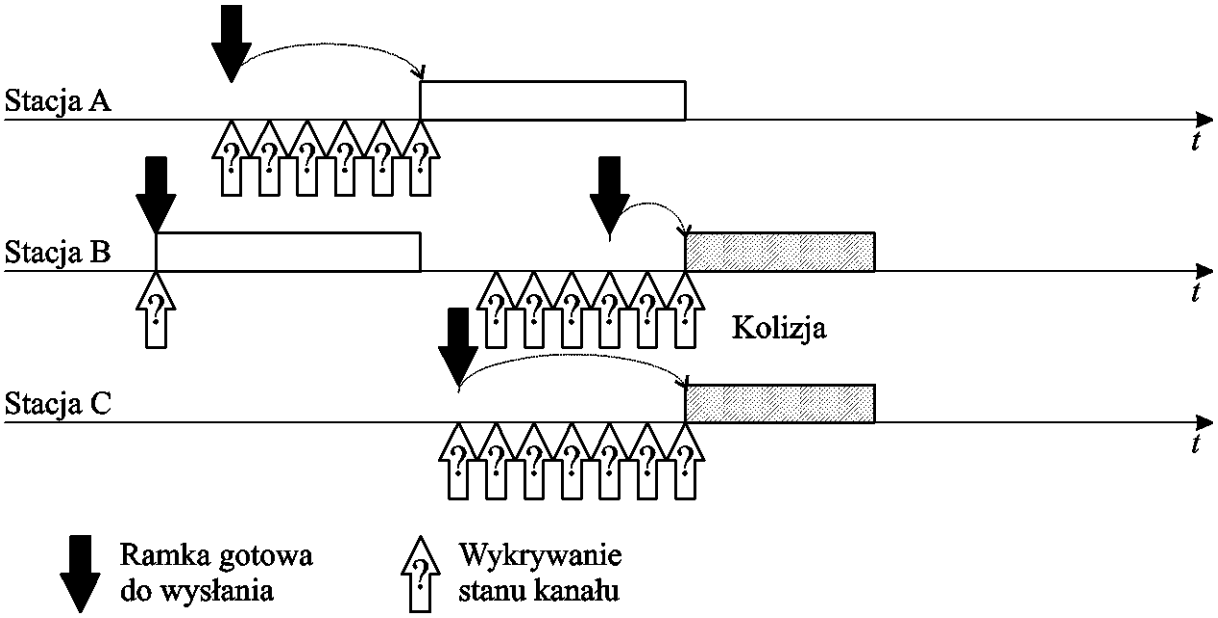
Rys. 1.24. Zasada działania nietrwałej odmiany protokołu CSMA
Fig. 1.24. Non-persistent CSMA protocol operating rules

1.4.2.2. Protokół CSMA w wersji trwałej

W trwałej (ang. *persistent*) odmianie protokołu CSMA stacja, po stwierdzeniu zajętości łącza, kontynuuje badanie jego stanu aż do uzyskania dostępu. Mechanizm ten dobrze sprawdza się przy niskich obciążeniach łącza, pozwala bowiem zminimalizować czas jego bezczynności – stacja może rozpocząć transmisję, gdy tylko zakończy się poprzednia transmisja. Jeżeli jednak w czasie zajętości łącza więcej niż jedna stacja dokona próby dostępu, po zwolnieniu łącza nastąpi kolizja. W porównaniu z nietrwałą odmianą protokołu można zatem spodziewać się wyższej przepustowości i mniejszych opóźnień przy małych obciążeniach łącza,

przy większych natomiast może nastąpić utrata stabilności wskutek licznych kolizji. Zasadę działania protokołu ilustruje rys. 1.25. Stacja B uzyskuje dostęp w pierwszej próbie. Podczas tej transmisji stacja A rozpoczyna próbę dostępu i kontynuuje ją aż do chwili zwolnienia łącza przez stację B. Ponieważ tylko stacja A miała ramkę przygotowaną do wysłania, transmisja przebiega bezkolizyjnie. Podczas tego przesyłu stacje B i C rozpoczynają próby dostępu, co – zgodnie z opisanymi powyżej zasadami – powoduje kolizję tuż po zakończeniu transmisji przez stację A. Stacje te będą zatem ponawiać próbę transmisji w późniejszym, losowo wyznaczonym, czasie.

Trwała odmiana protokołu CSMA wydaje się bardziej sprawiedliwa, ponieważ losowanie opóźnienia ma miejsce dopiero po wystąpieniu kolizji. Ponadto – o ile nie wystąpią kolizje – transmisja jest prowadzona w takiej samej kolejności, w jakiej stacje rozpoczynały próby uzyskania dostępu do łącza.

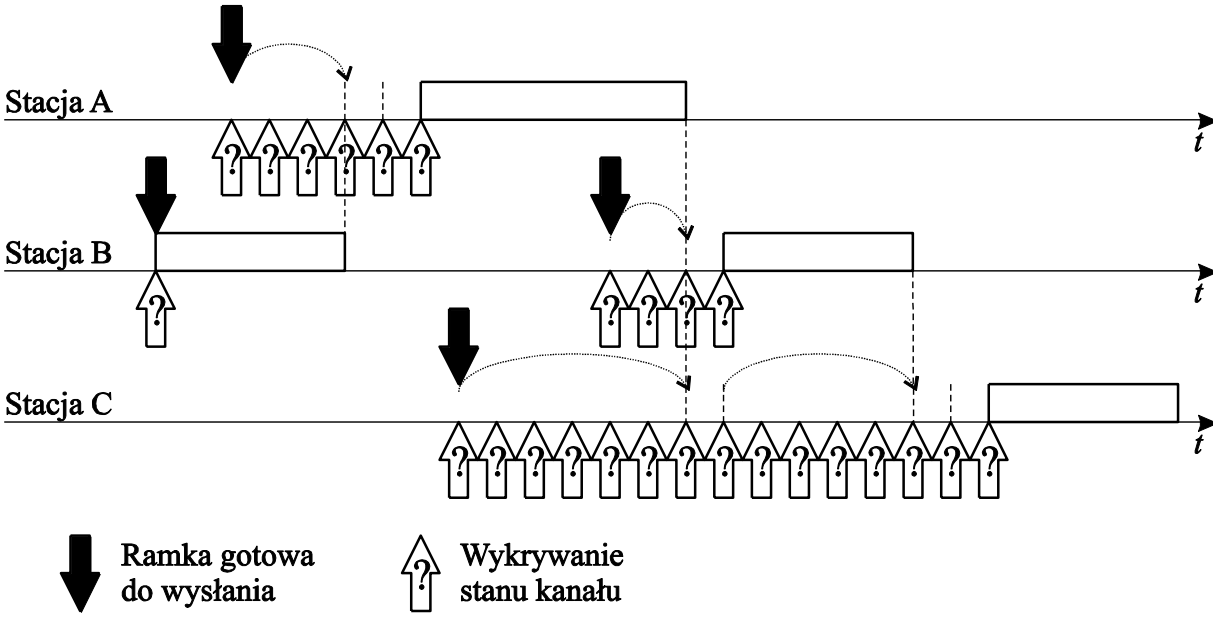


Rys. 1.25. Zasada działania trwałej odmiany protokołu CSMA
Fig. 1.25. Persistent CSMA protocol operating rules

1.4.2.3. Protokół CSMA w wersji p-trwałej

W *p*-trwałej (ang. *p-persistent*) odmianie protokołu CSMA obowiązują podobne zasady jak w wersji trwałej. Oba warianty zachowują się tak samo w przypadku stwierdzenia zajętości łącza. O ile jednak w odmianie trwałej transmisja rozpoczyna się natychmiast po zwolnieniu łącza, wersja *p*-trwała wykorzystuje nieco bardziej złożony mechanizm. Mianowicie, czas jest dzielony na szczeliny o długości równej podwojonemu maksymalnemu opóźnieniu propagacyjnemu w kanele transmisyjnym. Stacje, które mają ramki przygotowane do wysłania, dokonują prób transmisji z prawdopodobieństwem p w kolejnych szczelinach, pod warunkiem że żadna inna stacja nie rozpocznie transmisji wcześniej. Można zatem powiedzieć, że moment rozpoczęcia nadawania przez poszczególne stacje, próbujące uzyskać dostęp do łą-

cza, jest rozłożony w czasie. Rozwiązanie to ma tę zaletę, że jeśli w czasie zajętości łącza więcej niż jedna stacja dokona próby dostępu, po zwolnieniu łącza – w przeciwieństwie do odmiany trwałej – nie musi nastąpić kolizja. Jeżeli bowiem pierwsza ze szczelin, w której rozpocznie się transmisja, zostanie wylosowana przez tylko jedną stację, to ramka powinna zostać przesłana bezkolizyjnie. Jest to możliwe, ponieważ kolejne próby dostępu, dokonane przez inne stacje, mogą mieć miejsce dopiero w następnej szczelinie, a wtedy nośna pochodząca z nadającej stacji powinna być już odbierana w całej sieci. Zatem, stacje próbujące rozpocząć transmisję w następnej szczelinie stwierdzają, iż łącze jest już zajęte, i ponownie przechodzą w stan nasłuchu łącza. Zasada działania protokołu jest wyjaśniona na rys. 1.26. Stacja B rozpoczyna próbę dostępu do łącza w chwili, gdy jest ono wolne – może zatem zacząć transmisję natychmiast. W czasie tego przesyłu stacja A próbuje uzyskać dostęp do łącza, prowadzi zatem nasłuch do chwili, gdy zostanie ono zwolnione. Wówczas dokonuje losowego wyboru szczeliny, w której może rozpocząć nadawanie, o ile łącze nie zostanie zajęte; ponieważ jednak żadna inna stacja nie ubiega się o dostęp do łącza, transmisja dochodzi do skutku. W czasie jej trwania stacje B i C rozpoczynają próbę dostępu do łącza. Po zakończeniu transmisji przez stację A stacje te losują szczeliny, w których mogą rozpocząć transmisję. Rywalizację tę wygrywa stacja B, ponieważ wybrała wcześniejszą szczelinę; stacja C musi natomiast ponowić próbę dostępu po zakończeniu transmisji przez stację B.



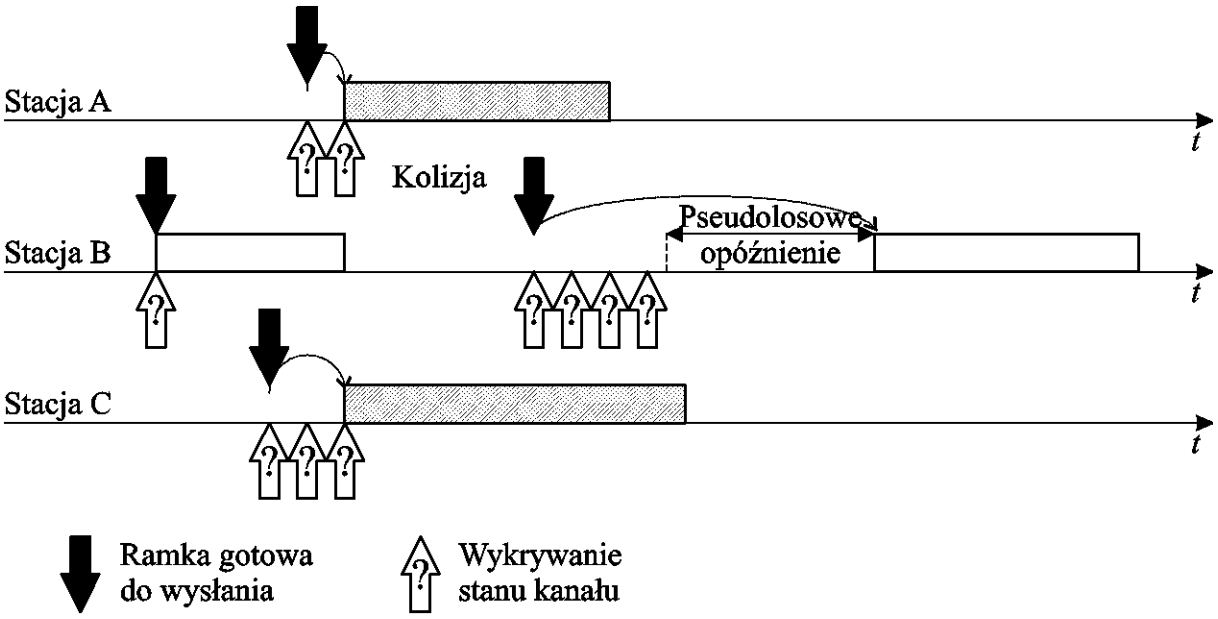
Rys. 1.26. Zasada działania *p*-trwałej odmiany protokołu CSMA
Fig. 1.26. *P*-persistent CSMA protocol operating rules

Wydaje się, że *p*-trwała odmiana protokołu CSMA jest rozsądnym kompromisem pomiędzy wersją trwałą i nietrwałą. Przy niskim obciążeniu łącza umożliwia bowiem uzyskanie dostępu niemal tak szybko jak wersja trwała; opóźnienie zwiększone jest jedynie o czas trwania pewnej liczby szczelin czasowych. Z kolei przy wysokim obciążeniu znacznie zmniejsza

się prawdopodobieństwo kolizji po okresie zajętości łącza, a to dzięki temu, że stacje próbujące uzyskać dostęp do łącza mogą wybierać różne szczeliny. Średnia liczba szczelin występujących w czasie rywalizacji jest odwrotnością prawdopodobieństwa transmisji (parametru *p*). Z tej pobieżnej analizy wynika, że prawdopodobieństwo to powinno być różne w zależności od obciążenia łącza – większe przy małym obciążeniu, mniejsze przy dużym.

1.4.2.4. Protokół CSMA o ograniczonej trwałości

Protokół CSMA o ograniczonej trwałości (ang. *limited persistence*) [38] można uznać za rozwiązanie pośrednie między odmianą trwałą i nietrwałą. Przy próbie dostępu do łącza stacja postępuje zgodnie z zasadami wariantu trwałego, jednak tylko przez pewien czas. Jeśli w tym czasie łącze pozostaje zajęte, stacja wyznacza losowe opóźnienie, po którym ponawia próbę dostępu, podobnie jak w wariancie nietrwałym. Przy małych obciążeniach łącza rozwiązanie takie pozwala na redukcję opóźnień przy dostępie do łącza i lepsze wykorzystanie przepustowości kanału w porównaniu z odmianą nietrwałą. Z drugiej strony, przy dużych obciążeniach istnieje możliwość zmniejszenia ryzyka kolizji tuż po okresie zajętości łącza, jak ma to miejsce w protokole trwałym. Zasadę działania protokołu pokazano na rys. 1.27.



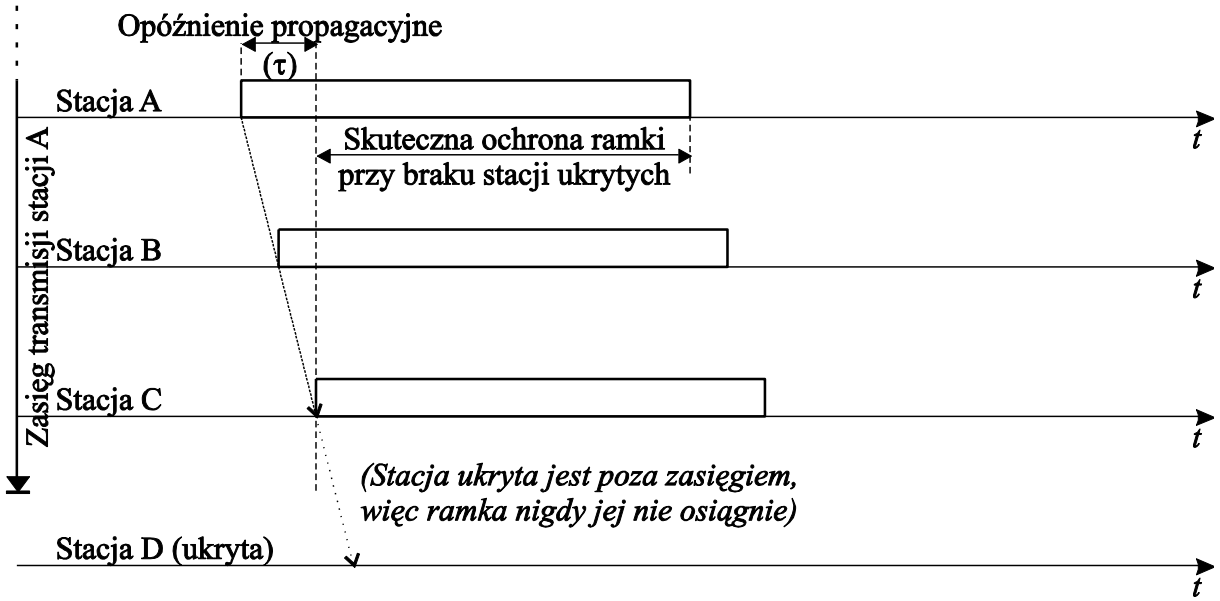
Rys. 1.27. Zasada działania protokołu CSMA o ograniczonej trwałości
Fig. 1.27. Limited persistent CSMA protocol operating rules

Podczas transmisji prowadzonej przez stację B dwie kolejne stacje (A i C) próbują uzyskać dostęp do łącza. Ponieważ od rozpoczęcia tych prób do zwolnienia łącza mija czas krótszy niż ustalony czas trwałości protokołu, po zakończeniu transmisji przez stację B występuje kolizja między ramkami wysłanymi przez stacje A i C. Z kolei stacja B próbuje uzyskać dostęp do łącza; ponieważ jednak czas trwałości protokołu mija przed zwolnieniem łącza przez stację C, stacja B wyznacza pseudolosowe opóźnienie, po którym ponawia próbę dostępu, podobnie jak w protokole nietrwałym. Ponieważ jednak tylko stacja B próbowała uzyskać

dostęp do łącza, przez czas równy wyznaczonemu opóźnieniu łącze może albo pozostać niewykorzystane, albo zajęte przez inną stację, która rozpoczęła próbę dostępu do łącza w korzystniejszej chwili. Z tej pobieżnej analizy wynika, że protokół o ograniczonej trwałości może przynieść korzyści tylko w ściśle określonych sytuacjach. Prawdopodobnie dlatego jest on stosowany jedynie jako protokół pomocniczy przy przesyłaniu ramek sterujących. Czas trwania tych ramek jest wówczas zbliżony do czasu trwałości protokołu [38].

1.4.2.5. Zachowanie protokołu CSMA w sieciach bezprzewodowych

Protokoły rodziny CSMA są znacznie bardziej wydajne od obu odmian protokołu Aloha, utrzymują także wysoką wydajność przy wysokich obciążeniach sieci. Jest to jednak prawda tylko w sieciach przewodowych i niektórych sieciach bezprzewodowych, w których wszystkie stacje mają możliwość bezpośredniej komunikacji i mogą rozpoznać każdą transmisję korzystając z mechanizmu wykrywania nośnej. Niestety, w wielu sieciach bezprzewodowych – ze względu na zależną od lokalizacji stacji zdolność wykrywania nośnej (ang. *location dependent carrier sensing*) – mechanizm ten chroni ramkę w zasięgu nadawcy, ale nie odbiorcy. Występują zatem stacje ukryte i odkryte, które nie mogą prawidłowo ocenić stanu łącza. W tym przypadku stacja nie będzie mogła prawidłowo ocenić zajętości łącza nawet wówczas, gdy upłynie już czas opóźnienia propagacyjnego. Wykrywanie nośnej nie jest zatem w pełni skuteczne w obecności stacji ukrytych, a wydajność protokołu spada wówczas do poziomu porównywalnego z protokołem Aloha. Zjawisko to ilustruje rys. 1.28.



Rys. 1.28. Wpływ opóźnienia propagacyjnego i stacji ukrytej na możliwość wykrycia nośnej
Fig. 1.28. Influence of propagation delay and hidden station upon carrier detection possibility

Ogólnie, protokoły rodziny CSMA nie są trudne do zaimplementowania, wymagają bowiem pewnych układów sprzętowych i małego narzutu programowego. Ich wydajność zależy od relacji między opóźnieniem propagacyjnym a czasem transmisji ramki. Zależność ta, zilu-

strowana na rys. 1.28, pokazuje, jak duży fragment ramki nie jest chroniony przed kolizją. Najwyższą wydajność protokół uzyskuje zatem wówczas, gdy sieć zajmuje stosunkowo niewielki obszar, a przesyłane ramki są długie. Pomimo że w sieciach bezprzewodowych CSMA nie zawsze wykazuje dostatecznie wysoką efektywność, wykrywanie nośnej można wspomóc przez wymianę ramek sterujących, aby zmniejszyć wpływ stacji ukrytych i odkrytych.

1.4.3. Protokoły rodziny BTMA

Protokoły, należące do rodziny BTMA (ang. *Busy Tone Multiple Access*) [107] wykorzystujące mechanizm wykrywania tonu zajętości zamiast wykrywania nośnej, można uznać za pierwsze próby rozwiązania problemu stacji ukrytej. Przyjmuje się, że kanał transmisyjny jest podzielony na dwie części:

- kanał komunikatów (danych), zajmujący większość pasma częstotliwości,
- kanał tonu zajętości (sterujący) o względnie małej szerokości pasma.

Każda stacja z ramką gotową do wysłania musi sprawdzić, czy łącze jest wolne. Może to uczynić, prowadząc nasłuch kanału sterującego przez pewien czas. Jeśli łącze jest zajęte, transmisję odkłada się na później. Podobnie jak w protokołach CSMA, kolejna próba dostępu także rozpoczyna się od sprawdzenia zajętości łącza.

Ton zajętości jest zazwyczaj falą sinusoidalną, którą może wytworzyć:

- każda stacja, wykrywająca aktywność w kanale danych (jak w protokole BTMA [107]),
- tylko adresat przesyłanej ramki danych (jak w protokole RI-BTMA [114]),
- początkowo każda stacja wykrywająca transmisję, a po rozpoznaniu adresu docelowego – tylko adresat (jak w protokole WCD [42]).

1.4.3.1. Protokół BTMA

Pierwsza metoda, użyta w protokole BTMA [107], jest najprostsza, a jednocześnie bardzo wydajna w zakresie zmniejszania liczby stacji ukrytych. Właściwość ta jest jednak uzyskana kosztem niepotrzebnie zwiększonej liczby stacji odkrytych. Można zatem powiedzieć, iż obszar zajęty przez określoną transmisję jest dużo większy, niż jest to naprawdę konieczne (por. rozdział 1.3.2).

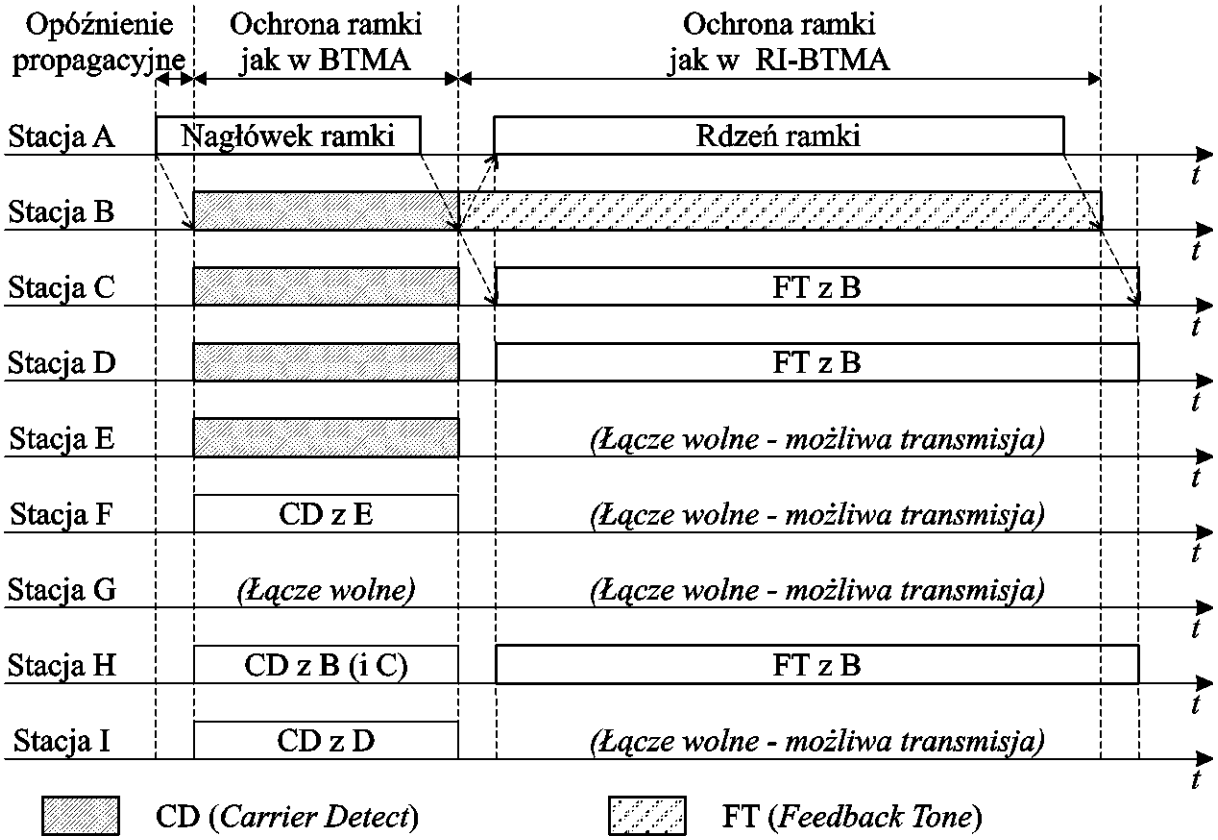
1.4.3.2. Protokół RI-BTMA

Aby uniknąć zwiększania liczby stacji odkrytych, wytwarzanie tonu zajętości można ograniczyć do adresata przesyłanej informacji jak w protokole RI-BTMA (ang. Receiver Initiated BTMA) [114]. W ten sposób ramka jest chroniona przed kolizją w pobliżu odbiornika, a więc w jedynym miejscu, w którym naprawdę powinna być chroniona (nie ma potrzeby ochrony w pobliżu nadajnika lub stacji, do których ramka nie jest adresowana). Ta metoda także jest prosta, jednak nie chroni ramki do chwili rozpoznania jej adresu przeznaczenia.

Tak więc na początku transmisji ramka jest narażona na kolizje, wynikające z braku sygnału ochronnego.

1.4.3.3. Protokół WCD

Protokół WCD (ang. *Wireless Collision Detect*) [42] łączy zalety obu metod wytwarzania tonu zajętości. Używa się tu dwóch tonów zajętości. Wszystkie stacje, znajdujące się w zasięgu nadawcy, wytwarzają ton CD (ang. *Carrier Detected*), gdy tylko wykryją aktywność w kanale danych. Podobnie jak w protokole BTMA ton CD zajmuje stosunkowo duży obszar, pokazany na rys. 1.11. Proces ten trwa do chwili rozpoznania adresu odbiorcy, wówczas odbiorca zmienia ton zajętości na FT (ang. *Feedback Tone*), jak w protokole RI-BTMA, zajmując obszar, pokazany na rys. 1.12. Obecność tonu FT oznacza, iż transmisja początkowej części zakończyła się sukcesem. Pozostałe stacje wyłączają tony zajętości, gdy tylko okaże się, że nie są adresatami ramki. Zasady działania protokołu WCD, w sieci podobnej do pokazanej na rys. 1.12, wyjaśniono na rys. 1.29.



Rys. 1.29. Zasada działania protokołu WCD
Fig. 1.29. WCD protocol operating rules

Warto zauważyć, iż protokół WCD posiada pewne zdolności wykrywania kolizji [42]. Jeśli bowiem ramka jest podzielona na nagłówek (zawierający przynajmniej adres docelowy) i rdzeń (zawierający dane), można wykryć kolizję, zanim zostanie wysłany rdzeń ramki. Po wysłaniu nagłówka nadawca czeka na pojawienie się tonu FT. Brak tego tonu może oznaczać

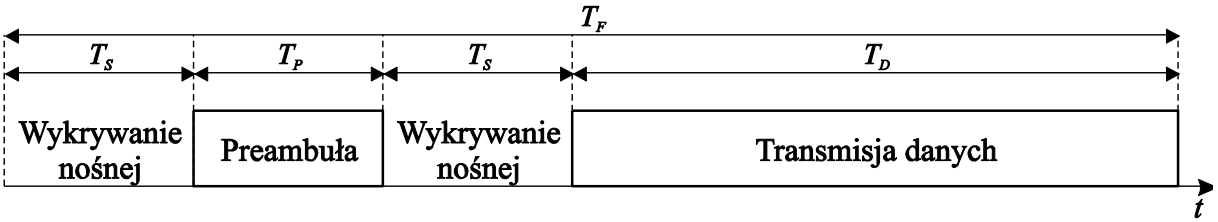
kolizję, błąd transmisji lub nawet brak stacji docelowej. W każdym z tych przypadków transmisję należy przerwać przed wysłaniem rdzenia ramki, gdyż jest ona niecelowa. Z tego punktu widzenia WCD zachowuje się także podobnie do protokołów wykorzystujących wymianę ramek sterujących [41].

Główną wadą protokołów opartych na wykrywaniu tonu zajętości jest konieczność zmniejszenia szerokości kanału danych. Kolejnym problemem jest możliwa różnica w zasięgu transmisji w kanale sterującym i danych [107]. Ponadto, zagłuszając kanał sterujący, można łatwo zablokować działanie całej sieci. Metoda ta nie jest tak łatwa do zaimplementowania jak np. CSMA, gdyż wymaga nadajników-odbiorników radiowych, mających możliwość prowadzenia transmisji przynajmniej w dwóch kanałach jednocześnie. Zapewne z tego powodu omawiane protokoły nie znalazły, jak dotąd, szeroko znanych zastosowań.

1.4.4. Protokoły z wykrywaniem kolizji

W sieciach przewodowych można łatwo zaimplementować nie tylko mechanizmy unikania kolizji oparte na wykrywaniu nośnej, lecz także wykrywanie kolizji, podnoszące wydajność protokołu przez znajdowanie kolizji występujących podczas transmisji ramki. Wymaga to jednak możliwości nasłuchu podczas nadawania. Warunku tego nie można jednak spełnić w sieciach radiowych ze względu na występowanie efektu przechwytywania [104]. Z tego powodu stacja nie może wykryć transmisji pochodzącej z innych stacji. Nie ma zatem sensu używanie jednokanałowych, dwukierunkowych nadajników-odbiorników, gdyż wyższy koszt ich produkcji nie zwróci się przez podniesienie wydajności sieci.

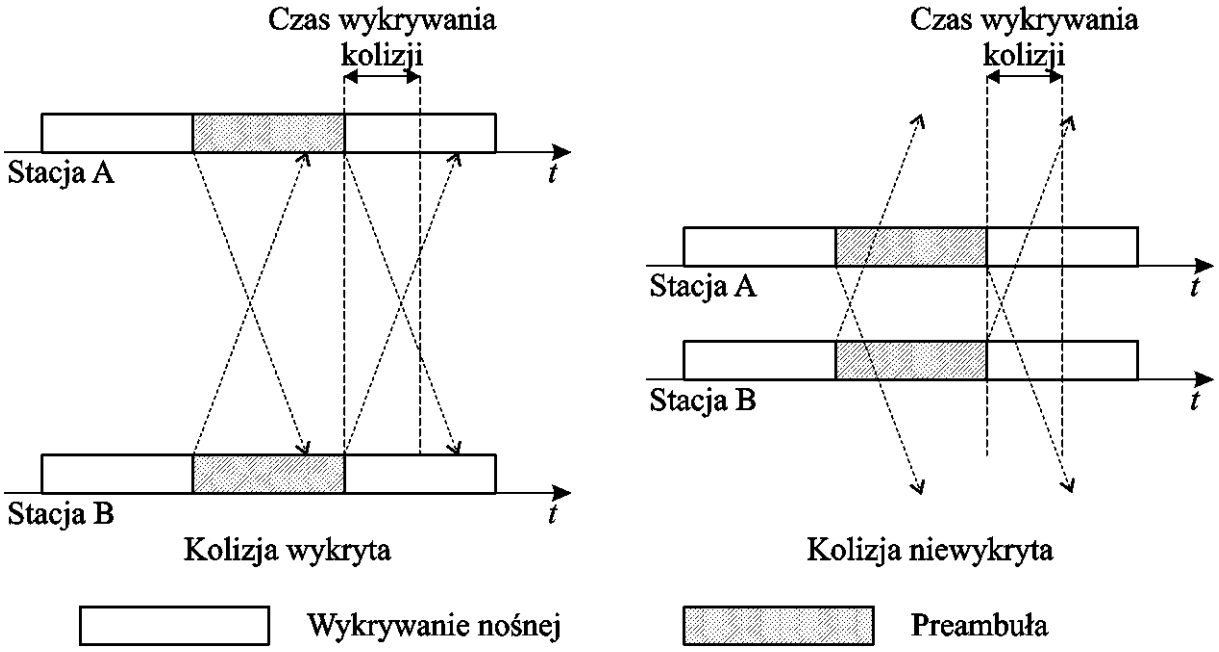
Jedną z możliwych metod wczesnego – tj. przed zakończeniem transmisji ramki – wykrywania kolizji jest podział ramki na dwie części w celu umożliwienia dwukrotnego wykrywania nośnej. Metodę tę zastosowano w protokole CSMA-TCD (ang. *CSMA with Time-Split Collision Detection*) [69], a jej działanie wyjaśnia rys. 1.30.



Rys. 1.30. Zasada działania protokołu CSMA-TCD
Fig. 1.30. CSMA-TCD protocol operation rules

Gdy stacja zamierza nadać ramkę, musi najpierw przeprowadzić wykrywanie nośnej, aby stwierdzić, czy łącze jest wolne przez czas T_s. Jeśli tak, nadaje preambułę przez czas T_p, a następnie przerywa transmisję, aby ponowić wykrywanie nośnej. Gdy drugie sprawdzenie również wykaże, że łącze jest wolne (co może oznaczać brak kolizji), wysyłanie ramki można kontynuować; w przeciwnym przypadku transmisja danych jest zatrzymywana, gdyż wystąpi-

ła kolizja podczas nadawania preambuły. Aby zapewnić prawidłowe wykrywanie kolizji, zarówno czas wykrywania nośnej (T_s) jak i czas trwania preambuły (T_p) muszą być dłuższe niż opóźnienie propagacyjne. Niestety, w pewnych przypadkach protokół nie gwarantuje wykrycia kolizji. Przypadek udanego i nieudanego wykrycia kolizji pokazano na rys. 1.31. W drugim przypadku kolizji nie wykryto, ponieważ stacje są zbyt blisko siebie i transmisja preambuły zakończyła się, zanim upłynął czas wykrywania kolizji w drugiej stacji.

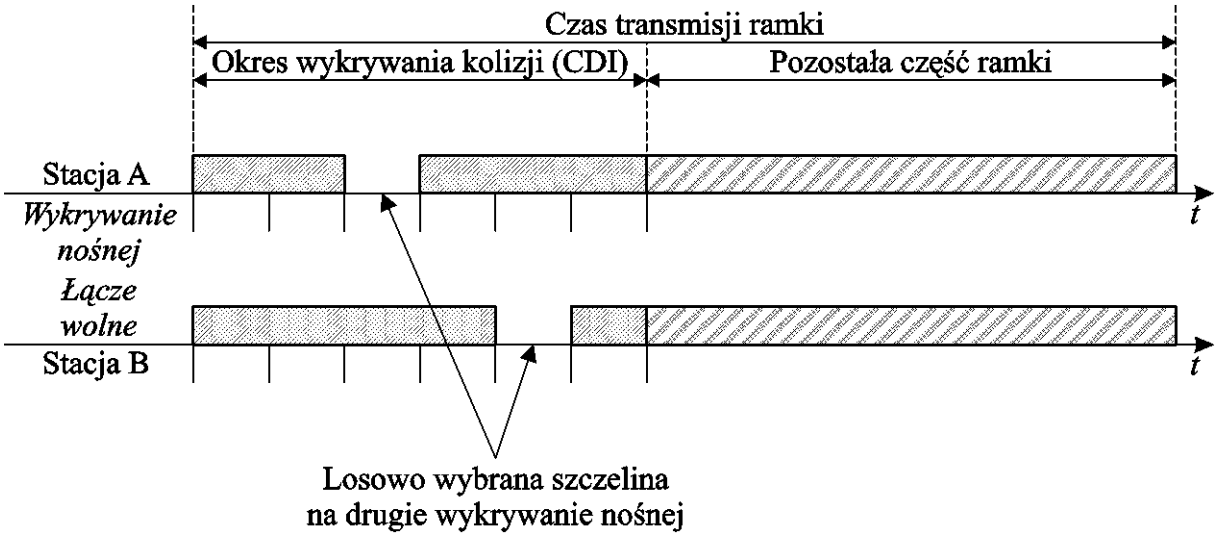


Rys. 1.31. Udane i nieudane wykrycie kolizji w protokole CSMA-TCD
Fig. 1.31. Successful and unsuccessful collision detection i CSMA-TCD protocol

Podwójne wykrywanie kolizji można także zrealizować w inny sposób [87]. Po pierwszym wykrywaniu czas dzieli się na okres wykrywania kolizji (CDI, ang. *Collision Detection Interval*) i właściwą transmisję danych. Okres CDI dzieli się na szczeliny o długości odpowiadającej opóźnieniu propagacyjnemu. Stacja, która uznała łącze za wolne podczas pierwszego sprawdzenia i rozpoczęła nadawanie, wybiera losowo jedną spośród szczelin okresu CDI, tymczasowo zatrzymuje transmisję i ponawia wykrywanie nośnej. Niezależnie od wyniku tego sprawdzenia transmisja musi trwać do końca okresu CDI – gwarantuje to, że wszystkie stacje, których ramki uległy zniszczeniu, wykryją kolizję. Zasadę działania tego protokołu zilustrowano na rys. 1.32.

Jedyny przypadek, kiedy kolizja może pozostać niewykryta, występuje wówczas, gdy wszystkie nadające stacje wybiorą tę samą szczelinę na powtórne sprawdzenie stanu łącza. W efekcie błędnie uznają one, iż kolizja nie wystąpiła, a transmisja ramki będzie kontynuowana. Jest to jednak mniej prawdopodobne niż w protokole CSMA-TCD, a to ze względu na zmienne i pseudolosowe położenie szczeliny powtórnego wykrywania nośnej.

W rzeczywistości okres CDI nie musi być dzielony na szczeliny – czas drugiego wykrywania nośnej można także wybrać losowo bez takiego podziału. W obu jednak przypadkach informacja o położeniu przerwy na drugie wykrywanie nośnej musi być zawarta na początku ramki – w przeciwnym przypadku stacja może nie odebrać ramki poprawnie.



Rys. 1.32. Wykrywanie kolizji z losowym wyborem szczeliny dla drugiego wykrywania nośnej
Fig. 1.32. Collision detection with random selection of the slot for second carrier sense

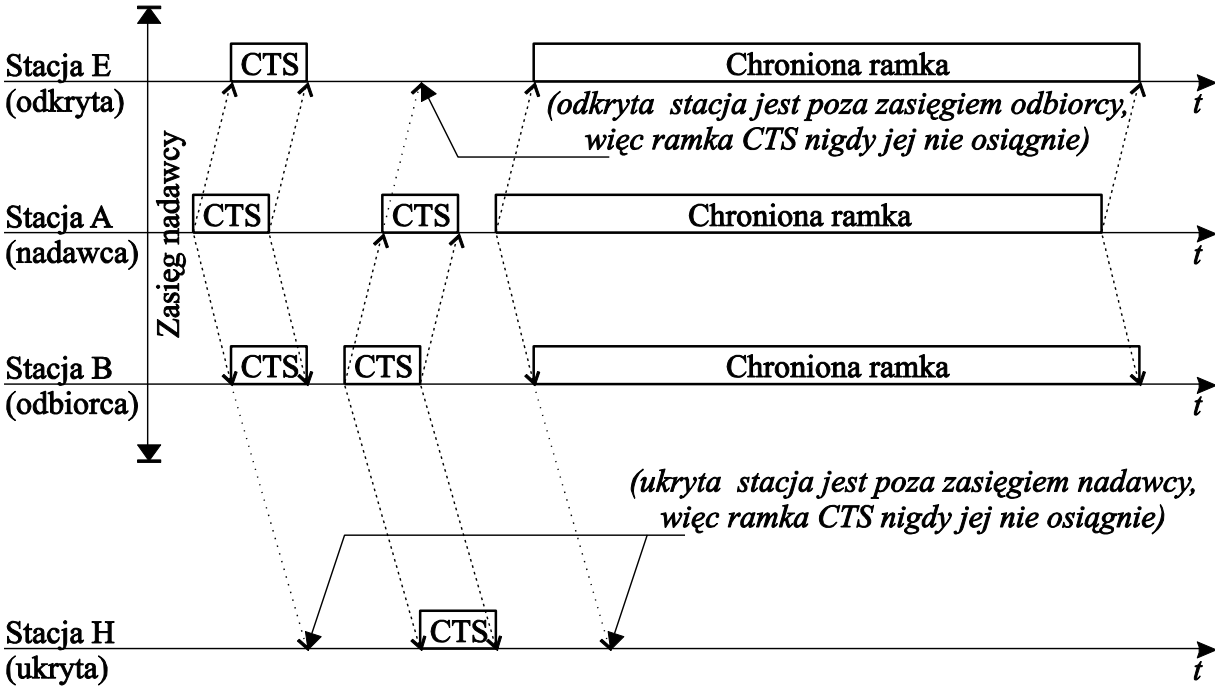
Liczbę szczelin CDI można dostrajać stosownie do parametrów sieci, jednak później musi ona pozostać niezmieniona w całej sieci. Jest oczywiste, iż wraz ze skracaniem okresu CDI rośnie prawdopodobieństwo wybrania tej samej szczeliny przez różne stacje. W szczególności, gdy okres CDI składa się tylko z jednej szczeliny, protokół redukuje się do „zwykłego" CSMA. Z drugiej jednak strony, gdy liczba szczelin jest zbyt duża, okres CDI wydłuża się, a zatem, nawet w przypadku wykrycia kolizji, łącze pozostaje zajęte przez odpowiednio dłuższy czas. Można zatem wyliczyć optymalną długość okresu CDI, dla której, przy zadanych parametrach sieci, wydajność protokołu jest największa [87]. Cennym pomysłem byłaby także możliwość dynamicznej zmiany tej długości stosownie do obciążenia sieci i liczby kolizji, jednak konieczność utrzymania jednej wartości w całej sieci istotnie utrudnia wprowadzenie takiego mechanizmu ze względu na brak centralnej stacji sterującej.

1.4.5. Protokoły MACA i MACAW

Analiza zachowania sieci bezprzewodowych zawierających stacje ukryte i odkryte może prowadzić do spostrzeżenia, iż wykrywanie nośnej nie jest efektywną metodą dostępu do łącza dla tych sieci. Pomimo że można ją zastąpić przez zmodyfikowane wykrywanie tonu zajętości (jak np. w protokole WCD), implementacja takiej metody nie jest łatwa, gdyż wymaga ona wydzielonego kanału sterującego przeznaczonego jedynie dla celów sterowania. Protokół WCD wymaga nawet dwóch takich kanałów dla przesyłania dwóch różnych tonów zajętości.

Obserwacja ta zaowocowała powstaniem protokołu MACA (ang. *Medium Access with Collision Avoidance*) [62].

W protokole tym w ogóle nie prowadzi się wykrywania nośnej. Zamiast niego wprowadzono wymianę ramek sterujących poprzedzającą przesył danych – nadawca wysyła ramkę RTS, na którą odbiorca powinien odpowiedzieć ramką CTS. Ideę tego protokołu wyjaśnia rys. 1.33.

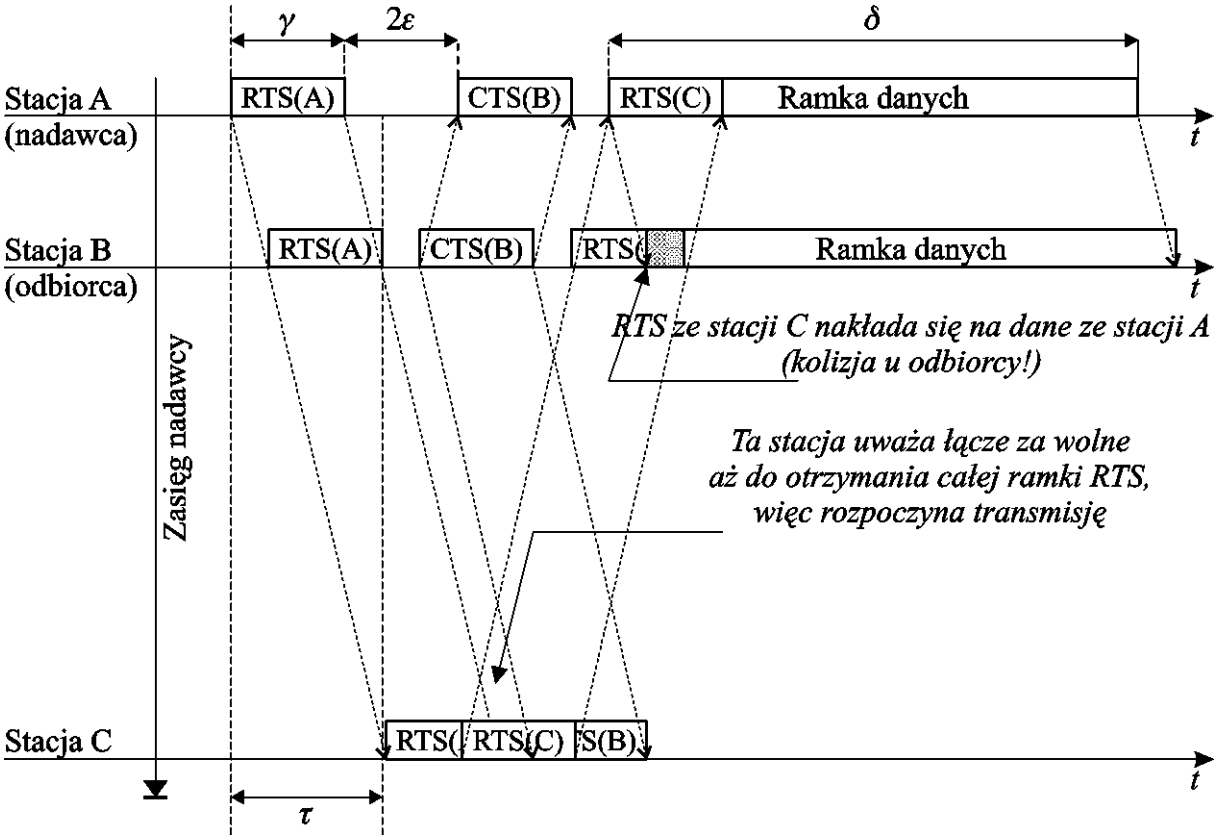


Rys. 1.33. Zasada wymiany ramek sterujących w protokole MACA
Fig. 1.33. The rules of control frames exchange in MACA protocol

Wymiana ramek sterujących zapobiega negatywnym skutkom występowania stacji ukrytych i odkrytych. Stacja ukryta bowiem może odebrać ramkę CTS od odbiorcy informacji, podczas gdy stacja odkryta – ramkę RTS od nadawcy. Co więcej, na podstawie analizy otrzymanych ramek stacje mogą rozpoznać, czy są ukryte czy odkryte względem przygotowywanej transmisji danych (rys. 1.33). Po prawidłowej wymianie ramek RTS-CTS łącze zostaje uznane za bezkolizyjne i zarezerwowane dla określonej transmisji. Czas rezerwacji można łatwo określić, o ile ramki RTS i CTS niosą informację o przewidywanym czasie transmisji. Aby uniknąć kolizji, stacja ukryta powinna powstrzymać się od nadawania przez cały czas rezerwacji, natomiast stacja odkryta może rozpocząć transmisję, gdy tylko nadawca otrzyma ramkę CTS. Transmisja stacji odkrytej powinna jednak zakończyć się wraz z końcem rezerwacji na wypadek, gdyby odbiorca zamierzał przesłać potwierdzenie.

Niestety, protokół MACA nie zapobiega kolizjom między ramkami sterującymi. Każda stacja przyjmuje bowiem łącze za wolne do chwili otrzymania całej ramki RTS lub CTS. Nie mając informacji o rozpoczęciu transmisji ramki sterującej (wskutek braku wykrywania no-

śnej), stacja może rozpocząć nadawanie własnej ramki RTS, powodując kolizję z inną ramką RTS lub nawet – w szczególnie niekorzystnym przypadku, np. w sieciach ruchomych – z ramką danych. Jeśli kolizja obejmuje tylko ramki RTS, łącze nie zostanie przypisane żadnej stacji, a nieudana rezerwacja zostanie utracona, ale przynajmniej nie trzeba ponawiać transmisji znacznie dłuższych ramek danych. Jeśli jednak kolizja obejmie także ramki danych, muszą one być retransmitowane, mimo że rezerwacja była przeprowadzona poprawnie. Przypadek taki pokazano na rys. 1.34. Można wykazać [34], że protokół może skutecznie chronić ramki sterujące i danych przed kolizjami z innymi ramkami sterującymi, o ile tylko czas trwania ramek RTS i CTS jest dłuższy od podwojonego opóźnienia propagacyjnego w danej sieci ($\gamma > 2\tau$).



Rys. 1.34. Kolizja między ramką danych i RTS w protokole MACA [34]
Fig. 1.34. A collision between RTS and data frames in MACA protocol

Efektywność protokołu MACA w obecności licznych stacji ukrytych i odkrytych jest wyższa niż CSMA, szczególnie przy długich ramkach danych.

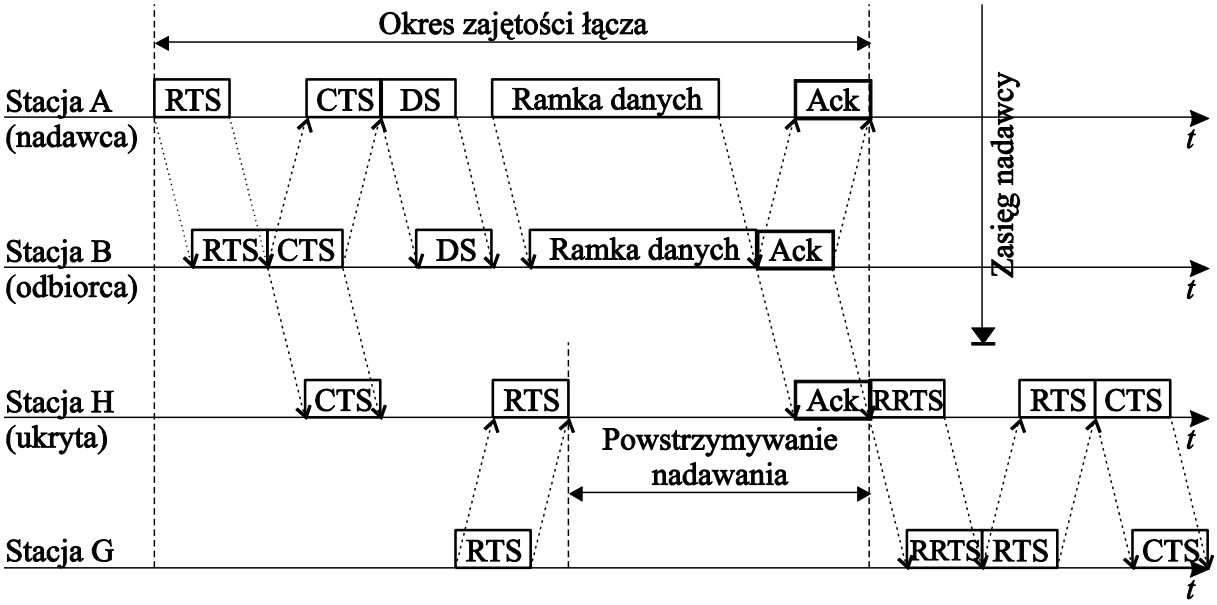
W sieciach radiowych przepustowość kanału jest ograniczona przez czas przełączania między nadawaniem a odbiorem (ang. *turnaround time*). Aby zmniejszyć liczbę takich przełączeń, a tym samym podnieść wydajność protokołu, zaproponowano uproszczony wariant protokołu MACA, MACA-BI (ang. *MACA By Invitation*) [101]. W tej odmianie nie stosuje się wymiany ramek RTS-CTS. Zamiast niej odbiorca wysyła ramkę RTR (ang. *Ready To Re-*

ceive) do nadawcy. Mechanizm taki wystarcza dla ochrony przed kolizjami, ponieważ RTR, podobnie jak i CTS, informuje stacje w zasięgu odbiorcy o nadchodzącej transmisji. Tym niemniej, odbiorca nie ma informacji o stanie kolejki ramek u nadawcy, ta informacja zatem musi być przesłana wraz z danymi. Jedyną wadą tego protokołu jest konieczność przewidywania, który nadawca zamierza przesłać dane i jaka (przynajmniej w przybliżeniu) jest ich objętość.

W protokole MACAW (ang. *Medium Access with Collision Avoidance for Wireless*) [11] rozwinięto koncepcję ramek sterujących, wprowadzoną w protokole MACA. Między innymi, określono nowe typy ramek sterujących:

- DS (ang. *Data Sending*), informującą stacje o udanym zakończeniu negocjacji RTS-CTS, a co za tym idzie, o rozpoczęciu transmisji danych,
- Ack (ang. *Acknowledge*), potwierdzającą prawidłowy odbiór ostatniej ramki danych,
- RRTS (ang. *Request for RTS*), informującą o gotowości do przyjęcia ramki RTS; ramka ta jest używana, jeśli w czasie wstrzymywania transmisji stacja ukryta odebrała ramkę RTS; po zakończeniu okresu wstrzymywania umożliwia przeprowadzenie prawidłowej negocjacji RTS-CTS.

Zasady działania protokołu MACAW wyjaśniono na rys. 1.35.



Rys. 1.35. Zasada wymiany ramek sterujących w protokole MACAW
Fig. 1.35. The rules of control frames exchange in MACAW protocol

Zgodnie z obliczeniami [11], schemat przesyłu ramek RTS-CTS-DS-Dane-Ack jest znacznie bardziej wydajny w sieciach z dużym poziomem zakłóceń, nieznacznie zmniejszając przepustowość sieci, gdy zakłócenia nie występują. Ramka DS zapewnia bardziej sprawiedliwy podział czasu łącza. Dodatkowe ulepszenia w tym zakresie są efektem zastosowania zmienionego algorytmu wycofywania (ang. *backoff*) – zamiast znanej z sieci Ethernet metody

BEB (ang. *Binary Exponential Backoff*) zastosowano MILD (ang. *Multiplication Increase Linear Decrease*). W sieciach z metodą BEB można zaobserwować efekt zawłaszczenia łącza przez stację, która wygrała rywalizację w sytuacji nasycenia łącza [93]. Stacja ta zmniejsza okno rywalizacji do wartości minimalnej. Pozostałe stacje natomiast zwiększają je dwukrotnie po każdej nieudanej próbie dostępu aż do osiągnięcia wartości maksymalnej. Jak nietrudno zauważyć, jedna tylko wygrana rywalizacja może prowadzić do całkowicie niesprawiedliwego podziału czasu łącza. Zaproponowana w [11] metoda MILD w przypadku przegranej rywalizacji także zwiększa okno, ale już nie dwu-, tylko 1,5-krotnie. Ponadto, wygrana rywalizacja zmniejsza okno, ale nie do wartości minimalnej – dotychczasowa wielkość zmniejsza się o 1. Ponadto, stacje mogą rozsyłać stan swojego licznika wycofywania (ang. *backoff counter*) we wszystkich ramkach, z wyjątkiem RTS, co zapewnia wszystkim stacjom zbliżone szanse uzyskania dostępu do łącza.

1.4.6. Protokoły rodziny FAMA

FAMA (ang. *Floor Acquisition Multiple Access*[3]) [34] stanowi grupę protokołów, wykorzystujących zarówno wykrywanie nośnej, jak i wymianę ramek sterujących poprzedzającą transmisję danych, wprowadzoną w protokole MACA. Istotą protokołu jest dynamiczne zezwalanie poszczególnym stacjom na sterowanie łączem. Podobne mechanizmy są używane także w protokołach z dynamiczną rezerwacją (np. SRMA [108], MSAP [64] czy BRAM [18]), jednak FAMA nie wykorzystuje ani osobnego kanału sterującego, ani centralnej stacji koordynującej pracę sieci.

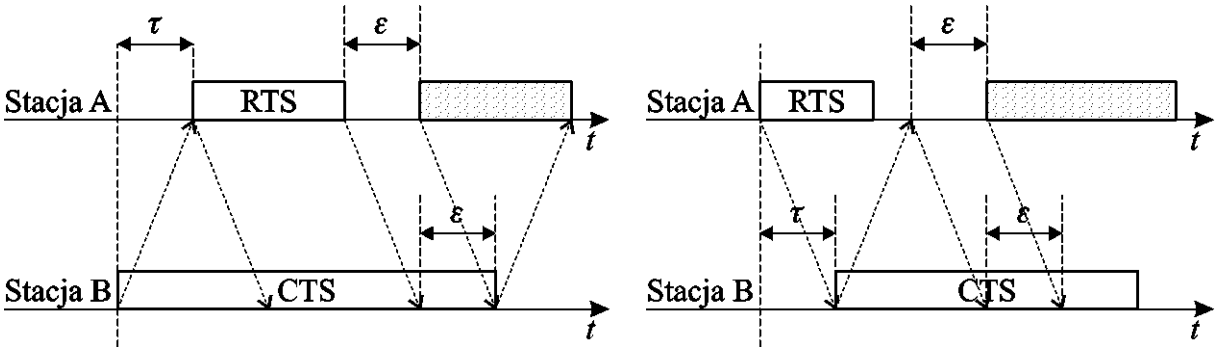
Przed rozpoczęciem transmisji stacja musi uzyskać prawo sterowania łączem. Mechanizm przekazywania sterowania odbywa się na zasadzie wymiany informacji sterującej, którą przesyła się w tym samym kanale co dane w taki sposób, że mimo iż mogą nastąpić kolizje między ramkami sterującymi, dane są przesyłane zawsze bez kolizji. Jest to możliwe, o ile przestrzega się określonych zależności czasowych, między innymi czas transmisji ramek sterujących nie może być krótszy niż podwojony czas propagacji w kanale. W przeciwnym razie ramki danych mogłyby ulec kolizji z ramkami sterującymi [34].

Można wyróżnić kilka odmian protokołu FAMA. Bez wykrywania nośnej FAMA odpowiada protokołowi MACA. Inna odmiana, FAMA-NTR (ang. *Non-persistent Transmission Request*) łączy nietrwałe wykrywanie nośnej z wymianą ramek sterujących RTS-CTS. Gdy stacja ma ramkę do wysłania, musi najpierw uzyskać prawo nadawania, tzn. przejąć sterowanie łączem. Początkowo stacja prowadzi nasłuch łącza używając wykrywania nośnej. Jeśli łącze jest wolne, stacja wysyła ramkę RTS i czeka przez pewien czas na ramkę CTS. Jeśli CTS nadejdzie w ściśle określonym czasie, stacja uzyskuje prawo sterowania łączem i może

[3] w języku angielskim zwrot *to acquire the floor* oznacza „zabrać głos”.

wysłać określoną liczbę ramek danych. Po ich wysłaniu stacja zwalnia łącze i, jeśli to konieczne, ubiega się o ponowny przydział łącza. Jeśli przed wysłaniem ramki RTS łącze jest zajęte, lub jeśli ramka CTS nie nadeszła w określonym czasie, stacja uznaje próbę dostępu za nieudaną i ponawia ją po upływie losowego czasu.

W odmianie FAMA-NCS (ang. *Non-persistent Carrier Sense*) [35] długość ramki RTS jest większa niż maksymalne opóźnienie propagacyjne w kanale ($T_{RTS} > \tau$, rys. 1.36). Warunek ten pozwala uniknąć sytuacji, gdy stacja znajdująca się blisko nadawcy zakończyła już odbiór ramki RTS, natomiast stacja odległa jeszcze nie rozpoczęła odbioru. Mechanizm ten zmniejsza zatem ryzyko kolizji. Długość ramki CTS musi być jeszcze większa. Można wykazać [35], że powinna ona przekraczać długość ramki RTS o co najmniej podwojony czas propagacji w kanale, czas przełączania układu radiowego oraz czas przetwarzania ramki ($T_{CTS} > T_{RTS} + 2\tau + \varepsilon$). Można zatem powiedzieć, że CTS dominuje nad RTS. Każda stacja, która rozpoczęła nadawanie ramki RTS przed odebraniem CTS, będzie mogła odebrać przynajmniej koniec ramki CTS po wysłaniu własnej ramki RTS. Co prawda, nie zdekoduje ona prawidłowo ramki CTS, ale uzyska informację o kolizji ramek sterujących. Mechanizm ten pozwala bezkolizyjnie odebrać ramkę danych. Ramka CTS pełni także rolę podobną do tonu zajętości w protokołach rodziny BTMA. Opisane wymagania co do długości ramek sterujących wyjaśniono na rys. 1.36.



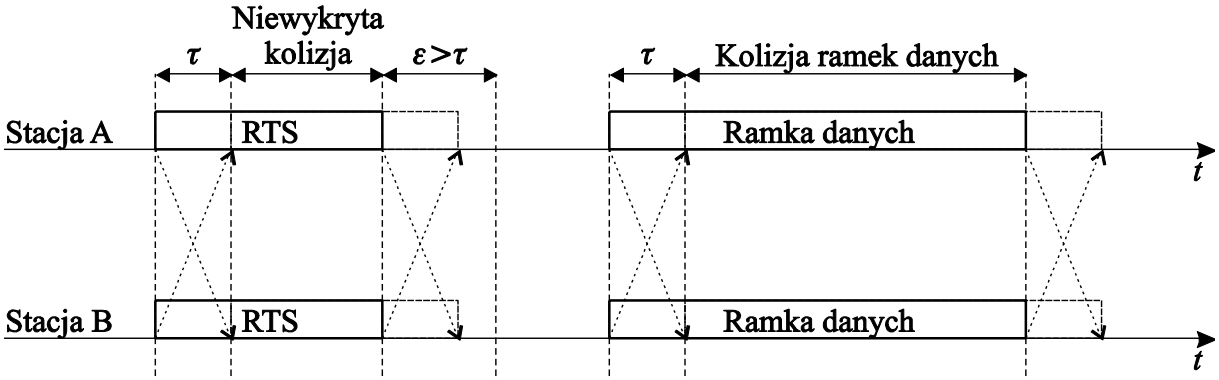
Rys. 1.36. Wymagania co do długości ramek sterujących w protokole FAMA [35]
Fig. 1.36. Control frame length requirements in FAMA protocol

Aby zwiększyć efektywność protokołu, stacja posiadająca prawo nadawania może przesłać większą liczbę ramek. Jeśli nadawca zamierza wysłać kolejną ramkę, informuje odbiorcę o swym zamiarze, ustawiając odpowiedni znacznik w ramce poprzedniej, zamiast wykonywać kolejną negocjację RTS-CTS. Po odebraniu ramki z ustawionym znacznikiem odbiorca wysyła ramkę CTS, informując sąsiednie stacje (w tym ukryte) o dodatkowej transmisji.

Protokół FAMA-NPS (ang. *Non-persistent Packet Sensing*) [35] jest zbliżony do FAMA-NCS, ale nie wykrywa nośnej. Prawidłowe przejęcie sterowania łączem jest zatem możliwe tylko wówczas, gdy sieć nie zawiera stacji ukrytych bądź gdy odbiorca prześle więcej niż

jedną ramkę CTS. Wynika to z możliwości kolizji pomiędzy ramkami sterującymi, gdyż nie są one chronione wykrywaniem nośnej.

Odmiana FAMA-PJ (ang. *Passive Jamming*) [33] łączy protokół FAMA i wykrywanie kolizji uzyskane metodą przerw w nadawaniu. W przeciwieństwie do poprzednich protokołów wykrywających kolizje [68, 87], FAMA-PJ używa tzw. pasywnego zagłuszania zamiast aktywnego. W zagłuszaniu aktywnym biorą udział tylko stacje bezpośrednio zaangażowane w kolizję (stacje aktywne). Natomiast w zagłuszaniu pasywnym biorą udział jedynie stacje niezaangażowane w kolizję (pasywne). Zagłuszanie aktywne – po wykryciu kolizji – może nie być wystarczające, gdy czas propagacji sygnału w kanale (ε) jest większy niż czas przełączania układu radiowego (τ). W tym przypadku stacje wysyłające ramki RTS (aktywne) mogą prowadzić nasłuch łącza dopiero po całkowitym ich wysłaniu. Stacje aktywne nie mogą wykryć nośnej pochodzącej od innej stacji, a zatem błędnie przyjmują, iż kolizja nie nastąpiła i rozpoczynają transmisję ramek danych, które także ulegają kolizji. Opisane zjawisko wyjaśnia rys. 1.37.



Rys. 1.37. Kolizja ramek danych wskutek niewykrytej kolizji ramek RTS
Fig. 1.37. Data frames collision as a result of undetected RTS frames collision

Jeśli stacja pasywna (niezaangażowana w jakąkolwiek transmisję) wykryje nośną, lecz po upływie określonego czasu nie zdekoduje ramki RTS, zakłada wystąpienie kolizji i przesyła sygnał zagłuszający przez czas $\varepsilon + 2\tau$. Czas ten wystarcza, aby poinformować wszystkie stacje o kolizji między ramkami RTS.

Pewne elementy protokołów FAMA i MACAW można zauważyć m. in. w protokole DFWMAC (ang. *Distributed Foundation Wireless Medium Access Control*), zdefiniowanym w standardzie IEEE 802.11 [52]. Pomimo iż ogólna zasada działania jest zbliżona – w DFWMAC stosuje się zarówno wykrywanie nośnej, jak i wymianę ramek sterujących – protokoły te różnią się pewnymi szczegółami. Przykładowo, w protokole FAMA-NCS przyjęto długość ramki CTS większą niż RTS; w DFWMAC natomiast jest odwrotnie. W protokołach FAMA wymiana ramek sterujących jest obowiązkowa, wykrywanie nośnej zaś można uznać za opcjonalne, gdyż nie występuje w każdej odmianie protokołu. W DFWMAC natomiast wykrywanie nośnej jest obowiązkowe, wymiana ramek sterujących zaś – opcjonalna; instrukcje

użytkownika wielu urządzeń zgodnych ze standardem IEEE 802.11 odradzają nawet włączanie tej metody unikania kolizji. O ile podejście takie jest uzasadnione w prostych sieciach, o tyle w złożonych konfiguracjach, szczególnie rozległych terytorialnie, włączenie wymiany ramek sterujących może jednak przynieść pewne korzyści. W porównaniu z protokołem FAMA, w DFWMAC dodano ramkę potwierdzenia (Ack), podobnie jak w protokole MACAW. Pomimo wymienionych różnic można stwierdzić, iż protokoły rodziny FAMA mają obecnie kolosalne znaczenie w bezprzewodowych sieciach lokalnych.

1.4.7. Protokół BAPU

Opisane powyżej protokoły, pomimo wymiany informacji sterującej poprzedzającej transmisję danych, rozwiązują jedynie zagadnienie ukrytego i odkrytego nadajnika [9, 10]. Uniknięcie problemu ukrytego i odkrytego odbiornika wymaga natomiast, aby:

- Ukryty odbiornik mógł wysłać informację o powstrzymywaniu transmisji; aby uniknąć kolizji, informacja ta musi być przesłana w osobnym kanale;
- Odkryty odbiornik mógł odebrać informację sterującą nawet wówczas, gdy trwa transmisja danych; aby to zapewnić, jest potrzebny osobny kanał sterujący;
- Stacje zagłuszające były poinformowane o trwającej transmisji danych; można to uzyskać przez zwiększenie zasięgu transmisji w kanale sterującym.

Mając na uwadze powyższe stwierdzenia, w protokole BAPU (ang. *Basic Access Protocol solUtions*) [9] wprowadzono dwa oddzielne kanały: danych i sterujący. Dodatkowo kanał sterujący ma powiększony zasięg transmisji. Dzięki takiemu rozwiązaniu stacje, mogące interferować w kanale danych, stają się stacjami ukrytymi lub odkrytymi w kanale sterującym, a zatem istnieje możliwość poinformowania ich o nadchodzącej transmisji.

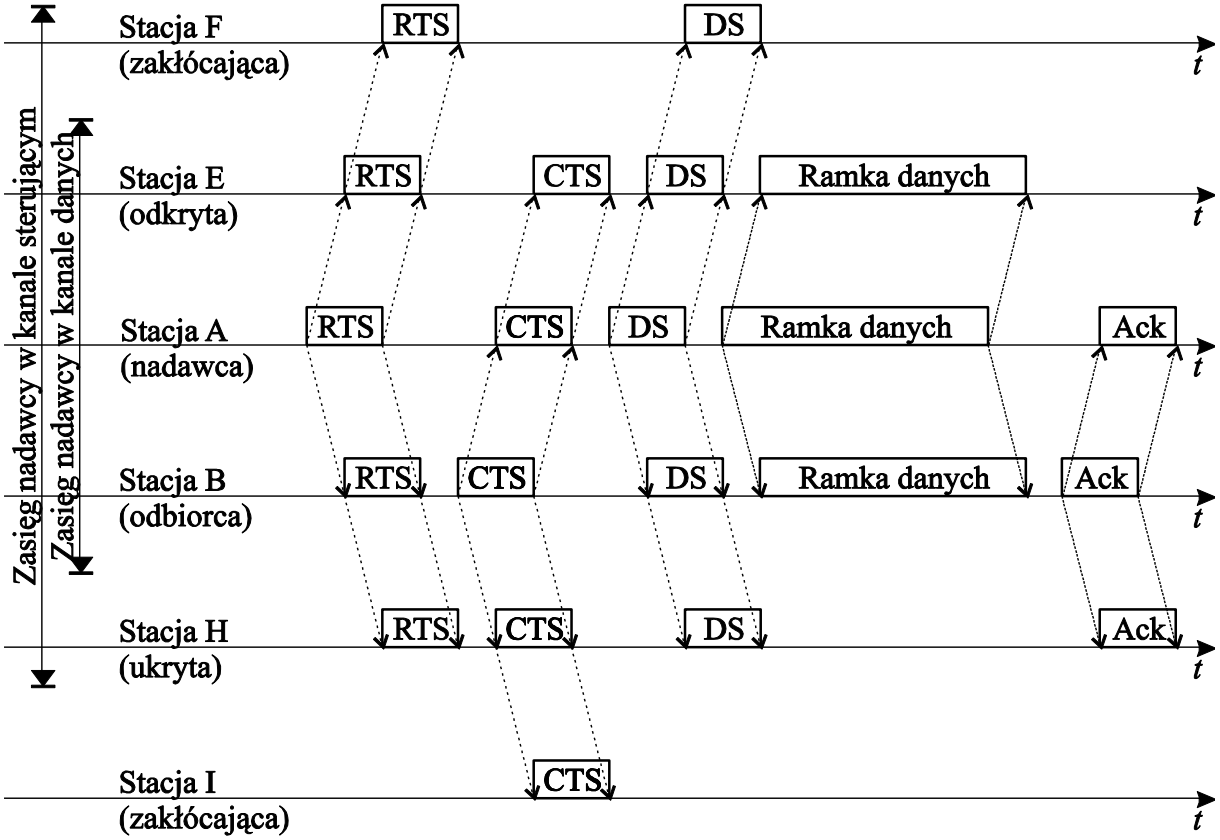
W protokole używa się pięciu typów ramek sterujących:

- RTS, informującą o gotowości stacji do transmisji,
- CTS, informującą o gotowości stacji do odbioru danych,
- DS, informującą o poprawnym zakończeniu negocjacji RTS-CTS i rozpoczęciu transmisji danych,
- Ack, potwierdzającą prawidłowy odbiór ostatniej ramki danych,
- NCTS, informującą o braku gotowości stacji do odbioru danych, np. gdy znajduje się ona w zasięgu innej transmisji.

Ramki danych i potwierdzenia (Ack) przesyła się w kanale danych, podczas gdy wszystkie pozostałe – w kanale sterującym. Ideę protokołu wyjaśnia rys. 1.38.

Obszar zajęty przez określoną transmisję (ze stacji A do B) przedstawiono na rys. 1.16. Jest oczywiste, iż obszar ten jest większy niż w przypadku innych podobnych protokołów (np. MACA lub FAMA), ponieważ zasięg kanału sterującego jest większy. Ponieważ jednak

o transmisji informuje się dodatkowe stacje (w tym zakłócające), zmniejsza się ryzyko kolizji. Co więcej, cała wymiana ramek zachodzi w lepszych warunkach, ponieważ zmniejsza się także poziom zakłóceń.



Rys. 1.38. Zasada działania protokołu BAPU
Fig. 1.38. BAPU protocol operating rules

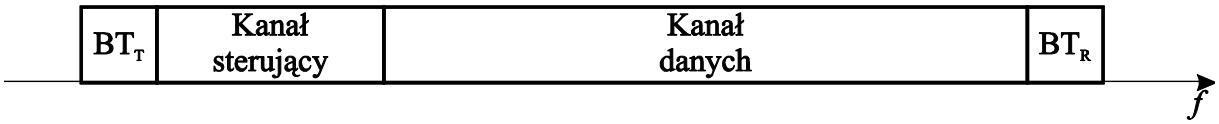
Wadą protokołu BAPU jest duża złożoność urządzeń transmisyjnych – muszą one zapewniać możliwość jednoczesnej i niezależnej transmisji w obu kanałach. Z tego zapewne powodu protokół ten nie ma, jak dotąd, praktycznych zastosowań.

1.4.8. Protokoły rodziny DBTMA

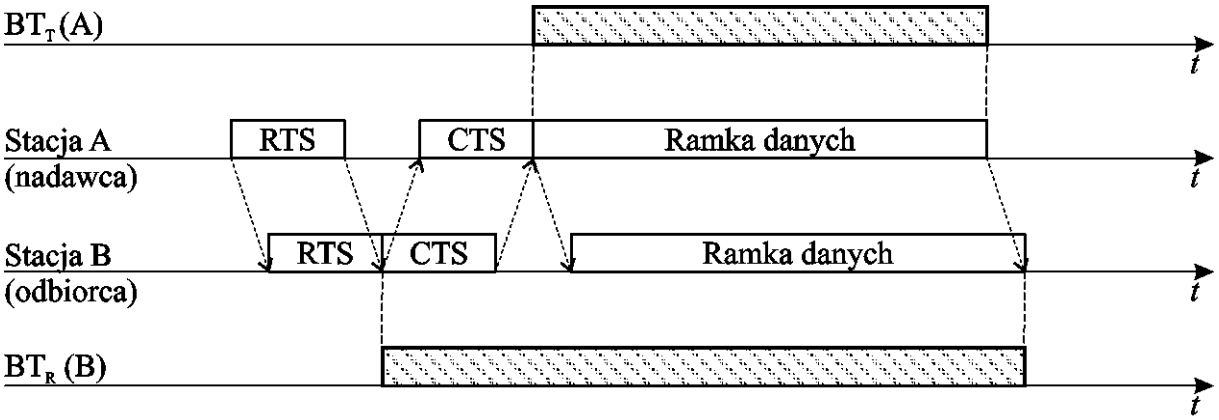
Protokół DBTMA (ang. *Dual Busy Tone Multiple Access*) jest kombinacją wykrywania tonu zajętości i wymiany ramek sterujących. Kanał transmisyjny jest podzielony na kanał danych, w którym przesyła się tylko ramki danych, oraz kanał sterujący, używany do wymiany ramek sterujących. Dodatkowo w kanale sterującym jest prowadzona transmisja tonu zajętości. W protokole DBTMA wyróżnia się dwa takie tony: BT_T (ang. *Transmit Busy Tone*), wytwarzany przez nadajnik, oraz BT_R (ang. *Receive Busy Tone*), wytwarzany przez odbiorcę informacji. Dokładne zasady wytwarzania tonów zajętości różnią się w zależności od odmiany protokołu. Organizację kanału transmisyjnego przedstawia rys. 1.39.

W pierwszej odmianie protokołu DBTMA [27] używa się zarówno obu ramek sterujących (RTS i CTS), jak i obu tonów zajętości. Gdy nadawca zamierza nadać ramkę, musi najpierw

sprawdzić obecność sygnału BT_R, aby upewnić się, że żadna inna stacja w pobliżu nie nadaje. Jeśli łącze jest wolne, stacja wysyła ramkę RTS, kontynuując jednak nasłuch ewentualnego sygnału BT_R do końca tej transmisji. Gdy taki sygnał się pojawi, stacja porzuca transmisję, nawet jeśli negocjacja RTS-CTS była udana. Po prawidłowym odebraniu ramki RTS odbiorca odpowiada ramką CTS i rozpoczyna wytwarzanie tonu BT_R. Z kolei, nadawca wysyła ramkę danych, chronioną dodatkowo sygnałem BT_T. Tony zajętości są wytwarzane w sposób ciągły przez obie porozumiewające się stacje aż do zakończenia transmisji ramki danych. Zasadę działania tej odmiany protokołu przedstawiono na rys. 1.40.



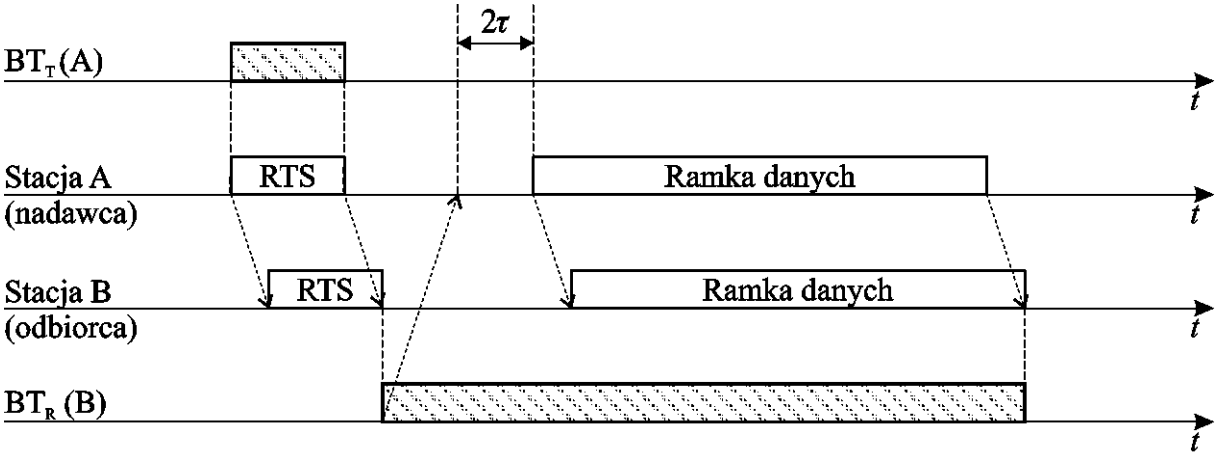
Rys. 1.39.Organizacja kanału transmisyjnego w protokole DBTMA
Fig. 1.39. Transmission channel organisation in DBTMA protocol



Rys. 1.40. Zasada działania początkowej wersji protokołu DBTMA
Fig. 1.40. The rules of the initial variant of DBTMA protocol

Nietrudno zauważyć, że tony zajętości BT_T i BT_R w pewnym sensie dublują ochronę wynikającą z faktu przesłania ramek sterujących, odpowiednio RTS i CTS. Tym niemniej, w przeciwieństwie do owych ramek, tony zajętości zapewniają ciągłą ochronę przez cały czas transmisji danych nawet wówczas, gdy ramki RTS lub CTS nie zostały prawidłowo odebrane przez pewne stacje. Wykrywanie tonów zajętości jest także wystarczające dla określenia położenia stacji względem prowadzonej transmisji – stacje ukryte nie odbierają bowiem tonu BT_T, odkryte natomiast – BT_R.

Druga odmiana protokołu jest znacznie uproszczona [28]. Wymaga ona tylko pojedynczego kanału dla transmisji ramek – zarówno sterujących, jak i danych – a także dwóch sygnałów zajętości łącza. Znaczenie tonów zajętości jest jednak nieco zmienione – ton BT_T chroni ramkę RTS zamiast ramki danych jak w poprzednim wariancie. Zasady działania protokołu pokazano na rys. 1.41.



Rys. 1.41. Zasada działania zmodyfikowanej wersji protokołu DBTMA
Fig. 1.41. The rules of the modified variant of DBTMA protocol

Gdy stacja zamierza rozpocząć nadawanie, musi najpierw sprawdzić występowanie w sieci tonów BT_T lub BT_R. Jeśli nie wykryje ona żadnego z nich, przyjmuje, że w jej zasięgu żadna inna stacja nie wysyła ramki RTS ani nie odbiera danych. Stacja może zatem wysłać ramkę RTS wraz z sygnałem ochronnym BT_T. Gdy ramka ta dotrze do adresata, musi on sprawdzić, czy jakaś stacja w pobliżu prowadzi transmisję – wskazuje na to obecność tonu BT_T. Jeśli łącze jest wolne, odbiorca włącza sygnał BT_R, powiadamiając sąsiednie stacje o zamiarze odbierania danych, a nadawcę – o możliwości bezkolizyjnego nadawania. Po usłyszeniu tonu BT_R nadawca czeka jeszcze przez podwojone opóźnienie propagacyjne w kanale (2τ), aby upewnić się, iż wszystkie ewentualne transmisje ramek RTS zostały porzucone. Następnie nadawca rozpoczyna transmisję ramki danych. Po jej zakończeniu odbiorca wyłącza sygnał BT_R. Można zatem powiedzieć, że BT_T pełni rolę zbliżoną do ramki DS w protokole MACAW, podczas gdy BT_R – do ramek CTS lub Ack. Pojedynczy ton zajętości wystarcza dla ochrony ramki danych, ponieważ jest słyszany przez stacje ukryte, natomiast przez odkryte – nie.

Zaletą używania ciągłych sygnałów ochronnych zamiast ramek sterujących jest zmniejszenie liczby przełączeń układów radiowych (uwaga ta nie w pełni dotyczy protokołu DBTMA – szczególnie pierwszej odmiany – ponieważ używa on także ramek sterujących). Metoda ta jest także skuteczniejsza w sieciach zawierających liczne stacje ruchome – stacja mogła nie odebrać prawidłowo ramki sterującej, gdy była poza zasięgiem odbiorcy, ale po zbliżeniu się do niego zostaje powiadomiona o przebiegającej transmisji ciągłym sygnałem zajętości. Możliwy jest także nasłuch łącza podczas nadawania, co umożliwia wykrywanie kolizji podobnie, jak w sieciach przewodowych. Niestety, protokół ten wymaga bardzo złożonych układów nadawczo-odbiorczych i zapewne z tego powodu – mimo interesujących właściwości – nie jest, jak dotąd, stosowany w praktyce.

1.5. Porównanie wydajności protokołów dostępu do łącza

Można przyjąć, iż spośród wielu miar wydajności protokołów dostępu do łącza najważniejszymi i najczęściej wyznaczanymi są przepustowość i opóźnienie dostępu. Najczęściej oblicza się te wartości w zależności od obciążenia łącza. Pozwala to ustalić, czy w określonych warunkach protokół osiąga wystarczającą dla danej aplikacji wydajność, a także czy zachowuje w tych warunkach stabilność.

Zarówno przepustowość, jak i opóźnienie można wyznaczyć na kilka sposobów. Pierwszym z nich jest analiza właściwości protokołu, wykorzystująca rachunek prawdopodobieństwa. Pozwala ona na stosunkowo szybkie oszacowanie żądanych charakterystyk protokołu, chociaż przy bardziej złożonych protokołach uzyskiwane zależności są często złożone i wymagają skomplikowanych obliczeń. Inną metodą uzyskiwania wspomnianych zależności jest symulacja komputerowa. Jest to bardzo popularna obecnie metoda, przy użyciu odpowiednich narzędzi pozwala bowiem oszacować charakterystykę całego stosu protokołów. Przykładowo, użycie modelu opisującego warstwę fizyczną, liniową, sieciową i transportową pozwala na określenie zachowania protokołu TCP/IP w danej sieci. Podejście takie umożliwia uzyskanie wyników, które można porównać z rzeczywistymi osiągami sieci.

Opisane metody mają pewną wadę – mogą bowiem nie odzwierciedlać rzeczywistości z wystarczającą dokładnością. Przykładowo, przy symulacji komputerowej użyte modele mogą niedoskonale opisywać symulowaną sieć. Oczywiście, nie pozostaje to bez wpływu na jakość uzyskanych wyników. Z kolei przy wykorzystaniu rachunku prawdopodobieństwa często przyjmuje się pewne założenia, które w ogóle umożliwiają uzyskanie wyniku w postaci równania analitycznego, ale – podobnie jak model symulacyjny – mogą odbiegać od rzeczywistości. Z tego powodu warto dokonać przynajmniej próbnej implementacji protokołu, aby skonfrontować wyniki obliczeń lub symulacji z pomiarami uzyskanymi w rzeczywistej sieci.

1.5.1. Oszacowanie teoretyczne

Jedną z częściej używanych miar jakości protokołu dostępu do łącza jest stopień wykorzystania kanału transmisyjnego, przedstawiony jako funkcja obciążenia wprowadzonego do kanału. Obciążenie to określa się jako liczbę ramek, wytworzonych przez wszystkie stacje sieci w czasie transmisji pojedynczej ramki. Wykorzystując tę wielkość, można oszacować maksymalne wykorzystanie kanału z uwzględnieniem kolizji i narzutu protokołu. Warto zauważyć, że uwzględnia się tu jedynie sam mechanizm dostępu do łącza, pomijając jednocześnie narzut protokołu wynikający z formatu używanych ramek i zasad ich wymiany oraz z obecności wyższych warstw sieci.

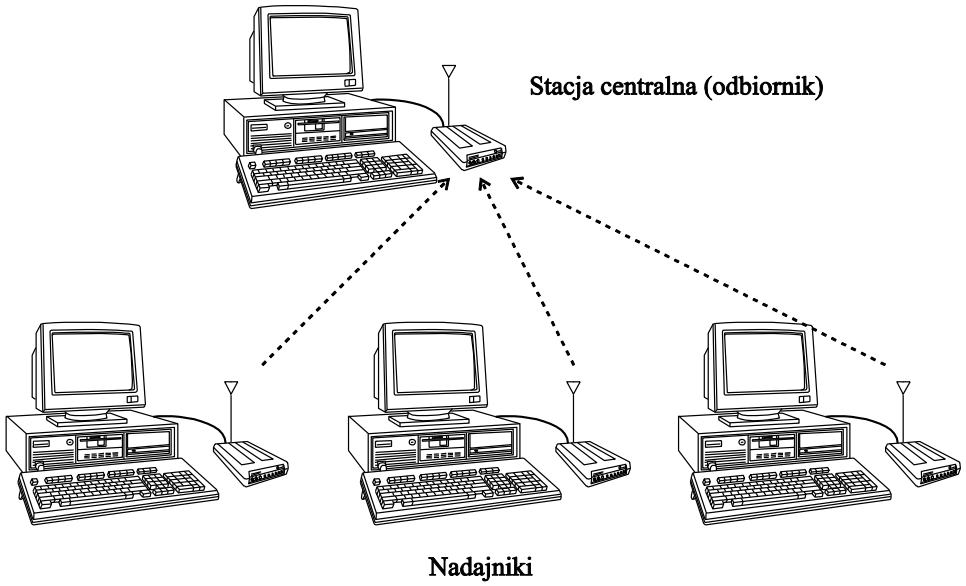
Przedstawione dalej wzory wyprowadzono przy następujących założeniach [65]:

- proces wytwarzania ramek jest procesem Poissona z intensywnością g ramek na sekundę,

- czas transmisji ramki jest stały i równy T sekund,
- potwierdzenia prawidłowo przesłanych (nieutraconych w wyniku kolizji) ramek przesyłane są tak, że nie wnoszą dodatkowego obciążenia kanału – np. w osobnym pasmie,
- liczba stacji w sieci jest nieskończenie duża,
- w sieci nie występują błędy transmisji, zatem jedyną przyczyną utraty ramki mogą być kolizje,
- nie występuje efekt przechwytywania, zatem żadna ramka – spośród biorących udział w kolizji – nie zostanie odebrana poprawnie.

Dla uproszczenia wzorów używa się znormalizowanej intensywności wytwarzania ramek ($G = gT$), określającej liczbę ramek (g) wytworzonych w całej sieci w czasie przesyłania pojedynczej ramki (T).

Model sieci przyjęty dla opisanych dalej rozważań zakłada istnienie jednego odbiornika, do którego są kierowane ramki z pozostałych stacji. Jest to typowa sytuacja w sieciach przewodowych, w których występuje pojedynczy kanał komunikacyjny współdzielony przez wszystkie stacje, niezależnie od tego, która z nich jest nadawcą, a która odbiorcą ramki. W sieciach bezprzewodowych, szczególnie zdecentralizowanych (np. w sieciach ad hoc), można prowadzić wiele jednoczesnych transmisji bez kolizji. Jest to możliwe, gdy adresaci tych ramek znajdują się w zasięgu ich nadawców, ale poza zasięgiem innych nadajników (adresaci są zatem stacjami ukrytymi przed nadajnikami innymi niż nadawcy). Spostrzeżenie to jest szczególnie istotne podczas określania bezkolizyjnego harmonogramu transmisji w sieciach ad hoc wykorzystujących protokoły rodziny TDMA, np. [78], ponieważ dzięki temu można znacząco podnieść przepustowość sieci, nawet do kilkuset procent. Obliczeń uzyskanych w takiej sieci nie można by jednak porównać z analogicznymi wynikami dla sieci przewodowych, w których tylko jedna stacja może prowadzić w danej chwili bezkolizyjną transmisję. Jedyny właściwy model sieci bezprzewodowej zakłada zatem sieć scentralizowaną, taką jak pokazaną na rys. 1.42.



Rys. 1.42. Model sieci przyjęty dla rozważań
Fig. 1.42. The network model accepted for considerations

1.5.1.1. Protokoły rodziny Aloha

Ponieważ protokoły rodziny Aloha nie wykorzystują żadnej z opisanych w rozdziale 1.3 metod unikania kolizji, opóźnienie propagacyjne nie wpływa na ich wydajność. Stopień wykorzystania kanału można określić jako [65]:

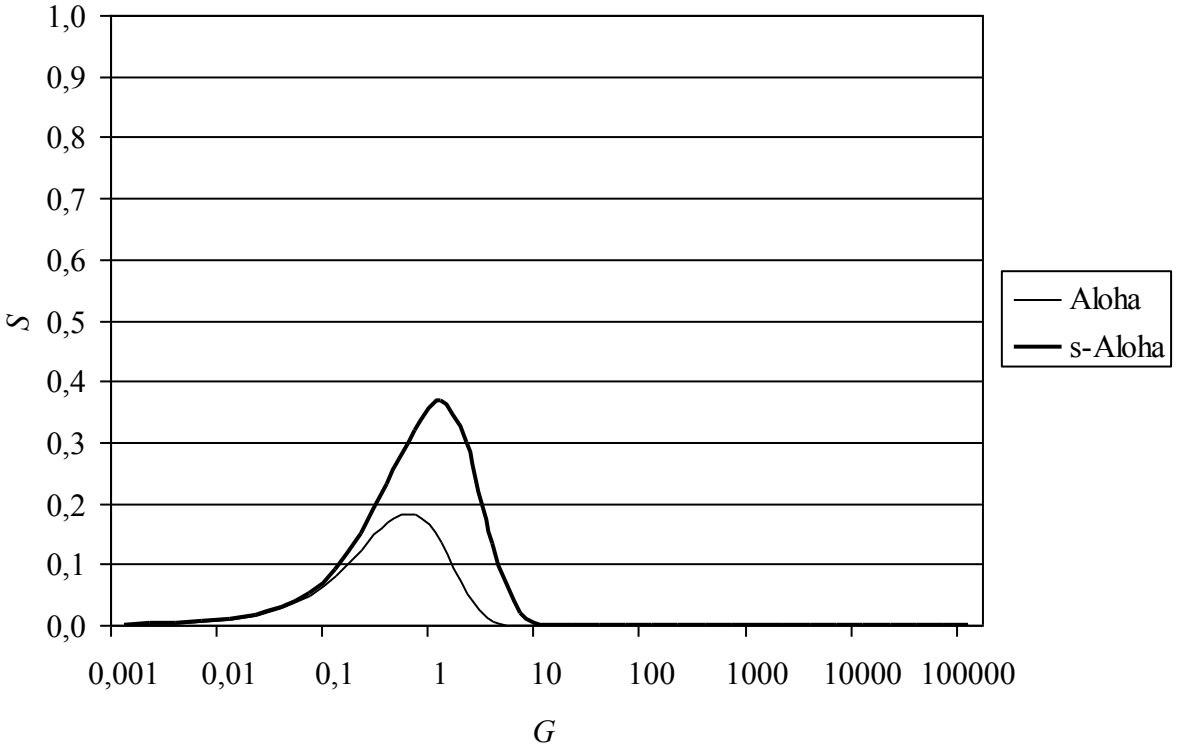
$$S_{\text{Aloha}} = Ge^{-2G} \tag{1.6}$$

dla podstawowej wersji protokołu oraz

$$S_{\text{s-Aloha}} = Ge^{-G} \tag{1.7}$$

dla wersji szczelinowej. Ze względu na częste kolizje, nawet przy małym obciążeniu sieci, wykorzystanie kanału nie przekracza około 18,5% dla protokołu Aloha i około 37% dla wersji szczelinowej. Powyższe zależności zilustrowano na rys. 1.43.

Jak widać na wykresie, niskie wykorzystanie kanału nie jest jedyną wadą protokołów rodziny Aloha. Drugą – nie mniej istotną – jest ryzyko niestabilności protokołu. Najwyższa przepustowość występuje dla $G = 0{,}5$ w przypadku protokołu Aloha i dla $G = 1$ w przypadku wersji szczelinowej. Biorąc pod uwagę poprzednie rozważania, nietrudno zauważyć, iż warunki te mogą bardzo łatwo wystąpić. Przykładowo, $G = 0{,}5$ oznacza, że jedna nowa ramka zostaje wytworzona w czasie transmisji aż dwóch ramek, jest to zatem stosunkowo niewielkie obciążenie sieci. Po przekroczeniu tej wartości zwiększona liczba kolizji znacząco zmniejsza przepustowość. Sytuacja taka utrzymuje się do chwili, gdy obciążenie spadnie poniżej wartości, dla której protokół osiąga najwyższą wydajność. Jest zatem oczywiste, że oba protokoły są bardzo czułe na zmiany obciążenia łącza, a ich stabilność w typowych warunkach działania sieci jest niewystarczająca.



Rys. 1.43. Wydajność protokołów rodziny Aloha
Fig. 1.43. The efficiency of Aloha-family protocols

1.5.1.2. Protokoły rodziny CSMA

Protokoły rodziny CSMA wprowadzono w celu zmniejszenia prawdopodobieństwa kolizji ramek przez poprzedzający rozpoczęcie nadawania nasłuch łącza. W przypadku zajętości łącza transmisja powinna być odłożona na później. Żadna stacja nie potrafi jednak wykryć nośnej natychmiast, a to ze względu na właściwości układów transmisyjnych oraz opóźnienie propagacyjne. Wydajność protokołów rodziny CSMA zależy zatem od niektórych parametrów transmisji, jak odległość i długość ramki. Zależność ta jest opisana następująco:

$$a = \frac{\tau}{\delta} = \frac{D_{\max} / c}{l_d / v} = \frac{v \cdot D_{\max}}{c \cdot l_d}, \tag{1.8}$$

gdzie: δ – czas transmisji ramki [s], τ – opóźnienie propagacyjne w kanale [s], v – prędkość transmisji [b/s], $D_{\max}$ – zasięg transmisji [m], l_d – długość ramki [b], zaś c – prędkość propagacji sygnału [m/s]. Prędkość propagacji sygnału zazwyczaj uznaje się za równą prędkości światła, choć jest to prawda tylko dla propagacji wolnoprzestrzennej. Obliczoną wartość parametru a można interpretować jako zależność między czasem transmisji ramki a czasem potrzebnym na wykrycie kolizji przez wszystkie stacje. Parametr ten określa, jak duża część ramki zostanie wysłana, zanim wszystkie stacje wykryją zajętość kanału (innymi słowy – jak duża część ramki nie jest chroniona przez mechanizm wykrywania kolizji). Wartości parametru mogą zmieniać się w zakresie <0; 1>. Wartość $a = 0$ oznacza optymalne warunki – brak

opóźnienia propagacyjnego, cała ramka chroniona przez mechanizm wykrywania nośnej. Wartość $a = 1$ oznacza najgorsze warunki – całkowity brak ochrony ramki. Wartość $a = 1$ przyjmuje się także dla sieci zawierających stacje ukryte, ponieważ mechanizm wykrywania nośnej nie jest wówczas skuteczny. Parametr ten jest istotny dla wydajności protokołu, ponieważ im większa część ramki jest chroniona przez mechanizm wykrywania nośnej, tym mniejsze jest prawdopodobieństwo kolizji i – co za tym idzie – wyższa wydajność.

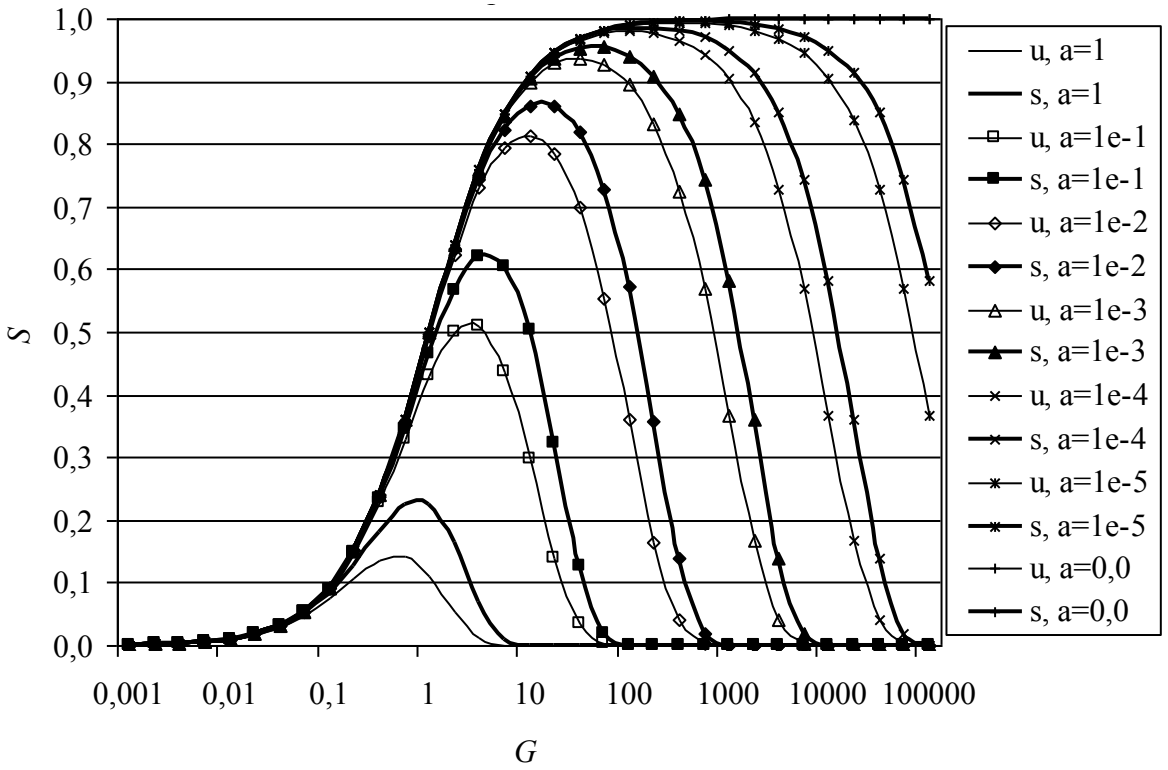
Dla nietrwałego protokołu CSMA stopień wykorzystania kanału można określić następująco [65]:

$$S_{\text{CSMA}} = \frac{Ge^{-aG}}{G(1+2a)+e^{-aG}}, \tag{1.9}$$

natomiast dla wersji szczelinowej

$$S_{\text{s-CSMA}} = \frac{aGe^{-aG}}{1+a-e^{-aG}}. \tag{1.10}$$

Zależności te, dla kilku wartości parametru *a*, ilustruje rys. 1.44.



Rys. 1.44. Wydajność nieszczelinowego (u) i szczelinowego (s) nietrwałego protokołu CSMA
Fig. 1.44. The efficiency of unslotted (u) and slotted (s) nonpersistent CSMA protocol

Jak widać na wykresie, dla małych wartości parametru *a* protokół uzyskuje znacznie większą wydajność niż dla wartości wyższych. Jest to zgodne z oczekiwaniami, ponieważ im mniejsza wartość parametru *a*, tym większa część ramki jest chroniona przez mechanizm wykrywania nośnej. W obecności ukrytych stacji – tj. dla $a = 1$ – stopień wykorzystania kana-

łu spada do około 15÷25%, jest zatem porównywalny z osiągami protokołu Aloha. Jest to zrozumiałe, ponieważ obecność stacji ukrytych uniemożliwia skuteczne wykrywanie nośnej, więc w rzeczywistości CSMA ulega degradacji do protokołu Aloha. Co więcej, czas poświęcony na wykrywanie nośnej jest wówczas stracony, tak więc osiągi CSMA są nawet gorsze niż protokołu Aloha, w którym w ogóle nie wykrywa się nośnej.

Pomimo niskiej wydajności nietrwałego CSMA w obecności stacji ukrytych, protokół ten ujawnia swe zalety przy małych wartościach a. Maksymalne wykorzystanie kanału osiąga 80÷90% przy zapewnieniu wystarczającej stabilności. Inna ciekawa właściwość jest taka, że przy $a = 0$ (tj. gdy wszystkie stacje wykrywają nośną natychmiast) i wyjątkowo dużym obciążeniu sieci (np. $G \geq 100$) protokół osiąga praktycznie stuprocentowe wykorzystanie kanału i nigdy nie traci stabilności.

Protokół CSMA w wersji trwałej z prawdopodobieństwem 1 wprowadzono w celu podniesienia wydajności przez eliminację okresów bezczynności, powstałych w wyniku odkładania kolejnej próby transmisji na później, jak w wariancie nietrwałym. Eliminację taką można uzyskać przez wymuszenie transmisji natychmiast po zwolnieniu łącza. Stopień wykorzystania kanału dla protokołu trwałego z prawdopodobieństwem 1 można określić jako [65]:

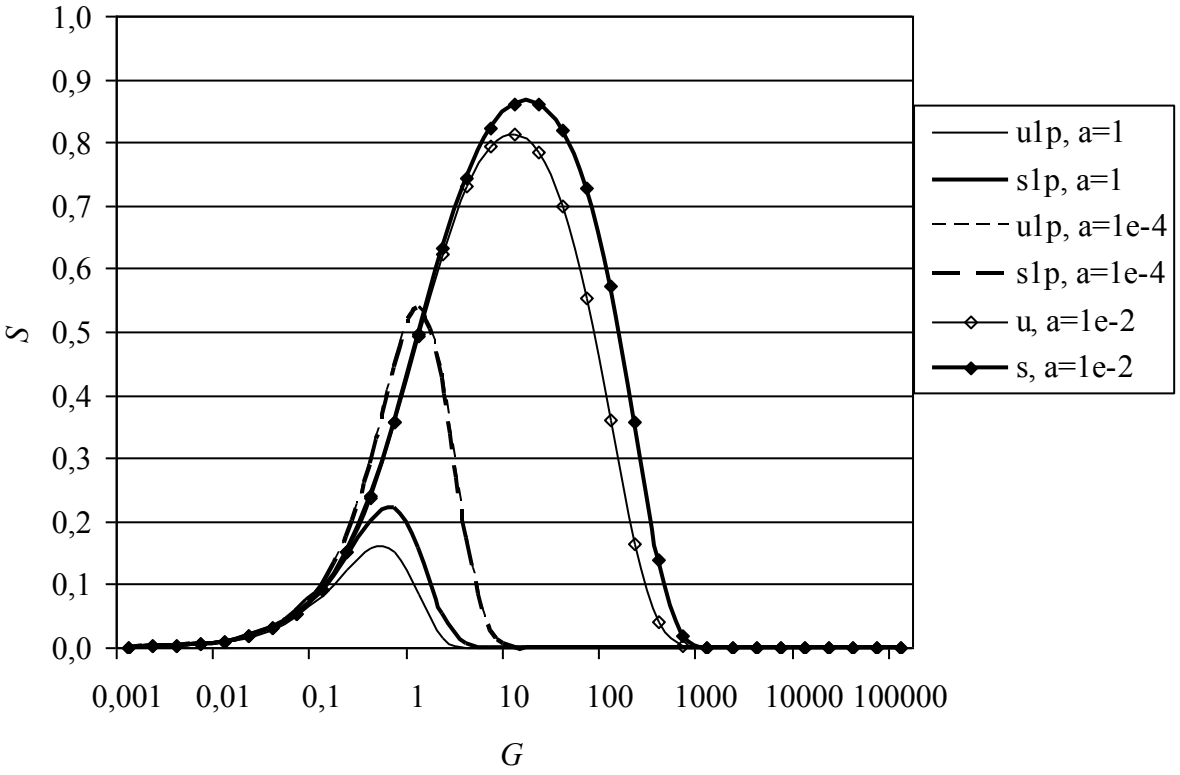
$$S_{1-p\text{CSMA}} = \frac{Ge^{-(2a+1)G}\left(1+G+\frac{aG}{2}\right)}{G(1+2a)-(1-e^{-aG})+(1+aG)e^{-aG}}, \tag{1.11}$$

natomiast dla wersji szczelinowej:

$$S_{\text{s}-1-p\text{CSMA}} = \frac{Ge^{-(1+a)G}(1+a-e^{-aG})}{(1+a)(1-e^{-aG})+ae^{-(1+a)G}}. \tag{1.12}$$

Powyższe zależności, dla kilku wartości parametru a, zilustrowano na rys. 1.45.

Zarówno w wersji szczelinowej, jak i nieszczelinowej wydajność trwałego protokołu CSMA znaczniej mniej, niż w przypadku protokołu nietrwałego, zależy od wartości parametru a i dla $a \leq 0{,}1$ osiąga wartość maksymalną – około 55%. Mniejsza jest także różnica między odmianą szczelinową i nieszczelinową protokołu. Tym niemniej, w obecności stacji ukrytych stopień wykorzystania kanału dalej jest niewielki. Stabilność protokołu także nie jest zadowalająca, gdyż – nawet w najlepszych warunkach – jest porównywalna ze stabilnością protokołu Aloha. W przeciwieństwie do nietrwałego CSMA, nawet dla $a = 0$ i przy bardzo wysokich obciążeniach, przepustowość łącza spada do zera. Wynika to z kolizji występujących tuż po zwolnieniu łącza, gdy więcej niż jedna stacja czeka na możliwość wysłania ramki. Tym niemniej, przy małych obciążeniach łącza, tj. dla $G \leq 1$, stopień wykorzystania kanału jest o około 10% wyższy niż dla protokołu nietrwałego. Wynika to z eliminacji okresów bezczynności, co przy małej intensywności ruchu w sieci przynosi poprawę wydajności. Podsumowując, protokół ten jedynie częściowo spełnia oczekiwania.



Rys. 1.45. Porównanie wydajności protokołów CSMA: nietrwałego (u), nietrwałego szczelinowego (s), 1-trwałego (u1p) i 1-trwałego szczelinowego (s1p)
Fig. 1.45. A comparison of CSMA protocols efficiency: nonpersistent (u), nonpersistent slotted (s), 1-persistent (u1p) and 1-persistent slotted (s1p)

Protokół CSMA w wersji trwałej z dowolnym prawdopodobieństwem transmisji wprowadzono w celu zmniejszenia ryzyka wystąpienia kolizji tuż po zwolnieniu łącza, co było główną bolączką protokołu trwałego z prawdopodobieństwem 1. Zmniejszenie prawdopodobieństwa takiej kolizji można uzyskać przez zmniejszenie prawdopodobieństwa transmisji bezpośrednio po zwolnieniu łącza. Pozwala to zwiększyć przepustowość przy wysokim obciążeniu łącza bez obniżania jej przy niskim. Stopień wykorzystania kanału dla trwałego protokołu z prawdopodobieństwem p (dla $p<1$) można wyrazić jako [65]:

$$S_{p-\mathrm{CSMA}} = \frac{(1-e^{-aG})\cdot\left[P'\pi_0 + P(1-\pi_0)\right]}{(1-e^{-aG})\cdot\left[a\bar{t}'\pi_0 + a\bar{t}(1-\pi_0)+1+a\right]+a\pi_0}. \tag{1.13}$$

Wartości P, P' można obliczyć nastepująco:

$$P = \sum_{n=1}^{\infty} P(n)\frac{\pi_n}{1-\pi_0}, \quad P' = \sum_{n=1}^{\infty} P(n)\pi'_n, \tag{1.14}$$

gdzie

$$P(n) = \sum_{l=n}^{\infty} \frac{lpq^{l-1}}{1-q^l}\left[\sum_{k=1}^{\infty}\frac{(kaG)^{l-n}}{(l-n)!}q^{kn}\left[1-q^n e^{-aG(1-q^k)}\right]\cdot e^{\left(\frac{q(1-q^{k-1})}{p}\right)aG} + (1-q^n)\delta_{l,n}\right] \tag{1.15}$$

oraz

$$\begin{aligned} \pi_n &= \frac{[(1+a)G]^n}{n!} \cdot e^{-(1+a)G} \quad (n \geq 0), \\ \pi'_n &= \frac{(aG)^n e^{-aG}}{(1-e^{-aG})n!} \quad (n \geq 1). \end{aligned} \tag{1.16}$$

Wartości $\bar{t}, \bar{t'}$ można obliczyć następująco:

$$\bar{t} = \sum_{n=1}^{\infty} \bar{t}_n \frac{\pi_n}{1-\pi_0}, \quad \bar{t'} = \sum_{n=1}^{\infty} \bar{t}_n \pi'_n, \tag{1.17}$$

gdzie

$$\bar{t}_n = \sum_{k=0}^{\infty} q^{(k+1)n} \cdot e^{\left(\frac{q(1-qk)}{p} - k\right)aG}. \tag{1.18}$$

W powyższych wzorach przyjęto, że $q = 1-p,$ zaś $\delta_{l,n}$ jest deltą Kroneckera.

Jeśli prawdopodobieństwo transmisji $p \leq 0{,}1,$ konieczne wartości można obliczyć w przybliżeniu w znacznie prostszy sposób [65]. W tym przypadku

$$P = \frac{\pi_0^p - \pi_0}{q(1-\pi_0)} - \frac{(1-e^{-Gap})(\pi_0^{1-p^2} - \pi_0)}{q(1-\pi_0) - qe^{-2Gap}(\pi_0^p - \pi_0)} \tag{1.19}$$

oraz

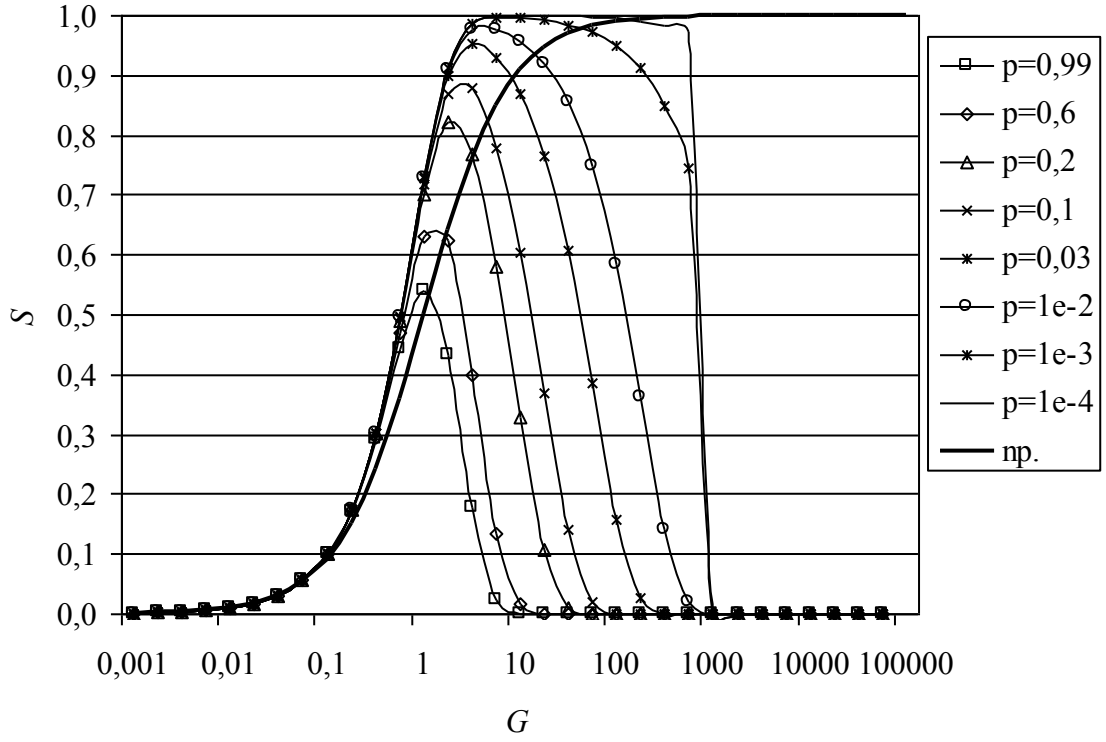
$$\bar{t} = \frac{\pi_0^p - \pi_0}{1-\pi_0 - (\pi_0^p - \pi_0)e^{-Gap}}. \tag{1.20}$$

Wartości $P'', \bar{t}$ nadal można obliczyć według wzorów (1.14) i (1.17), przyjmując dla przypadku uproszczonego, że $\pi_0 = e^{-aG}$. Pomimo że – według [65] – przybliżone obliczenia dają zadowalającą dokładność tylko dla $p \leq 0{,}1,$ można na ich podstawie uzyskać wykresy z wystarczającą dokładnością – różnica nie jest widoczna.

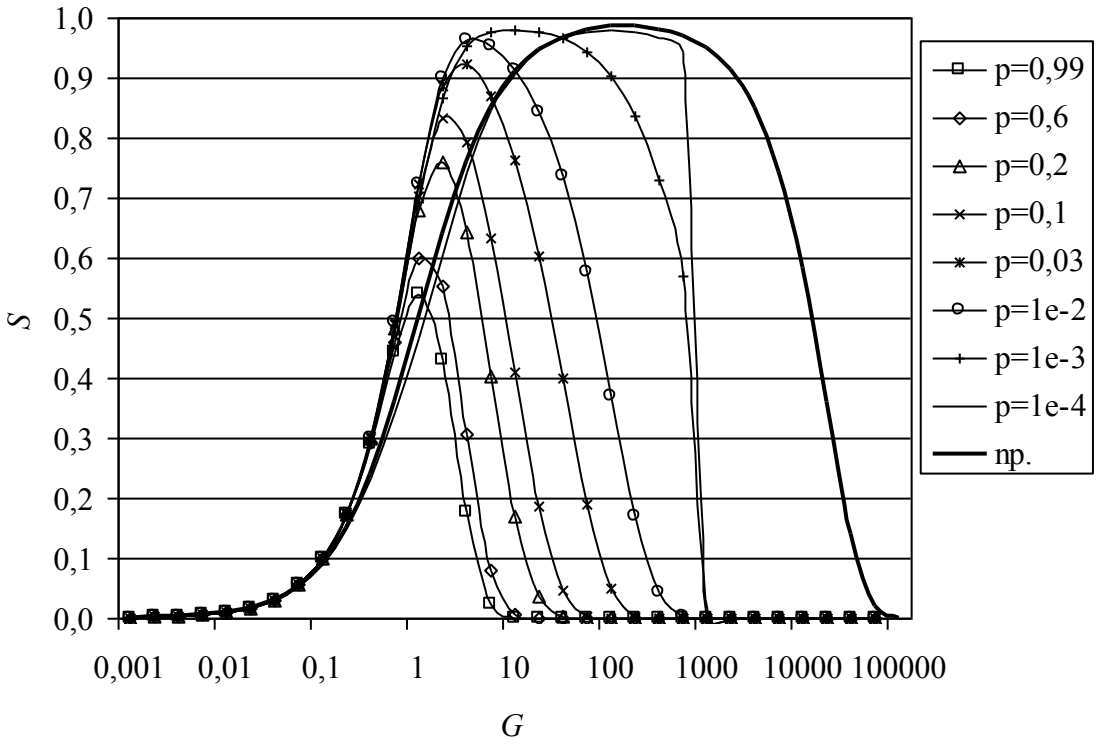
Dla $a = 0$ obliczenia znacznie się upraszczają. Stopień wykorzystania łącza można zatem wyrazić jako [65]:

$$S_{p-\mathrm{CSMA}(a=0)} = \frac{Ge^{-G}\left[1 + pG\sum_{k=0}^{\infty} \frac{((1-p)G)^k}{(1-(1-p)^{k+1})k!}\right]}{G + e^{-G}}. \tag{1.21}$$

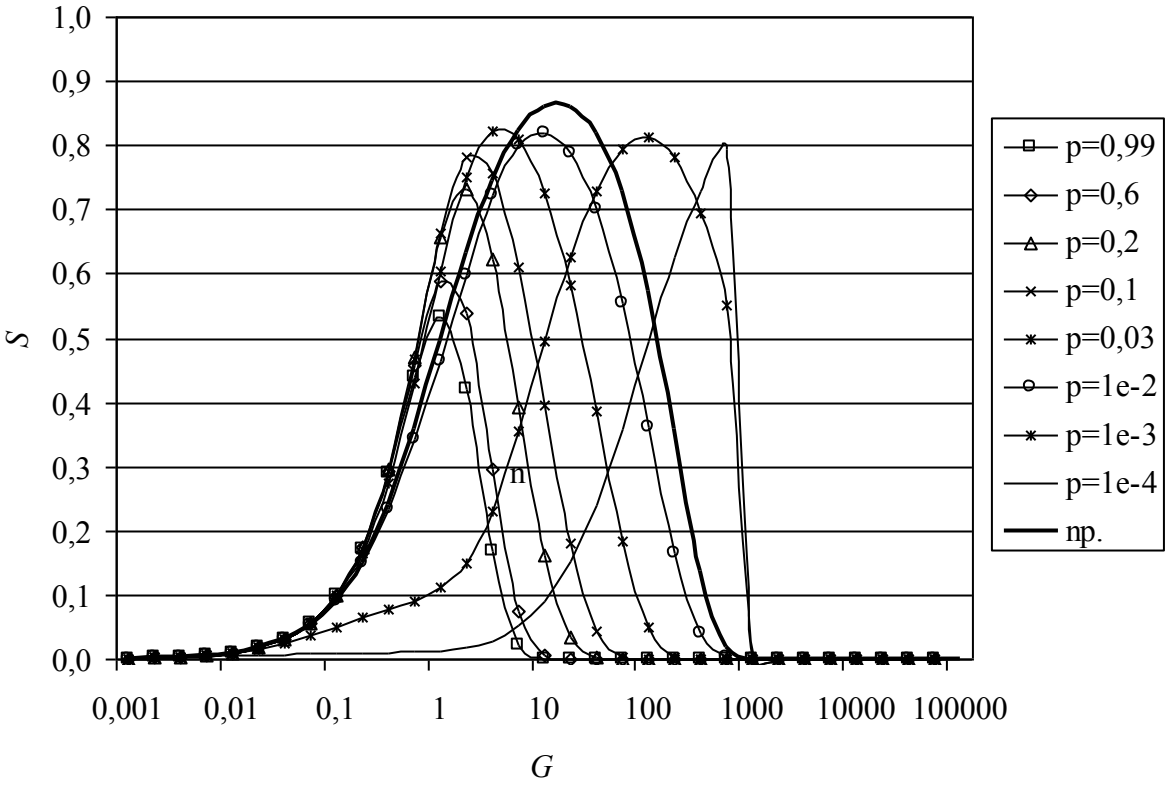
Wykresy wydajności trwałego protokołu CSMA dla różnych wartości parametrów p i a przedstawiono na rys. 1.46÷1.50.



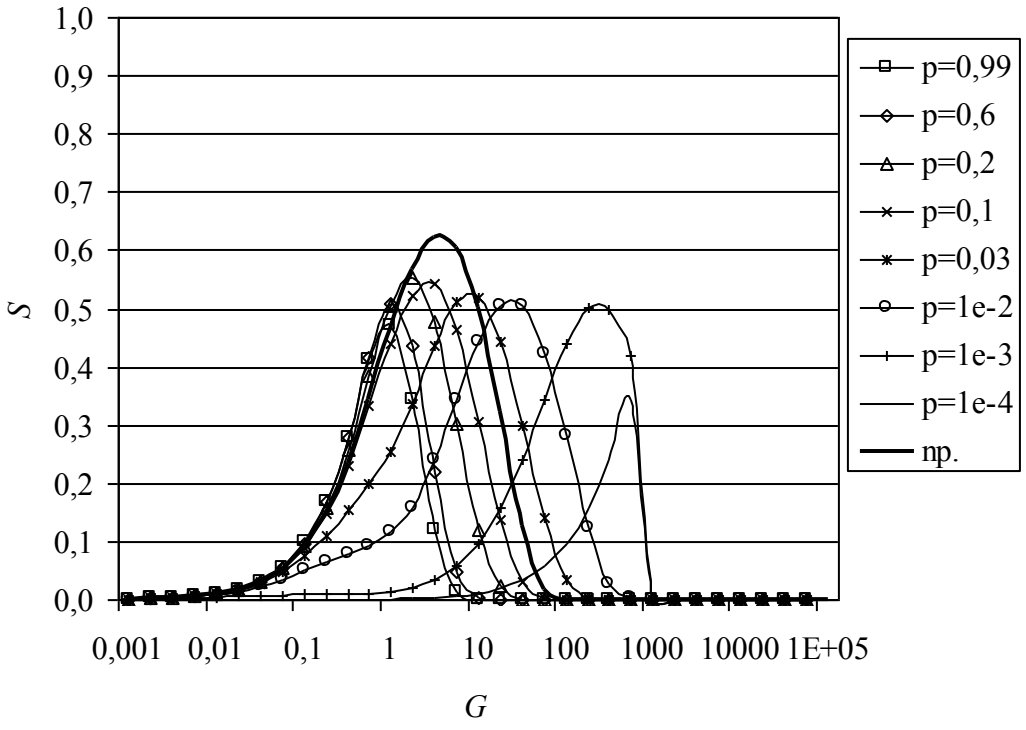
Rys. 1.46. Porównanie *p*-trwałego i nietrwałego (np.) protokołu CSMA dla *a*=0
Fig. 1.46. A comparison of *p*-persistent and nonpersistent (np.) CSMA for *a*=0



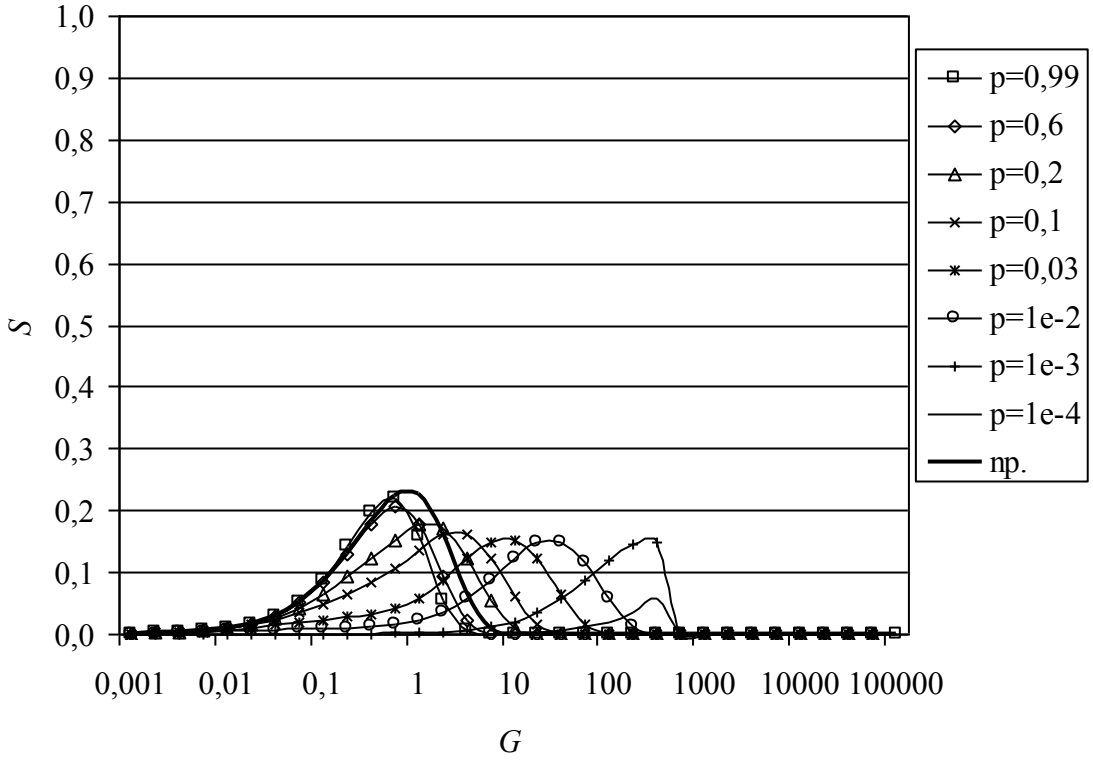
Rys. 1.47. Porównanie *p*-trwałego i nietrwałego (np.) protokołu CSMA dla *a*=0,0001
Fig. 1.47. A comparison of *p*-persistent and nonpersistent (np.) CSMA for *a*=0.0001



Rys. 1.48. Porównanie *p*-trwałego i nietrwałego (np.) protokołu CSMA dla *a*=0,01
Fig. 1.48. A comparison of *p*-persistent and nonpersistent (np.) CSMA for *a*=0.01



Rys. 1.49. Porównanie *p*-trwałego i nietrwałego (np.) protokołu CSMA dla *a*=0,1
Fig. 1.49. A comparison of *p*-persistent and nonpersistent (np.) CSMA for *a*=0.1



Rys. 1.50. Porównanie *p*-trwałego i nietrwałego (np.) protokołu CSMA dla *a*=1,0
Fig. 1.50. A comparison of *p*-persistent and nonpersistent (np.) CSMA for *a*=1.0

Jak widać na przedstawionych rysunkach, stopień wykorzystania łącza dla *p*-trwałego protokołu CSMA zależy nie tylko od wartości *a*, lecz także od prawdopodobieństwa transmisji *p*. Dla każdej sieci można zatem wyznaczyć optymalną wartość *p*, zapewniającą największy stopień wykorzystania kanału w określonych warunkach. Ogólnie, mniejsze wartości *p* pozwalają na uzyskanie wyższej przepustowości, szczególnie przy małych wartościach *a*, jednak przy $a = 1$, najlepsze wyniki uzyskuje się dla $p = 0{,}99$ (praktycznie jest to przypadek protokołu 1-trwałego). Z wykresów można wywnioskować, że najlepsze osiągi protokół uzyskuje dla $p \approx a$ i wówczas są one porównywalne z wynikami nietrwałego protokołu CSMA. Wraz ze zmniejszaniem wartości *a* uwidacznia się przewaga protokołu *p*-trwałego nad nietrwałym dla niskiego obciążenia sieci ($G \leq 1$), dając w niektórych przypadkach wzrost wydajności o 25%. Przy małym obciążeniu wyższą wydajność uzyskuje się dla większych wartości *p*. Wraz ze zwiększaniem obciążenia maksymalna przepustowość dla danej wartości *a* przesuwa się w stronę mniejszych wartości *p*. Wskazane byłoby zatem dynamiczne dostosowywanie prawdopodobieństwa transmisji *p* do chwilowego obciążenia sieci. Możliwość taka byłaby szczególnie korzystna w sieciach ad hoc, które zmieniają parametry znacznie częściej niż inne typy sieci. Wydaje się jednak, iż odpowiedni mechanizm jest trudny do realizacji w praktyce.

1.5.1.3. Protokoły rodziny CSMA/CD

Głównym celem protokołów z wykrywaniem kolizji jest skrócenie okresu zajętości łącza po wystąpieniu kolizji przez porzucenie transmisji ramki w tej sytuacji. Jak pokazano na przykładzie trwałego protokołu CSMA, skracanie okresu bezczynności nie przynosi oczekiwanych efektów.

Chociaż protokoły rodziny CSMA/CD nie mogą być używane w większości bezprzewodowych sieci komputerowych z powodu niemożności wykrywania kolizji w tych sieciach [144], warto oszacować ich osiągi w określonych warunkach. Pozwala to bowiem na określenie pewnego rodzaju kosztu używania bezprzewodowych mediów transmisyjnych.

W protokołach rodziny CSMA/CD po wykryciu kolizji stacje zagłuszają przez pewien czas łącze, aby poinformować inne stacje o zaistniałym konflikcie. Konieczne jest zatem wprowadzenie parametru, opisującego czas zajętości kanału w przypadku kolizji. Czas ten jest równy $\gamma = 2aT + t_{CD} + t_J$, gdzie t_{CD} – czas niezbędny na wykrycie kolizji, podczas gdy t_J – czas zagłuszania łącza po wykryciu kolizji. Zależność między γ i T jest określona jako $\gamma' = \gamma / T$ i wskazuje, jak duża część ramki jest stracona w przypadku kolizji. Wraz ze zmniejszaniem wartości γ' wydajność protokołu rośnie. Parametr ten można określić jako współczynnik skrócenia okresu zajętości łącza w przypadku kolizji w porównaniu z protokołem CSMA bez wykrywania kolizji. Dla $\gamma' = 1$ protokół CSMA/CD zachowuje się zatem jak CSMA – pomimo wykrycia kolizji ramka przesyłana jest w całości.

Stopień wykorzystania łącza nietrwałej (bez wymuszania transmisji) odmiany protokołu CSMA/CD można określić jako [88, 100]:

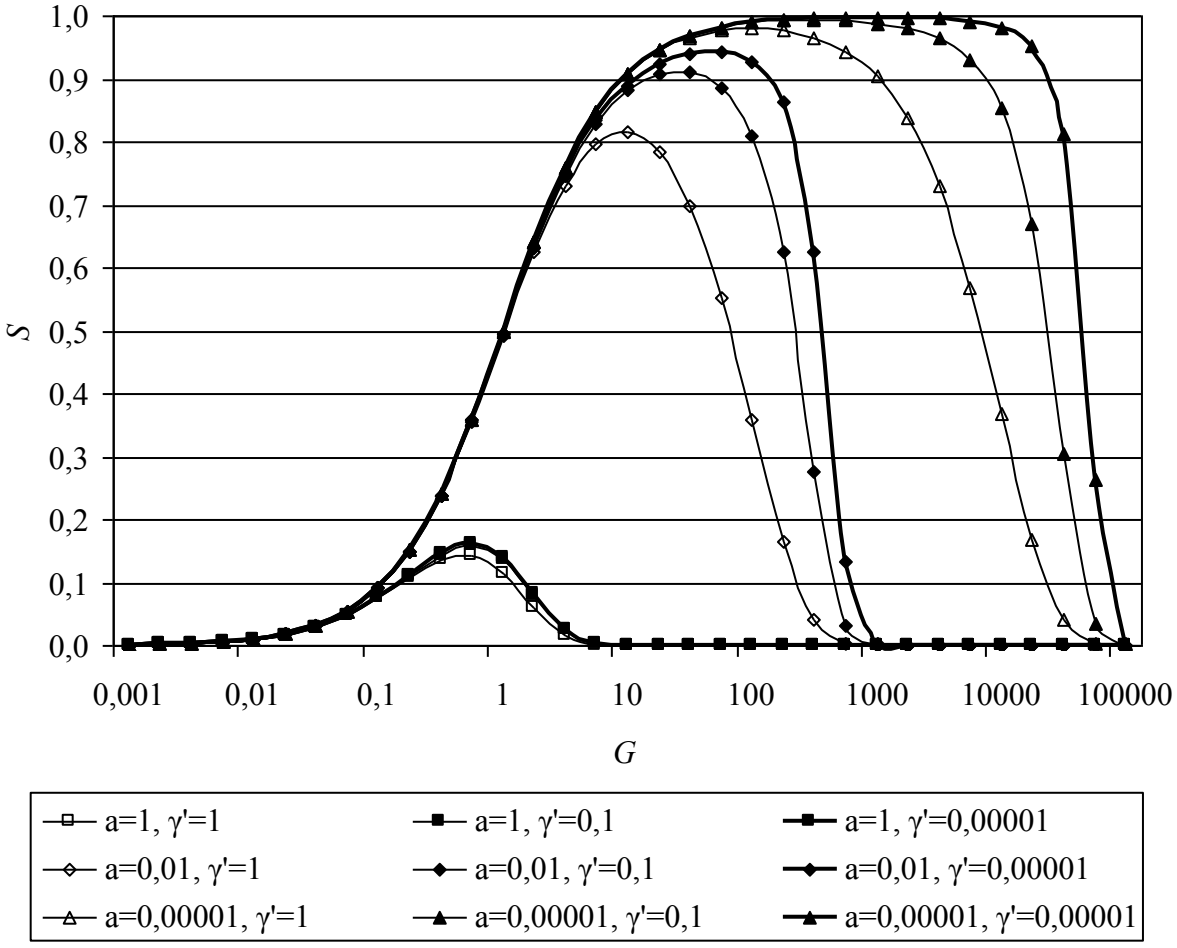
$$S_{CSMA/CD} = \frac{Ge^{-aG}}{2 + (G-1)e^{-aG} + (a+\gamma')G(1-e^{-aG})}, \tag{1.22}$$

natomiast dla wersji szczelinowej:

$$S_{s-CSMA/CD} = \frac{aGe^{-aG}}{aGe^{-aG} + (1-e^{-aG} - aGe^{-aG})\gamma' + a}. \tag{1.23}$$

Wydajność protokołu dla różnych wartości parametru *a* zilustrowano na rys. 1.51 (wersja nieszczelinowa) i rys. 1.52 (wersja szczelinowa). Można na nich zauważyć, że dla $\gamma' = 1$ protokół CSMA/CD ulega degradacji do „zwykłego” CSMA. Uwaga ta dotyczy zarówno szczelinowych, jak i nieszczelinowych odmian obu protokołów. Jest to zachowanie zgodne z oczekiwaniami, ponieważ $\gamma' = 1$ oznacza, że wykrycie kolizji zajmuje cały czas transmisji ramki, a więc nie występuje skrócenie okresu zajętości łącza w przypadku kolizji.

Stopień wykorzystania łącza szczelinowego protokołu CSMA/CD z wymuszaniem transmisji z prawdopodobieństwem 1 można wyprowadzić, korzystając z wyjaśnień zamieszczonych w [88]. W wyniku odpowiednich przekształceń można otrzymać następujące zależności:



Rys. 1.51. Wydajność nieszczelinowego, nietrwałego protokołu CSMA/CD
Fig. 1.51. Efficiency of unslotted nonpersistent CSMA/CD protocol

$$S_{s-1p-CSMA/CD} = \frac{U(\tau)}{B(\tau)\cdot I(\tau)}, \tag{1.24}$$

gdzie:

$$\begin{aligned} I(x) &= \frac{t}{1-a_0(x)}, \\ B(x) &= \frac{a_1(x)}{1-a_0(x)}\left[\tau_T + \left(1-a_0(\tau_T)\right)B(\tau_T)\right] + \left[1-\frac{a_1(x)}{1-a_0(x)}\right]\left[\tau_\gamma + \left(1-a_0(\tau_\gamma)\right)B(\tau_\gamma)\right], \\ U(x) &= \frac{a_1(x)}{1-a_0(x)}\left[T + \left(1-a_0(\tau_T)\right)U(\tau_T)\right] + \left[1-\frac{a_1(x)}{1-a_0(x)}\right]\left[\left(1-a_0(\tau_\gamma)\right)U(\tau_\gamma)\right] \end{aligned} \tag{1.25}$$

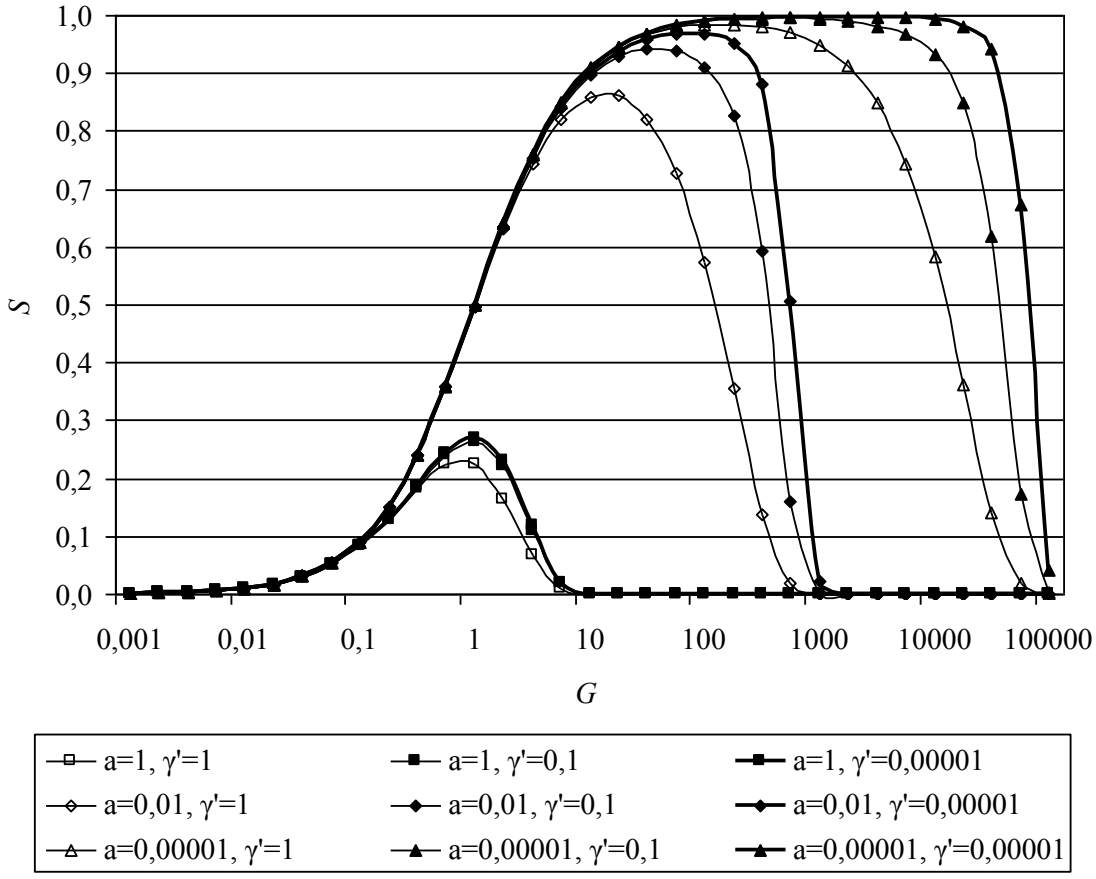
W powyższych wzorach przyjmuje się, że

$$\begin{aligned} B(\tau_T) &= \frac{\left(1-a_0(\tau_T)-a_1(\tau_T)\right)\left(2-a_0(\tau_\gamma)-a_1(\tau_\gamma)\right)\tau_\gamma + \left[a_1(\tau_T)+a_1(\tau_\gamma)\left(1-a_0(\tau_T)-a_1(\tau_T)\right)\right]\tau_T}{\left(1-a_0(\tau_T)\right)\left[a_0(\tau_\gamma)\left(1-a_1(\tau_T)\right)+a_0(\tau_T)a_1(\tau_\gamma)\right]}, \\ B(\tau_\gamma) &= \frac{\left(a_1(\tau_T)+1\right)a_1(\tau_\gamma)\tau_T + \left(1-a_0(\tau_T)-a_1(\tau_T)\right)\left(a_1(\tau_\gamma)+1\right)\tau_\gamma}{\left(1-a_0(\tau_\gamma)\right)\left[a_0(\tau_\gamma)\left(1-a_1(\tau_T)\right)+a_0(\tau_T)a_1(\tau_\gamma)\right]} \end{aligned} \tag{1.26}$$

oraz

$$U(\tau_T) = \frac{\left[a_1(\tau_\gamma)\left(1-a_0(\tau_T)\right)+a_1(\tau_T)a_0(\tau_\gamma)\right]T}{\left(1-a_0(\tau_T)\right)\left[a_0(\tau_\gamma)\left(1-a_1(\tau_T)\right)+a_0(\tau_T)a_1(\tau_\gamma)\right]},$$
$$U(\tau_\gamma) = \frac{a_1(\tau_\gamma)T}{\left(1-a_0(\tau_\gamma)\right)\left[a_0(\tau_\gamma)\left(1-a_1(\tau_T)\right)+a_0(\tau_T)a_1(\tau_\gamma)\right]}. \tag{1.27}$$

gdzie $a_i(x) = \dfrac{\left(\frac{Gx}{T}\right)^i e^{\left(-\frac{Gx}{T}\right)}}{i!}$ $(i = 0, 1)$, $\tau = aT$, $\tau_T = \tau + T = (a+1)T$, $\tau_\gamma = \tau + \gamma = (a+\gamma')T$.

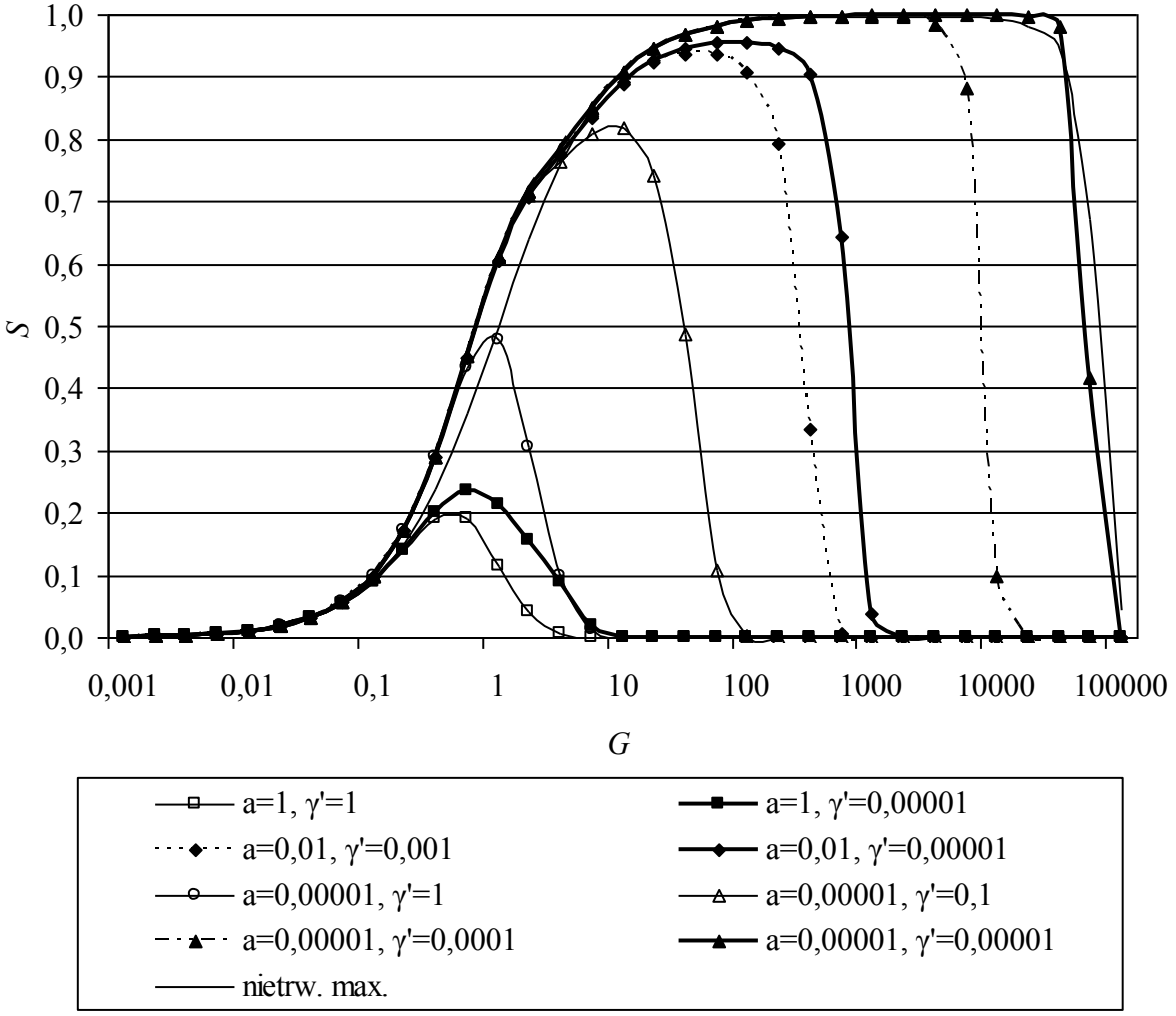


Rys. 1.52. Wydajność szczelinowego, nietrwałego protokołu CSMA/CD
Fig. 1.52. Efficiency of slotted nonpersistent CSMA/CD

Wydajność szczelinowego protokołu CSMA/CD z wymuszaniem transmisji zilustrowano na rys. 1.53.

Na przedstawionych wykresach (rys. 1.51÷1.53) widać, że różnice między poszczególnymi odmianami protokołu CSMA/CD nie są tak wielkie jak w przypadku CSMA. Zakładając – niezgodnie z prawdą – że protokół CSMA/CD w formie znanej z sieci przewodowych mógłby być użyty w typowej sieci bezprzewodowej, można zauważyć wysoką jego odporność na wpływ stacji ukrytych. Widać to szczególnie na rys. 1.52 dla szczelinowej, nietrwałej odmiany CSMA/CD. Nawet gdy $a = 1$, protokół osiąga wyższy stopień wykorzystania kanału niż protokół Aloha, co było nieosiągalne dla wszystkich odmian CSMA w opisywanych warun-

kach. CSMA/CD wykazuje także wysoką stabilność, nawet dla dużych – i nieczęsto spotykanych w praktyce – wartości parametru *a*. Z kolei dla niskich wartości *a* protokół osiąga wysoką wydajność bliską 100% oraz bardzo dobrą stabilność nawet przy bardzo dużym obciążeniu łącza. Szczelinowa, trwała odmiana protokołu jest także wydajniejsza od nietrwałej przy niskim obciążeniu łącza $(G \le 1)$. W przeciwieństwie także do CSMA, dla którego nadmierne zwiększanie prawdopodobieństwa transmisji zmniejsza wydajność przy dużym obciążeniu łącza, trwały protokół CSMA/CD w tych warunkach pozostaje porównywalny z nietrwałym. Żaden inny protokół spośród omawianych nie uzyskuje takich osiągów.



Rys. 1.53. Porównanie szczelinowego trwałego i nietrwałego protokołu CSMA/CD
Fig. 1.53. Comparison of slotted 1-persistent and nonpersistent CSMA/CD protocols

Chociaż protokół CSMA/CD w postaci znanej z sieci przewodowych nie może być używany w większości typowych sieci bezprzewodowych, znane są rozwiązania, które pozwalają na wykrywanie kolizji także i w tych sieciach [68, 69, 87]. Mogą one być także użyte w kanałach dwukierunkowych naprzemiennych. Ich wydajność – między innymi ze względu na możliwość wystąpienia niewykrytych kolizji oraz konieczność chwilowego przerywania

transmisji, co zwiększa narzut protokołu szczególnie w sieciach radiowych z powodu relatywnie długiego czasu przełączania nadajnika-odbiornika radiowego – jest jednak niższa niż typowego protokołu CSMA/CD.

1.5.1.4. Protokół MACA

Protokoły rodziny CSMA i CSMA/CD, wystarczająco wydajne w sieciach przewodowych i niektórych bezprzewodowych, pracują nieefektywnie w obecności stacji ukrytych lub odkrytych. Z tego powodu w niektórych protokołach dla sieci bezprzewodowych wykrywanie nośnej zastąpiono wymianą ramek sterujących poprzedzającą transmisję danych. Pierwszym takim protokołem jest MACA.

Ze względu na użycie ramek sterujących wydajność protokołów wykorzystujących ten mechanizm jest uzależniona od stosunku czasu transmisji ramki sterującej i ramki danych. Zależność tę można opisać następująco:

$$b = \frac{\gamma}{\delta}, \tag{1.28}$$

gdzie γ i δ oznaczają odpowiednio czasy transmisji ramek sterujących i danych [s]. Wartość obliczonego w ten sposób parametru b mieści się zazwyczaj w przedziale $<10^{-1}; 10^{-4}>$; zmniejszenie jej podnosi wydajność protokołu przez zmniejszenie jego narzutu. Ponieważ wartość b zależy od właściwości protokołu, w niektórych sieciach może przekroczyć podane granice (przykładowe wartości parametrów a oraz b dla kilku przypadków sieci bezprzewodowych zebrano w tabeli 1.1). W niektórych przypadkach wartość b można obliczyć w sposób uproszczony:

$$b = \frac{l_c}{l_d}, \tag{1.29}$$

gdzie l_c i l_d oznaczają odpowiednio długość ramek sterujących i danych wyrażoną w bitach lub bajtach. Zależność powyższa jest jednak prawdziwa tylko wówczas, gdy ramki sterujące i danych są przesyłane z tą samą prędkością transmisji. Warunek ten nie zawsze jest spełniony. Przykładowo, w standardzie bezprzewodowych sieci lokalnych IEEE 802.11 ramki sterujące przesyła się zazwyczaj z mniejszymi prędkościami niż dane. Zależność powyższa zakłada także jednakową długość ramek sterujących RTS i CTS, podczas gdy w standardzie IEEE 802.11 RTS jest o kilka bajtów dłuższy. Z drugiej strony, można wykazać, iż dla poprawy skuteczności unikania kolizji metodą wymiany ramek sterujących ramka CTS powinna być dłuższa od RTS [34].

Stopień wykorzystania kanału dla protokołu MACA można wyrazić następująco [32]:

$$S_{\mathrm{MACA}} = \frac{1}{e^{G(2b+a)}(b+a+\frac{1}{G}+F)+e^{Gb}(b+\frac{a}{2}+P(a-F))+1+\frac{3a}{2}+F+P(a-F)}, \tag{1.30}$$

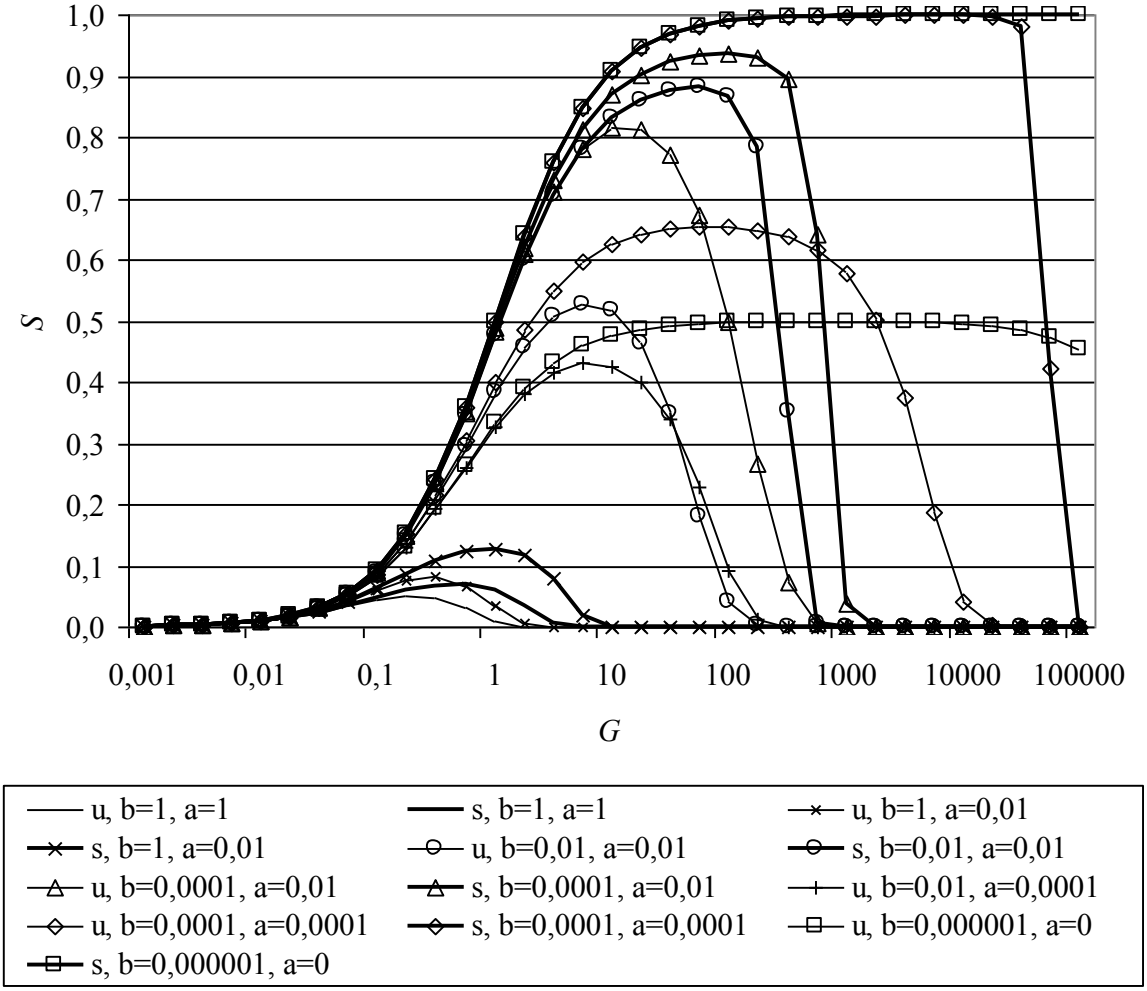
a dla wersji szczelinowej:

$$S_{\text{s-MACA}} = \frac{1}{1+4(a+b)+\frac{1}{G}e^{G(b+a)}}, \tag{1.31}$$

gdzie F oraz P zdefiniowano jako:

$$F = \frac{e^{Gb}-1-Gb}{Gb(1-e^{-Gb})}, \quad P = \frac{e^{-Gb}-e^{-G(b+a)}}{1-e^{-G(b+a)}}. \tag{1.32}$$

Zależności powyższe – dla różnych wartości parametrów a oraz b – zilustrowano na rys. 1.54.



Rys. 1.54. Wydajność nieszczelinowego (u) i szczelinowego (s) protokołu MACA
Fig. 1.54. The efficiency of unslotted (u) and slotted (s) MACA protocol

Na przedstawionym wykresie widać, że osiągi protokołu MACA w wersji szczelinowej i nieszczelinowej znacznie się różnią, przy czym wersja szczelinowa zawsze zachowuje wyższą wydajność. Podczas gdy dla wersji szczelinowej zmniejszanie wartości a i b zawsze przynosi poprawę wydajności, odmiana nieszczelinowa osiąga najlepsze wyniki, gdy $a \approx b$. Istotnie, dla danej wartości a (lub b) zarówno zwiększenie, jak i zmniejszenie wartości drugiego parametru powoduje zmniejszenie wydajności. Odmiana szczelinowa wykazuje najwyższą efektywność – zarówno przepustowość, jak i stabilność – dla najmniejszych wartości

obu parametrów. Odmiana nieszczelinowa natomiast przeciwnie – najwyższą przepustowość uzyskuje dla innego zestawu parametrów $(a = 10^{-2}, b = 10^{-4})$ niż najlepszą stabilność $(a = 0, b = 10^{-6})$. Z wykresu można także odczytać, że dla wersji nieszczelinowej zmniejszanie wartości b jest znacznie ważniejsze niż a. Dla wersji szczelinowej zależność ta nie jest taka oczywista. Dla najgorszych parametrów $(a = 1, b = 1)$ wydajność obu odmian protokołu spada poniżej osiągów protokołu Aloha. Gdy tylko jeden z parametrów ma wartość 1, wydajność nieznacznie się polepsza, ale i tak nie przekracza 15%, pozostaje zatem nadal poniżej osiągów protokołu Aloha. Można zatem powiedzieć, że w obecności ukrytych stacji protokół MACA – pomimo zastosowanych mechanizmów unikania kolizji, które powinny wydajnie pracować w takich warunkach – nie spełnia oczekiwań.

1.5.1.5. Protokół FAMA

Protokół FAMA można rozpatrywać jako złożenie protokołów CSMA i MACA, ponieważ używa on dwóch metod unikania kolizji – zarówno wykrywania nośnej, jak i wymiany ramek sterujących.

Stopień wykorzystania kanału dla protokołu FAMA można wyrazić jako [32]:

$$S_{\text{FAMA}} = \frac{1}{b+1+\frac{1}{G}(2-e^{-aG})+e^{aG}(b+4a)}, \tag{1.33}$$

natomiast w wersji szczelinowej:

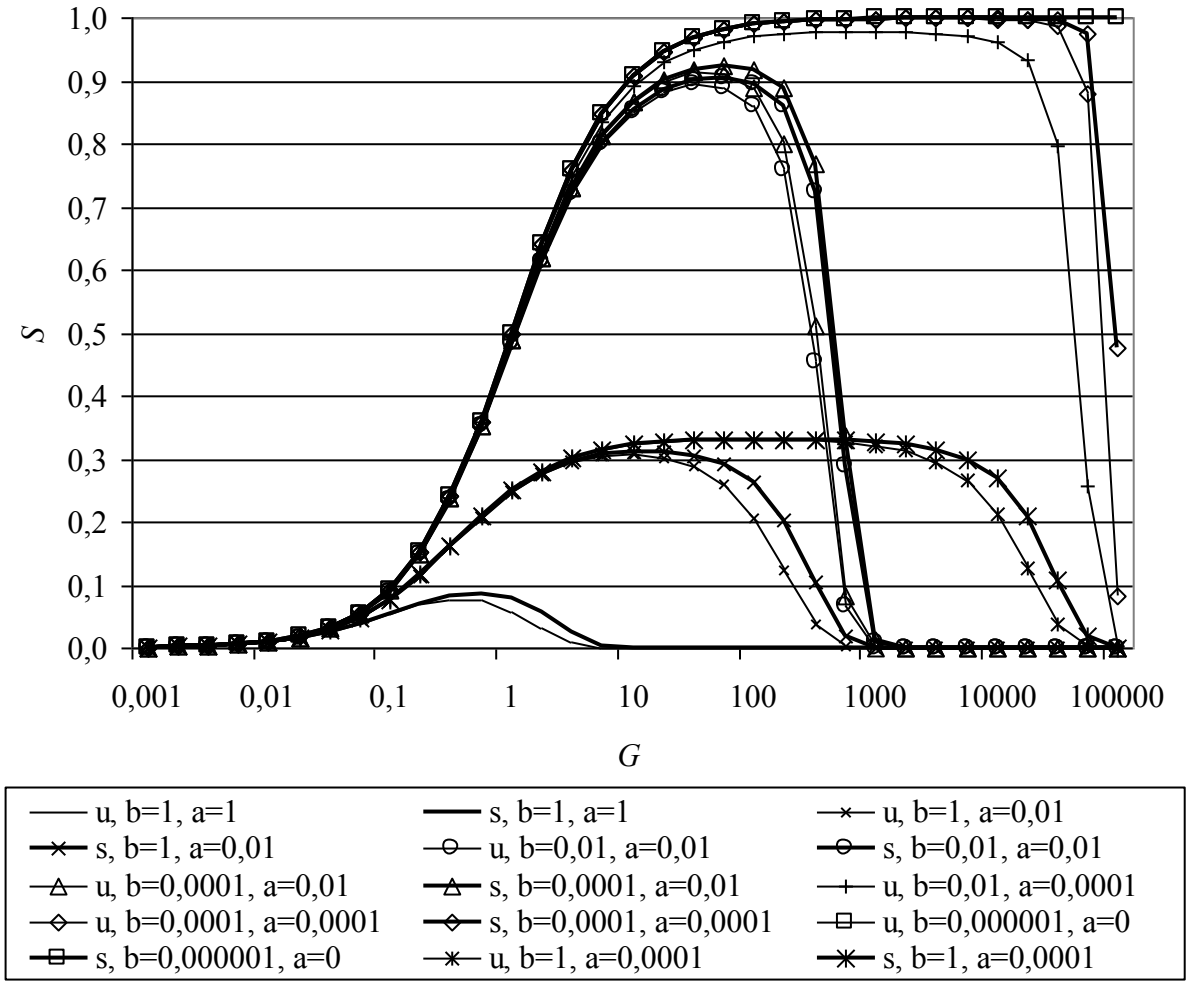
$$S_{\text{s-FAMA}} = \frac{Gae^{-Ga}}{Gae^{Ga}(b+a+1)+(1-e^{-Ga})(b+3a)+a}. \tag{1.34}$$

Zależności te, dla różnych wartości parametrów a oraz b, zilustrowano na rys. 1.55.

Na przedstawionych wykresach widać, że wsparcie wymiany ramek sterujących wykrywaniem nośnej przynosi korzyści – wynikowy protokół jest bardziej wydajny i stabilny niż bez wykrywania nośnej. Ochrona ramek sterujących, uzyskana dzięki wykrywaniu nośnej, pomaga zatem zmniejszyć liczbę kolizji, co z kolei przekłada się na wzrost wydajności protokołu ze względu na większe prawdopodobieństwo udanej transmisji. Chociaż MACA i FAMA – ze względu na użycie tej samej metody wymiany ramek sterujących – wydają się zbliżone do siebie, różnica między szczelinową i nieszczelinowa odmianą protokołu FAMA jest znacznie mniejsza niż w protokole MACA.

Dla każdej wartości parametru b zmniejszanie wartości a przynosi zarówno wzrost stopnia wykorzystania kanału, jak i poprawę stabilności protokołu. Właściwość ta jest szczególnie widoczna dla $b = 1$. Gdy $a = 1$, protokół zachowuje się gorzej niż Aloha – największa wydajność nie przekracza wówczas 10%. Zmniejszenie wartości a do 0,01 – przy niezmienionej wartości b – pomaga uzyskać lepszą wydajność (około 32%), co jednak pozostaje poniżej osiągów znacznie prostszego protokołu Aloha w odmianie szczelinowej. W tych warunkach

FAMA zachowuje jednak lepszą stabilność. Dalsze zmniejszanie wartości *a*, np. do 0,0001, nie zwiększa znacząco maksymalnej przepustowości, ale polepsza stabilność. Gdy wartości obu parametrów nie przekraczają 0,01, protokół – zarówno w odmianie szczelinowej, jak i nieszczelinowej – może osiągnąć ponad 90% wykorzystanie kanału przy odpowiednio dużym obciążeniu łącza. Ogólnie można powiedzieć, że zmniejszanie wartości parametru *b* przynosi wzrost wydajności, natomiast *a* – poprawę stabilności protokołu.



Rys. 1.55. Wydajność nieszczelinowego (u) i szczelinowego (s) protokołu FAMA
Fig. 1.55. The efficiency of unslotted (u) and slotted (s) FAMA protocol

1.5.1.6. Protokół DBTMA

Protokół DBTMA można rozpatrywać jako złożenie protokołów BTMA oraz MACA, ponieważ używa on zarówno wykrywania tonu zajętości, jak i wymiany ramek sterujących w celu zwiększenia skuteczności unikania kolizji. Protokół DBTMA występuje w dwóch odmianach (por. rozdz. 1.5.1.6). Stopień wykorzystania kanału dla pierwotnej odmiany można opisać jako [46]:

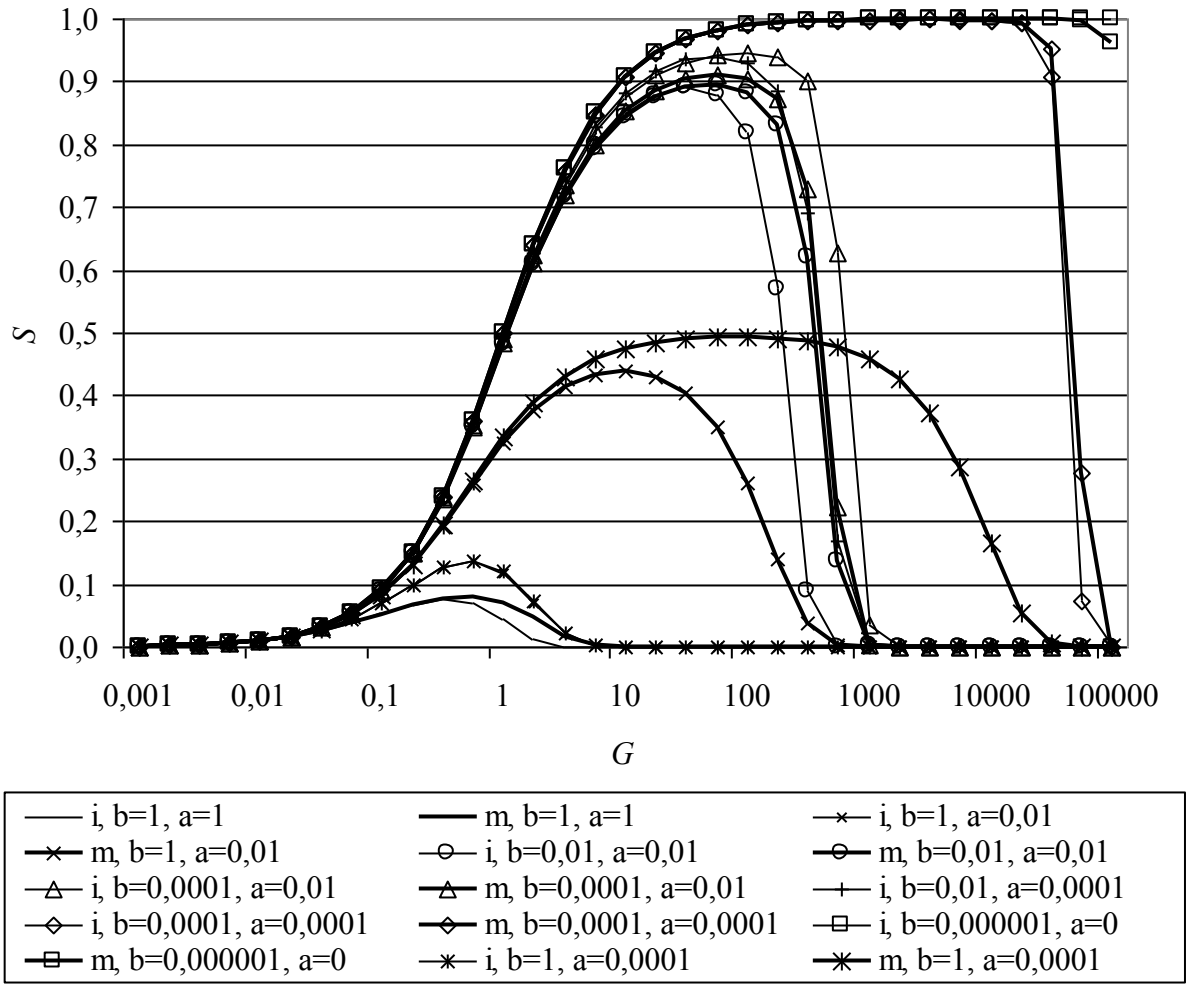
$$S_{\mathrm{DBTMAi}} = \frac{P_s \delta}{P_s(2\gamma + 3\tau + \delta) + (1 - P_s) \cdot 1.5\gamma + 1/g}, \tag{1.35}$$

a dla wersji zmodyfikowanej [28]:

$$S_{\mathrm{DBTMAm}} = \frac{P_s \delta}{P_s(\delta + \gamma + t_d + 6\tau) + (1 - P_s) \cdot 1.5(\gamma + \tau + t_d/2) + 1/g}, \tag{1.36}$$

gdzie $P_s = e^{-g(t_d+\tau)}$. W powyższych wzorach γ oznacza czas transmisji ramki sterującej [s], δ – czas transmisji ramki danych [s], τ – opóźnienie propagacyjne [s], t_d – czas wykrywania tonu zajętości [s], g – liczba ramek wytworzonych w sieci podczas transmisji jednej ramki. Warto zauważyć, że wykrycie nośnej w protokole CSMA może nastąpić znacznie szybciej niż t_d, ponieważ nośna zajmuje całą szerokość pasma, ton zajętości natomiast – tylko wąski podkanał. Przyjmuje się zatem, że wykrycie tonu zajętości wymaga więcej czasu niż wykrycie nośnej. Dlatego też w analizie efektywności protokołów czas wykrywania nośnej można pominąć, natomiast czas wykrycia tonu zajętości należy uwzględnić.

Zależności powyższe, dla różnych wartości *a* oraz *b*, zilustrowano na rys. 1.56.



Rys. 1.56. Wydajność pierwotnego (i) i zmodyfikowanego (m) protokołu DBTMA
Fig. 1.56. The efficiency of initial (i) and modified (m) DBTMA protocol

Na przedstawionych wykresach widać, że – podobnie jak wykrywania nośnej – wykrywanie tonu zajętości znacząco wspomaga wymianę ramek sterujących jako skuteczny mecha-

nizm unikania kolizji. Tym niemniej, jeżeli $a = 1,$ obie wersje protokołu DBTMA zachowują się gorzej niż Aloha – stopień wykorzystania kanału nie przekracza 15%. W tych warunkach wersja zmodyfikowana wykazuje jednak nieco lepszą wydajność niż wersja początkowa. Wraz ze zmniejszaniem się wartości a wydajność protokołu szybko rośnie nawet wówczas, gdy $b = 1.$ To polepszenie osiągów protokołu jest szczególnie widoczne dla wariantu zmodyfikowanego, gdy $b = 1,$ natomiast $a = 0.01$ lub $a = 0.0001.$ Gdy wartości zarówno parametru *a*, jak i *b* są bardzo małe, wydajność obu odmian jest porównywalna i nie można wówczas określić, która z nich jest lepsza – zależy to od konkretnych wartości. Przykładowo, gdy $a = b = 0.01,$ wersja zmodyfikowana wykazuje wyższość. Dalsze zmniejszanie wartości *a* podnosi wydajność tej wersji bardziej niż wersji początkowej. Jednak wraz ze zmniejszaniem wartości *b* szybciej poprawia się efektywność wersji początkowej. Można zatem powiedzieć, że odmiana początkowa protokołu DBTMA jest czulsza na wartość parametru *b*, zmodyfikowana natomiast zależy bardziej od wartości *a*.

1.5.2. Porównanie wydajności protokołów

Wykorzystując zależności przedstawione powyżej, można porównać właściwości protokołów w różnych warunkach. Szczególnie interesujące jest ich zachowanie w typowym środowisku bezprzewodowym [132]. Dla celów takiego porównania wybrano kilka przykładowych sieci:

- Packet Radio z ramkami danych zawierającymi 32 lub 256 bajtów danych, prędkością transmisji 9,6 kb/s i zasięgiem 20 km,
- bezprzewodową sieć lokalną zbliżoną do IEEE 802.11, z ramkami zawierającymi 256 lub 2312 bajtów danych, prędkością transmisji 2, 11 lub 54 Mb/s i zasięgiem transmisji odpowiednio 50, 20 i 10 m.

Parametry sieci użytych dla celów porównania oraz obliczone na ich podstawie wartości parametrów a i *b* zebrano w tabeli 1.1.

Tabela 1.1
Parametry sieci przyjęte dla oszacowania wydajności protokołów dostępu do łącza

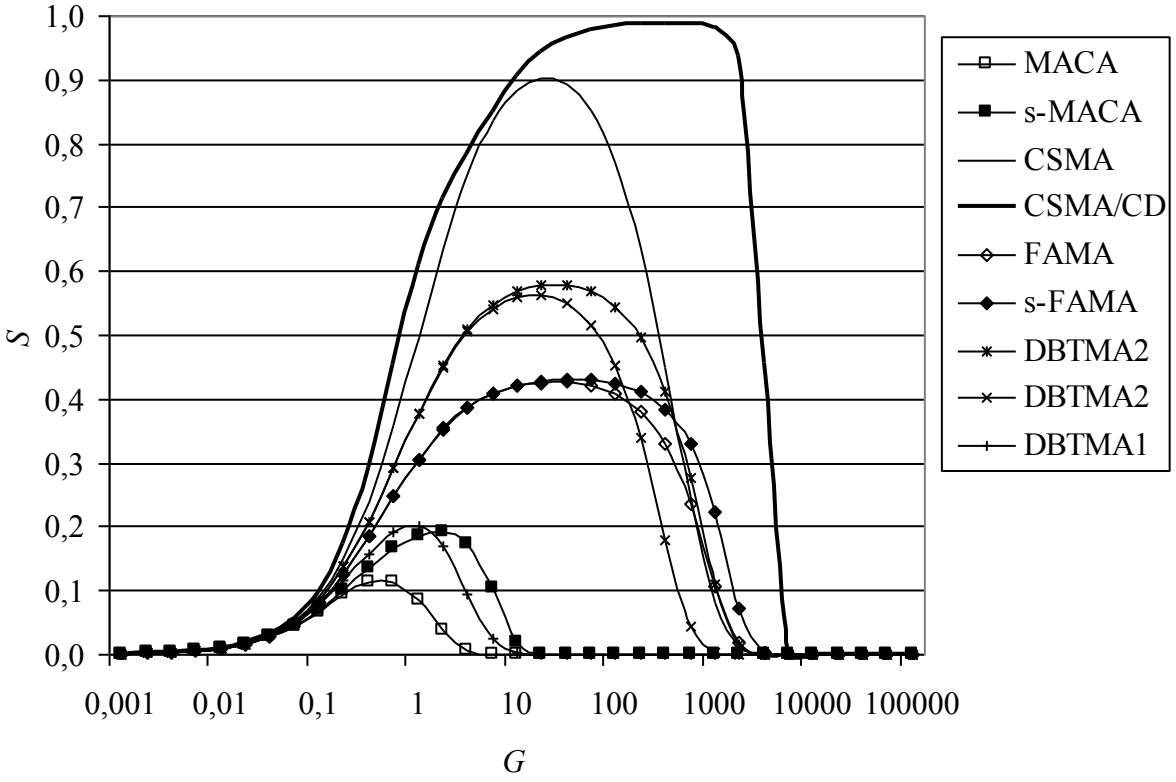
Typ sieci	Prędkość transmisji [kb/s]	Zasięg transmisji [m]	Długość ramki danych [B]	*a*	*b*
Packet Radio	9,6	20000	52	0,0015385	0,3846
Packet Radio	9,6	20000	276	0,0002899	0,0725
WLAN	2000,0	50	276	0,0001510	0,0725
WLAN	2000,0	50	2346	0,0000178	0,0085
WLAN	11000,0	30	2346	0,0000586	0,0085
WLAN	54000,0	10	2346	0,0000977	0,0085

Warto zauważyć, iż przyjęte wartości parametrów nie odbiegają zbytnio od wartości rzeczywistych. Przykładowo, w standardzie IEEE 802.11 ramka RTS zawiera 20 B, natomiast CTS – jedynie 14. Maksymalna pojemność pola danych ramki wynosi 2312 B, podczas gdy najdłuższa ramka zawiera 2346 B. W sieci Packet Radio nie używa się ramek RTS i CTS; dla potrzeb niniejszego oszacowania przyjęto jednak rozmiar typowy dla ramek sterujących zdefiniowanych w protokole AX.25 czy HDLC. Przy typowym rozmiarze pola adresowego ramki protokołu AX.25, wynoszącym 7 B, rozmiar ramki sterującej wynosi 20 B. Ramka danych zawiera dodatkowo nie więcej niż 256 B danych. W analizie pominięto narzut wnoszony przez elementy warstwy fizycznej, jak np. preambuły, czasy przełączania odbiór-nadawanie itp. Dla protokołu DBTMA przyjęto czas wykrywania tonu zajętości $t_d = 10^{-4}$ and $t_d = 10^{-6}$.

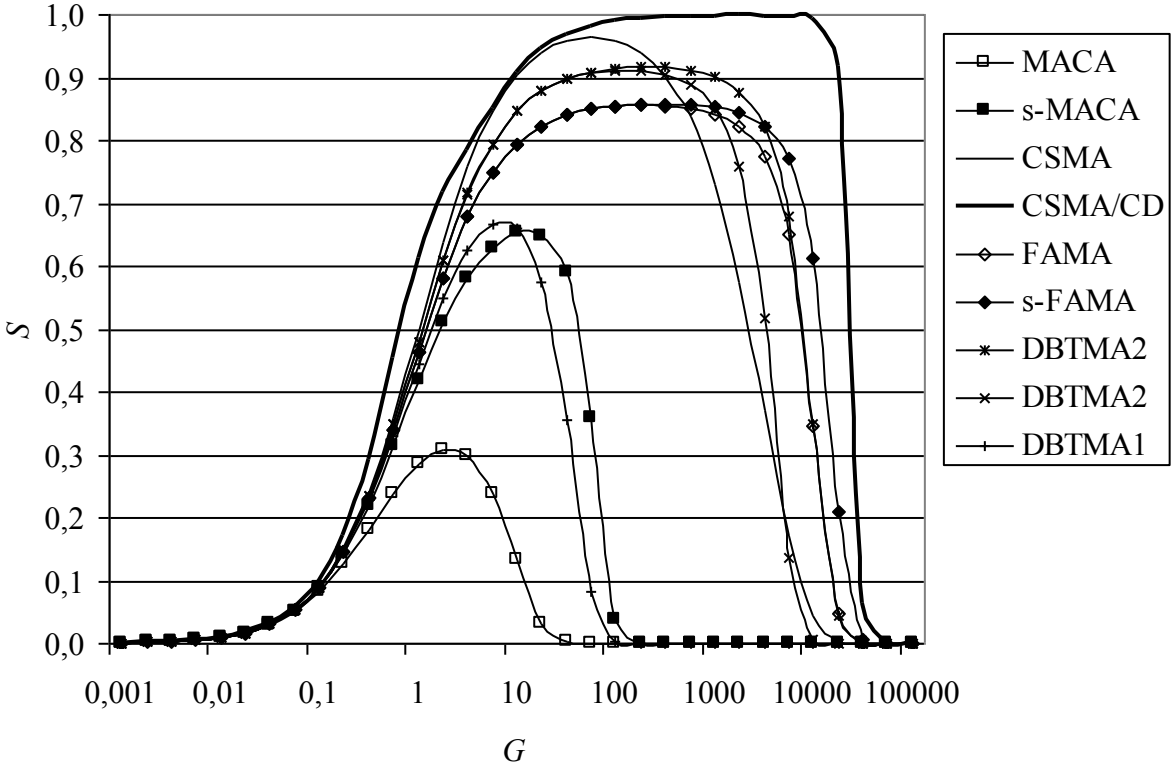
Wyniki uzyskane dla podanego zbioru parametrów i sieci Packet Radio przedstawiono na rys. 1.57 (ramki 32 B) i 1.58 (ramki 256 B). Wyniki uzyskane dla sieci lokalnej o prędkości 2 Mb/s przedstawiono na rys. 1.59 (ramki 256 B) i 1.60 (ramki 2312 B). Wyniki dla sieci lokalnych o wyższej prędkości transmisji i z ramkami 2312 B przedstawiono z kolei na rys. 1.61 (prędkość 11 Mb/s) i 1.62 (prędkość 54 Mb/s).

Na przedstawionych rysunkach łatwo zauważyć, że dla każdego zestawu parametrów protokół CSMA/CD wykazuje najwyższą efektywność. Niestety, metody tej nie można stosować w większości sieci bezprzewodowych. Protokoły FAMA i MACA osiągają największą wydajność dla długich ramek danych. Jest to oczywiste, ponieważ przy krótszych ramkach rośnie narzut protokołu wynikający z użycia ramek sterujących RTS i CTS. Zależność tę opisuje parametr *b*. Z kolei, wpływ opóźnienia propagacyjnego (parametr *a*) jest większy dla wyższych prędkości transmisji. Pomimo to dla przyjętych wartości parametrów opisujących konfigurację sieci Packet Radio, opóźnienie propagacyjne także odgrywa istotną rolę ze względu na duży zasięg transmisji. Protokół CSMA bez wykrywania kolizji wykazuje niemal taką samą wydajność jak CSMA/CD i może być użyty w rozważanych sieciach, jednak dla sieci lokalnej o wysokiej prędkości transmisji (11 i 54 Mb/s) protokoły FAMA i, częściowo, DBTMA mają wyższą efektywność. W przypadku sieci o prędkości 54 Mb/s, dla obciążenia sieci *G* między 100 i 1000, szczelinowy protokół MACA także osiąga wyższą wydajność niż CSMA.

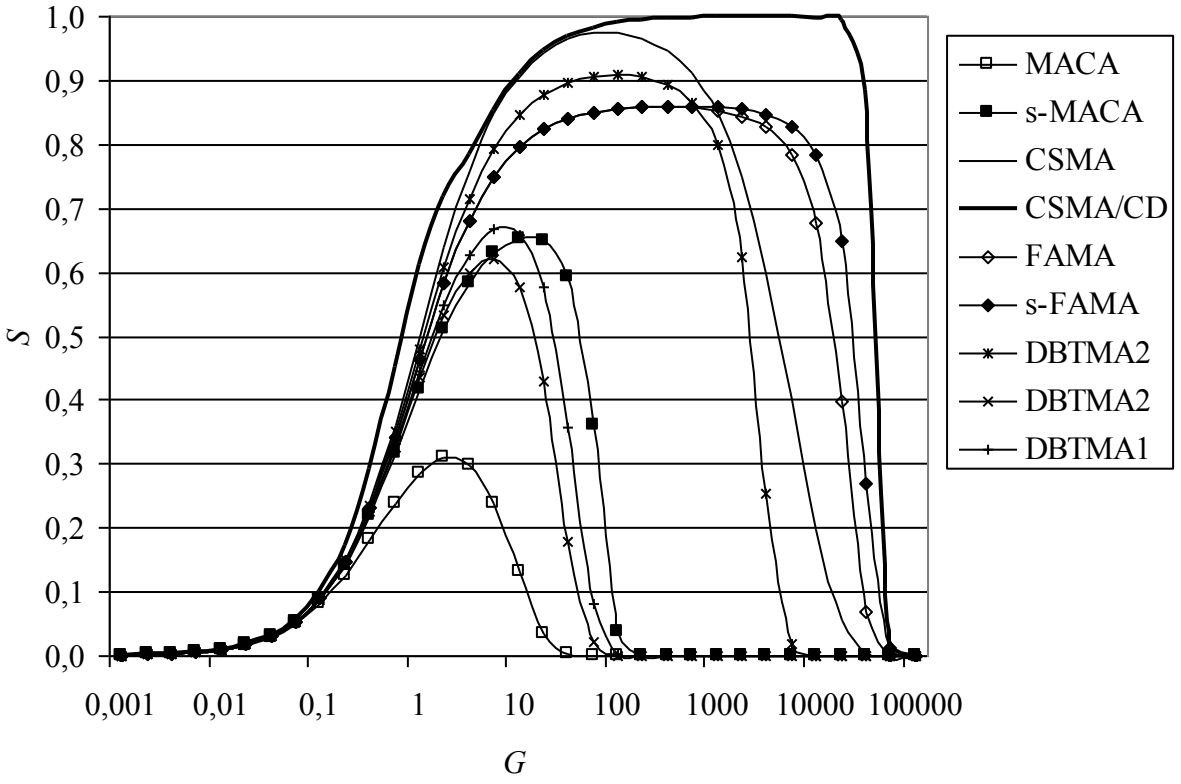
Spośród rozważanych protokołów najniższą wydajność wykazuje zawsze nieszczelinowa odmiana protokołu MACA. Podczas transmisji krótkich ramek osiągi tego protokołu są nawet gorsze niż szczelinowego protokołu Aloha, ale wraz ze zwiększaniem długości ramki stopień wykorzystania kanału rośnie nawet do 45%. W tych warunkach stabilność tego protokołu jest także wyższa niż protokołu Aloha. Inną ciekawą właściwością protokołu MACA jest niezależność od prędkości i zasięgu transmisji, a to dlatego, że nie wykorzystuje się tu wykrywania nośnej ani tonu zajętości.



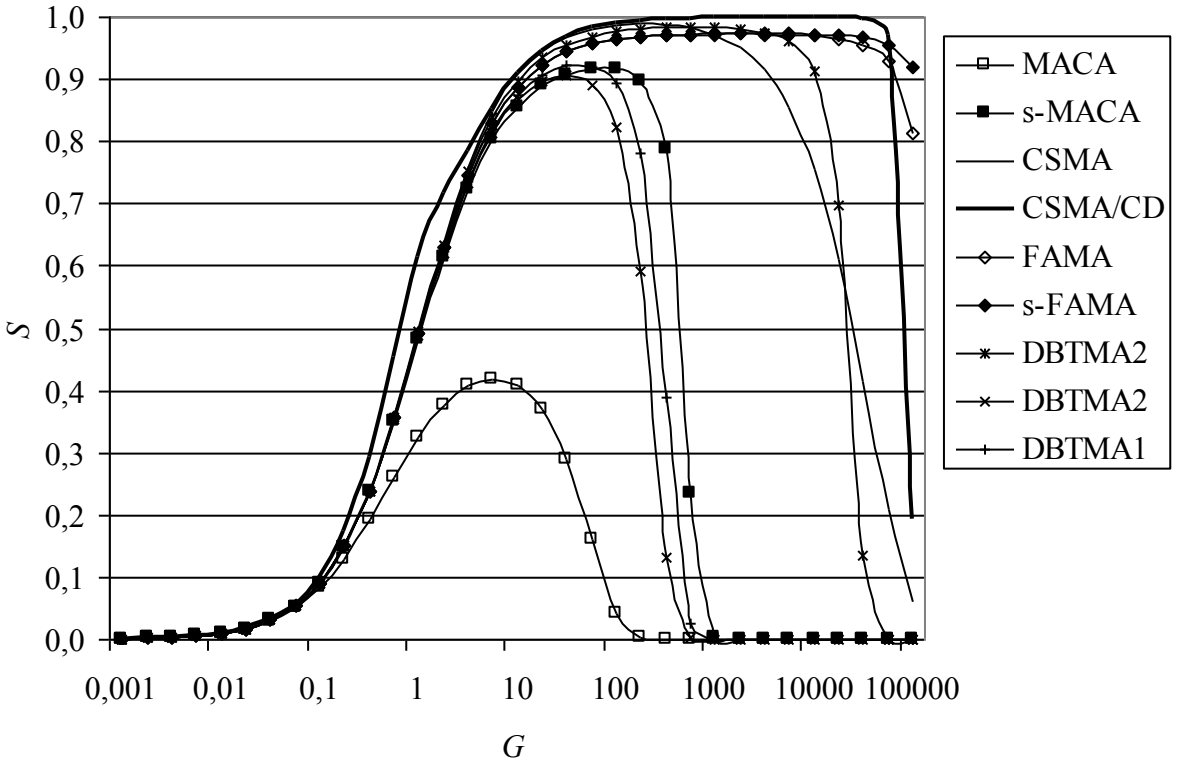
Rys. 1.57. Porównanie wydajności protokołów dla sieci Packet Radio z krótkimi ramkami
Fig. 1.57. Protocols efficiency comparison for Packet Radio with short frames



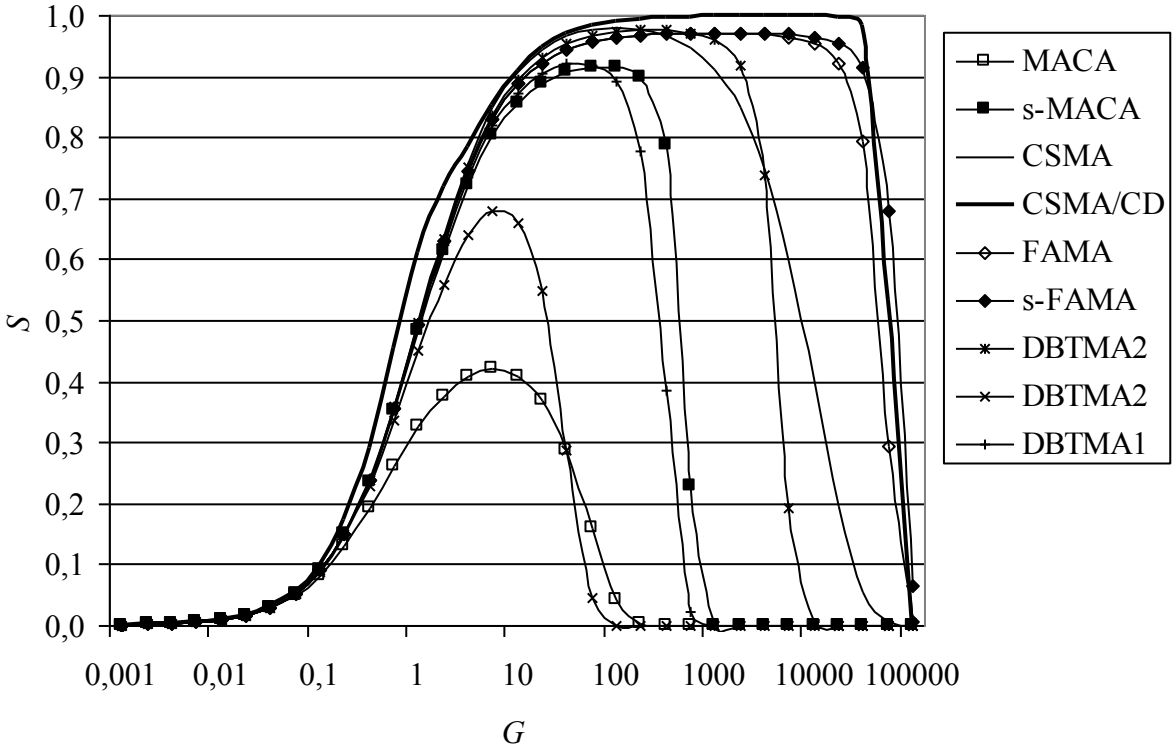
Rys. 1.58. Porównanie wydajności protokołów dla sieci Packet Radio z długimi ramkami
Fig. 1.58. Protocols efficiency comparison for Packet Radio with long frames



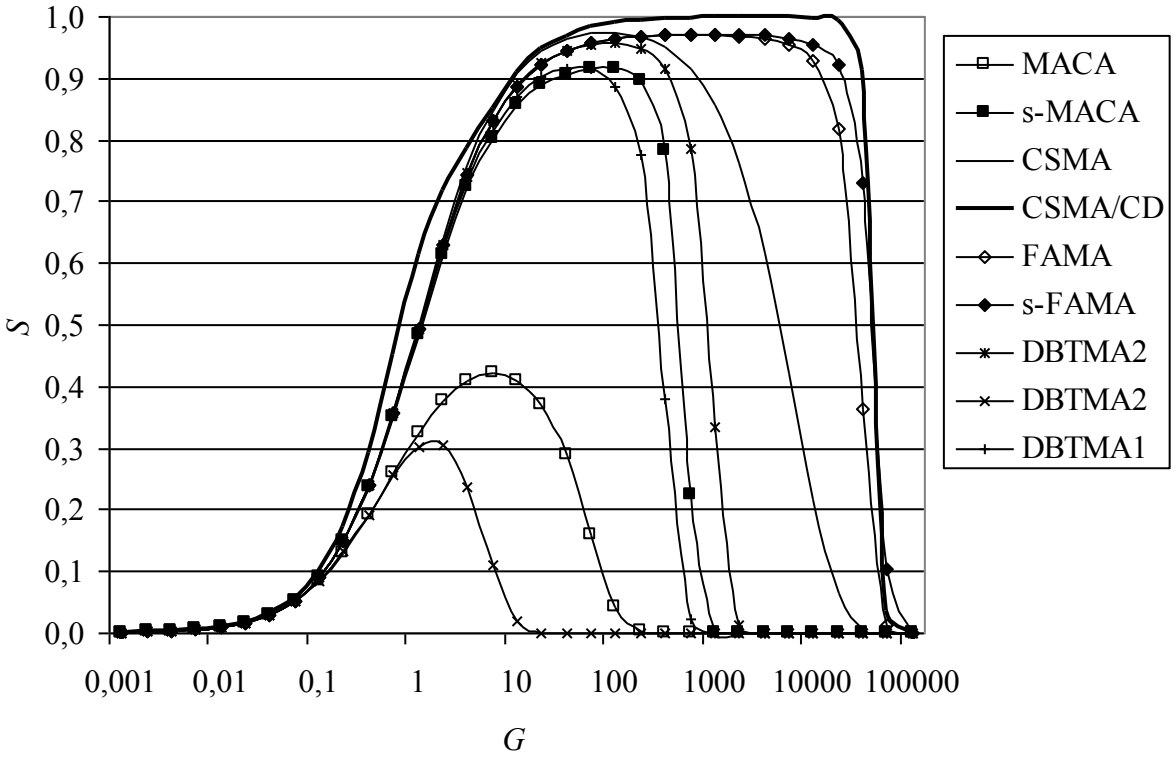
Rys. 1.59. Porównanie wydajności protokołów dla sieci lokalnej z krótkimi ramkami (2 Mb/s)
Fig. 1.59. Protocols efficiency comparison for 2 Mbps WLAN with short frames



Rys. 1.60. Porównanie wydajności protokołów dla sieci lokalnej z długimi ramkami (2 Mb/s)
Fig. 1.60. Protocols efficiency comparison for 2 Mbps WLAN with long frames



Rys. 1.61. Porównanie wydajności protokołów dla sieci lokalnej z długimi ramkami (11 Mb/s)
Fig. 1.61. Protocols efficiency comparison for 11 Mbps WLAN with long frames



Rys. 1.62. Porównanie wydajności protokołów dla sieci lokalnej z długimi ramkami (54 Mb/s)
Fig. 1.62. Protocols efficiency comparison for 54 Mbps WLAN with long frames

Ogólnie rzecz ujmując, najlepsze wyniki uzyskano dla sieci lokalnej o prędkości transmisji 2 Mb/s, w której przesyła się długie ramki. Jest to wynik zgodny z oczekiwaniami, gdyż sieć ta ma mały zasięg transmisji, tak więc opóźnienie propagacyjne jest małe. Z kolei mniejsza prędkość transmisji dodatkowo zmniejsza wartość parametru *a*. Długość ramek natomiast wpływa na zmniejszenie wartości parametru *b*. Dalsze zwiększanie prędkości transmisji powoduje zatem zmniejszenie wydajności protokołu. Tym niemniej, w większości przypadków, stopień wykorzystania kanału dla protokołów FAMA i DBTMA jest niemal taki sam jak dla CSMA i CSMA/CD. Ponadto, FAMA wykazuje wyższą stabilność niż CSMA/CD, nawet dla wyjątkowo dużego obciążenia sieci (G>1000). Jest to szczególnie widoczne dla sieci lokalnej 2 Mb/s z długimi ramkami (rys. 1.60).

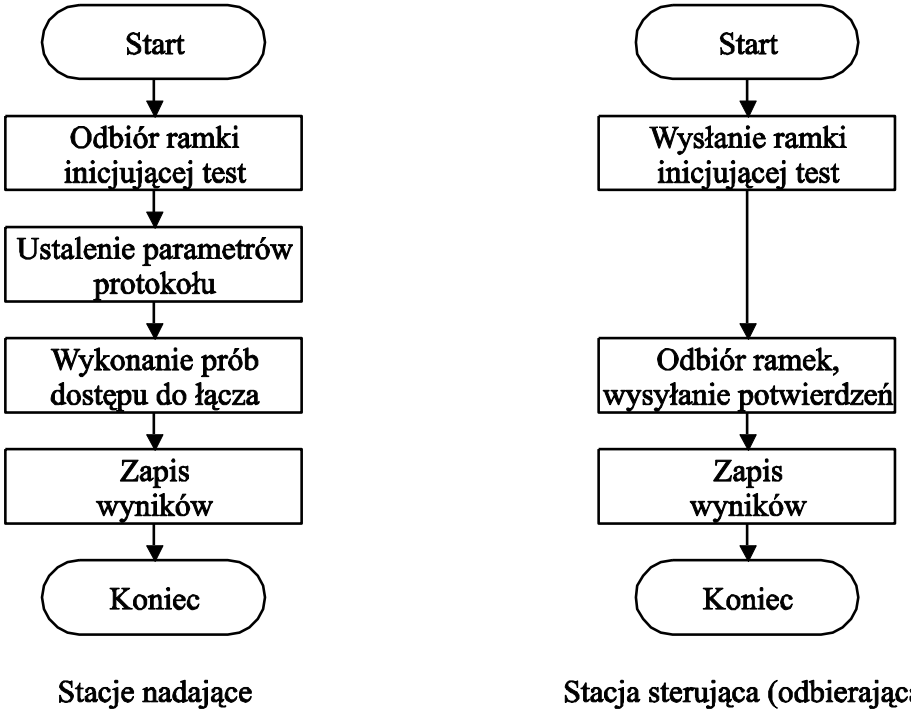
1.5.3. Pomiary w doświadczalnej sieci bezprzewodowej

W celu zweryfikowania analitycznych metod oceny wydajności protokołów dostępu do łącza zrealizowano doświadczalną sieć bezprzewodową, w której następnie przeprowadzano badania wydajności kilku wybranych protokołów [130].

1.5.3.1. Realizacja segmentu doświadczalnej sieci komputerowej

Segment doświadczalnej bezprzewodowej sieci komputerowej składa się z czterech węzłów. Każdy z węzłów zawiera komputer klasy PC, do którego poprzez interfejs RS-232C podłączono modem radiowy. Jeden z węzłów sieci pełni wyłącznie funkcję odbiornika, do którego są adresowane wszystkie ramki w sieci. Jego zadaniem jest odbiór, zliczanie i potwierdzanie wszystkich ramek. Takie rozwiązanie pozwala na analizę efektywności sieci na podstawie wyników zebranych przez tylko jeden węzeł. Poprawność transmisji ramki jest weryfikowana na podstawie sumy kontrolnej, umieszczonej na końcu ramki. W przypadku gdy odbiornik otrzyma ramkę przekłamaną, potwierdzenie nie jest przesyłane. Dodatkowo węzeł będący odbiornikiem ma za zadanie synchronizować oraz inicjować inne węzły. Dokonuje się tego poprzez wysłanie do nadajników krótkiej ramki, zawierającej numer protokołu dostępu do łącza, rozmiar pola danych oraz liczbę ramek, jaką ma wysłać pojedynczy nadajnik w czasie pojedynczego testu. Ramka ta jest także sygnałem do rozpoczęcia nadawania ramek. Pozostałe węzły pracują wyłącznie jako nadajniki, generując w losowo wybranych chwilach ramki i wysyłając je do odbiornika według określonego protokołu dostępu do łącza. Każda ze stacji nadających wytwarza w ciągu sekundy g ramek o stałej, zadanej długości. Kolejno wysyłane przez dany węzeł ramki są numerowane w celu zapewnienia możliwości wykrycia zgubienia ramki, np. wskutek kolizji lub innego błędu transmisji.

Na rys. 1.63 przedstawiono zasady działania oprogramowania stacji centralnej (odbierającej ramki) i stacji nadających ramki.



Rys. 1.63. Zasady działania stacji nadającej i odbierającej
Fig. 1.63. Operating rules of transmitter and receiver

1.5.3.2. Wyznaczenie charakterystyki przepływu danych

Wyznaczenie charakterystyki przepływu polegało na wykonaniu serii n testów. Pojedynczy test wykonywano przez czas T_{pom}. W kolejnych testach każda z trzech stacji generowała, w losowo wybranych chwilach, od 1 do n ramek o czasie transmisji T. Dostęp do łącza odbywał się zgodnie z regułami badanego protokołu. Każdy test rozpoczynano pojedynczą ramką inicjującą, wysłaną przez stację odbierającą do pozostałych węzłów. W czasie testu stacja ta odbiera ramki i wysyła potwierdzenia. Na podstawie adresu stacji źródłowej i numeru ramki w stacji odbierającej można było określić liczbę prawidłowo przesłanych ramek od poszczególnych nadawców, z czego suma stanowiła całkowitą liczbę ramek przesłanych w pojedynczym teście. Pojedynczą serię wykonywano dla zadanej długości ramki, której odpowiadał czas transmisji T, przy zmiennym obciążeniu łącza. Czas trwania pojedynczego testu T (rys. 1.64) dobierano w zależności od długości ramki danych, przy czym założono, że:

$$T = t_d + t_w + t_{ack}, \tag{1.37}$$

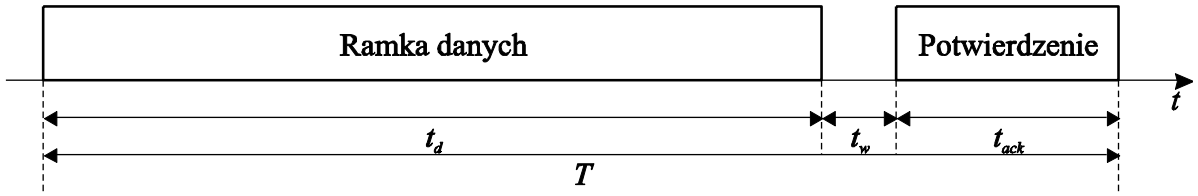
gdzie: t_d – czas transmisji ramki z danymi, t_w – czas generacji potwierdzenia przez stację centralną, t_{ack} – czas transmisji potwierdzenia.

Czasy transmisji ramki można obliczyć ze wzoru:

$$t = \frac{10r}{v}, \tag{1.38}$$

gdzie: r – rozmiar ramki w bajtach, v – prędkość transmisji. Czas generacji potwierdzenia przez stację odbierającą ustalono doświadczalnie i przyjęto, że wynosi 2 ms. Przyjęto także,

że rozmiar ramki w bajtach uwzględnia nagłówek oraz preambułę. Pominięto w ten sposób narzut związany z transmisją informacji sterującej.



Rys. 1.64. Zależności czasowe przy poprawnej transmisji ramki
Fig. 1.64. Time dependencies of error-free frame transmission

Na podstawie wyniku pojedynczego testu można było obliczyć zajętość kanału G oraz efektywność jego wykorzystania S:

$$G = \frac{l_s T}{T_{pom}}, \quad S = \frac{l_r T}{T_{pom}}, \tag{1.39}$$

gdzie: l_r – całkowita liczba ramek odebranych przez stację centralną, l_s – sumaryczna liczba ramek wysłanych przez poszczególne stacje, T_{pom} – czas pomiaru.

Po wykonaniu całej serii testów wyznaczono charakterystykę przepływu w kanale w funkcji zajętości dla zadanego protokołu, przy określonej długości ramki i prędkości transmisji. W przeprowadzonych badaniach wykonywano trzy jednakowe serie, a wyniki uśredniano. Maksymalna liczba prób dostępu do kanału przez każdą ze stacji wynosiła 120. W zależności od długości ramki czas pomiaru T_{pom} dobierano tak, aby uzyskać maksymalną zajętość kanału 2,2. Badanie efektywności dla większego obciążenia sieci uniemożliwiała mała liczba stacji nadających.

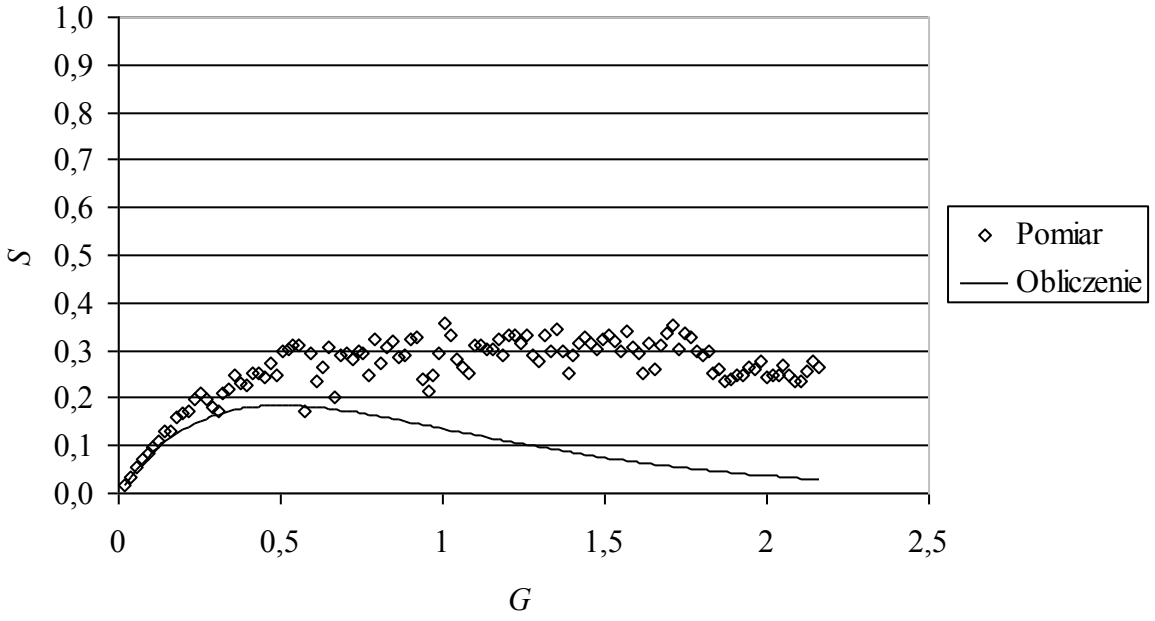
1.5.3.3. Efekt przechwytywania

Przed przystąpieniem do wyznaczenia charakterystyk wybranych protokołów zbadano występowanie w sieci efektu przechwytywania oraz oszacowano jego wpływ na uzyskane wyniki. W tym celu oprogramowanie węzłów przygotowano tak, by węzły generujące ruch w sieci zapamiętywały czasy generacji ramek, a stacja odbierająca ramki zapamiętywała czasy ich odbioru. Otrzymane wyniki zapisano do pliku. Test wykonano dla protokołu Aloha w wersji prostej i szczelinowej. W obu testach zajętość kanału wynosiła 1. W przypadku Aloha uzyskano przepływ w wysokości 0,47, dla Aloha szczelinowego – 0,49. Analizując czasy dostępu stacji nadawczych i czasy odbioru ramek przez stacje centralną, można było stwierdzić, że w sieci występuje efekt przechwytywania. Czasy transmisji niektórych z nich nakładały się na siebie. W niektórych przypadkach, pomimo wystąpienia kolizji, transmisja jednej z kolidujących ramek kończyła się sukcesem. Po odrzuceniu ramek, które pomimo kolizji zostały odebrane bezbłędnie, ustalono, że przepływ w sieci dla protokołu Aloha wyniósłby 0,22, a dla Aloha szczelinowego – 0,44. Jak widać, w przypadku Aloha efekt prze-

chwytywania podniósł efektywność ponaddwukrotnie, dla Aloha w wersji szczelinowej efekt ten poprawił ogólną efektywność w znacznie mniejszym stopniu.

1.5.3.4. Badanie protokołu Aloha

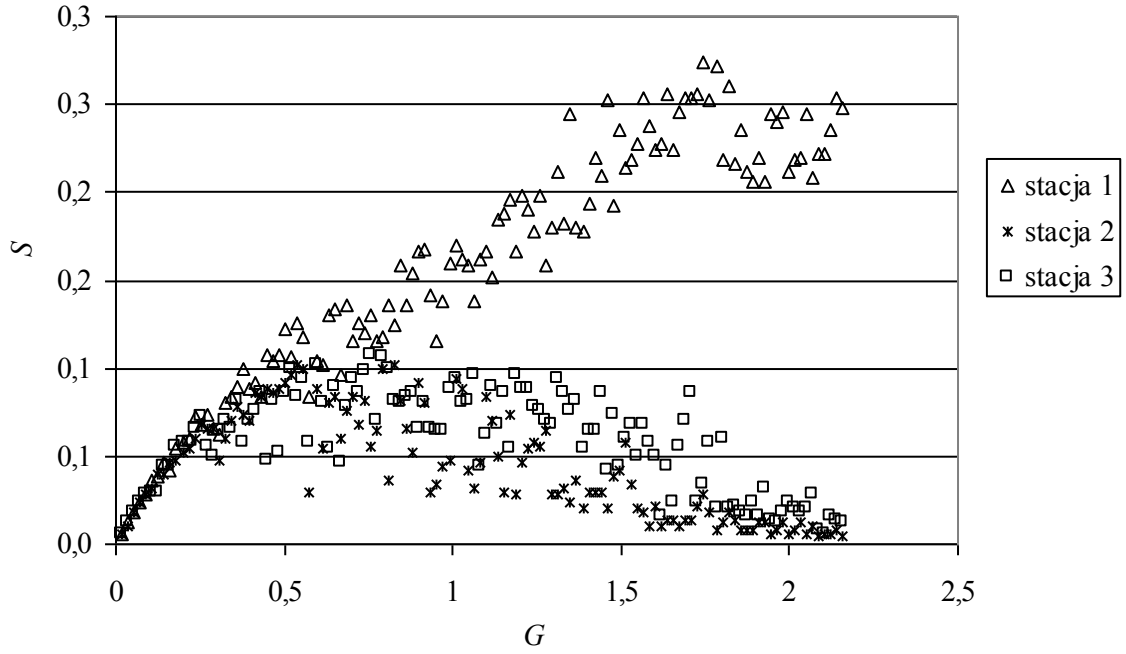
Zgodnie z oszacowaniem analitycznym (rozdz. 1.5.1.1) maksymalna efektywność protokołu Aloha wynosi około 0,18. Protokół badano przy prędkości transmisji 9,6 kb/s i ramce o czasie transmisji T=30 ms. Stacje wysyłające ramki umieszczono jedna obok drugiej w odległości 1 m od stacji odbierającej. Na rys. 1.65 przedstawiono otrzymaną wydajność protokołu Aloha dla całej sieci, a także – dla porównania – obliczoną według wzoru (1.6). Z kolei rys. 1.66 przedstawia wydajność protokołu w rozbiciu na poszczególne nadajniki.



Rys. 1.65. Obliczona i zmierzona efektywność protokołu Aloha
Fig. 1.65. Calculated and measured Aloha protocol efficiency

Chociaż otrzymana charakterystyka przepływności (rys. 1.65) ma kształt zbliżony do charakterystyki wyznaczonej za pomocą modelu matematycznego, to uzyskana maksymalna efektywność tego protokołu w badanym segmencie sieci jest znacznie wyższa od teoretycznej. Na podstawie obserwacji efektu przechwytywania stwierdzono, że właśnie to jest powodem zwiększenia maksymalnego całkowitego przepływu. Następny wykres (rys. 1.66) pokazuje, jak ogólny przepływ w kanale rozłożył się na poszczególne stacje. Z wykresu wyraźnie widać, że stacja 1 przesyła znacznie więcej ramek od pozostałych. Mamy więc do czynienia z niesprawiedliwym podziałem łącza, co jest jednym ze skutków występowania efektu przechwytywania. W czasie badań zauważono, że większe oddalenie stacji 1 od stacji odbierającej, w stosunku do pozostałych, powoduje zmniejszenie liczby poprawnie przesłanych ramek przez tę stację. W przypadku gdyby efekt ten nie występował, wykres przepływności dla sta-

cji 1 byłby podobny do wykresów pozostałych stacji, co dałoby po zsumowaniu całkowitą wydajność zbliżoną do teoretycznej.



Rys. 1.66. Zmierzona efektywność protokołu Aloha dla poszczególnych stacji
Fig. 1.66. Measured Aloha protocol efficiency for individual stations

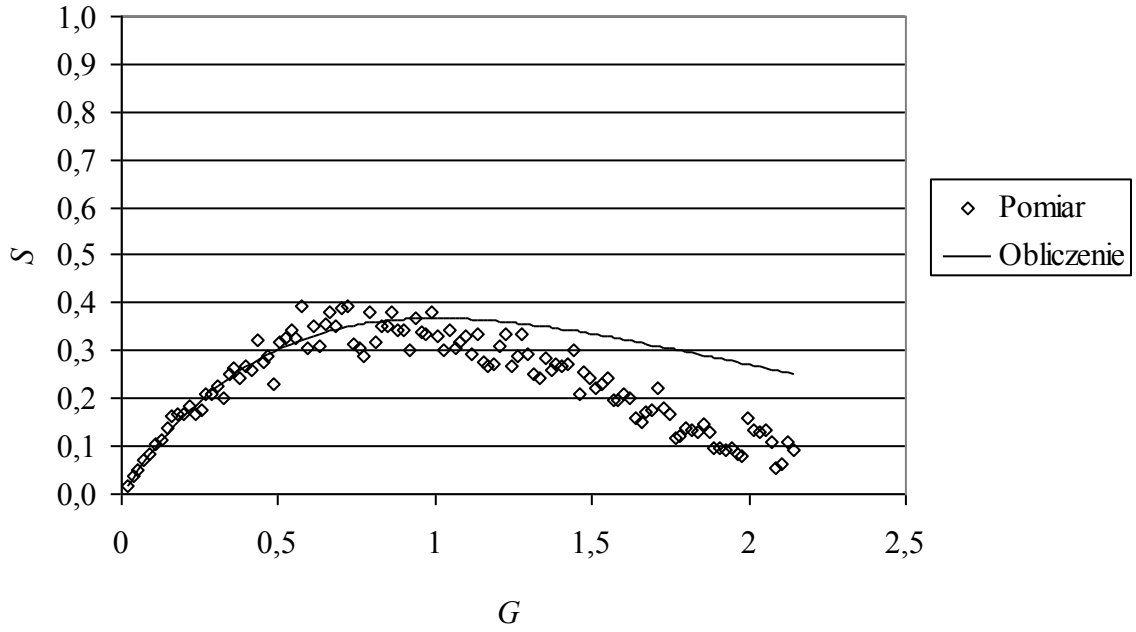
1.5.3.5. Badanie protokołu s-Aloha

Zgodnie z oszacowaniem analitycznym (rozdz. 1.5.1.1) maksymalna efektywność protokołu Aloha w wersji szczelinowej jest dwukrotnie większa niż dla protokołu Aloza i wynosi około 0,37. Oczekiwano, że w przypadku tego protokołu efekt przechwytywania będzie występował w mniejszym stopniu. W protokole Aloha kolidujące ramki mogą nakładać się na siebie częściowo (np. suma kontrolna jednej ramki i preambuła drugiej), podczas gdy w przypadku s-Aloha ramki takie nakładają się na siebie całkowicie. W drugim przypadku przekłamaniu może ulec większa liczba bitów ramki. Można zatem dojść do wniosku, że w protokole s-Aloha prawdopodobieństwo prawidłowego odbioru jednej z kolidujących ramek jest mniejsze, niż w protokole Aloha.

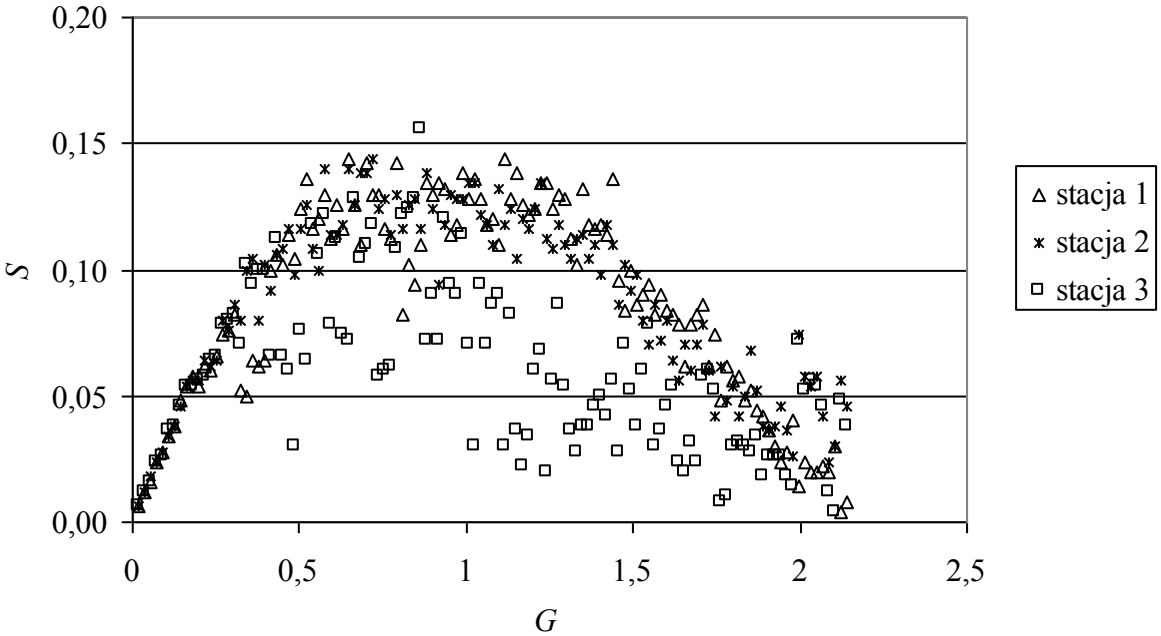
Badanie protokołu s-Aloha przeprowadzono w identyczny sposób jak protokołu Aloha. Na rys. 1.67 przedstawiono otrzymaną wydajność protokołu s-Aloha dla całej sieci, a także – dla porównania – obliczoną według wzoru (1.7). Z kolei rys. 1.68 przedstawia wydajność protokołu w rozbiciu na poszczególne nadajniki.

Otrzymana charakterystyka przepływności dla protokołu s-Aloha (rys. 1.67) w dużym stopniu odpowiada charakterystyce teoretycznej. Uzyskana wydajność maksymalna S wynosi około 0,35. Nie jest ona jednak dwukrotnie większa niż otrzymana dla protokołu Aloha. Zgodnie z oczekiwaniem, efekt przechwytywania w tym przypadku występował w mniejszym

stopniu, nie powodując znaczącego zwiększenia efektywności protokołu. Na podstawie kolejnego wykresu (rys. 1.68) można stwierdzić, że podział kanału był bardziej sprawiedliwy. Potwierdza to tezę, że synchronizacja stacji i podział czasu na szczeliny jest szczególnie opłacalny w przypadku protokołu Aloha.



Rys. 1.67. Obliczona i zmierzona efektywność protokołu s-Aloha
Fig. 1.67. Calculated and measured s-Aloha protocol efficiency



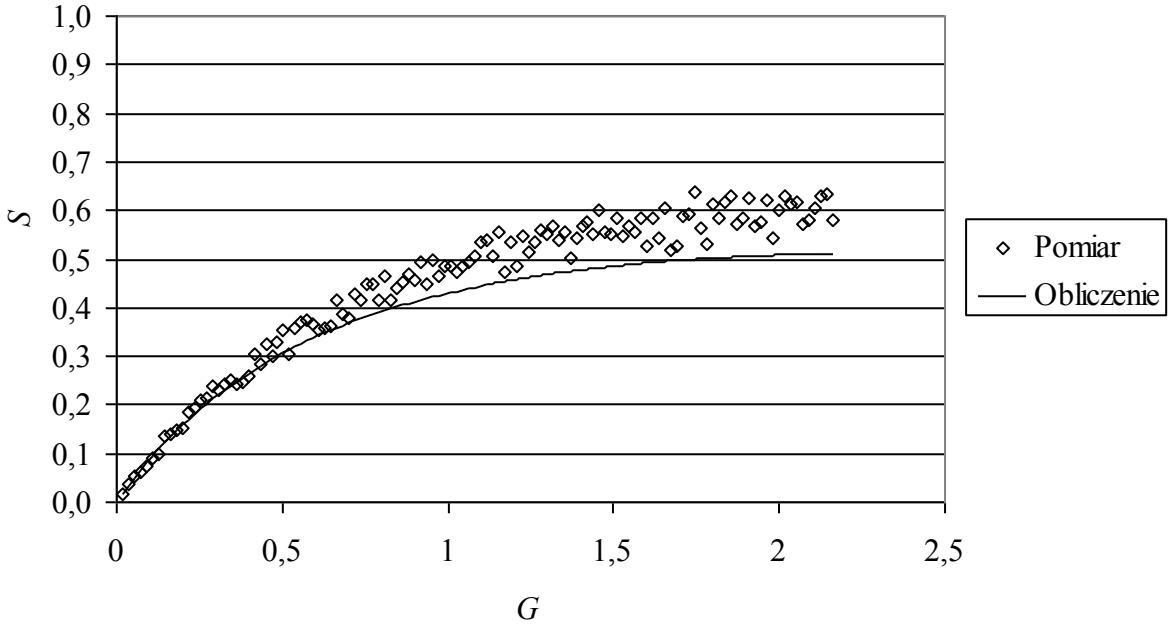
Rys. 1.68. Zmierzona efektywność protokołu s-Aloha dla poszczególnych stacji
Fig. 1.68. Measured s-Aloha protocol efficiency for individual stations

1.5.3.6. Badanie protokołu CSMA

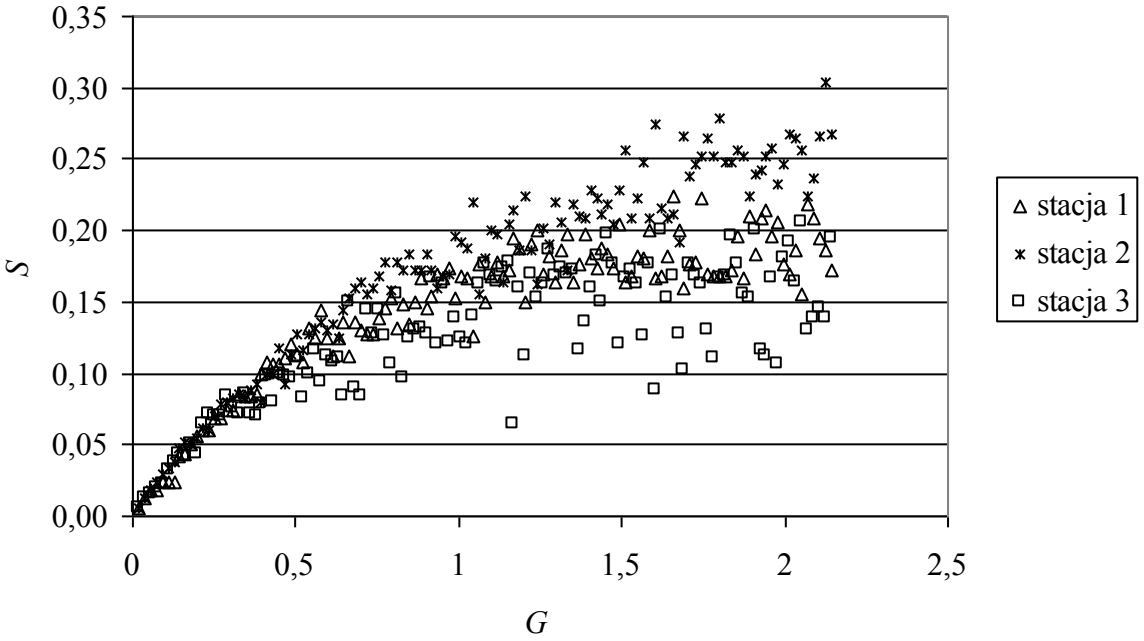
Efektywność protokołu CSMA, zgodnie z obliczeniami teoretycznym (rozdz. 1.5.1.2), zależy od parametru *a*, który podaje związek między czasem wykrycia nośnej sygnału i czasem transmisji ramki. W zastosowanych urządzeniach do transmisji bezprzewodowej czas wykrycia nośnej sygnału wynosił 3 ms. Efektywność protokołu zbadano dla dwóch długości ramek: 28 B i 178 B. Badania przeprowadzono przy prędkości transmisji 9,6 kb/s. Czasy transmisji ramek wynosiły zatem odpowiednio 29 ms i 185 ms. Znając czasy transmisji ramek i czas wykrycia nośnej sygnału, można obliczyć współczynniki *a* dla obu przypadków. Na rys. 1.69 przedstawiono wyniki uzyskane dla a_1=0,1 dla całej sieci, a na rys. 1.70 – dla poszczególnych stacji. Na rys. 1.71 przedstawiono wyniki uzyskane dla a_1=0,016 dla całej sieci, a na rys. 1.72 – dla poszczególnych stacji.

Otrzymane wyniki potwierdziły słuszność stosowania mechanizmu wykrywania nośnej sygnału. Mechanizm ten – zgodnie z oczekiwaniami – zapewnia znacznie mniejszą liczbę kolizji w sieci i efektywniejsze wykorzystanie łącza. Otrzymane wyniki są zbliżone do wyników teoretycznych. Potwierdza się też zależność efektywności protokołu CSMA od parametru *a*. Jak widać na wykresach, stosowanie dłuższych ramek sprzyja uzyskaniu wyższej wydajności. Z wykresów pokazujących, w jaki sposób przepływ rozkładał się na poszczególne węzły (rys. 1.70 i 1.72), można powiedzieć, że efekt przechwytywania przy zastosowaniu protokołu CSMA również występuje. Efektywna przepustowość kanału nie rozkłada się bowiem równomiernie na wszystkie stacje. W sieci z protokołem CSMA liczba występujących kolizji jest znacznie mniejsza, dlatego efekt przechwytywania występuje w najmniejszym stopniu spośród badanych protokołów.

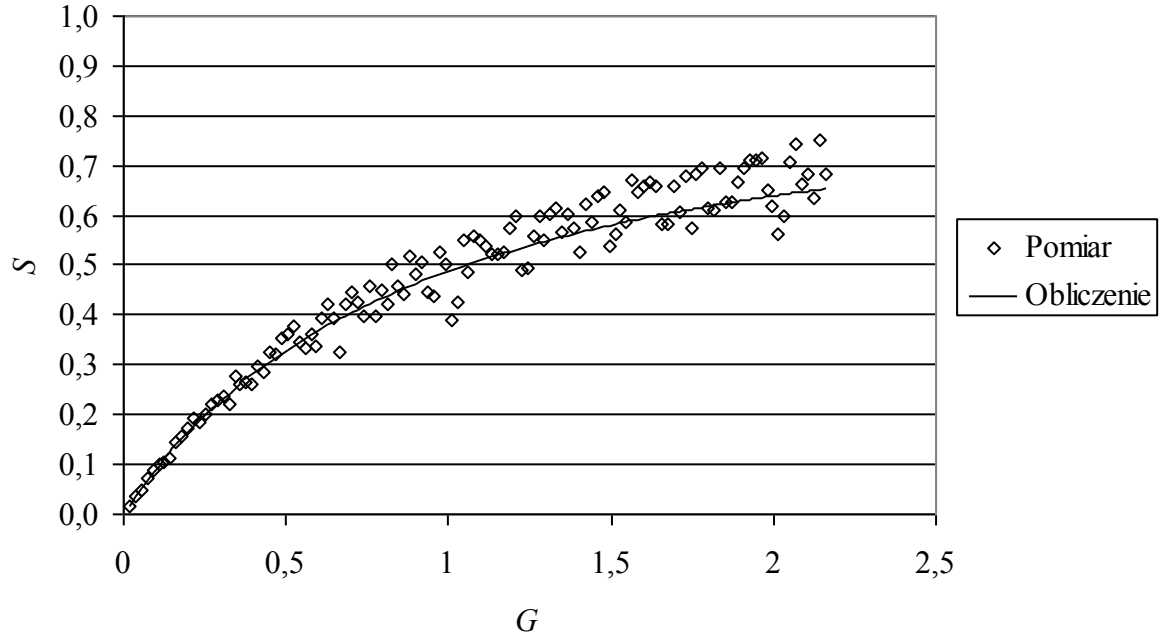
Czas propagacji w badanym segmencie sieci jest dużo mniejszy niż czas wykrycia nośnej sygnału. Można zatem stwierdzić, że jakość tego protokołu jest związana z jakością zastosowanych urządzeń, a w szczególności układu wykrywania nośnej sygnału. Jest to szczególnie widoczne w sieciach o niewielkim zasięgu transmisji, gdyż wtedy czas propagacji jest stosunkowo mały. W sieciach o dużym zasięgu wpływ czasu wykrywania nośnej na wydajność protokołu może być pomijalny.



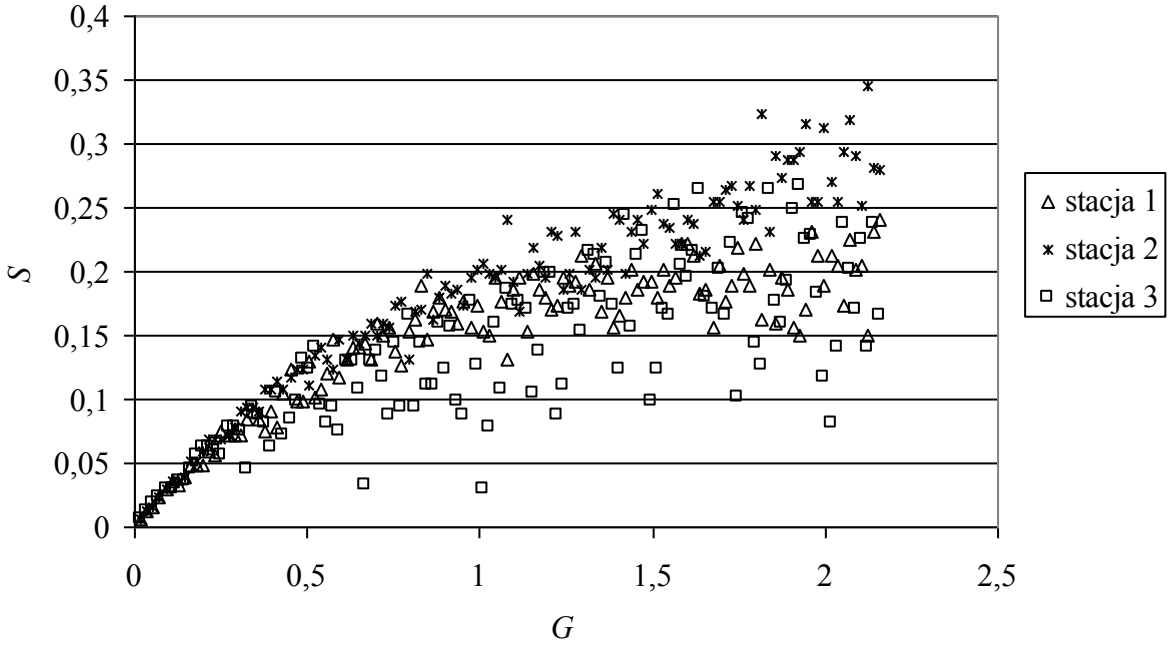
Rys. 1.69. Obliczona i zmierzona efektywność protokołu CSMA (a=0,1)
Fig. 1.69. Calculated and measured CSMA protocol efficiency (a=0,1)



Rys. 1.70. Zmierzona efektywność protokołu CSMA dla poszczególnych stacji (a=0,1)
Fig. 1.70. Measured CSMA protocol efficiency for individual stations (a=0,1)



Rys. 1.71. Obliczona i zmierzona efektywność protokołu CSMA (a=0,016)
Fig. 1.71. Calculated and measured CSMA protocol efficiency (a=0,016)



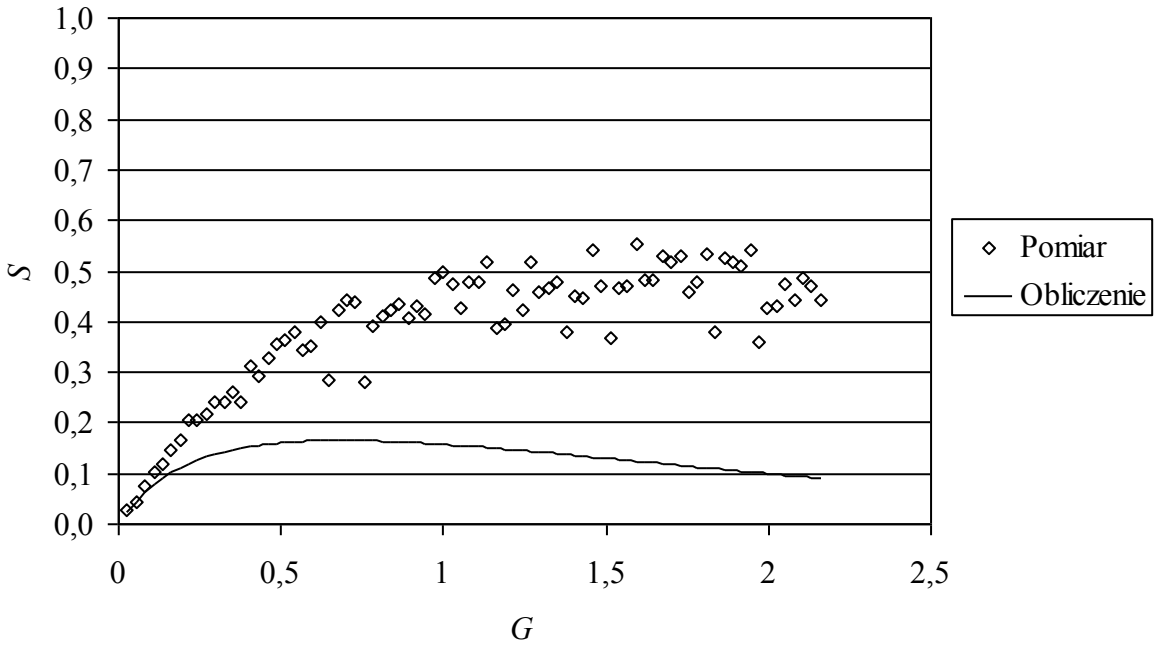
Rys. 1.72. Zmierzona efektywność protokołu CSMA dla poszczególnych stacji (a=0,016)
Fig. 1.72. Measured CSMA protocol efficiency for individual stations (a=0,016)

1.5.3.7. Badanie protokołu MACA

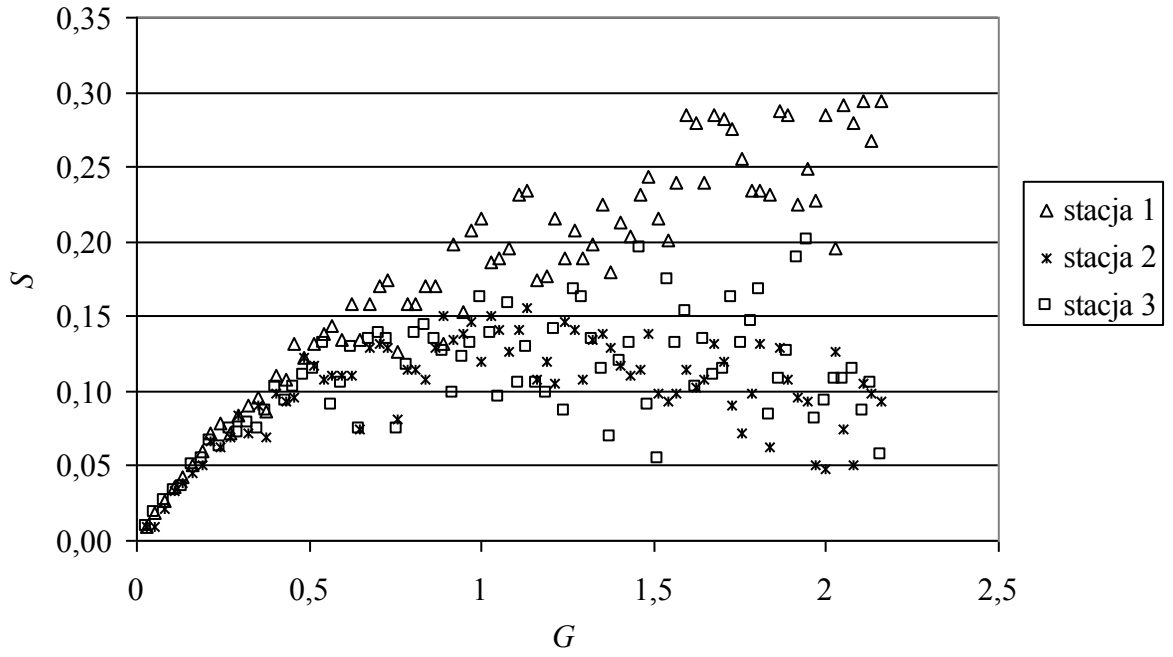
Efektywność protokołu MACA, zgodnie z obliczeniami teoretycznym (rozdz. 1.5.1.4), zależy od parametru *a*, który określa związek między czasem wykrycia nośnej sygnału i czasem transmisji ramki, a także od parametru *b*, określającego zależność między czasem transmisji ramki sterującej (RTS i CTS) i ramki danych. W zastosowanych urządzeniach do transmisji bezprzewodowej czas wykrycia nośnej sygnału wynosił 3 ms. Przyjęto długość ramki sterującej równą 14 B. Efektywność protokołu zbadano dla dwóch długości ramek danych: 28 B i 178 B. Badania przeprowadzono przy prędkości transmisji 9,6 kb/s. Czasy transmisji ramek danych wynosiły zatem odpowiednio 29 ms i 185 ms, natomiast ramek sterujących – ok. 15 ms. Znając czasy transmisji ramek i czas wykrycia nośnej sygnału, można obliczyć współczynniki *a* i *b* dla obu przypadków. Na rys. 1.73 przedstawiono wyniki uzyskane dla a_1=0,1 i b_1=0,5 dla całej sieci, a na rys. 1.74 – dla poszczególnych stacji. Na rys. 1.75 przedstawiono wyniki uzyskane dla a_2=0,016 i b_2=0,08 dla całej sieci, a na rys. 1.76 – dla poszczególnych stacji.

Badania wykazały, że protokół MACA dla krótkich ramek danych wykazuje właściwości zbliżone do protokołu Aloha. Większa efektywność protokołu MACA, w stosunku do uzyskanych wyników z badań Aloha, wynika z faktu, że ramki RTS i CTS są traktowane w pewnym sensie jak część ramki danych. Ponieważ badania przeprowadzono w warunkach laboratoryjnych, gdzie nie ma stacji ukrytych, stacja po otrzymaniu ramki RTS wstrzymywała transmisję. Ramki CTS i danych są przesyłane zatem praktycznie bezkolizyjnie. Podobnie jak w przypadku Aloha zaobserwowano duży wpływ efektu przechwytywania, poprawiającego całkowitą efektywność i powodującego niesprawiedliwy podział łącza. Badania potwierdziły, że zastosowanie długich ramek danych znacznie poprawia efektywność MACA. Zachowanie takie można uznać za typowe dla wszystkich protokołów rywalizacyjnych, które wykorzystują jakikolwiek mechanizm unikania kolizji. Protokół zaczyna wykazywać właściwości zbliżone do protokołu CSMA, zapewniając też bardziej sprawiedliwy podział łącza, nawet przy występującym efekcie przechwytywania.

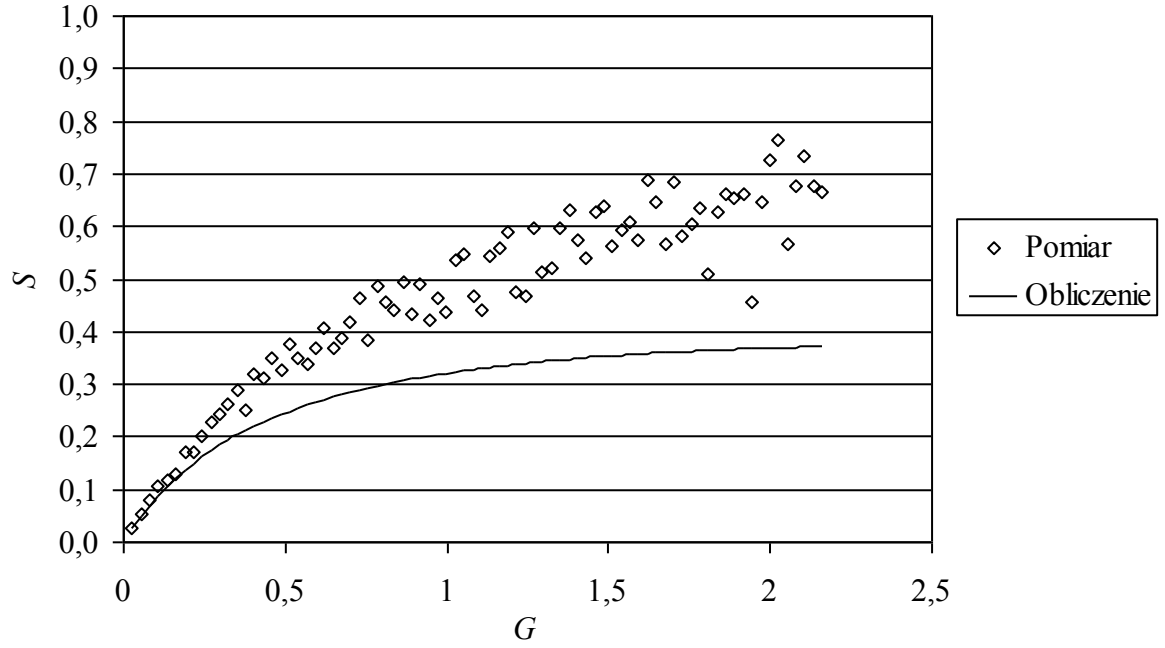
Zbudowanie segmentu sieci bezprzewodowej umożliwiło zbadanie i porównanie czterech protokołów dostępu do łącza bezprzewodowego. Otrzymane wyniki różniły się, w zależności od protokołu, w większym lub mniejszym stopniu od wyników teoretycznych. W badanym segmencie sieci występował efekt przechwytywania, którego nie uwzględniają podstawowe modele matematyczne. Efekt przechwytywania powodował zwiększenie efektywności wykorzystanie łącza, które niestety nie rozkładało się równomiernie na wszystkie węzły. Możliwe jest uwzględnienie tego zjawiska, jak również innych występujących w sieciach bezprzewodowych w modelach matematycznych. Wymaga to jednak poznania ich i określenia, jak duży wpływ mają na efektywność sieci.



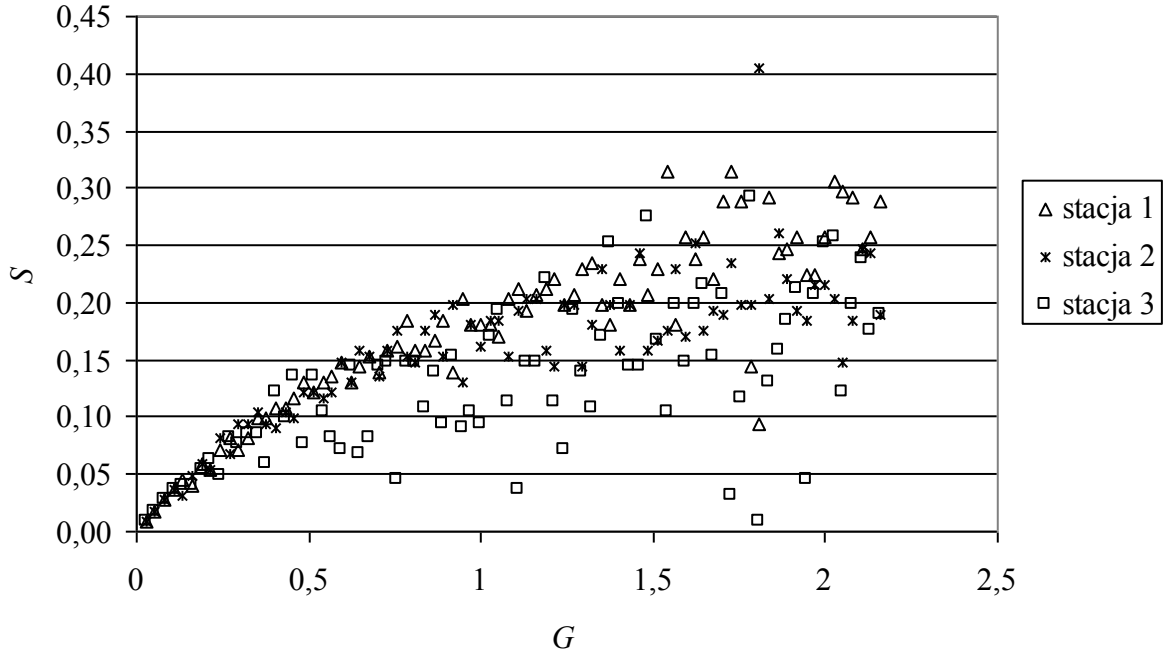
Rys. 1.73. Obliczona i zmierzona efektywność protokołu MACA (a=0,1, b=0,5)
Fig. 1.73. Calculated and measured MACA protocol efficiency (a=0,1, b=0,5)



Rys. 1.74. Zmierzona efektywność protokołu MACA dla poszczególnych stacji (a=0,1, b=0,5)
Fig. 1.74. Measured MACA protocol efficiency for individual stations (a=0,1, b=0,5)



Rys. 1.75. Obliczona i zmierzona efektywność protokołu MACA (a=0,016, b=0,08)
Fig. 1.75. Calculated and measured MACA protocol efficiency (a=0,016, b=0,08)



Rys. 1.76. Zmierzona efektywność protokołu MACA dla poszczególnych stacji (a=0,016, b=0,08)
Fig. 1.76. Measured MACA protocol efficiency for individual stations (a=0,016, b=0,08)

1.6. Podsumowanie rozdziału

W niniejszym rozdziale opisano zjawiska, występujące w sieciach bezprzewodowych, a mające istotny wpływ na działanie protokołu dostępu do łącza. Następnie przedstawiono metody unikania i wykrywania kolizji, możliwe do zrealizowania w sieciach bezprzewodowych. Na podstawie tej analizy określono warunki, w których wykrywanie kolizji – pomimo występowania efektu przechwytywania – jest możliwe w sieciach wykorzystujących promieniowanie podczerwone z wiązką rozproszoną. Porównano także zachowanie dwóch metod unikania kolizji w sieci ad hoc zawierającej stacje ruchome i określono kryterium skuteczności unikania kolizji metodą wymiany ramek sterujących. Kolejny fragment rozdziału przedstawia rywalizacyjne protokoły dostępu do łącza, zaprojektowane dla sieci bezprzewodowych. W protokołach tych są stosowane opisane wcześniej metody unikania i wykrywania kolizji. Dla wybranych protokołów przeprowadzono analizę wydajności w różnych warunkach pracy sieci, włączając warunki typowe dla kilku przypadków istniejących sieci bezprzewodowych. Wybrane protokoły zostały także zaimplementowane w małej, doświadczalnej sieci bezprzewodowej, w której dokonano pomiaru ich wydajności dla kilku wybranych konfiguracji. Uzyskane wyniki doświadczalne odbiegają nieco od wyników analitycznych, co może świadczyć o występowaniu w sieci zjawisk, które nie zostały uwzględnione w modelu, np. efektu przechwytywania.

Do najważniejszych, oryginalnych fragmentów rozdziału można zaliczyć:

- określenie warunków, w których wykrywanie kolizji w sieci z podczerwienią rozproszoną może być możliwe,
- zdefiniowanie kryterium skuteczności działania unikania kolizji metodą wymiany ramek sterujących poprzedzającej transmisję informacji,
- dogłębne porównanie wydajności protokołów dostępu do łącza w różnych warunkach, w tym warunkach typowych dla wybranych sieci bezprzewodowych, na podstawie dostępnych w literaturze modeli analitycznych,
- pomiar wydajności wybranych protokołów dostępu do łącza w małej, doświadczalnej sieci bezprzewodowej.

2. PROTOKÓŁ AX.25

Protokół AX.25 [6] jest stosowany jako warstwa liniowa w radioamatorskiej sieci Packet Radio. Sieć ta powstała na przełomie lat 70. i 80. XX wieku [63] i jest do dziś używana w wielu krajach, także w Polsce. Co prawda, popularność telefonii komórkowej i Internetu powoduje znaczny spadek zainteresowania siecią Packet Radio, jednak stwarza też możliwość powstania nowych zastosowań. Przykładowo, sieć ta może stanowić bazę dla przesyłu danych o charakterze telemetrycznym przy użyciu protokołu APRS (ang. *Automatic Position Reporting System*) [109]. Niezależnie jednak od zastosowań, sieć Packet Radio – a w szczególności stosowany w niej protokół oraz wyposażenie transmisyjne – charakteryzuje się wieloma interesującymi właściwościami, może także stanowić przykład realizacji pewnych rozwiązań [127, 154].

Sieć Packet Radio pracuje w kilkunastu pasmach częstotliwości radioamatorskich z zakresu fal krótkich i ultrakrótkich [25, 26]. Fale krótkie – ze względu na swoje właściwości fizyczne [7, 99] – pozwalają na osiągnięcie większego zasięgu transmisji kosztem jej jakości; prędkość transmisji jest niska (0,3 kb/s), a częste błędy transmisji utrudniają efektywny przesył danych [25]. Fale ultrakrótkie pozwalają na uzyskanie znacznie lepszej jakości transmisji kosztem zasięgu. W przypadku fal ultrakrótkich początkowo stosowano prędkość 1,2 kb/s, co było podyktowane zarówno właściwościami nadajników-odbiorników radiowych (większe prędkości transmisji wymagały ich modyfikacji) oraz dostępnością układów scalonych (obecne wówczas na rynku modemy nie pozwalały na wyższe prędkości transmisji). Dla obu prędkości transmisji zastosowano różne typowe warianty modulacji częstotliwości – Bell 202 dla 1,2 kb/s oraz Bell 103 dla 0,3 kb/s. Jednocześnie prowadzono próby zwiększenia prędkości transmisji przez zmianę metody modulacji oraz modyfikację układów nadajników-odbiorników radiowych. Dzięki czterowartościowej modulacji fazy (QPSK, ang. *Quadrature Phase Shift Keying*) udało się uzyskać prędkość 2,4 kb/s. Pomimo iż rozwiązanie to może pracować z typowym nadajnikiem-odbiornikiem, nie jest obecnie stosowane, ponieważ dostępne układy radiowe umożliwiają połączenie z pominięciem filtrów wejściowych, ograniczających przenoszone pasmo częstotliwości do ok. 3 kHz, a co za tym idzie, także prędkość transmisji. W takim przypadku, przy zastosowaniu odpowiedniego modemu, można uzyskać prędkości

rzędu 9,6 kb/s, a nawet wyższe, sięgające kilkuset kb/s, o ile pozwala na to szerokość pasma radiowego oraz wydajność sprzętu używanego do transmisji [26].

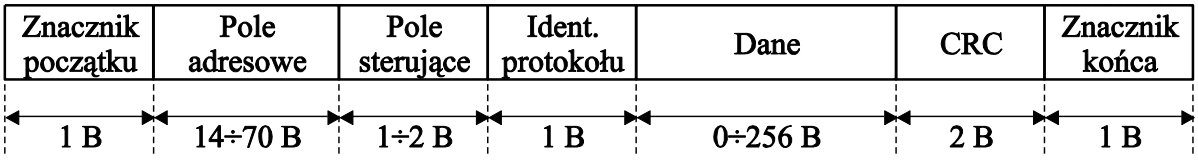
2.1. Opis protokołu AX.25

Protokół AX.25 należy do rodziny protokołów HDLC [56]. Można go uznać za modyfikację odmiany LAPB (ang. *Link Access Protocol Balanced*) [57], ponieważ wykorzystuje te same typy ramek i rodzaje połączeń. Różnice obejmują elementy charakterystyczne dla łączności radioamatorskiej. W szczególności, znacznie zwiększono rozmiar pola adresowego, a także wprowadzono identyfikator protokołu oraz dodatkowe typy ramek. Tak uzyskany protokół AX.25, pomimo że stanowi warstwę liniową sieci Packet Radio, zawiera pewne elementy charakterystyczne dla warstwy sieciowej (oznaczenie trasy przesyłu ramek) i transportowej (sterowanie przepływem). Należy jednak mieć na uwadze, iż mechanizmy te są realizowane w sposób znacznie uproszczony, a ich występowanie wywodzi się z protokołu HDLC. Mimo to protokół AX.25 może być jedynym protokołem w sieci (bez wyższych warstw), może także stanowić warstwę liniową w sieciach bardziej złożonych, np. wykorzystujących stos protokołów TCP/IP.

Protokół AX.25 opracowano w celu zapewnienia niezawodnej transmisji pomiędzy stacjami sieci, niezależnie od sposobu realizacji warstwy fizycznej, protokołów wyższych warstw, rodzaju łącza i liczby połączeń logicznych. Przesył informacji użytkownika jest możliwy niezależnie od nawiązania połączenia logicznego. Poniższy opis protokołu oparty jest na wersji 2.2 [6] ze wskazaniem ważniejszych różnic w stosunku do wersji 2.0 [25]. Warto zauważyć, że znaczna liczba odmian stosowanego obecnie oprogramowania nie ma możliwości wykorzystania mechanizmów wprowadzonych w wersji 2.2.

2.1.1. Format ramki

Ramki używane w protokole AX.25 oparte są na formacie stosowanym w protokole HDLC, z uwzględnieniem charakteru łączności radioamatorskiej. Ogólny format ramki przedstawiono na rys. 2.1.



Rys. 2.1. Ogólny format ramki protokołu AX.25
Fig. 2.1. General format of AX.25 protocol frame

Wszystkie pola ramki są przesyłane począwszy od bitu najmniej znaczącego, z wyjątkiem pola CRC, przesyłanego w kolejności odwrotnej.

Niektóre ramki są traktowane jako polecenia (ang. *command*), na które należy udzielić odpowiedzi (ang. *response*).

2.1.1.1. Ograniczniki ramki

Ramka rozpoczyna się i kończy 8-bitowymi ogranicznikami o wartości 7E szesnastkowo (01111110 dwójkowo). Jest to jedyne miejsce w ramce, w którym może wystąpić kolejno sześć bitów o wartości '1'. Dla zapewnienia przezroczystości protokołu, ciąg bitów zawarty między ogranicznikami jest poddawany tzw. szpikowaniu zerami (ang. *bit stuffing*). Polega ono na automatycznym wprowadzaniu bitu o wartości '0' po każdym ciągu pięciu bitów o wartości '1'. Operacja ta, przeprowadzana w nadajniku, zwiększa długość ramki średnio o 1/63 (dodaje się średnio jeden bit na 62 bity oryginalnej informacji) [59]. Po stronie odbiorczej natomiast każdy tak wprowadzony dodatkowy bit jest automatycznie usuwany.

Ograniczniki ramki są także przesyłane wówczas, gdy zachodzi konieczność utrzymania nadawania, a nie ma gotowej do wysłania ramki.

2.1.1.2. Pole adresowe

Pole adresowe ramki jest stosunkowo długie. Wynika to zarówno z długości pojedynczego adresu (7 bajtów), jak i liczby adresów (2÷10). Protokół AX.25 dopuszcza bowiem występowanie aż ośmiu stacji pośredniczących, które umożliwiają przesył informacji między stacjami, znajdującymi się poza swoim zasięgiem, pozbawionych zatem możliwości komunikowania się bezpośrednio. Obecna wersja protokołu (2.2) redukuje liczbę stacji pośredniczących do dwóch, co jest wystarczające w przypadku korzystania z tzw. stacji węzłowych.

Adresem stacji sieci Packet Radio jest znak wywoławczy (ang. *call sign*) operatora, nadawany po uzyskaniu uprawnień radioamatorskich. Znak ten może zawierać do 6 bajtów i składa się z dużych liter i cyfr. Jeśli znak wywoławczy stacji jest krótszy niż 6 bajtów, uzupełnia się go znakami spacji. Znak wywoławczy koduje się przez przesunięcie każdego bajta o jeden bit w lewo. Siódmy bajt adresu zawiera informacje o charakterze sterującym, a jego format jest różny dla pola adresowego nadawcy lub odbiorcy i stacji pośredniczącej. Bajt ten zawiera:

- 4-bitowy identyfikator wtórny (SSID, ang. *Secondary Station Identifier*), pozwalający na rozróżnienie usług dostępnych w ramach jednej stacji (występuje tu pewne podobieństwo do numerów portów w protokole TCP/IP),
- w adresie nadawcy lub odbiorcy – bit C (ang. *Command/Response*), umożliwiający rozróżnienie ramki rozkazu (ang. *command*) i odpowiedzi (ang. *response*),
- w adresie stacji pośredniczącej – bit H (ang. *Has been repeated*), wskazujący, czy ramka była już retransmitowana przez daną stację,

- znacznik końca pola adresowego (X, ang. *address eXtension)*, konieczny ze względu na jego zmienną długość, zależną od liczby stacji uczestniczących w przesyle określonej ramki (2÷10).
- dwa bity zarezerwowane dla przyszłych zastosowań (R, ang. *Reserved*), które, w razie potrzeby, mogą być jednak używane lokalnie w danej sieci.

Strukturę bajta sterującego pola adresowego wyjaśniono w tabeli 2.1.

Tabela 2.1
Struktura bajta sterującego adresu w protokole AX.25

Typ adresu	Bit							
	7	6	5	4	3	2	1	0
Adres nadawcy lub odbiorcy	C	R	R	SSID				X
Adres stacji pośredniczącej	H	R	R	SSID				X

Każda stacja pośrednicząca przed nadaniem ramki zmienia bit H w polu adresowym zawierającym jej adres. Operacja taka wymaga oczywiście ponownego obliczenia sumy kontrolnej. W przypadku gdy w ramce podano więcej adresów stacji pośredniczących, każda kolejna stacja ustawia bit H w odpowiednim polu adresowym. Dzięki temu stacja, która odebrała ramkę już przez siebie wcześniej nadaną, nie będzie nadawać jej ponownie.

2.1.1.3. Pole sterujące ramki

Pole sterujące ramki określa typ ramki oraz ewentualnie jej numer kolejny. W początkowych wersjach protokołu długość tego pola wynosiła 8 bitów, obecnie jest możliwe także używanie pola o długości 16 bitów. Większa długość pola umożliwia stosowanie rozszerzonej numeracji ramek – modulo 128 wobec numeracji modulo 8 przy mniejszym rozmiarze pola. Wybór numeracji jest dokonywany podczas nawiązywania połączenia. Interpretacja poszczególnych bitów pola zależy od typu pola sterującego ramki – wyjaśnia to tabela 2.2 (pole sterujące 8-bitowe) oraz 2.3 (pole sterujące 16-bitowe).

Tabela 2.2
Struktura 8-bitowego pola sterującego protokołu AX.25

Typ pola sterującego	Bit pola sterującego							
	7	6	5	4	3	2	1	0
Ramka informacyjna	N(R)			P	N(S)			0
Ramka zarządzająca	N(R)			P/F	S	S	0	1
Ramka nienumerowana	M	M	M	P/F	M	M	1	1

W tabelach 2.2 i 2.3 przyjęto następujące oznaczenia:

- N(R) – kolejny numer odebrany (ang. *Received Sequence Number*);
- N(S) – kolejny numer nadawany (ang. *Send Sequence Number*);
- S – bity określające typ ramki zarządzającej;
- M – bity modyfikujące treść ramki nienumerowanej;

- P/F – bit pytanie-zakończenie (ang. *Poll/Final*), interpretowany w kontekście stanu bitu C w polu adresowym, a oznaczany i używany jako bit P (ang. *poll*) w ramkach rozkazu i jako F (ang. *final*) w ramkach odpowiedzi.

Tabela 2.3
Struktura 16-bitowego pola sterującego protokołu AX.25

Typ pola sterującego	Bit pola sterującego															
	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Ramka informacyjna	N(R)							P	N(S)							0
Ramka zarządzająca	N(R)							P/F	0	0	0	0	S	S	0	1

2.1.1.4. Identyfikator protokołu

Identyfikator protokołu (PID, ang. *Protocol Identifier*) umożliwia określenie, czy AX.25 jest jedynym stosowanym protokołem, czy też jest tylko nośnikiem informacji wymienianej przy użyciu protokołów wyższych warstw. W drugim z wymienionych przypadków PID określa typ protokołu warstwy 3 modelu ISO/OSI, tj. warstwy sieciowej. Identyfikator protokołu występuje tylko w ramkach informacyjnych (I) oraz informacyjnych nienumerowanych (UI), ponieważ pozostałe ramki nie przenoszą informacji z wyższych warstw.

2.1.1.5. Pole informacyjne

Pole informacyjne występuje tylko w ramkach: informacyjnych (I), informacyjnych nienumerowanych (UI), identyfikacyjnych (XID), testowych (TEST) oraz sygnalizujących odrzucenie ramki (FRMR). Maksymalna długość pola wynosi 256 bajtów, może jednak być dodatkowo ograniczona przez ustawienie odpowiednich parametrów.

Pole informacyjne ramek I oraz UI służy do wymiany informacji przychodzącej z wyższych warstw sieci. W ramkach XID pole służy do przekazania dopuszczalnych dla danej stacji wartości parametrów, zaś w ramkach FRMR niesie informację o przyczynach odrzucenia ramki. Znaczenie pola informacyjnego ramki TEST nie jest określone.

2.1.1.6. Suma kontrolna

Pole sumy kontrolnej (FCS, ang. *Frame Check Sequence*) pozwala sprawdzić, czy ramka została odebrana bezbłędnie. Ramka, której suma kontrolna obliczona i odebrana nie są równe, zostaje odrzucona, a do jej nadawcy zostaje wysłana prośba o retransmisję. Wartość pola jest obliczana na podstawie wielomianu $G(x) = x^{16} + x^{12} + x^5 + 1$, zgodnie z algorytmem CRC-CCITT [56].

2.1.2. Typy ramek

Wśród ramek w protokole AX.25 wyróżnia się ramki informacyjne, zarządzające i nienumerowane. Zasady ich użycia omówiono dokładniej w rozdziale 2.1.4.

2.1.2.1. Ramki informacyjne

Ramki informacyjne (**I**, ang. *Information*) przenoszą dane użytkownika oraz informacje wewnętrzne protokołów pracujących na poziomie wyższych warstw sieci. W polu sterującym ramki informacyjnej – niezależnie od jego rozmiaru – są przenoszone dwa numery. Numer N(S) oznacza kolejny numer ramki po stronie nadawczej, N(R) natomiast – po stronie odbiorczej. Numery te służą do identyfikacji ramek i ich potwierdzania. Za pomocą ramki informacyjnej można zatem potwierdzić prawidłowy odbiór niepotwierdzonych jeszcze odebranych wcześniej ramek informacyjnych.

2.1.2.2. Ramki zarządzające

Ramki zarządzające (ang. *supervisory*) służą do potwierdzania odebranych ramek, zgłaszania gotowości odbiorcy i żądania retransmisji.

Ramka **RR** (ang. *Receiver Ready*), zawierająca numer N(R), potwierdza prawidłowy odbiór wszystkich ramek informacyjnych o numerze nieprzekraczającym N(R)-1. Ponadto, zgłasza gotowość do odbioru kolejnych ramek informacyjnych, włączając w to gotowość osiągniętą po okresie zajętości, może także służyć do zapytania o stan innej stacji.

Ramka **RNR** (ang. *Receiver Not Ready*), zawierająca numer N(R), także potwierdza prawidłowy odbiór wszystkich ramek informacyjnych o numerze nieprzekraczającym N(R)-1. Ponadto, zgłasza brak gotowości do odbioru kolejnych ramek informacyjnych. Jeśli nadawca wysłał ramki o numerze przekraczającym N(R)-1, powinien je retransmitować po osiągnięciu gotowości odbiorcy.

Ramka **REJ** (ang. *Reject*), zawierająca numer N(R), potwierdza prawidłowy odbiór wszystkich ramek informacyjnych o numerze nieprzekraczającym N(R)-1 i żąda ponowienia transmisji ramek o numerach N(R) i większych. Dla danego kierunku transmisji dopuszczalny jest tylko jeden stan odrzucenia ramki; wyjście z tego stanu jest możliwe po bezbłędnym przesłaniu ramki o numerze N(R), której błędny odbiór spowodował wcześniejsze odrzucenie. Ramka REJ może także oznaczać nadejście duplikatu ramki.

Ramka **SREJ** (ang. *Selective Reject*), zawierająca numer N(R), jest używana w celu żądania ponownego przesłania pojedynczej ramki informacyjnej o numerze N(R). Jeśli przy tym w ramce SREJ ustawiono bit P/F, ramki o numerach mniejszych niż N(R) zostają potwierdzone. Odbiorca może wysłać kilka ramek SREJ (z wyzerowanym bitem P/F), o ile zawierają one różne numery N(R). Ramki SREJ nie są wysyłane w stanie odrzucenia ramki (REJ); podobnie ramki REJ nie wysyła się w stanie selektywnego odrzucenia (SREJ). W odpowiedzi na ramkę SREJ nadawca powinien ponowić przesłanie tylko ramki o numerze N(R), pomijając przy tym wysłane wcześniej ramki następne. Można jednak za retransmitowaną ramką przesłać kolejne nieprzesyłane wcześniej ramki. Ramkę SREJ wprowadzono w wersji 2.2 protokołu AX.25.

2.1.2.3. Ramki nienumerowane

Ramki nienumerowane służą do przekazywania dodatkowych informacji sterujących; większość z nich jest używana podczas nawiązywania i rozwiązywania połączeń.

Ramka **SABM** (ang. *Set Asynchronous Balanced Mode*) jest żądaniem nawiązania połączenia logicznego między nadawcą a adresatem. Adresat powinien odpowiedzieć ramką UA, jeśli połączenie jest możliwe lub DM, jeśli stan adresata nie pozwala na nawiązanie połączenia. Po nawiązaniu połączenia dopuszcza się przesyłanie ramek wyłącznie z 8-bitowym polem sterującym.

Ramka **SABME** (ang. *Set Asynchronous Balanced Mode Extended*) jest zbliżona do ramki SABM, jednak po nawiązaniu połączenia ramki – zależnie od ich typu – mogą zawierać 8- lub 16-bitowe pole sterujące. Jeśli adresat pracuje zgodnie z wcześniejszą niż 2.2 wersją protokołu, powinien odpowiedzieć ramką FRMR. Pozostałe zasady są takie same jak w przypadku ramki SABM.

Ramka **DISC** (ang. *Disconnect*) jest żądaniem rozwiązania połączenia logicznego. Przed wykonaniem tego polecenia odbiorca powinien potwierdzić jego odbiór ramką UA, nadawca ramki DISC rozwiązuje bowiem połączenie dopiero po odebraniu tej odpowiedzi.

Ramka **UA** (ang. *Unnumbered Acknowledge*) jest specjalnym potwierdzeniem, używanym jako odpowiedź na wysłane ramki SABM, SABME i DISC.

Ramka **DM** (ang. *Disconnected Mode*) jest wysyłana, gdy stacja pozostająca w stanie rozłączenia odbierze ramkę inną niż SABM, SABME lub UI. Ponadto, ramka DM może być odpowiedzią na próbę nawiązania połączenia, którego odbiorca nie może zaakceptować.

Ramka **UI** (ang. *Unnumbered Information*) służy do przenoszenia informacji, gdy nie występuje konieczność potwierdzania lub sterowania przepływem. Ramki tego typu są także używane do przesyłu informacji bez nawiązanego połączenia logicznego. Przesłanie ramki UI z ustawionym bitem P/F wywołuje przesłanie odpowiedzi ramką DM, jeśli adresat jest rozłączony lub ramkami RR albo RNR, jeśli adresat jest w stanie połączenia z nadawcą.

Ramka **XID** (ang. *Exchange Identification*), wprowadzona w wersji 2.2 protokołu AX.25, wymusza identyfikację i przesłanie cech odbiorcy (także za pomocą ramki XID). Przesłanie odpowiedzi jest wstrzymywane, gdy adresat nawiązuje lub rozwiązuje połączenie lub odebrał nieznaną ramkę (stan FRMR). Stacja odbierająca ramkę XID może odpowiedzieć ramką FRMR, jeśli długość pola informacyjnego przekracza możliwości odbiorcy lub pracuje on z protokołem w wersji nie późniejszej niż 2.0.

Pole informacyjne ramki XID zawiera elementy informacyjne. Każdy z nich rozpoczyna się identyfikatorem formatu (FI, ang. *Format Identifier*), po którym następuje identyfikator grupy (GI, ang. *Group Identifier*) oraz określenie długości grupy (GL, ang. *Group Length*). Pola identyfikatorów są jedno-, a długości – dwubajtowe. Pozostałą część elementu informa-

cyjnego stanowią struktury, zawierające identyfikator parametru (PI, ang. *Parameter Identifier*), długość parametru (PL, ang. *Parameter Length*) oraz wartość parametru (PV, ang. *Parameter Value*). Pola PI oraz PL są jednobajtowe. Pole PV ma długość określoną przez wartość pola PL i może być interpretowane jako liczba lub pole bitowe. Wśród negocjowanych parametrów można znaleźć m. in:

- sposób uzyskania dwukierunkowości łącza (naprzemienne i jednoczesne),
- sposób retransmisji (klasyczna, selektywna i mieszana),
- sposób numeracji ramek (podstawowa i rozszerzona),
- maksymalną długość pola informacyjnego ramki (parametr N_1),
- maksymalną wielkość okna (parametr k),
- maksymalny czas czekania na potwierdzenie (parametr T_1),
- maksymalną liczbę powtórzeń (parametr N_2).

Ramka **TEST** służy do sprawdzenia połączenia. Stacja, która odebrała taką ramkę, powinna odpowiedzieć także ramką TEST. W ramce dopuszczalne jest pole informacyjne, jednak jak dotąd nie określono jego znaczenia.

Ramka **FRMR** (ang. *Frame Reject*) jest wysyłana, gdy adresat stwierdził błąd, którego nie można usunąć przez retransmisję. Zazwyczaj sytuacja taka występuje, gdy odebrano ramkę [25]:

- o nieznanej wartości pola sterującego (nieznanego typu),
- o niewłaściwej długości pola informacyjnego,
- informacyjną z nieważną wartością N(S) lub N(R),
- zawierającą niedozwolone dla jej typu pole informacyjne,
- zarządzającą z ustawionym bitem P/F, gdy jest ona odpowiedzią na ramkę z wyzerowanym bitem P/F,
- typu UA lub DM w nieoczekiwanym momencie.

Ramka FRMR zawiera 3-bajtowe pole informacyjne, określające przyczynę odrzucenia ramki. Pole to zawiera m. in. pole sterujące odrzuconej ramki, stan liczników ramek nadanych i odebranych (V(R) oraz V(S)) stacji nadającej FRMR oraz bity, wskazujące dokładny powód odrzucenia ramki (np. nieznany typ ramki, niewłaściwy numer ramki lub niedopuszczalna długość pola informacyjnego). Strukturę tego pola przedstawiono w tabeli 2.4 [6].

W tabeli 2.4 przyjęto następujące oznaczenia:

- V(R) – zmienna stanu odbioru (por. rozdział 2.1.3.1),
- V(S) – zmienna stanu nadawania,
- C – bit określający, czy odrzucona ramka była rozkazem czy odpowiedzią,
- W – wskazuje, czy rozkaz był nieważny lub niewykorzystywany,
- X – wskazuje wystąpienie niedozwolonego pola informacyjnego,

- Y – wskazuje niedopuszczalną długość pola informacyjnego,
- Z – wskazuje odbiór ramki o niedopuszczalnej wartości N(R).

Tabela 2.4
Struktura pola informacyjnego ramki FRMR

Bajt	Bit							
	7	6	5	4	3	2	1	0
0	pole sterujące odrzuconej ramki							
1	V(R)			C	V(S)			0
2	0	0	0	0	Z	Y	X	W

Wersja 2.2 protokołu – w przeciwieństwie do wcześniejszych – nie dopuszcza wysyłania ramki FRMR, potrafi jednak rozpoznać ten typ ramki. Błędy wymienione powyżej można bowiem skuteczniej rozwiązać przez ponowne ustanowienie (ang. *reset*) połączenia, wymieniając ramki SABM lub SABME oraz UA.

2.1.3. Liczniki, zegary i parametry protokołu

Dla właściwego działania protokołu zdefiniowano kilka liczników, służących do odmierzania czasu bądź numerowania ramek. Niektóre z nich przeniesiono z definicji protokołu LAPB, pozostałe natomiast wprowadzono dla uwzględnienia specyfiki łączności radioamatorskiej.

2.1.3.1. Liczniki ramek

W protokole występują liczniki ramek, wprowadzone w celu realizacji ich numeracji oraz odnajdywania ramek zagubionych, np. wskutek błędów transmisji.

Zmienna stanu nadawania V(S) (ang. *Send State Variable*) zawiera kolejny numer, który będzie przypisany najbliższej wysyłanej ramce informacyjnej. Stan zmiennej jest aktualizowany wraz z wysłaniem każdej ramki. Zmienna ta nie jest przesyłana do innych stacji.

Kolejny numer nadawania N(S) (ang. *Send Sequence Number*) zawiera numer kolejny ramki informacyjnej, która jest właśnie nadawana. Tuż przed rozpoczęciem transmisji zmiennej N(S) przypisuje się wartość V(S). Aktualna wartość N(S) znajduje się w polu sterującym każdej ramki informacyjnej.

Zmienna stanu odbioru V(R) (ang. *Receive State Variable*) zawiera kolejny numer najbliższej oczekiwanej ramki informacyjnej. Wartość zmiennej jest aktualizowana wraz z odbiorem bezbłędnej ramki informacyjnej o numerze równym aktualnej wartości V(R) i nie jest przesyłana do innych stacji.

Kolejny numer odebrany N(R) (ang. *Received Sequence Number*) jest aktualizowany do wartości V(R) przed wysłaniem ramki informacyjnej lub zarządzającej, a następnie wysyłany w polu sterującym takiej ramki. W ten sposób niejawnie potwierdza się prawidłowy odbiór ramek o numerach do N(R)-1 włącznie.

Zmienna stanu potwierdzeń V(A) (ang. *Acknowledge State Variable*) zawiera numer kolejny ostatniej ramki potwierdzonej przez drugą stację uczestniczącą w połączeniu. Przyjmuje się, że V(A)-1 jest równe wartości N(S) ostatniej potwierdzonej ramki informacyjnej. Wartość zmiennej V(A) nie jest przesyłana do innych stacji.

2.1.3.2. Zegary

Zegar potwierdzeń (ang. *Acknowledgement Timer*), oznaczany jako T_1, odlicza czas od chwili wysłania ramki, dzięki czemu nadawca ramki nie czeka nieskończenie długo na odpowiedź. Początkowa wartość licznika powinna zależeć od prędkości transmisji i innych parametrów warstwy fizycznej. Wpływ na nią ma także fakt użycia stacji pośredniczących i ich liczba pomiędzy nadawcą a adresatem ramki.

Zegar opóźnienia odpowiedzi (ang. *Response Delay Timer*), oznaczany jako T_2, ogranicza czas upływający od odebrania ramki informacyjnej do wysłania potwierdzenia. Umożliwia to stacji odbierającej odczekanie na ewentualne pojawienie się kolejnych ramek informacyjnych, dzięki czemu odbiorca może jednocześnie potwierdzić większą liczbę takich ramek. Stosowanie zegara T_2 nie jest wymagane przez opis protokołu, ale jest zalecane dla zwiększenia wydajności. Wydaje się jednak, że zegar T_2 nie jest w ogóle potrzebny, jeśli nadawca oznacza ostatnią wysyłaną ramkę informacyjną przez ustawienie bitu P/F, co wymusza na odbiorcy natychmiastowe przesłanie odpowiedzi. Ponieważ opis protokołu również nie wymaga takiego mechanizmu – a nawet nie wymienia wprost jego stosowania – rozwiązanie tego zagadnienia zależy od rodzaju użytego oprogramowania.

Zegar nieaktywnego łącza (ang. *Inactivity Link Timer*), oznaczany jako T_3, umożliwia stwierdzenie używalności łącza. Po czasie T_3 od otrzymania ostatniej ramki stacja wysyła zapytanie do stacji, z którą ma nawiązane połączenie logiczne. Jeśli stacja odpowie – ramką RR lub RNR, stosownie do stanu – świadczy to o sprawności łącza i braku danych do przesłania. Brak odpowiedzi natomiast można traktować jako „zniknięcie" stacji, co – po wyczerpaniu określonej liczby prób – może spowodować anulowanie połączenia.

Zegar czasu trwania szczeliny (ang. *Slot Time Timer*), oznaczany jako T_{102}, określa długość szczeliny czasowej, używanej przy rywalizacyjnym dostępie do łącza.

Zegar włączenia nadajnika (ang. *Transmitter Startup Timer*), oznaczany jako T_{103}, określa czas, jaki musi upłynąć od chwili włączenia nadajnika do momentu, kiedy jest on gotowy do nadawania informacji. Czas ten pozwala m. in. na ustabilizowanie parametrów nadajnika oraz wykrycie nośnej przez odbiornik i jest ściśle uzależniony od właściwości używanego sprzętu nadawczo-odbiorczego.

2.1.3.3. Parametry protokołu

Licznik znaków w ramce informacyjnej (ang. *Maximum Number of Octets in an I Field*), oznaczany jako N_1, określa maksymalną pojemność pola danych ramki informacyjnej. Parametr ten może przyjmować wartości z zakresu 1÷256.

Licznik powtórzeń (ang. *Maximum Number of Retries*), oznaczany jako N_2, określa maksymalną liczbę prób transmisji w przypadku braku odpowiedzi z wywoływanej stacji. Jest on używany wraz z zegarem T_1.

Licznik niepotwierdzonych ramek informacyjnych (ang. *Maximum Number of I Frames Outstanding*), oznaczany jako *k*, określa maksymalną wielkość okna protokołu i może przyjmować wartości z zakresu 1÷7 w wersjach 2.0 i wcześniejszych, natomiast w wersji 2.2 – jeśli połączenie wykorzystuje rozszerzone pole sterujące i numerowanie ramek – 1÷127.

2.1.4. Zasady wymiany ramek

Podczas działania sieci stacje przesyłają między sobą liczne ramki. Niejako przy okazji – w sposób niewidoczny dla użytkownika – są używane zegary i liczniki, a także są modyfikowane wewnętrzne zmienne protokołu. Wśród zachodzących wymian ramek można wyróżnić nawiązanie połączenia, negocjację parametrów, transmisję danych po nawiązaniu połączenia, ponowne nawiązanie połączenia, rozwiązanie połączenia i transmisję danych w stanie rozłączenia.

2.1.4.1. Nawiązanie połączenia

Procedura nawiązania połączenia rozpoczyna się od wysłania ramki SABM lub SABME, zależnie od pożądanego przez inicjatora połączenia sposobu numeracji ramek. Jednocześnie stacja ta włącza licznik T_1. Jeśli próba połączenia powiedzie się, stacja wywołana odpowiada ramką UA i zeruje wewnętrzne zmienne V(S), V(A) oraz V(R). Po odebraniu ramki UA także stacja wywołująca zeruje te zmienne i wyłącza licznik T_1. Jeśli stacja wywoływana nie odpowie przed upływem czasu T_1, stacja wywołująca ponawia próbę nawiązania połączenia w taki sam sposób aż do osiągnięcia N_2 prób.

Jeśli stacja wywoływana nie może nawiązać połączenia, odpowiada ramką DM. Po odebraniu takiej ramki stacja wywołująca porzuca próbę nawiązania połączenia.

Stacja wywołująca ignoruje wszystkie ramki prócz ramek SABM, DISC, UA oraz DM, wysłanych przez stację wywoływaną.

2.1.4.2. Negocjacja parametrów

Negocjacja parametrów łącza może wystąpić w dowolnym czasie, wydaje się jednak, iż przeprowadzenie jej tuż po zestawieniu połączenia jest najbardziej uzasadnione. W tym celu wykorzystuje się ramki XID. Stacja, która pracuje zgodnie z wcześniejszą niż 2.2 wersją protokołu i nie obsługuje tych ramek, odpowiada ramką FRMR. W takim przypadku stacja, która

wysłała ramkę XID, przyjmuje domniemane wartości parametrów dla współpracy z wcześniejszymi wersjami protokołu. Natomiast odebranie w odpowiedzi ramki XID oznacza, że obie stacje pracują z wersją 2.2 lub późniejszą. Implikuje to m. in. możliwość wykorzystania ramek SREJ oraz segmentacji.

W wersji 2.2 protokołu AX.25 zaimplementowano zarówno negocjację, jak i powiadamianie o wartościach parametrów. O ile przy tym negocjacja wymaga wspólnych ustaleń pomiędzy zaangażowanymi stacjami i wymaga przesłania co najmniej dwóch ramek XID, o tyle powiadamianie jedynie informuje odbiorcę ramki XID o maksymalnych wartościach parametrów. Proces negocjacji rozpoczyna się od wysłania ramki rozkazu XID. Ramka ta zawiera wartości parametrów akceptowalne przez stację wysyłającą. Stacja odbierająca taką ramkę dobiera wartości parametrów akceptowane przez siebie, ale niewykraczające poza limity narzucone przez treść odebranej ramki XID. Wartości te są następnie odsyłane do stacji inicjującej negocjację ramką odpowiedzi XID.

Negocjowane parametry są podzielone na kilka grup, zgodnie z podziałem na poszczególne pola informacyjne.

Pole klasy procedury (ang. *Classes of Procedure*) pozwala ustalić sposób uzyskania dwukierunkowości łącza. Łącze dwukierunkowe jednoczesne można zestawić tylko wówczas, gdy obie zaangażowane stacje obsługują ten rodzaj transmisji. W przypadku braku pola przyjmuje się przez domniemanie łącze dwukierunkowe naprzemienne (ang. *half-duplex*).

Pole funkcji opcjonalnych (ang. *HDLC Optional Functions*) pozwala ustalić możliwość użycia funkcji retransmisji REJ, SREJ lub SREJ-REJ oraz numerację ramek modulo 8 lub 128. Przyjęte wartości parametrów odpowiadają możliwościom obu stacji. W przypadku braku pola przyjmuje się przez domniemanie użycie retransmisji selektywnej (SREJ) oraz numerację ramek modulo 8.

Pole długości pola informacyjnego odbioru (ang. *I Field Length Receive*) informuje inne stacje o maksymalnej długości pola informacyjnego ramki (N_1). Pozostałe stacje nie mogą przekroczyć tej wartości, ale mogą przesyłać ramki krótsze. W przypadku braku pola przyjmuje się N_1=256 B.

Pole wielkości okna odbioru (ang. *Window Size Receive*) informuje inne stacje o maksymalnej wielkości okna (k). Jeśli wynegocjowano numerację modulo 128, można negocjować zmniejszenie wielkości okna w celu ograniczenia zużycia pamięci. Jeżeli używa się funkcji retransmisji selektywnej, wymaga się, by stacja odbierająca mogła buforować k ramek w każdej sytuacji. W przypadku braku pola przyjmuje się k=7.

Pole zegara potwierdzenia (ang. *Acknowledge Timer*) pozwala ustalić wspólny dla obu stacji czas czekania na potwierdzenie (T_1). Jako wspólną przyjmuje się większą z wartości

zadeklarowanych w ramce XID rozkazu i odpowiedzi. W przypadku braku pola przyjmuje się T_1=3000 ms.

Pole retransmisji (ang. *Retries*) pozwala ustalić wspólną dla obu stacji liczbę prób retransmisji (N_2). Jako wspólną przyjmuje się większą z wartości zadeklarowanych w ramce XID rozkazu i odpowiedzi. W przypadku braku pola przyjmuje się N_2=10.

Jeśli adresat ramki rozkazu XID pracuje z wcześniejszą niż 2.2 wersją protokołu, przyjmuje się praktycznie takie same wartości parametrów, jak podano powyżej w opisie poszczególnych pól informacyjnych. Jedyne różnice dotyczą wówczas metody retransmisji (przyjmuje się retransmisję nieselektywną przy użyciu wyłącznie ramek REJ) oraz wielkości okna (przyjmuje się k=4).

Warto zauważyć, iż stosowane w protokole AX.25 elementy negocjacji zostały zaadaptowane z procedur protokołu HDLC. Część z nich nie jest konieczna dla negocjacji cech protokołu AX.25.

2.1.4.3. Wymiana informacji

Po nawiązaniu połączenia stacje mogą wymieniać ramki dowolnego typu.

Stacja wysyłająca ramkę informacyjną nadaje jej numer N(S)=V(S), po czym inkrementuje V(S) i uruchamia zegar T_1. Jeśli ostatnia otrzymana wartość N(R) jest równa V(S)+k, wstrzymuje się transmisję ewentualnych kolejnych ramek w celu uniknięcia przekroczenia dozwolonej wielkości okna. Stacja w stanie zajętości, nieprzyjmująca (tymczasowo) ramek informacyjnych, może jednak wysyłać takie ramki, póki ich adresat nie jest także w stanie zajętości.

Po odebraniu prawidłowej ramki informacyjnej (nieprzekłamanej i z prawidłowym numerem kolejnym), jeśli adresat nie jest zajęty, przyjmuje ramkę i inkrementuje zmienną V(R). Następnie, jeśli ma przygotowaną ramkę informacyjną do wysłania, nadaje jej numer N(R)=V(R) i wysyła ją, potwierdzając tym odebraną ramkę. Wysłana ramka informacyjna może zostać poprzedzona ramką RR z numerem N(R)=V(R). Jeśli natomiast adresat nie ma ramki informacyjnej do wysłania, przesyła tylko potwierdzenie RR po ewentualnym odczekaniu pewnego krótkiego czasu dla upewnienia się, że nadawca zakończył już transmisję.

Jeśli adresat ramki informacyjnej jest zajęty, ignoruje wszystkie ramki informacyjne, odsyłając potwierdzenie RNR z odpowiednim numerem, wskutek czego przesył danych zostaje tymczasowo wstrzymany. Stacja, która zgłosiła zajętość, może być okresowo próbkowana ramkami RR lub RNR z ustawionym bitem P. Powinna ona odpowiadać również ramkami RR lub RNR, stosownie do sytuacji, z ustawionym bitem F.

Ramki potwierdzeń mają wyższy priorytet niż pozostałe ramki. Uzyskuje się to przez wysłanie potwierdzenia natychmiast po zwolnieniu łącza.

Podczas wymiany informacji może się zdarzyć, że zostanie odebrana bezbłędna ramka informacyjna o numerze N(S)≠V(R). Niezgodność tych wartości świadczy o utracie co najmniej jednej ramki. W tej sytuacji odbiorca podejmuje akcję, zależną od stosowanej procedury retransmisji.

W przypadku stosowania „klasycznej" retransmisji odbiorca wysyła ramkę REJ z numerem N(R)=V(R), czyli równym numerowi ostatnio odebranej prawidłowej ramki powiększonemu o 1. Operacja taka wymusza retransmisję, począwszy od ramki o tak uzyskanym numerze N(R), w tym być może ramek, które poprzednio były przesłane bezbłędnie. Oczywiście, zmniejsza to wydajność protokołu, ale upraszcza działanie stacji odbierającej.

W przypadku stosowania retransmisji selektywnej ramka o niewłaściwym numerze N(S) zostaje zachowana, a odbiorca wysyła żądanie ponownego przesłania tylko ramek zagubionych, wysyłając jedną lub kilka ramek SREJ. Ten sposób działania podwyższa wydajność protokołu, gdyż ramki raz przesłane prawidłowo nie są niepotrzebnie retransmitowane. Z drugiej strony wymaga się, aby odbiorca miał możliwość kolejkowania i odtwarzania kolejności ramek, co komplikuje działanie stacji odbierającej i zwiększa jej zapotrzebowanie na pamięć.

W przypadku stosowania retransmisji mieszanej działanie odbiorcy zależy od liczby zagubionych ramek. Jeśli utracono tylko jedną ramkę, stosuje się procedurę retransmisji selektywnej, jeśli więcej – „klasycznej".

Odbiór ramki informacyjnej lub zarządzającej, nawet w stanie zajętości, wymaga sprawdzenia jej numeru N(R). Jest to spowodowane tym, że ramka ta może potwierdzać odbiór wysłanych wcześniej ramek.

Odebrawszy ramkę REJ, stacja może rozpocząć retransmisję, jeśli tylko łącze nie jest zajęte. W czasie pracy na łączu dwukierunkowym jednoczesnym odbiór ramki REJ może nastąpić podczas nadawania; w takim przypadku można przerwać transmisję ramki informacyjnej w celu dokonania retransmisji. Jeśli pozwala na to ustawiona wielkość okna, po przesłaniu ramek retransmitowanych można wysłać kolejne ramki. Jeśli odebrana ramka REJ miała ustawiony bit P, przed retransmisją należy odpowiedzieć – stosownie do sytuacji – ramką RR lub RNR z ustawionym bitem F.

Odebrawszy ramkę SREJ z numerem N(R), stacja powinna retransmitować ramkę informacyjną o tym numerze przy najbliższej okazji. Następnie, w razie potrzeby, można przesyłać kolejne ramki informacyjne. Jeśli ramka SREJ miała ustawiony bit P, potwierdza prawidłowy odbiór ramek o numerze do N(R)-1 włącznie.

Odebrawszy ramkę RNR z numerem N(R), stacja tymczasowo kończy nadawanie, uruchamia licznik T_3 i czeka na sygnalizację zakończenia stanu zajętości adresata. Po upływie czasu T_3 stacja wysyła ramkę rozkazu RR lub RNR z ustawionym bitem P, na którą odbiorca

odpowiada – stosownie do sytuacji – ramką RR, RNR lub REJ z ustawionym bitem F i odpowiednim numerem N(R). Ramki o numerach do N(R)-1 włącznie uważa się za potwierdzone.

Jeśli potwierdzenie (RR, RNR lub REJ) nie nadejdzie w czasie T_1 od chwili wysłania ramki informacyjnej, nadawca wysyła ramkę rozkazu RR lub RNR z ustawionym bitem P i ponownie uruchamia zegar T_1. W przypadku braku odpowiedzi po wyczerpaniu N_2 prób podejmuje się procedurę ponownego nawiązania połączenia.

2.1.4.4. Ponowne nawiązanie połączenia

Ponowne nawiązanie połączenia (ang. *link reset*) jest konieczne w wyniku wystąpienia poważnego błędu transmisji i przywraca początkowy stan połączenia we wszystkich zaangażowanych w nie stacjach. Procedurę tę uruchamia się po odebraniu nieoczekiwanej ramki odpowiedzi UA lub ramki FRMR, pochodzącej ze starszych wersji protokołu.

Ponowne nawiązanie połączenia odbywa się podobnie jak nawiązanie połączenia, tj. przez wymianę ramek SABM lub SABME oraz UA. Jeśli zakończy się ona powodzeniem, stan stacji jest taki sam jak tuż po nawiązaniu połączenia. W przypadku braku odpowiedzi UA stosuje się także te same procedury co podczas nawiązywania połączenia.

2.1.4.5. Rozwiązanie połączenia

Rozwiązanie połączenia jest oczywiście możliwe jedynie w stanie połączenia i może być zainicjowane przez dowolną stację zaangażowaną w to połączenie. Stacja ta wysyła ramkę DISC i włącza zegar T_1. Odbiorca tej ramki odpowiada ramką UA i przechodzi w stan rozłączenia. Inicjator rozłączenia, otrzymawszy ramkę UA lub DM, zatrzymuje zegar T_1 i także przechodzi w stan rozłączenia.

W przypadku braku odpowiedzi po upływie czasu T_1 inicjator rozłączenia ponawia opisaną procedurę co najwyżej N_2 razy. Jeśli mimo to nie otrzyma odpowiedzi, to i tak przechodzi w stan rozłączenia.

2.1.4.6. Stan rozłączenia

W stanie rozłączenia stacja monitoruje otrzymane ramki, odpowiednio reagując na próby nawiązania połączenia rozkazami SABM lub SABME i odpowiadając ramką DM na rozkazy DISC. Może także zainicjować połączenie, a także nadawać i odbierać ramki nienumerowane UI. W przypadku odebrania ramki rozkazu innego typu niż SABM lub SABME, bądź ramki UI z ustawionym bitem P, odpowiada ramką DM z ustawionym bitem F, zaś odebrana ramka zostaje zignorowana.

Przesył ramek UI w stanie rozłączenia umożliwia komunikację między wieloma użytkownikami jednocześnie. Ramki te nie są jednak potwierdzane ani retransmitowane, istnieje zatem ryzyko utraty (części) informacji w wyniku kolizji lub błędów transmisji.

2.1.4.7. Segmentacja i składanie ramek

Segmentację i składanie ramek wprowadzono w wersji 2.2 protokołu AX.25 w celu umożliwienia przesyłu jednostek pochodzących z warstwy sieciowej, mających długość przekraczającą N_1=256 B, jako pojedynczych jednostek warstwy liniowej. Segmentacja polega na podziale jednostki warstwy sieciowej na mniejsze fragmenty (segmenty), z których każdy poprzedzony jest 2-bajtowym nagłówkiem. Nagłówek ten zawiera liczbę segmentów pozostających do wysłania. Znając długość ramki protokołu AX.25 oraz całkowitą liczbę segmentów, można obliczyć wielkość bufora potrzebnego do odebrania całej jednostki warstwy sieciowej. Odbiorca może rozpoznać ramkę zawierającą segment na podstawie zawartości identyfikatora protokołu (PID) ramki. Na tej podstawie potrafi on odtworzyć oryginalną jednostkę warstwy sieciowej.

W przypadku utraty segmentu, np. wskutek błędu transmisji, protokół AX.25 nie podejmuje próby retransmisji. Warstwa sieciowa otrzymuje wówczas komunikat o błędzie.

2.1.5. Stacje pośredniczące i przekaźnikowe

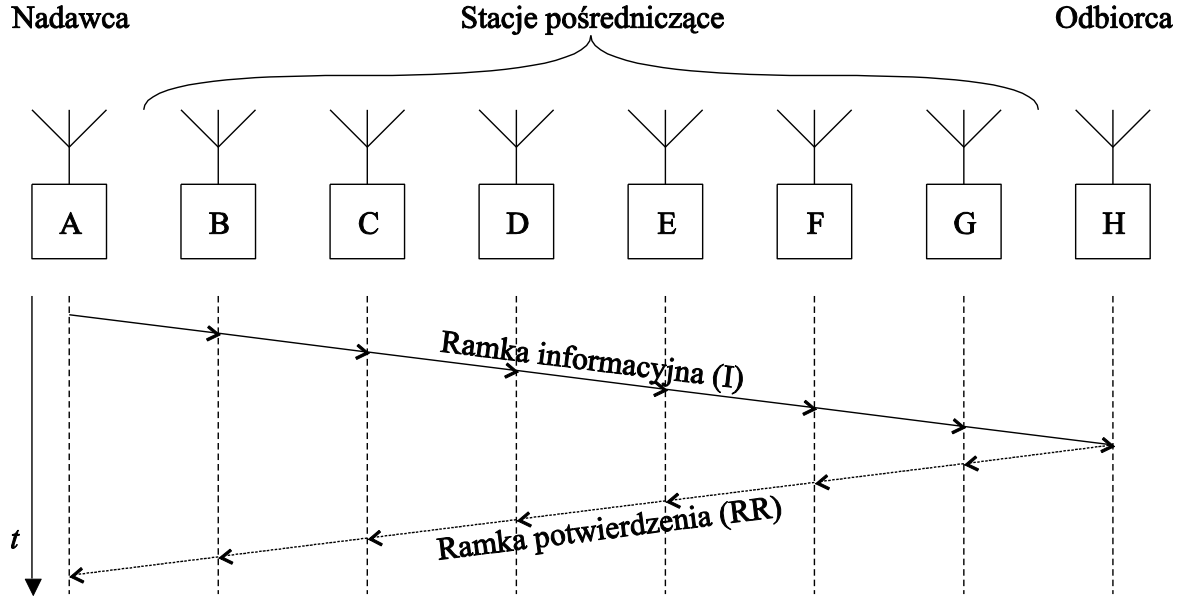
Każda stacja sieci Packet Radio może pełnić funkcję stacji pośredniczącej. Jej zadaniem jest przekazywanie odebranych ramek protokołu AX.25, a następnie ich retransmisja. Dzięki takiemu rozwiązaniu można przesyłać informacje pomiędzy stacjami, które znajdują się poza swoim zasięgiem i nie mogą komunikować się bezpośrednio. Stacje pośredniczące ułatwiają łączność także wówczas, gdy bezpośrednie połączenie między nadawcą i odbiorcą charakteryzuje się niską jakością, a co za tym idzie, dużą liczbą retransmisji. Pewnym mankamentem tego rozwiązania jest konieczność podania trasy ramki (listy stacji pośredniczących) już w chwili nawiązywania połączenia; operator stacji musi zatem znać położenie stacji i sam zdecydować, które z nich warto wykorzystać w celu nawiązania łączności. W zależności od liczby użytych stacji pośredniczących należy także odpowiednio zmodyfikować wartości początkowe niektórych liczników i zegarów, przede wszystkim licznika T_1.

Stacje pośredniczące nie zapamiętują ramek do chwili odbioru potwierdzenia od odbiorcy lub kolejnej stacji pośredniczącej. W przypadku przekłamania ramki, np. wskutek kolizji lub błędu transmisji, konieczna jest zatem retransmisja przez nadawcę, a ramka musi ponownie przebyć całą trasę od nadawcy do odbiorcy. Zakładając, iż znane jest prawdopodobieństwo bezbłędnej transmisji ramki pomiędzy dwiema stacjami p, prawdopodobieństwo bezbłędnej transmisji ramki informacyjnej i potwierdzenia przy użyciu n stacji pośredniczących wynosi

$$q = p^{2(n+1)}. \tag{2.1}$$

Przy stosunkowo niezłym łączu o prawdopodobieństwie bezbłędnego przesłania ramki wynoszącym p=0,9 i użyciu trzech stacji pośredniczących wynikowe prawdopodobieństwo bezbłędnej transmisji wynosi $q=(0,9)^8$=0,43, przy ośmiu stacjach pośredniczących – zaledwie

$q=(0,9)^{18}=0,15$. Z tego powodu stosowanie stacji pośredniczących jest zalecane jedynie w ostateczności. Zasadę transmisji z wykorzystaniem stacji pośredniczących ilustruje rys. 2.2.

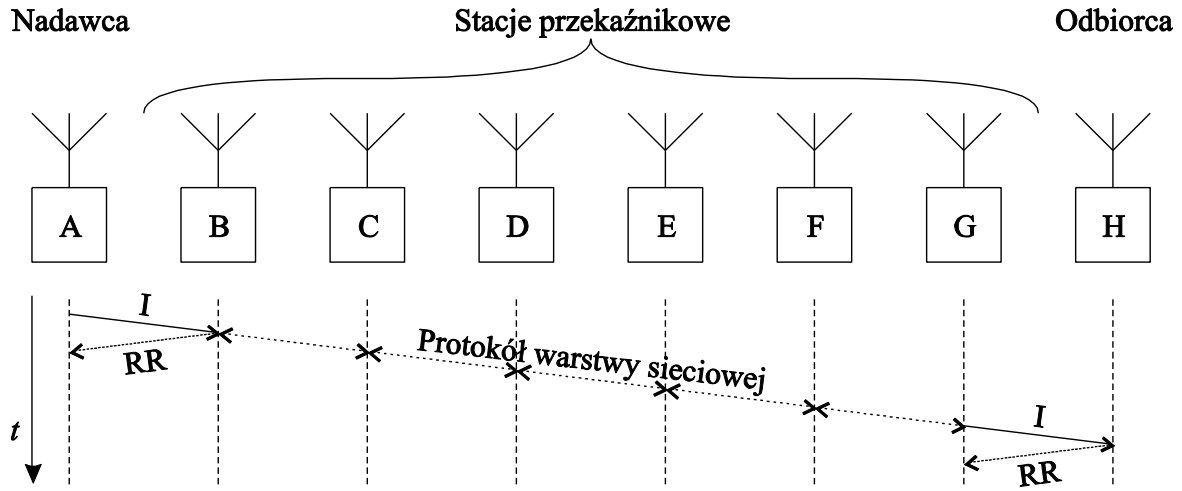


Rys. 2.2. Transmisja informacji z wykorzystaniem stacji pośredniczących
Fig. 2.2. Information transmission using intermediate stations

Pewną poprawę opisanej sytuacji można uzyskać, stosując specjalne stacje pośredniczące (*digipeater*, ang. *digital repeater*) o korzystnej lokalizacji i zwiększonym zasięgu [25]. W ten sposób zmniejsza się liczba stacji pośredniczących w transmisji, co zwiększa prawdopodobieństwo poprawnego przesłania informacji. Z drugiej strony, użytkownicy realizujący transmisję z wykorzystaniem takich stacji mogą nie słyszeć się wzajemnie (zjawisko stacji ukrytej), przez co na wejściu stacji mogą występować kolizje.

Znaczną poprawę warunków transmisji można uzyskać, stosując stacje przekaźnikowe, realizujące protokół warstwy sieciowej. Z punktu widzenia użytkownika pełnią one podobną funkcję jak stacje bazowe sieci telefonii komórkowej. Przy nawiązaniu połączenia nie podaje się już całej trasy ramki, tylko adres odbiorcy i najbliższej stacji przekaźnikowej. Stacja ta jest odpowiedzialna za wybór trasy ramki i bezbłędne przesłanie jej do adresata. Proces ten jest niewidoczny dla użytkownika. Stacja przekaźnikowa potwierdza odbiór ramki nadawcy niejako w imieniu odbiorcy. Zasadę transmisji z wykorzystaniem stacji przekaźnikowych ilustruje rys. 2.3.

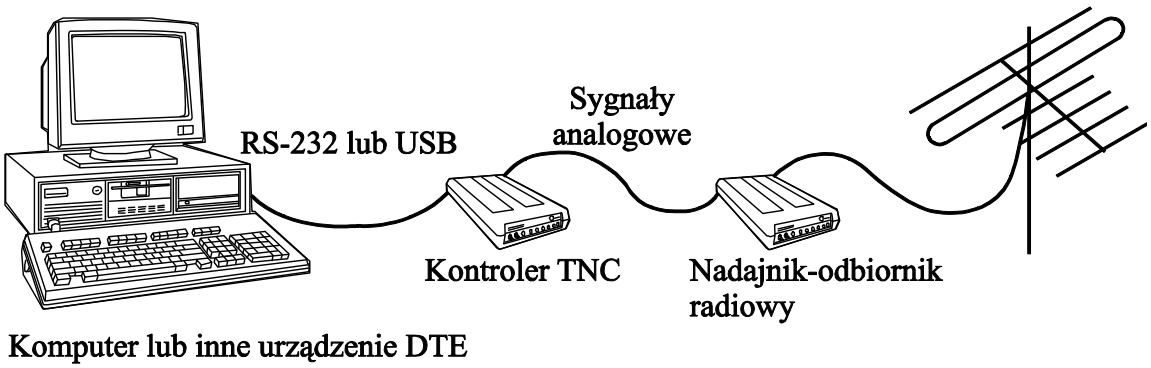
Budowa stacji przekaźnikowej wymaga co najmniej dwóch łączy radiowych. Łącza obsługujące stacje lokalne pracują zgodnie z protokołem AX.25, zaś łącza do innych stacji przekaźnikowych wykorzystują protokół warstwy sieciowej. Wskazane jest, aby te ostatnie były łączami dwukierunkowymi jednoczesnymi w celu uzyskania wyższej wydajności sieci. Istnieje kilka typów oprogramowania stacji przekaźnikowej, można także wykorzystać różne protokoły sieciowe. Jednym z nich może być TCP/IP.



Rys. 2.3. Transmisja informacji z wykorzystaniem stacji przekaźnikowych
Fig. 2.3. Information transfer using relay stations

2.2. Kontrolery TNC

Kontroler TNC (ang. *Terminal Node Controller*) jest układem mikroprocesorowym, zapewniającym możliwość podłączenia urządzenia DTE (np. komputera, sterownika przemysłowego, stacji meteorologicznej itp.) do sieci Packet Radio [25, 63]. Budowę typowej stacji sieci Packet Radio z wykorzystaniem TNC pokazano na rys. 2.4.



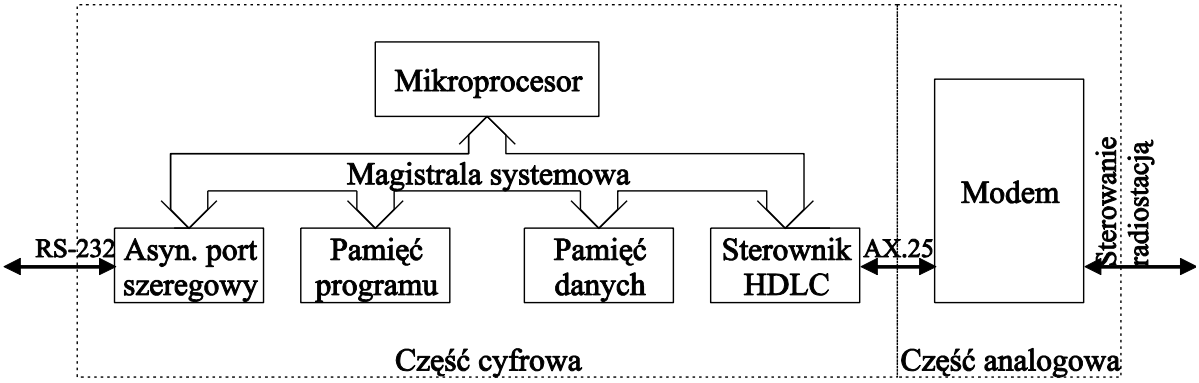
Rys. 2.4. Budowa stacji sieci Packet Radio zawierającej kontroler TNC
Fig. 2.4. Structure of Packet Radio station containing TNC controller

2.2.1. Budowa kontrolera TNC

Kontroler TNC [25] składa się z części cyfrowej, zapewniającej przetworzenie postaci danych napływających z komputera zgodnie z wymogami sieci i zasadami działania protokołu AX.25, oraz części analogowej, pełniącej funkcję modemu i umożliwiającej sterowanie radiostacją bezpośrednio z kontrolera. Ogólny schemat blokowy kontrolera przedstawiono na rys. 2.5. Może on sugerować, że każdy z wymienionych na rysunku elementów musi występować jako odrębny układ scalony; przy obecnych możliwościach technologicznych, przy

zastosowaniu mikrosterowników jednoukładowych, jest jednak możliwa realizacja całej części cyfrowej w postaci pojedynczego układu (niektóre mikrosterowniki, jak np. układy firmy Motorola/Freescale, wymagają jednak dołączenia zewnętrznych układów pamięci programu i danych). Zresztą nawet w starszych konstrukcjach, zawierających np. mikroprocesor Zilog Z80, port szeregowy i sterownik HDLC występują często w pojedynczym układzie scalonym (najczęściej Z80-SIO lub 8530). Prócz elementów pokazanych na rys. 2.5 kontroler może oczywiście zawierać dodatkowe wyposażenie, jak np.:

- zegar czasu rzeczywistego (RTC, ang. *Real Time Clock*),
- dodatkowe sterowniki HDLC, nierzadko umożliwiające jednoczesną pracę kilku łączy radiowych,
- dodatkowe łącza, jak np. USB czy Ethernet,
- pamięć EEPROM przechowującą parametry układu,
- przetworniki analogowo-cyfrowe, umożliwiające odczyt stanu nadzorowanego obiektu.



Rys. 2.5. Schemat blokowy kontrolera TNC
Fig. 2.5. Block diagram of TNC controller

2.2.1.1. Część cyfrowa kontrolera

Część cyfrowa kontrolera TNC zawiera kompletny układ mikroprocesorowy, pełniący w zasadzie funkcję sterownika protokołu AX.25. Układ ten komunikuje się z dołączonym komputerem przez asynchroniczny port szeregowy zgodny ze standardem RS-232 lub przez USB, natomiast z nadajnikiem-odbiornikiem radiowym – przez sterownik protokołu HDLC i modem, znajdujący się już w części analogowej. Pamięć danych służy do buforowania przesyłanych danych i przechowywania parametrów kontrolera i protokołu AX.25. Aby parametrów tych nie trzeba było ponownie ustawiać po okresie wyłączenia kontrolera, pamięć danych – przynajmniej w części – jest podtrzymywana bateryjnie. Niektóre nowsze rozwiązania, np. TNC7 czy DLC7, przechowują ustawione wartości parametrów w pamięci nieulotnej EEPROM. Wybrane parametry konstrukcyjne części cyfrowej niektórych kontrolerów TNC zebrano w tabeli 2.5.

Tabela 2.5
Wybrane parametry części cyfrowej niektórych kontrolerów TNC [120, 151]

Typ	Producent	Procesor	f_{clk} [MHz]	ROM [KB]	RAM [KB]	RTC
TNC2	?	Z80	2,4576	32	16-32	–
TNC2D [73]	Muel	Z80	4,9152	2×32	32	–
TNC2H [98]	Symek	Z80	9,8304	2×32	32	–
Spirit-2 Std. [82]	Paccomm	Z80	9,8304	2×32	32	*DS1216*
Spirit-2 High Speed	Paccomm	Z80	19,6608	2×32	32	*DS1216*
KPC-9612+ [61]	Kantronics	68HC11	16,0000	128	128-512	DS1315
PK-96 [106]	Timewave	Z180	12,2880	64	128	*DS1216*
KAM-XL [61]	Kantronics	68HC902	9,8304	512	512	X1203
DSP-232 [106]	Timewave	68340	3,6864	128	256	*DS1216*
PTC-II [92]	SCS	68360	24,5760	256-512	512-2048	DS1602
TNC3S [98]	Symek	68302	14,7456	256-1024	64-2048	RTC58232
TNC31S [98]	Symek	68302	14,7456	128-512	128-512	*RTC58232*
TNC4e [47]	HBTron	68EN302	19,6608	1024	4096	RTC58232
TNC7multi [77]	NtG	LPC2106	58,9824	128	64	–
DLC7 [77]	NtG	S3C4530	49,1520	4096	32768	PCF8563

Z danych tych wynika, że dostępne kontrolery TNC znacznie różnią się parametrami, takimi jak typ i częstotliwość taktowania (f_{clk}) mikroprocesora. Cechy te mają znaczny wpływ na moc obliczeniową kontrolera, co z kolei może rzutować na osiągane efektywne prędkości transmisji. Nieco mniejsze różnice występują w zakresie pojemności pamięci. Parametr ten ma wpływ na wygodę użytkowania układu – duża pojemność pamięci programu daje możliwość wyboru używanego oprogramowania bądź też wprowadzenia nowych funkcji. Z kolei, wielkość pamięci RAM ma znaczenie przy stosowaniu układów jako skrzynek pocztowych lub stacji węzłowych. Wydaje się także, iż pojemność pamięci RAM poświęconej na bufory nadawczo-odbiorcze może wpływać na wydajność układu – zbyt mała jej wielkość może uniemożliwiać stosowanie odpowiednio długich pojemności ramek danych (parametr N_1 protokołu AX.25) i dużych wielkości okna (parametr k).

Najczęściej są spotykane konstrukcje kompatybilne ze standardem TNC2, oparte na mikroprocesorze Zilog Z80. Reprezentantami tej grupy są kontrolery TNC2, TNC2D i TNC2H. Są one wyposażone w 16 lub 32 KB pamięci RAM oraz 32 lub 64 KB pamięci EPROM – w drugim przypadku pamięć ta jest przełączana, co umożliwia użycie dwóch różnych wersji oprogramowania sterującego pracą kontrolera. Kontrolery Spirit-2 są wyposażone w wiele rozszerzeń, można jednak skonfigurować je w taki sposób, że są w pełni kompatybilne ze standardem TNC2. Nawet wówczas zapewniają one jednak wyższe prędkości transmisji niż

pozostałe kontrolery tej klasy. Kontrolery Spirit-2 mają ponadto możliwość wyposażenia w zegar czasu rzeczywistego rodziny DS1216 [29], umieszczany w podstawce pod układem pamięci RAM.

Nieco nowocześniejsze są kontrolery PK-96 i KPC-9612+, w których zastosowano odpowiednio mikroprocesor Zilog Z180 oraz Motorola/Freescale 68HC11. W obu tych układach prędkości transmisji są wyższe niż dla typowych TNC2, ale niższe niż w przypadku kontrolerów Spirit-2. Interesującą właściwością oprogramowania jest zdolność automatycznego ustawienia prędkości transmisji portu szeregowego RS-232 (ang. *autobaud*). PK-96, podobnie jak Spirit-2, może być wyposażony w zegar czasu rzeczywistego DS1216, natomiast KPC-9612+ zawiera zegar DS1315. Interesującą i niekiedy przydatną cechą kontrolera KPC-9612+ jest możliwość jednoczesnej pracy dwóch łączy radiowych, przy czym jedno z nich może przesyłać informacje z prędkościami 0,3÷1,2 kb/s, drugie natomiast – 4,8÷38,4 kb/s.

Znacznie większe możliwości dają układy TNC3 oraz TNC31, zbudowane z wykorzystaniem 16-bitowych mikrosterowników Motorola/Freescale 68302. Układy te mają wbudowane sterowniki protokołu HDLC i port szeregowy, wymagają zatem dołączenia jedynie zewnętrznej pamięci. Program może być przechowywany w pamięci EPROM lub Flash; w tym drugim przypadku można aktualizować oprogramowanie przez port szeregowy z wykorzystaniem poleceń systemu operacyjnego kontrolera. System ten umożliwia zresztą także konfigurację niektórych parametrów kontrolera i uruchamianie różnych typów oprogramowania, zależnie od potrzeb. Oprogramowanie może przy tym być uruchamiane z pamięci RAM lub zapisane w pamięci stałej. Kontroler TNC31 jest wyposażony w tylko jeden modem, natomiast TNC3 – w dwa (możliwa jest jednoczesna praca obu łączy). W przypadku rezygnacji z łącza RS-232 kontroler TNC3 może obsługiwać aż trzy modemy jednocześnie. Niektóre wersje są wyposażone w zegar czasu rzeczywistego RTC58323, od wersji zależy także ilość i typ zainstalowanej pamięci. Zbliżoną strukturę ma kontroler TNC4e, który można dołączyć bezpośrednio do sieci Ethernet dzięki zastosowaniu odpowiednio wyposażonej wersji mikrosterownika. W przypadku takiego połączenia kontroler dysponuje aż trzema portami HDLC, do których można podłączać różne modemy. Ułatwia to budowę stacji węzłowych (przekaźnikowych) sieci Packet Radio i jej integrację z innymi sieciami, np. z Internetem.

Kontrolery KAM-XL, DSP-232 i PTC-II (w dowolnej wersji) są tzw. kontrolerami wielofunkcyjnymi, umożliwiają bowiem pracę nie tylko w sieci Packet Radio, lecz także w innych systemach łączności radioamatorskiej. Z tego powodu są wyposażone w procesory DSP, pełniące funkcje modemu. KAM-XL i DSP-232, a także PTC-IIpro, mogą obsługiwać dwa łącza radiowe jednocześnie. DSP-232 może być wyposażony – podobnie jak Spirit-2 i PK-96 – w zegar czasu rzeczywistego DS1216. Kontrolery te zawierają różne odmiany mikrosterow-

ników Motorola/Freescale – 68HC902, 68340 lub 68360. Każdy z tych układów wymaga dołączenia zewnętrznej pamięci programu i danych.

Najnowocześniejszymi układami są TNC7 oraz DLC7, oba zbudowane z wykorzystaniem 32-bitowych mikrosterowników zawierających jądro ARM7. TNC7 pełni funkcję typowego kontrolera TNC, a znajdujący się w nim mikrosterownik zawiera także 128 KB pamięci Flash i 64 KB pamięci RAM i realizuje programowo część funkcji modemu. Układ ten nie jest niestety wyposażony w zegar czasu rzeczywistego, brak także informacji o możliwości jego dołączenia. Połączenie z komputerem można zrealizować przez port szeregowy RS-232 lub USB. W drugim z wymienionych przypadków port USB może także zasilać układ. Kontroler DLC7 jest w pewnym sensie następcą TNC4e, ma bowiem możliwość podłączenia do sieci Ethernet. Kontroler ten jest wyposażony w dwa sterowniki HDLC o prędkości transmisji do 10 Mb/s i zegar czasu rzeczywistego, dysponuje także pamięcią o dużej pojemności – 4 MB Flash i 32 MB RAM. Aktualizacja oprogramowania w TNC7 wymaga uruchomienia układu w specjalnym trybie, podczas gdy w DLC7 – podobnie jak w TNC3 – wystarczy użycie poleceń systemu operacyjnego. Oprogramowanie można zapisać i uruchamiać z pamięci RAM lub Flash. DLC7 dysponuje również gniazdem pamięci typu CompactFlash (CF), którą także można wykorzystać do przechowywania oprogramowania oraz ustawień kontrolera. Zarówno TNC7, jak i DLC7 są wyposażone w przełącznik, umożliwiający zdefiniowanie 10 różnych konfiguracji układu, wybieranych podczas jego uruchamiania.

Niektóre kontrolery – np. KPC-9612+ czy KAM-XL – mają możliwość sterowania prostymi obiektami czy urządzeniami, są bowiem wyposażone w kilka wyjść dwustanowych i wejść analogowych. Oprogramowanie sterujące tymi kontrolerami umożliwia także zdalny dostęp do kontrolera z innego kontrolera poprzez łącze radiowe.

2.2.1.2. Część analogowa kontrolera

Część analogowa kontrolera TNC zawiera modem, który umożliwia transmisję ramek protokołu AX.25 przy użyciu nadajników-odbiorników radiowych, także tych, które są przeznaczone głównie do przesyłu głosu. Przy prędkościach niższych są stosowane modulacje z rodziny AFSK (ang. *Audio Frequency Shift Keying*), zgodne z zaleceniem ITU-T V.23 [23] (para częstotliwości 1300/2100 Hz) lub Bell 202 [4] (para częstotliwości 1200/2200 Hz), umożliwiające transmisję z prędkością 1,2 kb/s, oraz ITU-T V.21 [24] lub Bell 103 [4] dla prędkości transmisji 0,3 kb/s. Przy prędkościach od 4,8 kb/s wzwyż stosuje się modulację FSK, zwaną także DFSK (ang. *Direct Frequency Shift Keying)*, gdyż konieczne jest wówczas bezpośrednie połączenie z układem nadawczo-odbiorczym. Połączenie takie pomija filtry wejściowe, dostosowane do przenoszenia sygnałów mowy, ograniczające prędkość transmisji do ok. 1,2÷2,4 kb/s. Wybrane parametry komunikacyjne niektórych kontrolerów TNC zebrano w tabeli 2.6.

Tabela 2.6
Wybrane parametry komunikacyjne niektórych kontrolerów TNC [120, 151]

Typ	R_w [kb/s] (RS232)	R_w [kb/s] (inne)	Modem wewn.	R_{wl} [kb/s]	Modem zewn.	R_{wl} [kb/s]
TNC2	0,3÷9,6	–	Am7910	0,3÷1,2	○	≤9,6
TNC2D	0,6÷19,2	–	Am7910	0,3÷1,2	○	≤9,6
TNC2H	0,3÷38,4	–	EPROM	9,6	○	19,2
Spirit-2 Std.	4,8÷57,6	–	EPROM	4,8÷57,6	○	
Spirit-2 H.S.	4,8÷57,6	–	EPROM	4,8÷57,6	○	
KPC-9612+	1,2÷38,4	–	MX604	0,3÷1,2	–	
			MX589	4,8÷38,4		
PK-96	1,2÷38,4	–	TCM3105	0,3÷1,2	○	≤38,4
			EPROM	4,8÷38,4		
KAM-XL	0,3÷38,4	–	DSP	0,3÷9,6	–	
DSP-232	0,1÷19,2	–	DSP	0,3÷9,6	–	
PTC-II	1,2÷115,2	–	DSP	0,3÷19,2	○	≤614,4
TNC3S	0,15÷115,2	–	–		TCM3105	1,2
TNC31S	0,15÷115,2	–			EPROM	4,8÷614,4
TNC4e	0,15÷115,2	Eth 10Mb/s	–		EPROM	4,8÷1228,8
TNC7multi	1,2÷115,2	USB 1,2÷921,6	Program.	1,2÷115,2	–	
DLC7	1,2÷115,2	Eth 100Mb/s	–		µproc.	4,8÷307,2

Modem można zrealizować na kilka sposobów, np.:

- przy użyciu układu scalonego, pełniącego funkcje modemu,
- poprzez wytwarzanie przebiegu analogowego z próbek za pomocą pamięci stałej lub mikroprocesora,
- z wykorzystaniem procesorów sygnałowych DSP (ang. *Digital Signal Processing*).

Przykładami popularnych układów scalonych, pełniących funkcje modemu, są np. Am7910 [4] i pochodne oraz TCM3105 [105], umożliwiające transmisję z modulacją i prędkościami nieprzekraczającymi 1,2 kb/s. Stosuje się je w kontrolerach TNC2 (głównie Am7910) oraz PK-96 i TNC3 (TCM3105). Znacznie rzadziej można spotkać układ 73M223 [1], stosowany m. in. we wcześniejszych wersjach kontrolera KPC-9612+. Ponieważ modemy te nie są już produkowane, można spotkać także układy MX604 (V.23) [74], FX604 (V.23) [36] lub FX614 (Bell 202) [37] o prędkości transmisji do 1,2 kb/s, stosowane w kontrolerach KPC-9612+ (nowsze wersje) oraz KPC-3+. Dla prędkości z zakresu 1,2÷4,8 kb/s można stosować układ MX469 [75] lub zbliżony CMX469 [21], a dla prędkości 4÷200 kb/s – CMX589 [22]. Ostatni z wymienionych układów jest stosowany w kontrolerze KPC-9612+ i pozwala uzyskać prędkości z zakresu 4,8÷38,4 kb/s. Układy serii FX/MX/CMX można także spotkać w kontrolerach Kameleon polskiej produkcji, stanowiących wersję roz-

wojową układów TNC2 (niestety nie w pełni kompatybilną pomimo wykorzystania układów rodziny Z80), a także w rosyjskich kopiach układu TNC3, w których pozwala uzyskać prędkość z zakresu 4,8÷192 kb/s [66]. Przełączanie prędkości transmisji może odbywać się sprzętowo, tj. za pomocą zworek, lub programowo, za pomocą odpowiednich poleceń wydawanych przez operatora.

Interesującym rozwiązaniem są modemy FSK, wykorzystujące możliwość generacji przebiegu analogowego za pomocą odpowiednio zaprogramowanej pamięci stałej. Są one stosowane m. in. w kontrolerach TNC3, Spirit-2 i PK-96. W wielu układach tego typu pamięć zawiera próbki kilkunastu różnych przebiegów, co pozwala na dobranie kształtu fali optymalnej dla stosowanego nadajnika-odbiornika radiowego. Tor odbiorczy realizuje się wówczas najczęściej z wykorzystaniem filtrów dolnoprzepustowych, zrealizowanych z użyciem wzmacniaczy operacyjnych. Układy takie pozwalają na uzyskanie prędkości transmisji z zakresu 4,8÷614,4 kb/s, a konstrukcja modemu pozwala na łatwą jej zmianę poprzez zmianę stopnia podziału częstotliwości taktującej. Zmiana prędkości pociąga jednak za sobą konieczność modyfikacji parametrów filtrów analogowych, dlatego też modemy takie wytwarza się najczęściej jako gotowe moduły skonfigurowane dla konkretnej prędkości transmisji – do grupy tej zalicza się modemy dla kontrolerów TNC3. Można jednak zainstalować elementy pasywne filtru analogowego (oporniki i kondensatory) w postaci wymiennego podzespołu – rozwiązanie takie jest stosowane w kontrolerze Spirit-2. Modem dla TNC3 można także zastosować w kontrolerach TNC4e, DLC7 oraz PTC-IIpro, przy czym w dwóch ostatnich nie umożliwia on wykorzystania możliwości programowego ustawienia prędkości transmisji.

W niektórych konstrukcjach modemów pamięć stałą zastąpiono mikrosterownikiem jednoukładowym, nie zmienia to jednak metody wytwarzania przebiegu analogowego. Wykorzystanie mikrosterownika pozwala natomiast ustawić prędkość transmisji programowo, tym niemniej dostrojenie filtru analogowego nadal wymaga wymiany określonych elementów pasywnych. Reprezentantem tej grupy modemów jest DM307, zaprojektowany dla kontrolera DLC7 i umożliwiający transmisję z prędkościami 4,8÷307,2 kb/s. Mimo wykorzystania tego samego złącza co w modemach dla TNC3, użycie DM307 poza kontrolerem DLC7 jest wątpliwe ze względu na konieczność programowej konfiguracji modemu. Podobnej konstrukcji jest modem wykorzystany w kontrolerze TNC7, jednak jest on zrealizowany programowo w głównym mikrosterowniku kontrolera; zewnętrzne elementy pasywne filtrów są dostępne w postaci wymiennych podzespołów, przygotowanych dla konkretnych prędkości transmisji. Modem ten umożliwia transmisję z prędkościami 1,2÷102,2 kb/s (można uzyskać prędkość 115,2 kb/s, ale praca układu jest wówczas niestabilna i może prowadzić do zawieszenia się kontrolera).

Ostatnia grupa modemów jest zrealizowana z wykorzystaniem procesorów sygnałowych (DSP, ang. *Digital Signal Processor*). Modemy takie wykorzystuje się jedynie w tych kontrolerach TNC, które mają możliwość pracy nie tylko w sieci Packet Radio, lecz także w innych systemach łączności radioamatorskiej. Do grupy tej należą kontrolery DSP-232, KAM-XL oraz PTC-II (w dowolnej wersji). Zmiana trybu łączności wymaga w takim przypadku zmiany prędkości transmisji i sposobu modulacji w bardzo szerokim zakresie, a więc użycie programowo konfigurowanego procesora sygnałowego jest w pełni uzasadnione. Modemy te stanowią zwykle stały element kontrolera – jedynie PTC-IIpro umożliwia wymianę modemu i może współpracować np. z modemami dla TNC3.

Warto zauważyć, że wymienność modemu jest pożądaną cechą kontrolera, gdyż umożliwia stosowanie znacznie szerszego zakresu prędkości transmisji niż w przypadku modemu wbudowanego. Z drugiej strony, zmiana prędkości transmisji pociąga za sobą wówczas konieczność modyfikacji sprzętowej układu, podczas gdy w nowszych kontrolerach z modemem wbudowanym zmiana taka wymaga jedynie wydania określonego polecenia. Wiele obecnie produkowanych kontrolerów tej klasy zapewnia pracę z większością popularnych prędkości transmisji.

2.2.2. Oprogramowanie kontrolera TNC

Oprogramowanie kontrolera TNC jest odpowiedzialne za prawidłową realizację protokołu AX.25, używanego w sieci Packet Radio. Dodatkowe funkcje obejmują komunikację z komputerem (a za jego pośrednictwem – z operatorem) oraz konfigurację niektórych parametrów kontrolera. Warto przy tym zauważyć, że o ile realizacja protokołu AX.25 powinna być ściśle ustandaryzowana (występują tu jednak pewne drobne różnice), o tyle w przypadku pozostałych funkcji można pozwolić sobie na pewną dowolność. Z tego zapewne powodu można obecnie spotkać kilka sposobów komunikacji komputera z kontrolerem TNC, optymalizowanych pod kątem wykorzystania kontrolera do komunikacji zarówno między ludźmi (zbiory poleceń TAPR i TF), jak i urządzeniami (tryby KISS oraz HOST, a także zbiór poleceń Hayes AT w układach TNC3). Dostępność wymienionych trybów pracy zależy od warstwy sprzętowej kontrolera – najwięcej oprogramowania jest dla kontrolerów wykorzystujących mikroprocesor Z80. Jako ciekawostkę można podać to, że kontrolery produkcji europejskiej, głównie niemieckiej, wykorzystują najczęściej zbiór poleceń TF, amerykańskie natomiast – TAPR.

Od użytego rodzaju oprogramowania może zależeć sposób realizacji określonych funkcji, a szczególnie protokołu AX.25. Można zatem spodziewać się wpływu oprogramowania na wydajność układu.

Podstawowe oprogramowanie kontrolerów TNC może pracować w trybie poleceń (ang. *command*) i rozmowy (ang. *converse*). Pierwszy z trybów umożliwia konfigurowanie znacz-

nej liczby parametrów kontrolera – w tym parametrów protokołu AX.25 oraz nadajnika-odbiornika radiowego – oraz zarządzanie połączeniami. W trybie rozmowy natomiast jest możliwe przesyłanie informacji, np. w podobny sposób, jak w komunikatorach internetowych.

W oprogramowaniu ze zbiorem poleceń TAPR przełączanie trybów jest realizowane za pomocą odpowiednich poleceń operatora, a w trybie poleceń można pozostawać dowolnie długo, aż do wydania polecenia, którego efektem – głównym lub ubocznym – jest przełączenie do trybu rozmowy. W oprogramowaniu ze zbiorem poleceń TF po wydaniu polecenia następuje automatyczny powrót do trybu rozmowy.

Na szczególną uwagę zasługuje tryb monitorowania, który występuje zarówno w oprogramowaniu ze zbiorem poleceń TAPR, jak i TF. Można go włączyć za pomocą określonego polecenia. W zbiorze TAPR występują najczęściej dodatkowe polecenia modyfikujące zakres monitorowania. W zbiorze TF pojedyncze polecenie służy zarówno do włączania i wyłączania monitorowania, jak i do modyfikacji jego zakresu. Monitorowanie działa niezależnie od trybu, w jakim pracuje kontroler (tryb poleceń i rozmowy).

W trybie monitorowania kontroler TNC udostępnia zdekodowane informacje o każdej odebranej ramce. Zawierają one adresy nadawcy, odbiorcy i ewentualnie stacji pośredniczących, a także zawartości pola sterującego i identyfikatora protokołu. W przypadku ramek informacyjnych (numerowanych oraz nienumerowanych) ujawniana jest także zawartość pola danych. Analizując dane udostępniane w trybie monitorowania, w szczególności zaś zdekodowane nagłówki ramek, można poznać proces wymiany informacji między kontrolerami TNC. Szczególnie cenna wydaje się zdekodowana zawartość pola sterującego, ponieważ wskazuje ona nie tylko typ ramki, lecz także jej numer (N(R) i N(S)) oraz stan bitu P/F. Poniżej przedstawiono przykładowy raport wymiany ramek, uzyskany za pomocą trybu monitorowania dostępnego w oprogramowaniu kontrolera TNC3 (zbiór poleceń TF). Ze względu na czytelność tekstu usunięto z niego zawartość pola danych.

```
2:fm TNC2 to TNC3 ctl I00^ pid F0
2:fm TNC2 to TNC3 ctl I01^ pid F0
2:fm TNC2 to TNC3 ctl I02+ pid F0
2:fm TNC3 to TNC2 ctl RR3-
2:fm TNC2 to TNC3 ctl I05+ pid F0
2:fm TNC3 to TNC2 ctl REJ3-
2:fm TNC2 to TNC3 ctl I03^ pid F0
2:fm TNC2 to TNC3 ctl I04^ pid F0
2:fm TNC2 to TNC3 ctl I05+ pid F0
2:fm TNC3 to TNC2 ctl RR6-
```

W przedstawionej wymianie nadawca (TNC2) przesyła ramki do odbiorcy (TNC3). Przesłane zostają trzy ramki informacyjne (I) o numerach N(S) 0, 1 i 2, przy czym ostatnią oznaczono ustawionym bitem P. Z tego powodu odbiorca przesyła potwierdzenie prawidłowego odbioru ramek z ustawionym bitem F. Kolejna odebrana ramka informacyjna nosi numer 5,

który jest różny od oczekiwanego (3). Zgodnie z wymogami protokołu, zostaje ona odrzucona, a odbiorca informuje nadawcę o tym fakcie ramką REJ z numerem 3. Sytuacja taka mogła być spowodowana błędem transmisji i przekłamaniem dwóch ramek (o numerach 3 i 4). Odebrawszy ramkę REJ, nadawca ponawia przesłanie zgubionych ramek informacyjnych (3 i 4) oraz ramki o numerze 5, która była przesłana prawidłowo już za poprzednim razem. Ten fragment transmisji zachodzi już bez zakłóceń, zatem odbiorca przesyła potwierdzenie RR z numerem 6.

2.3. Ocena wydajności protokołu AX.25

Protokół AX.25 może pracować na łączu dwukierunkowym naprzemiennym (ang. *half-duplex*) lub jednoczesnym (ang. *full-duplex*). W obu przypadkach, przed rozpoczęciem transmisji, nadawca musi sprawdzić, czy łącze jest wolne. Procedurę dostępu do łącza przeprowadza się zgodnie z protokołem CSMA w odmianie trwałej z prawdopodobieństwem $p<1$ (ang. *p-persistent CSMA*). Długość szczeliny używanej przez ten mechanizm jest określona parametrem $T_{102.}$ Parametr p, niewystępujący wprawdzie w definicji protokołu, określa natomiast stopień trwałości (ang. *persistence*) protokołu. Najczęściej przyjmowana wartość wyjściowa tego parametru wynosi 63, co oznacza prawdopodobieństwo transmisji w szczelinie równe 25%. Średnie opóźnienie przy dostępie do wolnego łącza można wyrazić następująco:

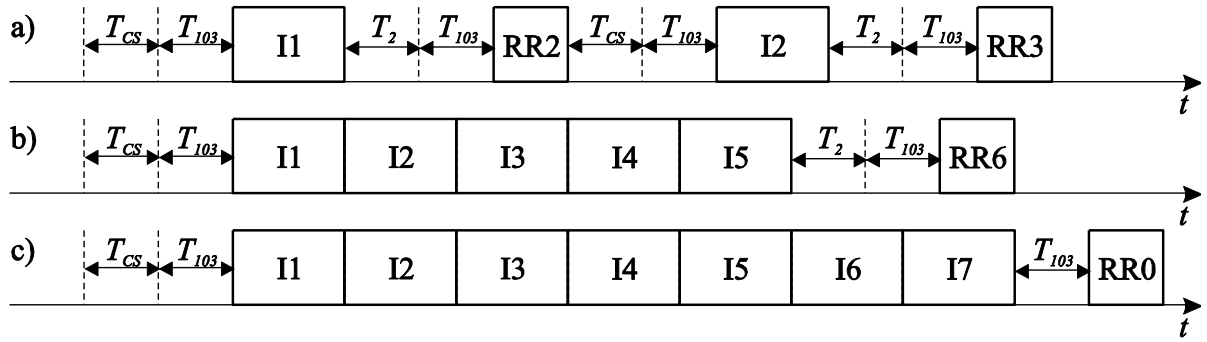
$$T_{CS} = \frac{256 \cdot T_{102}}{2 \cdot (p+1)}. \tag{2.2}$$

Jeśli łącze jest wolne, po upływie czasu T_{CS} nadawca włącza nadajnik radiowy. Z kolei musi on odczekać czas T_{103}, aby upewnić się, że nadajnik pracuje stabilnie, po czym rozpoczyna transmisję ramki informacyjnej (I). W przypadku łącza dwukierunkowego naprzemiennego, jeśli rozmiar okna (k) jest większy niż 1, a nadawca ma więcej ramek przygotowanych do transmisji, można wysłać większą liczbę ramek I w ciągu. Gdy transmisja się zakończy, a przesłano mniej niż 7 ramek, odbiorca czeka przez czas T_2, aby upewnić się, że nadawca nie wysyła kolejnej ramki informacyjnej. Po upływie tego czasu odbiorca włącza nadajnik, czeka przez czas T_{103}, a następnie wysyła ramkę potwierdzającą RR. Przebieg transmisji dla rozmiaru okna 1, 5 oraz 7 pokazano na rys. 2.6.

W przypadku łącza dwukierunkowego jednoczesnego każda ramka informacyjna jest potwierdzana indywidualnie, jednak potwierdzenia te są przesyłane w osobnym kanale częstotliwościowym, równolegle z transmisją kolejnych ramek informacyjnych. W takim przypadku nie uwzględnia się ustalonej wielkości okna (k), ponieważ wszystkie ramki informacyjne można przesłać, w miarę możliwości, w pojedynczej, długiej sekwencji.

Kilku słów komentarza wymaga użycie parametru T_2. Zależnie od rodzaju oprogramowania i szczegółów implementacji protokołu AX.25, nadawca może oznaczyć ostatnią ramkę in-

formacyjną w ciągu przez ustawienie bitu P/F w jej polu sterującym. Wymusza to natychmiastowe przesłanie potwierdzenia przez odbiorcę, bez konieczności odczekania czasu T_2 w celu upewnienia się, że nadawca zakończył transmisję ciągu ramek. Takie zachowanie protokołu jest szczególnie użyteczne na łączu dwukierunkowym naprzemiennym podczas transmisji długich fragmentów informacji pomiędzy dwiema stacjami. Tym niemniej, definicja protokołu nie nakazuje – jakkolwiek również nie zabrania – oznaczenia ostatniej ramki ciągu bitem P/F.



Rys. 2.6. Przebieg transmisji w protokole AX.25 przy rozmiarze okna równym: a) 1, b) 5 i c) 7
Fig. 2.6. Transmission course in AX.25 protocol when window size equals to: a) 1, b) 5 and c) 7

Dla potrzeb oceny wydajności protokołu załóżmy, że komunikacja odbywa się w warunkach idealnych, pozwalają one bowiem na uzyskanie najwyższych osiągów. Przyjmijmy zatem następujące założenia:

- sieć składa się z dwóch stacji – łącze jest przydzielone tylko jednej transmisji;
- nie występują kolizje ani błędy transmisji – nie ma potrzeby retransmisji;
- czas przetwarzania ramek jest pomijalny – nie występują dodatkowe opóźnienia.

Ponadto załóżmy, iż:

- prędkość transmisji łącza radiowego wynosi R_{wl}, przewodowego natomiast – R_w [b/s];
- całkowita wielkość danych użytkownika wynosi L;
- długość ramki potwierdzenia (RR) wynosi 20 bajtów;
- długość ramki informacyjnej (I) wynosi $20+N_1$ bajtów.

2.3.1. Czas przesyłu pojedynczych ramek

Biorąc pod uwagę wymienione powyżej długości ramek informacyjnych (I) i zarządzających (RR), można wyznaczyć czas ich przesyłu. Ze względu na szpikowanie zerami, używane dla uzyskania przezroczystości protokołu, długość ramki dodatkowo zwiększa się, średnio o 1/62 (tj. na każde 62 bity danych przypada średnio 1 bit dodatkowy) [59]. Zatem, biorąc pod uwagę, że przed faktycznym rozpoczęciem transmisji musi upłynąć czas T_{103}, czasy transmisji ramki danych i potwierdzenia można wyrazić następująco [137, 147]:

$$T_I = T_{103} + \frac{63}{62} \cdot \frac{160 + 8 \cdot N_1}{R_{wl}}, \tag{2.3}$$

oraz

$$T_{RR} = T_2 + T_{103} + \frac{63}{62} \cdot \frac{160}{R_{wl}}. \tag{2.4}$$

Powyższe zależności wyrażają czas zajętości łącza podczas transmisji określonych typów ramek, włączając czas stabilizacji nadajnika.

W przypadku stosowania protokołu w wersji 2.2 (z rozszerzoną numeracją ramek) narzut protokołu rośnie tylko o 8 bitów na ramkę. Przyrost ten można zatem pominąć bez zauważalnej utraty dokładności.

2.3.2. Łącze dwukierunkowe naprzemienne

W przypadku łącza dwukierunkowego naprzemiennego (ang. *half-duplex*), gdy stacje wymieniają tylko niewielkie porcje danych (nie większe niż N_1), ramki danych i potwierdzenia przesyłane są naprzemiennie, jak pokazano na rys. 2.6 a). Podczas przesyłu większej ilości danych nadawca może wysłać kolejno więcej ramek, nie przekraczając jednak ustalonej wielkości okna (k). Ramki te zostaną potwierdzone wspólnie, jak pokazano na rys. 2.6 b) i c).

2.3.2.1. Całkowity czas transmisji informacji

Zakładając, iż całkowity rozmiar danych wynosi $N_1 \times k$, długość cyklu transmisyjnego można wyrazić jako:

$$T_P = T_{103} + k \cdot \left(\frac{63}{62} \cdot \frac{160 + 8 \cdot N_1}{R_{wl}} \right) + T_2 + T_{103} + \frac{63}{62} \cdot \frac{160}{R_{wl}} \tag{2.5}$$

lub w postaci uproszczonej [137]

$$T_P = T_2 + 2 \cdot T_{103} + (1 + k) \cdot \left(\frac{63}{62} \cdot \frac{160}{R_{wl}} \right) + k \cdot \left(\frac{63}{62} \cdot \frac{8 \cdot N_1}{R_{wl}} \right). \tag{2.6}$$

Warto przy tym zauważyć, że pierwszy składnik równania (2.6) określa wyłącznie narzut protokołu, podczas gdy czas niezbędny do przesyłu samych danych jest określony przez drugi składnik równania.

Gdy całkowity rozmiar danych jest różny od iloczynu $N_1 \times k$, ostatni cykl transmisyjny będzie krótszy, niż pozwalają na to ustawione wartości tych parametrów. Można zatem zauważyć niewielką utratę wydajności, ponieważ końcowy fragment danych będzie przesłany z większym narzutem protokołu (mniej bitów danych na tę samą liczbę bitów sterujących). Dla całego pliku o rozmiarze L bajtów liczbę cykli transmisyjnych można określić jako:

$$n = \left\lceil \frac{L}{N_1 \cdot k} \right\rceil. \tag{2.7}$$

Przesył całego pliku będzie zatem zajmował czas [137]

$$T_D = \left\lceil \frac{L}{N_1 \cdot k} \right\rceil \cdot \left(T_2 + 2 \cdot T_{103} + (1+k) \cdot \frac{63}{62} \cdot \frac{160}{R_{wl}} \right) + \left\lceil \frac{L}{N_1} \right\rceil \cdot \left(\frac{63}{62} \cdot \frac{8 \cdot N_1}{R_{wl}} \right). \tag{2.8}$$

Podobnie jak poprzednio pierwszy składnik równania oznacza narzut protokołu, drugi zaś – czas potrzebny na przesył samych bitów danych. Efektywną przepustowość protokołu AX.25 można wyznaczyć następująco:

$$V_{wl} = \frac{8 \cdot L}{T_D}. \tag{2.9}$$

2.3.2.2. Wydajność protokołu

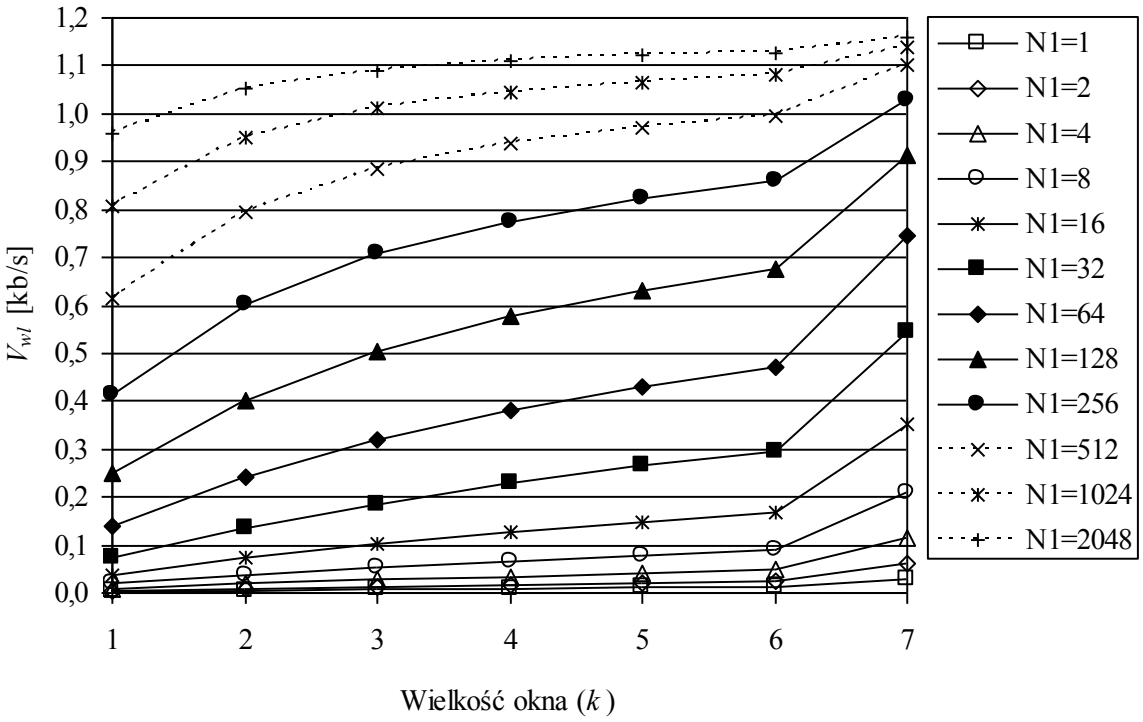
Dla dowolnego zestawu parametrów można oszacować wydajność protokołu, dzieląc czas potrzebny na przesłanie samej informacji przez całkowity czas transmisji, uwzględniający narzut protokołu. W przypadku protokołu AX.25, pracującego na łączu dwukierunkowym naprzemiennym, zależność tę można opisać następująco [137]:

$$\eta = \frac{k \cdot N_1 \cdot 8}{R_{wl}\left(T_2 + 2 \cdot T_{103}\right) + \frac{63}{62} \cdot 160 + k \cdot \frac{63}{62} \cdot \left(160 + 8 \cdot N_1\right)}. \tag{2.10}$$

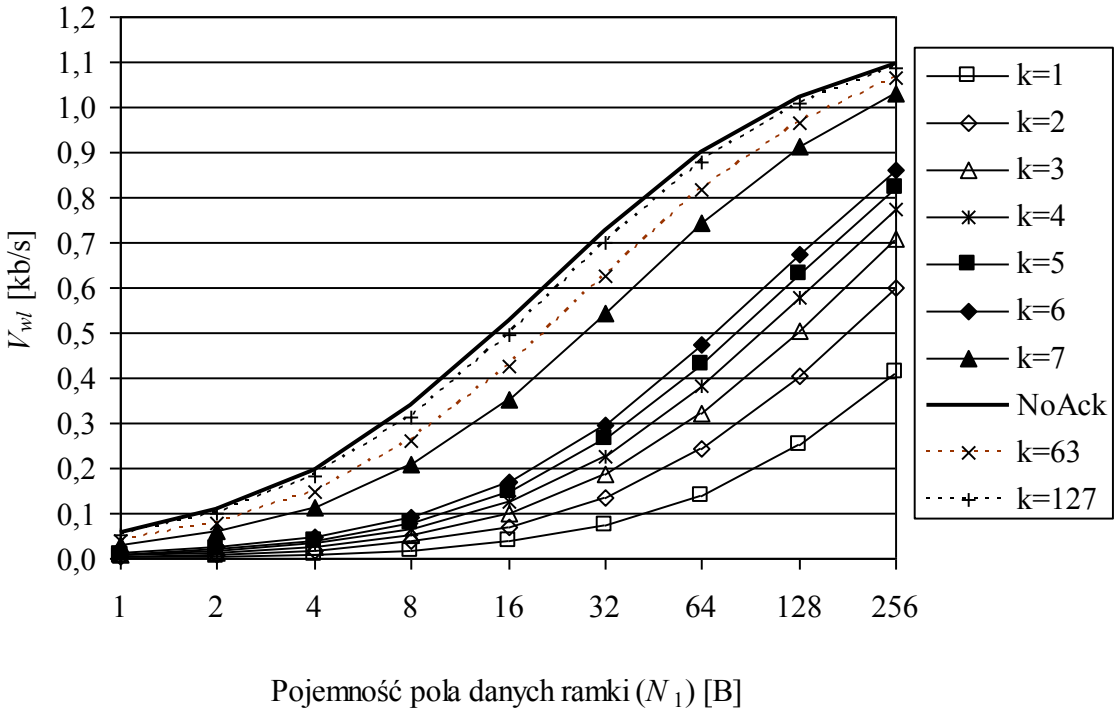
2.3.2.3. Wyniki obliczeń

Używając powyższych zależności, można oszacować efektywną przepustowość protokołu AX.25 dla różnych możliwych wartości N_1 (pojemność pola danych ramki informacyjnej) i k (wielkość okna). Jako parametry T_2 oraz T_{103} przyjęto domniemane wartości używane w kontrolerach TNC3 firmy Symek [98]. Wynoszą one: T_{103} – 250 ms, natomiast T_2 – 280 ms dla prędkości łącza radiowego 9,6 kb/s i 2247 ms dla 1,2 kb/s. Wyniki obliczeń pokazano na rys. 2.7 i 2.8 dla prędkości łącza radiowego 1,2 kb/s oraz rys. 2.9 i 2.10 dla prędkości 9,6 kb/s [137, 147].

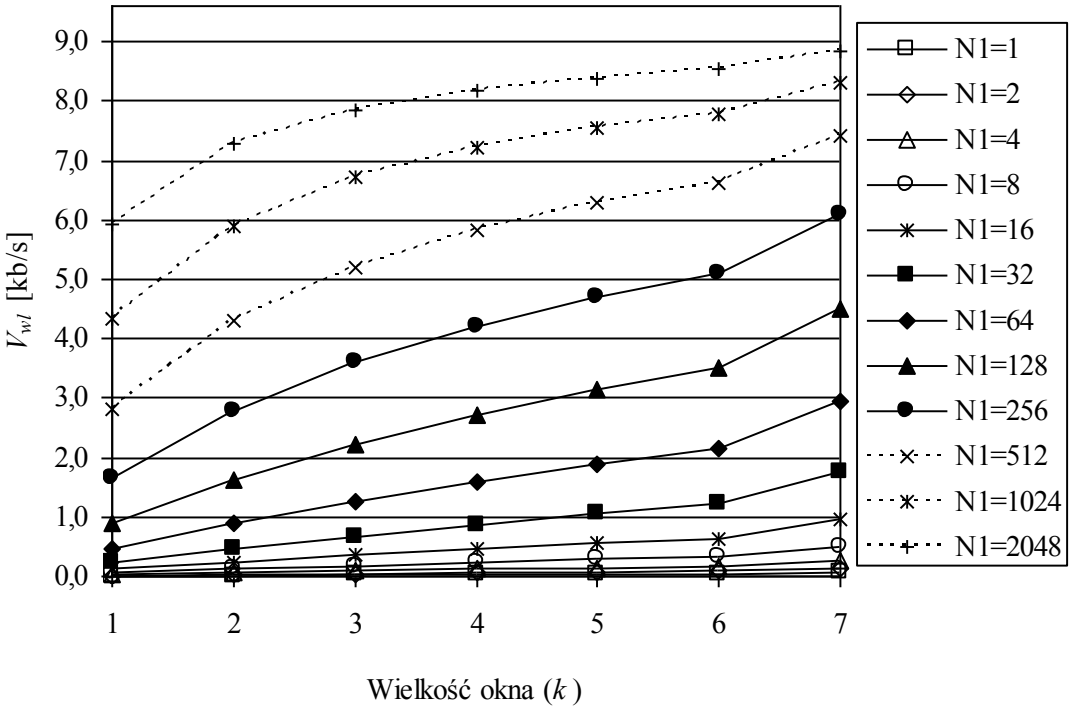
Na przedstawionych wykresach można zauważyć, że – zgodnie z oczekiwaniami – zwiększanie zarówno wielkości okna (k), jak i pojemności pola danych (N_1) przynosi wzrost efektywnej przepustowości łącza niezależnie od jego prędkości transmisji. Dla mniejszych rozmiarów oka, nieprzekraczających 4, podwojenie pojemności pola danych powoduje podwojenie przepustowości. Dla większych rozmiarów okna przyrost przepustowości jest największy dla pojemności pola danych 16 i 64 B. Dalsze zwiększanie tej pojemności nadal przynosi poprawę, ale wzrost przepustowości jest już wolniejszy. Zjawisko to jest szczególnie widoczne dla k=7 przy prędkości 1,2 kb/s dla transmisji z potwierdzeniem oraz przy prędkości 1,2 lub 9,6 kb/s dla transmisji bez potwierdzenia.



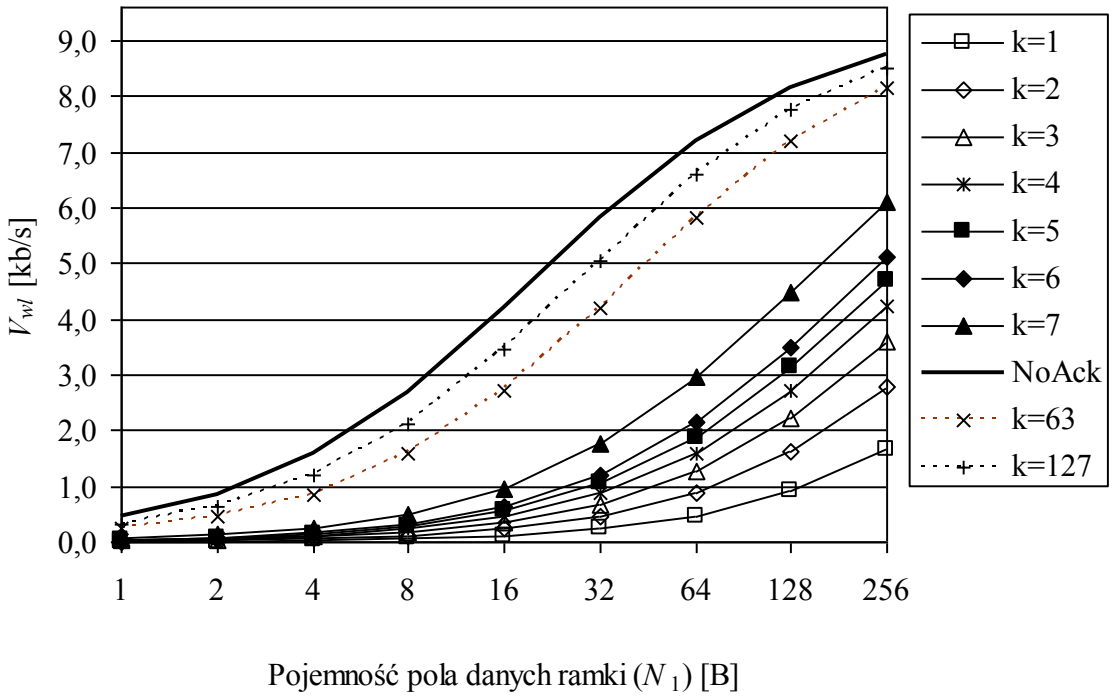
Rys. 2.7. Wpływ wielkości okna na efektywną prędkość transmisji dla łącza 1,2 kb/s
Fig. 2.7. Window size influence upon effective transmission speed for 1.2 kbps link



Rys. 2.8. Wpływ długości pola danych na efektywną prędkość transmisji dla łącza 1,2 kb/s
Fig. 2.8. Data field length influence upon effective transmission speed for 1.2 kbps link



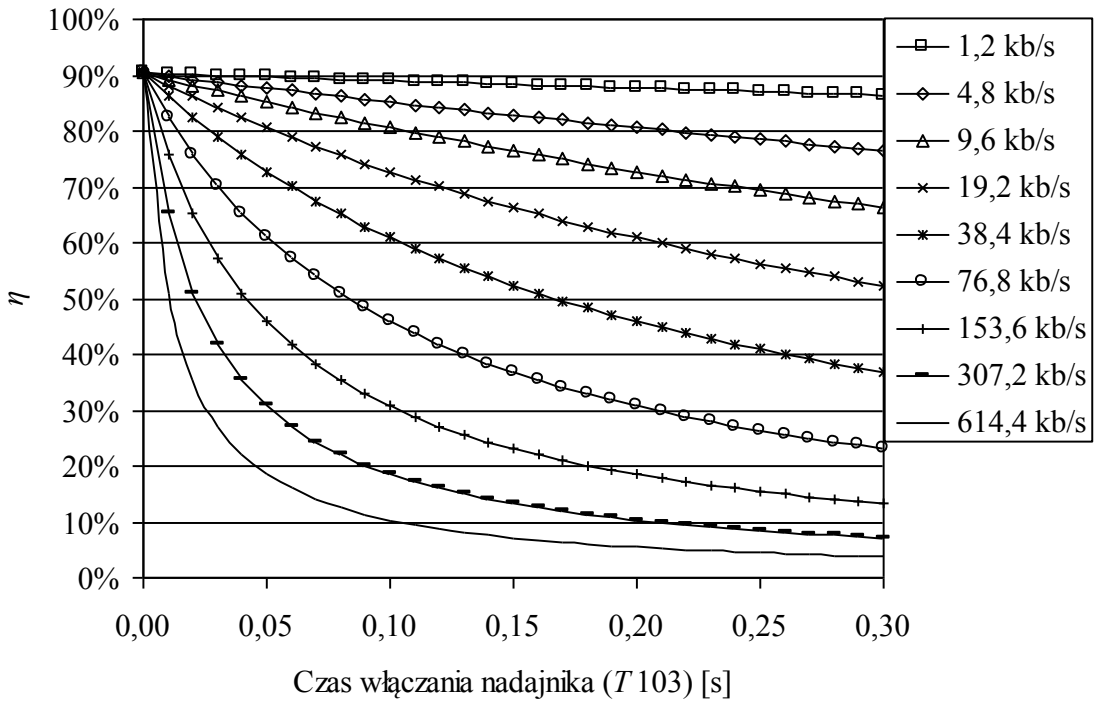
Rys. 2.9. Wpływ wielkości okna na efektywną prędkość transmisji dla łącza 9,6 kb/s
Fig. 2.9. Window size influence upon effective transmission speed for 9.6 kbps link



Rys. 2.10. Wpływ długości pola danych na efektywną prędkość transmisji dla łącza 9,6 kb/s
Fig. 2.10. Data field length influence upon effective transmission speed for 9.6 kbps link

Krzywa, reprezentująca zależność między rozmiarem okna a przepustowością, dzieli się na kilka części (rys. 2.7 i 2.9). Największy, niemal liniowy wzrost wydajności można zaobserwować dla rozmiaru z zakresu 1÷3. Kolejny segment, dla wielkości okna z zakresu 3÷5, cechuje się już mniejszym przyrostem wydajności. Można natomiast zauważyć dużą różnicę pomiędzy wynikami uzyskanymi dla k równego 6 i 7. Jest ona spowodowana czasem T_2, który nie występuje przy k równym 7, ponieważ protokół AX.25, podobnie jak i HDLC, nie dopuszcza większej liczby ramek informacyjnych w ciągu. Gdyby, przy mniejszej wielkości okna, nadawca mógł oznaczyć ostatnią wysyłaną ramkę w ciągu (przez ustawienie bitu sterującego P/F), różnica byłaby znacznie mniejsza, a wydajność protokołu – większa. Możliwość ta zależy jednak od użytego oprogramowania, a opis protokołu ani nie wymaga, ani nie zabrania oznaczania ostatniej ramki informacyjnej w ciągu, pozostawia więc pewną dowolność interpretacji.

Stosując zależność (2.10), można także oszacować maksymalną wydajność protokołu dla różnych prędkości łącza radiowego. Największą wydajność można zawsze uzyskać (teoretycznie) dla wartości parametrów N_1=256 oraz k=7. Jedynymi parametrami, których wartości będą się zmieniać, są zatem prędkość łącza radiowego (R_{wl}) oraz czas włączania nadajnika (T_{103}). Zależność wydajności protokołu AX.25 od wartości T_{103} przy różnych prędkościach transmisji pokazano na rys. 2.11.



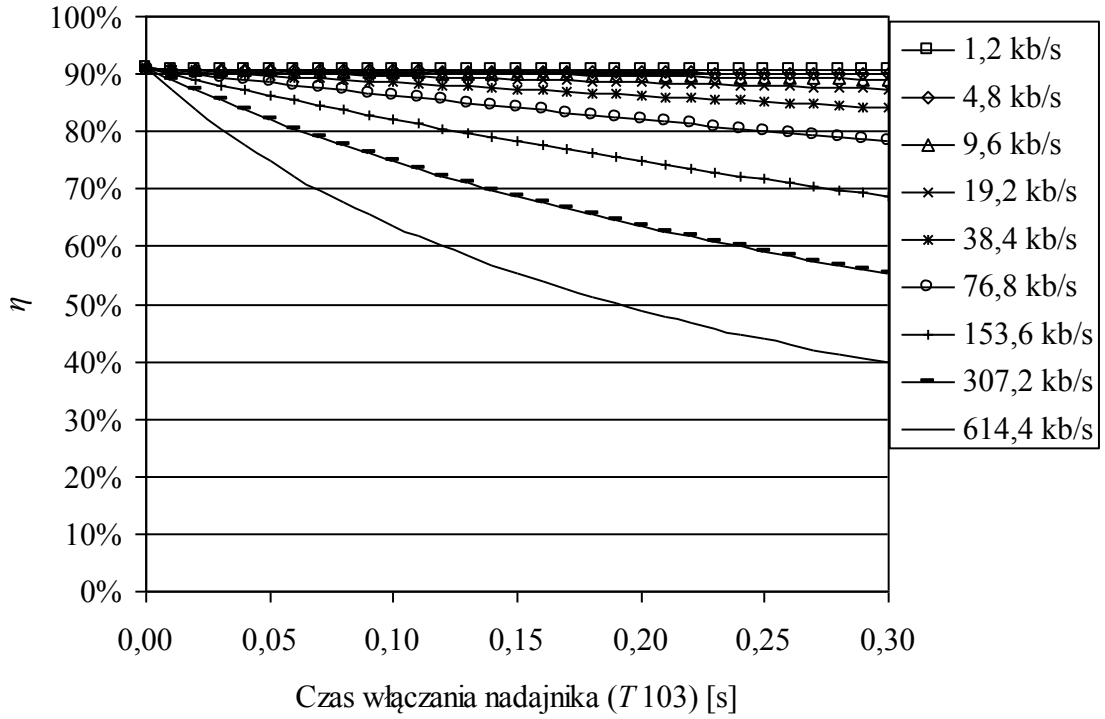
Rys. 2.11. Wpływ czasu włączania nadajnika na wydajność protokołu AX.25 – okno 7 ramek
Fig. 2.11. Transmitter startup time influence on AX.25 protocol performance – 7 frames window

Na przedstawionym wykresie widać, że zwiększanie T_{103} – zgodnie z oczekiwaniami – zmniejsza wydajność protokołu. Zależność ta jest jednak silnie uzależniona od prędkości transmisji łącza radiowego. Dla T_{103}=0, niezależnie od prędkości transmisji, protokół osiąga wydajność około 90%. Niestety, tak krótkiego czasu włączania nadajnika nie można wykorzystać w praktyce ze względu na właściwości nadajników-odbiorników radiowych używanych w sieci Packet Radio. Warto przy tym zauważyć, że niezerowy czas włączania nadajnika jest cechą wszystkich sieci radiowych, z tym że w nowoczesnych sieciach, jak np. IEEE 802.11 czy GSM, wynosi on kilka mikrosekund.

Dla niższych prędkości transmisji – powiedzmy, nieprzekraczających 9,6 kb/s – wydajność protokołu zmniejsza się liniowo wraz ze wzrostem czasu T_{103} i dla wartości 300 ms spada do ok. 87% przy prędkości 1,2 kb/s i ok. 66% przy 9,6 kb/s. Niestety, dla większych prędkości transmisji degradacja wydajności przebiega znacznie szybciej – przykładowo, dla prędkości 614,4 kb/s czas T_{103} równy tylko 50 ms powoduje spadek wydajności protokołu do wartości poniżej 20%. Warto przy tym zauważyć, że tylko niektóre urządzenia używane w sieci Packet Radio mogą pracować przy tak niskiej wartości T_{103}; typowe wartości zależą od właściwości nadajników-odbiorników radiowych i mieszczą się w przedziale 100÷300 ms. Z drugiej jednak strony, prędkości transmisji przekraczające kilkadziesiąt kb/s nie występują zbyt często – łącza naziemne pracują zwykle z prędkością 9,6 kb/s, satelitarne natomiast – 38,4 kb/s (na łączu „w dół").

Warto zauważyć, że degradacja wydajności protokołu zmniejsza także efektywną prędkość transmisji. Nawet dla najwyższych wartości parametrów (N_1=256, k=7), gdy T_{103} przyjmuje wartość 250 ms, efektywna prędkość transmisji wynosi od 6,6 kb/s (dla łącza radiowego 9,6 kb/s) do 27,3 kb/s (dla łącza 614,4 kb/s). Nietrudno zatem zauważyć, iż – wbrew obiegowej opinii – nawet bardzo wysoka, jak na możliwości sieci Packet Radio, prędkość transmisji, za jaką można uważać 614,4 kb/s, zapewnia znacznie mniejszą prędkość efektywną niż łącze RS-232, używane dla podłączenia kontrolera TNC do komputera.

Na rys. 2.12 przedstawiono wyniki analogicznych obliczeń, wykonanych dla protokołu w wersji 2.2. Wersja ta umożliwia stosowanie okien powiększonych do 127 ramek informacyjnych. Jak nietrudno zauważyć, spadek wydajności protokołu wraz ze wzrostem czasu włączania nadajnika jest dużo mniejszy. Przykładowo, dla najwyższej prędkości transmisji – 614,4 kb/s – przy T_{103}=0,3 s wydajność protokołu osiąga około 40%, podczas gdy przy oknie zawierającym 7 ramek spada do około 5%. Wyniki te są zgodne z oczekiwaniami, ponieważ przy mniejszej wielkości okna rośnie liczba zmian kierunku transmisji łącza naprzemiennego, a więc wpływ czasu T_{103} na całkowity czas transmisji jest większy.



Rys. 2.12. Wpływ czasu włączania nadajnika na wydajność protokołu AX.25 – okno 127 ramek
Fig. 2.12. Transmitter startup time influence on AX.25 protocol performance – 127 frames window

2.3.3. Łącze w pełni dwukierunkowe

Przedstawiona powyżej sytuacja zmienia się diametralnie, gdy łącze radiowe jest w pełni dwukierunkowe (ang. *full duplex*). Wprawdzie każda ramka informacyjna jest wówczas potwierdzana indywidualnie, jednak potwierdzenie przesyła się (w osobnym kanale częstotliwościowym) podczas transmisji kolejnej ramki informacyjnej.

2.3.3.1. Całkowity czas transmisji

Z punktu widzenia czasu transmisji istotne jest tylko ostatnie potwierdzenie, ponieważ będzie ono przesłane już po wysłaniu ostatniej ramki informacyjnej. Czas transmisji na łączu dwukierunkowym będzie zatem zbliżony do czasu transmisji prowadzonej bez potwierdzenia (rys. 2.8 i 2.10), jednak przy zachowaniu wiarygodności przesyłu.

Całkowity czas transmisji łączem dwukierunkowym można wyrazić wzorem [137, 147]:

$$T_D = T_{103} + \left\lceil \frac{L_D}{N_1} \right\rceil \cdot \left(\frac{63}{62} \cdot \frac{160 + 8 \cdot N_1}{R_{wl}} \right) + \left(\frac{63}{62} \cdot \frac{160}{R_{wl}} \right), \tag{2.11}$$

w którym środkowy składnik określa czas transmisji wszystkich ramek danych, ostatni zaś – czas transmisji ostatniego potwierdzenia.

2.3.3.2. Wydajność protokołu

W przypadku łącza dwukierunkowego jednoczesnego efektywność protokołu można określić jako:

$$\eta = \frac{8 \cdot L_D}{R_{wl} \cdot T_{103} + \frac{63}{62} \cdot \left\lceil \frac{L_D}{N_1} \right\rceil \cdot (160 + 8 \cdot N_1) + \frac{63}{62} \cdot 160}. \tag{2.12}$$

Przy odpowiednio długiej transmisji danych pierwszy i ostatni składnik czasu transmisji T_D jest pomijalny. W praktyce bardziej istotne jest, by całkowity rozmiar przesyłanej informacji był całkowitą wielokrotnością N_1. Wówczas efektywność protokołu jest niezależna od prędkości transmisji łącza radiowego i dla najdłuższych możliwych ramek (N_1=256) wynosi około 91,3%. Umożliwia to efektywne wykorzystanie wszystkich, nawet najwyższych prędkości transmisji – przy łączu radiowym o prędkości 614,4 kb/s otrzymuje się efektywną prędkość około 560 kb/s. W takiej sytuacji wykorzystanie pojemności sieci jest możliwe jedynie wówczas, gdy prędkość komunikacji z komputerem wzrośnie powyżej możliwości standardu RS-232 bądź też gdy większa liczba stacji będzie utrzymywać wysoką aktywność wymiany informacji. Z tego powodu niektóre kontrolery TNC wyposażono w możliwość podłączenia do sieci Ethernet (umożliwia to także integrację sieci Packet Radio i Internetu) lub z wykorzystaniem interfejsu USB. W obu przypadkach pozwala to na wystarczające zwiększenie prędkości transmisji.

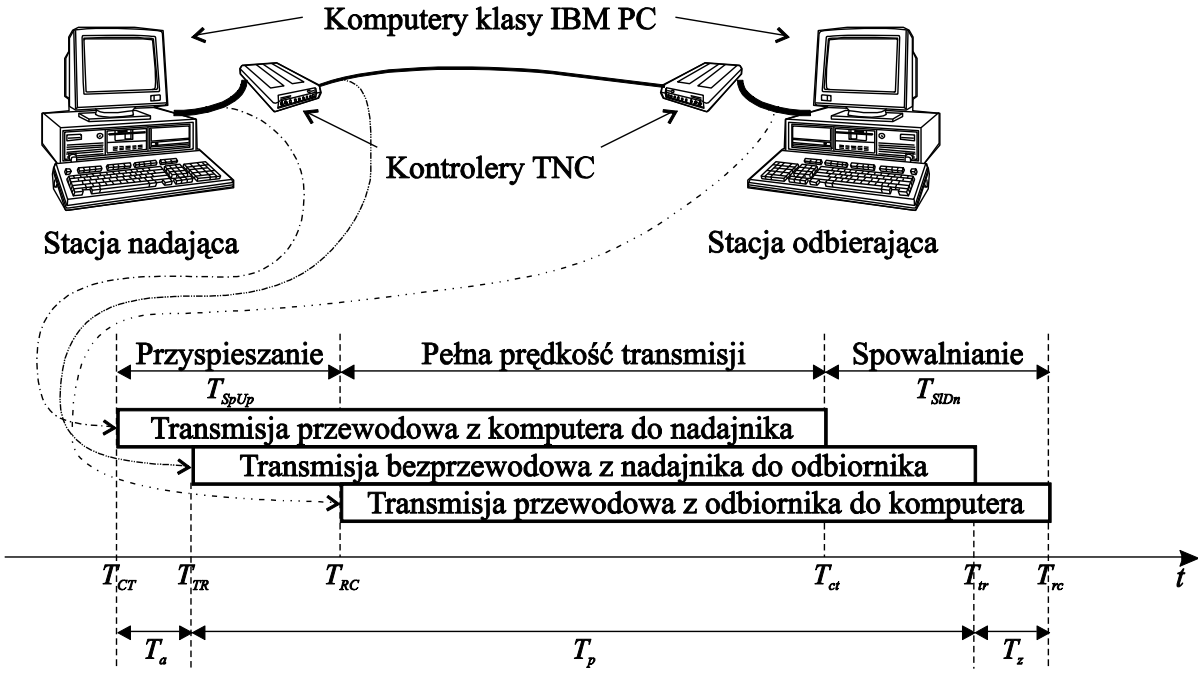
Hipotetyczne dłuższe ramki umożliwiają dalsze zwiększenie wydajności protokołu – dla N_1=2048 przekracza ona nieznacznie 97% [140]. Odpowiadająca temu przypadkowi efektywna przepustowość łącza dla prędkości 614,4 kb/s wynosi prawie 600 kb/s.

2.4. Model analityczny kontrolera TNC

Analiza pracy kontrolera TNC [122, 146] dotyczy przypadku przesyłania informacji między dwoma komputerami (lub innymi urządzeniami), połączonymi ze sobą za pośrednictwem dwóch kontrolerów TNC. W takim przypadku – ze względu na buforowanie przesyłanej informacji w pamięci kontrolera oraz przetwarzanie postaci informacji – transmisja przebiega w trzech etapach:

- transmisja przewodowa z nadającego komputera do kontrolera TNC,
- transmisja bezprzewodowa między kontrolerami TNC, przebiegająca zgodnie z wymogami protokołu AX.25,
- transmisja przewodowa z kontrolera TNC do odbierającego komputera.

Konfigurację rozważanej sieci wraz z wymienionymi etapami transmisji pokazano na rys. 2.13. Jak widać na przedstawionym rysunku, poszczególne etapy transmisji mogą przebiegać w dużym stopniu równolegle – zależy to między innymi od zależności między czasem transmisji bezprzewodowej (T_D) oraz czasami przyspieszania (T_{SpUp}) i spowalniania (T_{SlDn}). Nie bez znaczenia są także czasy transmisji przewodowej między komputerem a kontrolerem TNC po stronie nadającej (T_a) i odbierającej (T_z).



Rys. 2.13. Konfiguracja rozważanej sieci i etapy transmisji
Fig. 2.13. Considered network configuration and transmission stages

2.4.1. Efektywna prędkość transmisji

Przyjmijmy założenia i oznaczenia identyczne jak dla oszacowania efektywnej prędkości transmisji protokołu AX.25 (rozdział 2.3). Ze względu na sposób pracy kontrolera TNC, transmisja łączem radiowym może się rozpocząć, gdy w buforze zebranych zostanie przynajmniej N_1 znaków. Opóźnienie rozpoczęcia transmisji łączem radiowym wynosi zatem [122, 146]:

$$T_a = T_{TR} - T_{CT} = \frac{10 \cdot N_1}{R_w}, \tag{2.13}$$

gdzie R_w oznacza prędkość transmisji łącza przewodowego (RS-232).

Przy założeniu że prędkość efektywna łącza przewodowego znacznie przekracza prędkość efektywną łącza radiowego, po zakończeniu transmisji bezprzewodowej odbierający kontroler TNC wyśle jeszcze do komputera nie więcej niż N_1 znaków. Można więc przyjąć, że $T_z = T_{rc} - T_{tr} = T_a$.

Efektywna prędkość transmisji wynosi zatem

$$V_{ef} = \frac{8 \cdot L_D}{2 \cdot T_a + T_p}. \tag{2.14}$$

Warto zauważyć, że wpływ czasu T_a na efektywną prędkość transmisji maleje wraz ze wzrostem całkowitego rozmiaru przesyłanej informacji lub wraz ze zmniejszaniem się efektywności łącza bezprzewodowego. Przy odpowiednio długim czasie T_p czas T_a można zatem pomi-

nąć. Jest to równoważne założeniu, iż całkowity czas transmisji między komputerami jest równy czasowi transmisji między kontrolerami TNC z użyciem protokołu AX.25.

2.4.2. Opóźnienia transmisji

W pewnych zastosowaniach istotna może być nie tylko efektywna prędkość transmisji, lecz także opóźnienia wynikające z faktu buforowania i przetwarzania postaci przesyłanych danych. Szczególnie istotne są czasy, upływające:

- między rozpoczęciem transmisji przez nadawcę i rozpoczęciem odbioru przez adresata,
- między zakończeniem transmisji przez nadawcę i zakończeniem odbioru przez adresata.

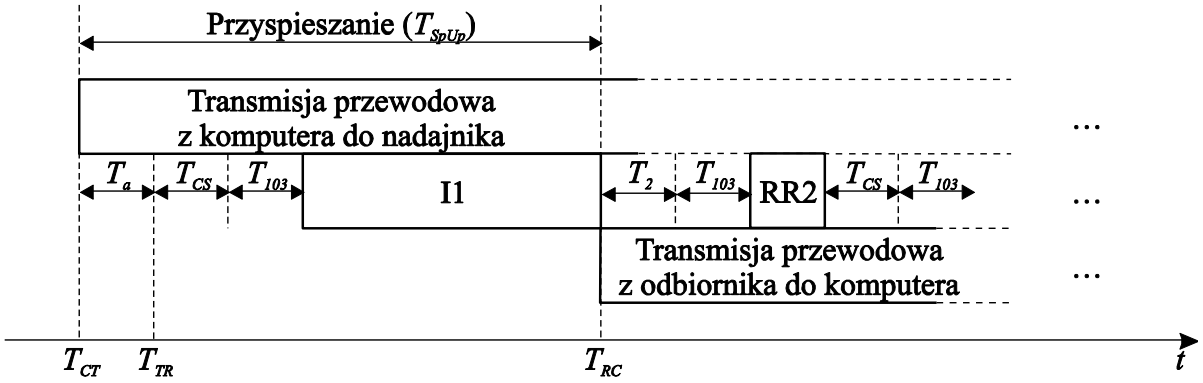
W dalszej części pracy czasy te będą nazywane odpowiednio opóźnieniami rozpoczęcia i zakończenia transmisji.

2.4.2.1. Opóźnienie rozpoczęcia transmisji

Opóźnienie rozpoczęcia transmisji, odpowiadające fazie „przyspieszania" na rys. 2.13, jest równe sumie czasu T_a wyliczonego według równania (2.13) oraz czasu transmisji jednej ramki łączem radiowym, wynosi zatem [122, 146]:

$$T_{SpUp} = \frac{10 \cdot N_1}{R_w} + T_{CS} + T_{103} + \frac{63}{62} \cdot \frac{160 + 8 \cdot N_1}{R_{wl}}. \tag{2.15}$$

Wyjaśnienie zależności (2.15) znajduje się na rys. 2.14 [122, 146]. Opóźnienie rozpoczęcia transmisji nie zależy od relacji między efektywną przepustowością łącza przewodowego i bezprzewodowego.



Rys. 2.14. Wyjaśnienie sposobu obliczenia czasu przyspieszania (T_{SpUp})
Fig. 2.14. Explanation of "speed-up" time (T_{SpUp}) calculation method

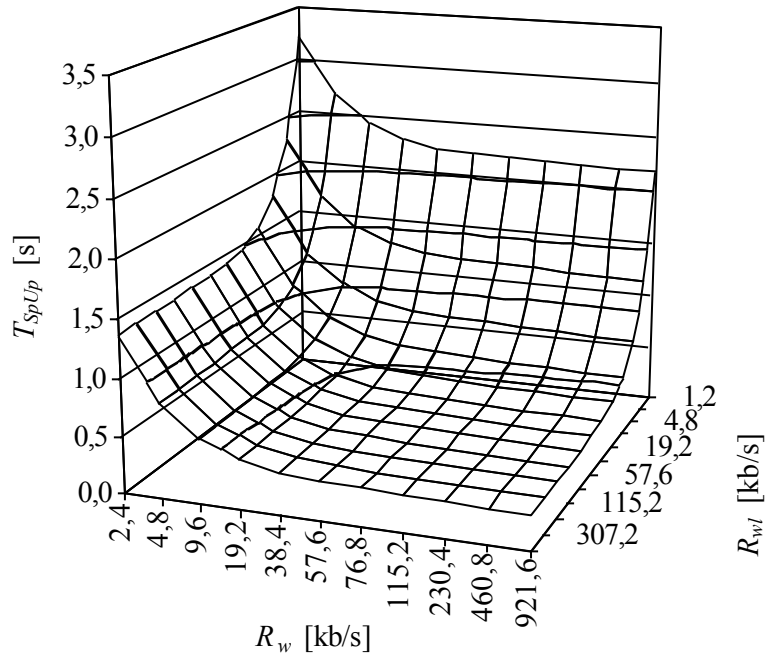
Na rys. 2.15 przedstawiono wyniki obliczeń czasu przyspieszania w zależności od prędkości transmisji łącza przewodowego i bezprzewodowego [122, 146]. Obliczenia wykonano dla najgorszego przypadku, tj. dla najdłuższych możliwych ramek (N_1=256), ponieważ wówczas wydłuża się czas T_a. Zgodnie z oczekiwaniami, największe opóźnienie występuje przy najniższych prędkościach transmisji (R_w=2,4 kb/s, R_{wl}=1,2 kb/s) i wynosi 3,4 s. Najmniejsze z kolei opóźnienie występuje przy najwyższych prędkościach (R_w=921,6 kb/s,

R_{wl}=614,4 kb/s) i wynosi ok. 0,46 s. Największe zmiany opóźnienia występują przy prędkościach nieprzekraczających R_w=38,4 kb/s i R_{wl}=19,2 kb/s; dalsze zwiększanie prędkości transmisji nie przynosi już znaczącej redukcji opóźnień. Wpływ na to ma składnik $T_{CS}+T_{103}$, którego wartość w rozważanym, typowym przypadku wynosi ok. 0,45 s. W celu dalszej redukcji opóźnienia można:

a) zastosować nadajnik-odbiornik radiowy o krótkim czasie włączania nadajnika (czas T_{103} może się wówczas skrócić z kilkuset do kilkudziesięciu milisekund),
b) zwiększyć prawdopodobieństwo transmisji protokołu *p*-CSMA (przy p=255 czas T_{CS} skraca się do 0,05 s),
c) zmniejszyć długość szczeliny protokołu *p*-CSMA (przy T_{102}=0 czas T_{CS} także wynosi 0, ale minimalna długość szczeliny może zależeć m. in. od rozległości terytorialnej sieci i właściwości zastosowanych urządzeń).

Po uwzględnieniu rozwiązań b) i c) suma czasów $T_{CS}+T_{103}$ wynosi już tylko około 0,1 s [122, 146].

Warto zauważyć, że opóźnienie rozpoczęcia transmisji nie zależy od sposobu uzyskania dwukierunkowości łącza, ponieważ uwzględnia czas transmisji tylko jednej ramki informacyjnej, niezależny od typu łącza.



Rys. 2.15. Czas przyspieszania (T_{SpUp}) dla łącza dwukierunkowego naprzemiennego
Fig. 2.15. "Speed-up" time (T_{SpUp}) for half-duplex link

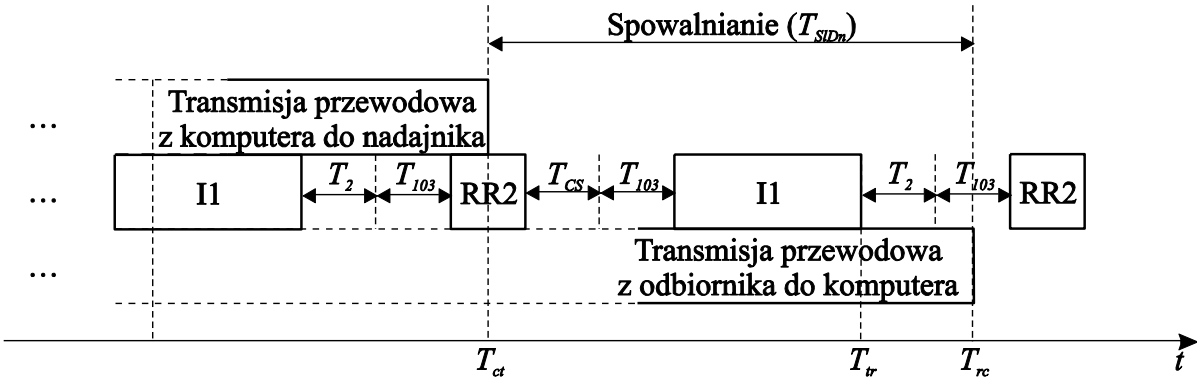
2.4.2.2. Opóźnienie zakończenia transmisji

Opóźnienie zakończenia transmisji, odpowiadające fazie „spowalniania" na rys. 2.13, jest znacznie trudniejsze do oszacowania [122, 146]. Wynika to z faktu buforowania informacji

w pamięci kontrolera TNC. Przy odpowiednio pojemnym buforze cała informacja może zostać przesłana po stronie nadającej w sposób ciągły, bez przerw spowodowanych zapełnieniem bufora. W tym czasie część danych jest już jednak przesyłana łączem radiowym. Ze względu na mniejszą prędkość efektywną tego łącza transmisja przewodowa po stronie odbiorczej odbywa się w sposób nieciągły. Aby zatem obliczyć moment zakończenia transmisji po stronie odbiorczej, należy uwzględnić moment ukończenia transmisji bezprzewodowej (T_{tr} na rys. 7) oraz czas transmisji nie więcej niż N_1 znaków łączem przewodowym. Transmisja bezprzewodowa jest opóźniona w stosunku do przesyłu po stronie nadawczej także o czas transmisji N_1 znaków łączem przewodowym ($T_{CT} - T_{TR} = T_a$). Jeśli przesył ten odbywa się w sposób ciągły, można łatwo wyznaczyć moment jego zakończenia ($T_{tr} = T_{TR} + T_p$). Biorąc powyższe pod uwagę, opóźnienie zakończenia transmisji można określić jako

$$T_{SlDn} = 2 \cdot T_a + T_p - \frac{10 \cdot L_D}{R_w}. \tag{2.16}$$

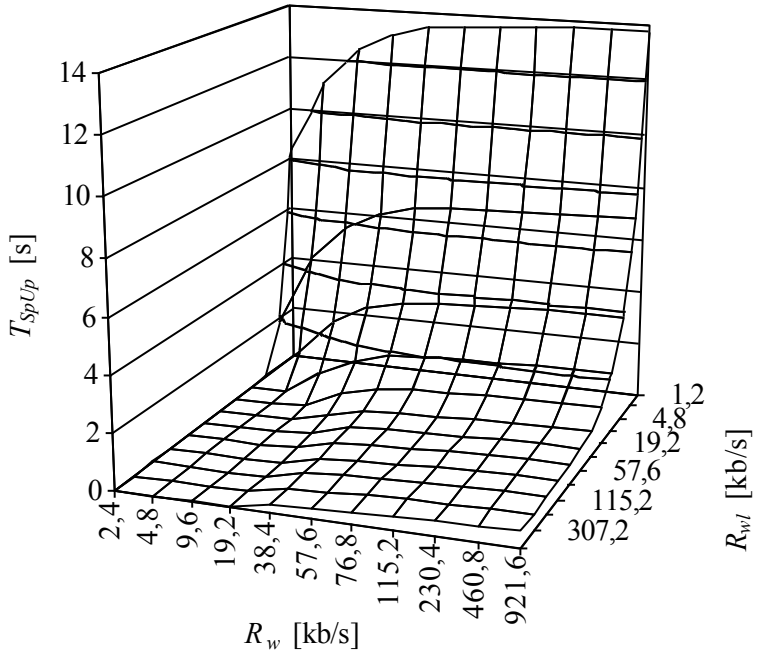
Wyjaśnienie zależności (2.16) pokazano na rys. 2.16 [122, 146]. Warto zauważyć, że transmisja bezprzewodowa kończy się już po upłynięciu czasu T_{SlDn}. Zjawisko to nie jest jednak istotne z punktu widzenia transmisji pomiędzy komputerami.



Rys. 2.16. Wyjaśnienie sposobu obliczenia czasu spowalniania (T_{SlDn})
Fig. 2.16. Explanation of "slow-down" time (T_{SlDn}) calculation method

Przedstawione rozważania są słuszne, gdy efektywna przepustowość łącza przewodowego jest wyższa niż analogiczna wielkość dla łącza bezprzewodowego, tj. $V_w > V_{wl}$. W przeciwnym bowiem przypadku równanie (2.16) może prowadzić do wartości ujemnych, które oczywiście są nieprawidłowe. W każdym jednak przypadku opóźnienie zakończenia transmisji powinno być nie mniejsze niż opóźnienie jej rozpoczęcia, ponieważ, niezależnie od prędkości transmisji, do przesłania łączem przewodowym pozostaje N_1 znaków. Wyjątkiem od tej reguły może być przypadek, w którym ostatnia ramka protokołu AX.25 zawiera niewielką liczbę bajtów danych, a łącze bezprzewodowe nie ustępuje przewodowemu pod względem efektywnej prędkości transmisji.

Na rys. 2.17 przedstawiono wyniki obliczeń czasu spowalniania w zależności od prędkości transmisji łącza przewodowego i bezprzewodowego dla łącza dwukierunkowego naprzemiennego [122, 146]. Przyjęto następujące wartości parametrów: N_1=256, k=7, L_D=1792. Ostatnia z tych wartości jest iloczynem dwóch poprzednich, co zapewnia maksymalną wydajność protokołu AX.25. Nie jest to zatem najgorszy przypadek – występuje on dla N_1=1 oraz k=1 – ale czas spowalniania i tak zależy głównie od rozmiaru przesyłanej informacji.



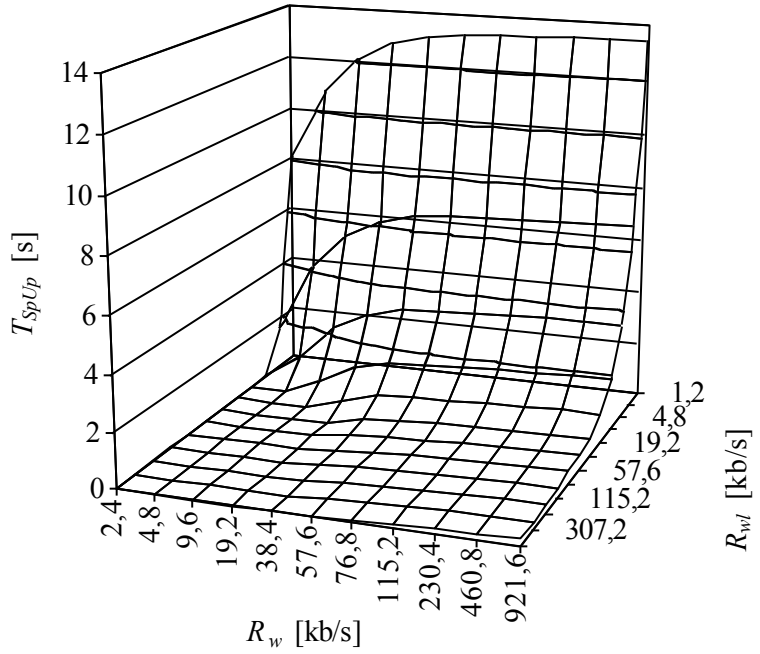
Rys. 2.17. Czas spowalniania (T_{SlDn}) dla łącza dwukierunkowego naprzemiennego
Fig. 2.17. "Slow-down" time (T_{SlDn}) for half-duplex link

Najniższe wartości opóźnienia występują przy niskich prędkościach łącza przewodowego (R_w≤19,2 kb/s) i wysokich – bezprzewodowego (R_w≥38,4 kb/s), ponieważ wówczas efektywna przepustowość łącza bezprzewodowego jest większa niż przewodowego. Dla takiego przypadku przyjęty model nie prowadzi do prawidłowych wartości, gdyż obliczone opóźnienia są ujemne. Należy wówczas założyć, że czas spowalniania nie powinien być krótszy od czasu przyspieszania, ponieważ także odpowiada on czasowi transmisji ramki informacyjnej zawierającej N_1 bajtów danych oraz transmisji tych danych łączem przewodowym.

Największe opóźnienia występują przy niskich prędkościach łącza bezprzewodowego (R_{wl}≤19,2 kb/s) i wysokich – przewodowego (R_w≥38,4 kb/s). O ile jednak wysoka prędkość łącza przewodowego powoduje wzrost opóźnienia co najwyżej do około 1 s, o tyle niska prędkość łącza bezprzewodowego znacznie bardziej sprzyja wzrostowi opóźnienia – przy R_{wl}=1,2 kb/s opóźnienie przekracza 10 s praktycznie dla wszystkich R_w.

W przypadku łącza dwukierunkowego jednoczesnego [122, 146] czas spowalniania wykazuje charakterystykę podobną jak dla łącza naprzemiennego, największe i najmniejsze

opóźnienia występują bowiem dla tych samych wartości parametrów (rys. 2.18). Nieco większy jest natomiast obszar, w którym przyjęty model wylicza ujemne wartości opóźnień. Jest to spowodowane większą przepustowością efektywną łącza dwukierunkowego jednoczesnego, w związku z czym większa, niż w przypadku łącza naprzemiennego, liczba prędkości transmisji łącza przewodowego charakteryzuje się niższą przepustowością efektywną.



Rys. 2.18. Czas spowalniania (T_{SlDn}) dla łącza dwukierunkowego jednoczesnego
Fig. 2.18. "Slow-down" time (T_{SlDn}) for full-duplex link

2.4.3. Dobór rozmiaru bufora

W niektórych zastosowaniach konieczne może być zapewnienie ciągłości transmisji po stronie nadawczej [122, 128]. Ponieważ w większości przypadków łącze przewodowe jest „szybsze" od radiowego (większa prędkość efektywna), przesyłane dane muszą być umieszczone w buforze. Minimalna pojemność bufora zależy od całkowitego rozmiaru przesyłanych danych L [B], a także od różnicy prędkości transmisji łącza przewodowego (V_w) i bezprzewodowego (V_{wl}) [b/s].

2.4.3.1. Pojemność bufora po stronie nadawcy

W celu obliczenia pojemności bufora po stronie nadawcy należy przyjąć, iż efektywna prędkość transmisji łącza przewodowego (po stronie nadawczej) odpowiada średniej prędkości napełniania bufora, zaś efektywna prędkość łącza radiowego – średniej prędkości jego opróżniania [128]. Mnożąc ich różnicę przez czas transmisji informacji łączem przewodowym, można otrzymać przybliżony rozmiar bufora, gwarantujący ciągłość transmisji na łączu przewodowym:

$$C = (V_w - V_{wl}) \cdot (T_{ct} - T_{CT}) = \left(\frac{8 \cdot R_w}{10} - \frac{8 \cdot L}{T_p} \right) \cdot \frac{10 \cdot L}{8 \cdot R_w} = L \cdot \left(1 - \frac{10 \cdot L}{R_w \cdot T_p} \right). \tag{2.17}$$

Powyższą zależność można wyjaśnić następująco. W czasie $T_w=L/V_w$ kontroler TNC otrzymuje od komputera L znaków. Jednakże, jeśli $V_{wl}<V_w$, może przesłać bezprzewodowo nie więcej niż $V_{wl}T_w=L \cdot V_{wl}/V_w$ znaków. Pozostałe dane zatem przekraczają możliwości transmisyjne kontrolera TNC i protokołu AX.25, muszą więc być umieszczone w buforze. Pojemność tego bufora można określić jako

$$C = L \cdot \left(1 - \frac{V_{wl}}{V_w} \right), \tag{2.18}$$

co, po kilku przekształceniach, prowadzi do zależności (2.17).

Warto zauważyć, że obie zależności są prawdziwe tylko, gdy $V_{wl}<V_w$. Jeśli warunek ten nie jest spełniony, mogą one prowadzić do uzyskania wartości ujemnych, co w przypadku pojemności bufora oczywiście nie ma sensu. Należy odrzucić także wyniki mniejsze niż N_1, gdyż tak obliczona pojemność bufora nie gwarantuje przechowania wystarczającej ilości danych dla ramek zawierających N_1 znaków.

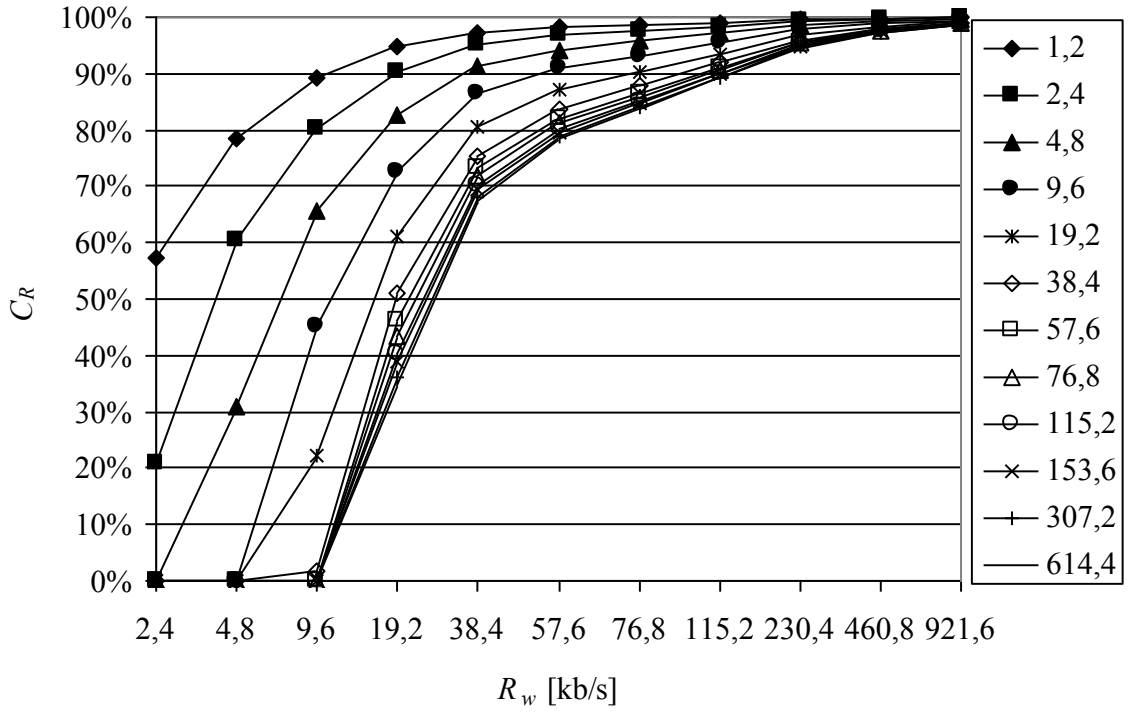
Pojemność bufora jest liniowo zależna od całkowitej wielkości przesyłanej informacji. Dlatego też bardziej ogólne wnioski można przedstawić na podstawie obliczeń względnej pojemności bufora. Można ją obliczyć, dzieląc pojemność bufora C przez całkowitą wielkość przesyłanej informacji L:

$$C_R = \frac{C}{L} = 1 - \frac{V_{wl}}{V_w}. \tag{2.19}$$

Względną pojemność bufora można zatem zinterpretować jako część przesyłanej informacji, która jest nadmiarowa w stosunku do możliwości transmisyjnych łącza bezprzewodowego przy danej prędkości łącza przewodowego.

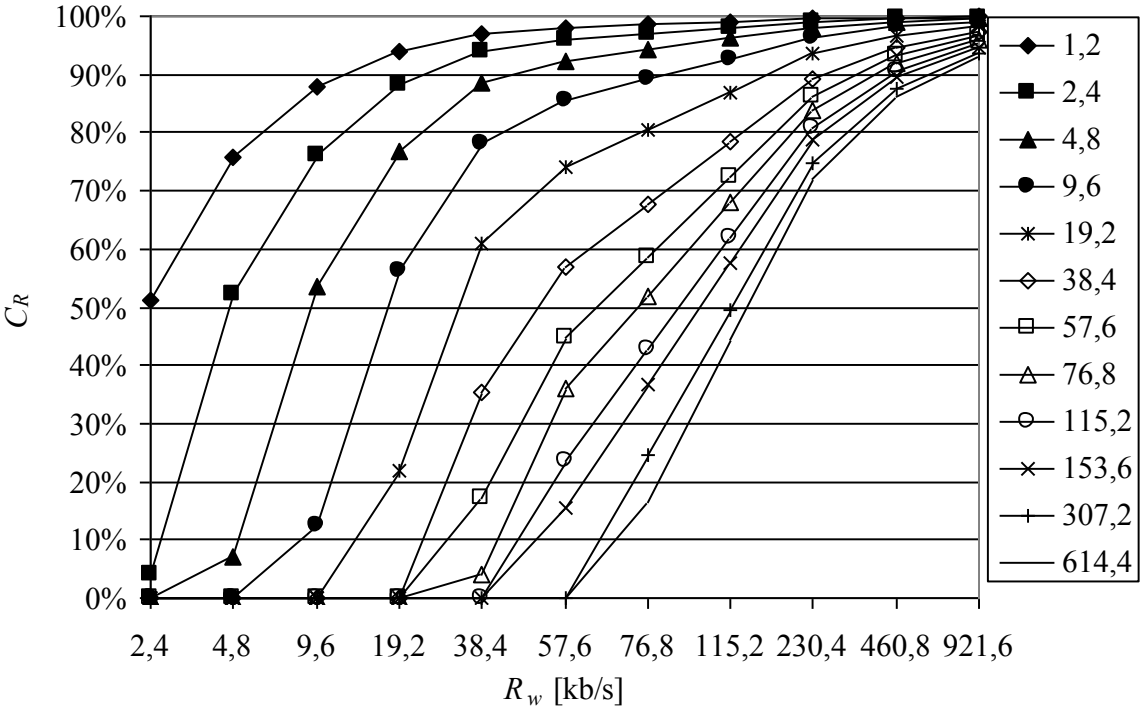
Na rys. 2.19 przedstawiono wyniki obliczeń względnej pojemności bufora dla łącza bezprzewodowego dwukierunkowego naprzemiennego [128]. Widać wyraźnie, że pojemność bufora rośnie bardzo szybko i z łatwością przekracza próg 90% objętości przesyłanej informacji. Przykładowo, gdy R_{wl}=1,2 kb/s i R_w=2,4 kb/s, zbuforować należy ponad połowę informacji (ok. 57%). Wraz ze wzrostem prędkości transmisji łącza przewodowego zapotrzebowanie na pojemność bufora rośnie. I tak, gdy R_w=4,8 kb/s, zbuforować należy niemal 80% informacji, gdy R_w=9,6 kb/s – około 90%. Dla najwyższej rozważanej prędkości transmisji łącza bezprzewodowego, tj. R_{wl}=614,4 kb/s, niezerowa pojemność bufora jest wymagana już przy R_w=19,2 kb/s i wynosi ona około 37% objętości informacji. Jest to zrozumiałe, ponieważ pomimo wysokiej prędkości transmisji efektywna przepustowość łącza bezprzewodowego naprzemiennego jest niewielka i dla przyjętych wartości parametrów wynosi około 18 kb/s.

Ogólnie rzecz ujmując, im wyższa prędkość transmisji łącza bezprzewodowego lub niższa przewodowego, tym później zaczyna się szybki wzrost względnej pojemności bufora. Zjawisko to odzwierciedla różnice między efektywną przepustowością obu łączy.



Rys. 2.19. Względna pojemność bufora dla łącza dwukierunkowego naprzemiennego
Fig. 2.19. Relative buffer capacity for half-duplex link

W przypadku łącza dwukierunkowego jednoczesnego (rys. 2.20) wymagana pojemność bufora rośnie niemal tak samo szybko jak dla naprzemiennego [128]. Wzrost ten występuje jednak przy wyższych prędkościach transmisji łącza przewodowego. Przykładowo, gdy R_{wl}=1,2 kb/s, pojemność bufora jest praktycznie taka sama jak dla łącza naprzemiennego. Natomiast gdy R_{wl}=614,4 kb/s, bufor nie jest wymagany dla $R_w \leq$57,6 kb/s, zaś gdy R_w=76,8 kb/s jego pojemność powinna wynosić co najmniej 16% objętości przesyłanych danych. Jednakże przy prędkościach gdy $R_w \geq$230,4 kb/s, pojemność ta wynosi ponad 70% objętości. Mniejsza wymagana pojemność bufora dla łącza jednoczesnego wynika z jego większej przepustowości efektywnej. Zjawisko to jest szczególnie widoczne dla mniejszych prędkości transmisji łącza przewodowego. Przy prędkościach większych, pomimo wyższej wydajności łącza bezprzewodowego, różnica prędkości jest mimo wszystko zbyt duża, skutkiem czego wymagana pojemność bufora jest praktycznie taka sama jak dla łącza naprzemiennego.



Rys. 2.20. Względna pojemność bufora dla łącza dwukierunkowego jednoczesnego
Fig. 2.20. Relative buffer capacity for full-duplex link

2.4.3.2. Pojemność bufora po stronie adresata

Zachowanie ciągłości transmisji po stronie odbierającej także jest możliwe, wymaga jednak znajomości całkowitej wielkości przesyłanych danych (L) w kontrolerze odbierającym [128]. Konieczne jest wówczas dodatkowe opóźnianie rozpoczęcia transmisji przewodowej po stronie odbierającej (T_{RC}). Należy także wziąć pod uwagę, iż odbierający kontroler przesyła informacje z łącza bezprzewodowego do przewodowego. Średnia prędkość napełniania bufora będzie zatem równa prędkości efektywnej łącza bezprzewodowego, zaś średnia prędkość opróżniania bufora – prędkości efektywnej łącza przewodowego.

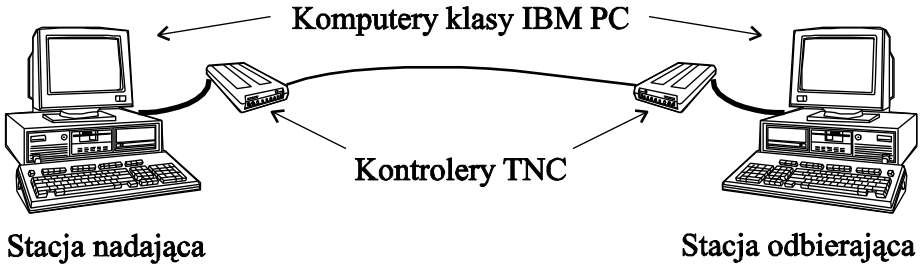
W praktyce, zachowanie ciągłości transmisji po stronie odbierającej wymaga, aby cała informacja była już odebrana przez kontroler TNC. W przeciwnym przypadku, ze względu na możliwe błędy transmisji, występujące podczas przesyłu bezprzewodowego pewne fragmenty danych mogłyby nie zostać dostarczone wystarczająco wcześnie, by zapewnić ciągłość transmisji na łączu przewodowym.

2.5. Porównanie wydajności kontrolerów TNC

Przedstawione w rozdziałach 2.3 i 2.4 obliczenia analityczne uwzględniają pracę sieci w warunkach idealnych, w szczególności zaś pomijają wpływ sprzętu i oprogramowania komunikacyjnego na użyteczne parametry transmisji. Biorąc pod uwagę znaczne różnice w pa-

rametrach konstrukcyjnych poszczególnych typów kontrolerów TNC, można się spodziewać, iż uzyskane wyniki będą zależały od użytego sprzętu. Jeśli ponadto dla danego typu kontrolera istnieje wiele różnych rodzajów i wersji oprogramowania sterującego, również może mieć to wpływ na uzyskane wyniki.

Badanie wydajności kontrolerów TNC wykonywano w doświadczalnej sieci, zawierającej jeden lub dwa komputery klasy IBM PC oraz dwa kontrolery [120, 151]. Pojedynczy komputer PC wystarcza, jeśli posiada dwa porty szeregowe lub USB, zależnie od użytych w danym teście kontrolerów. Transmisja między kontrolerami odbywała się przewodowo. Tę nieco egzotyczną – jak dla układów przeznaczonych dla sieci radiowych – konfigurację wybrano, aby uniknąć negatywnego wpływu zakłóceń radiowych na jakość transmisji. Ponadto, w tak zbudowanej sieci istnieje całkowita dowolność doboru parametrów transmisji, nieograniczona możliwościami nadajników-odbiorników radiowych, a jedynie możliwościami badanych kontrolerów (tabele 2.5 i 2.6 w rozdziale 2.2.1). Dzięki temu możliwe jest przeprowadzenie testów dla przypadków rzadko występujących lub trudnych do uzyskania w praktyce. Konfigurację sieci pokazano na rys. 2.21.



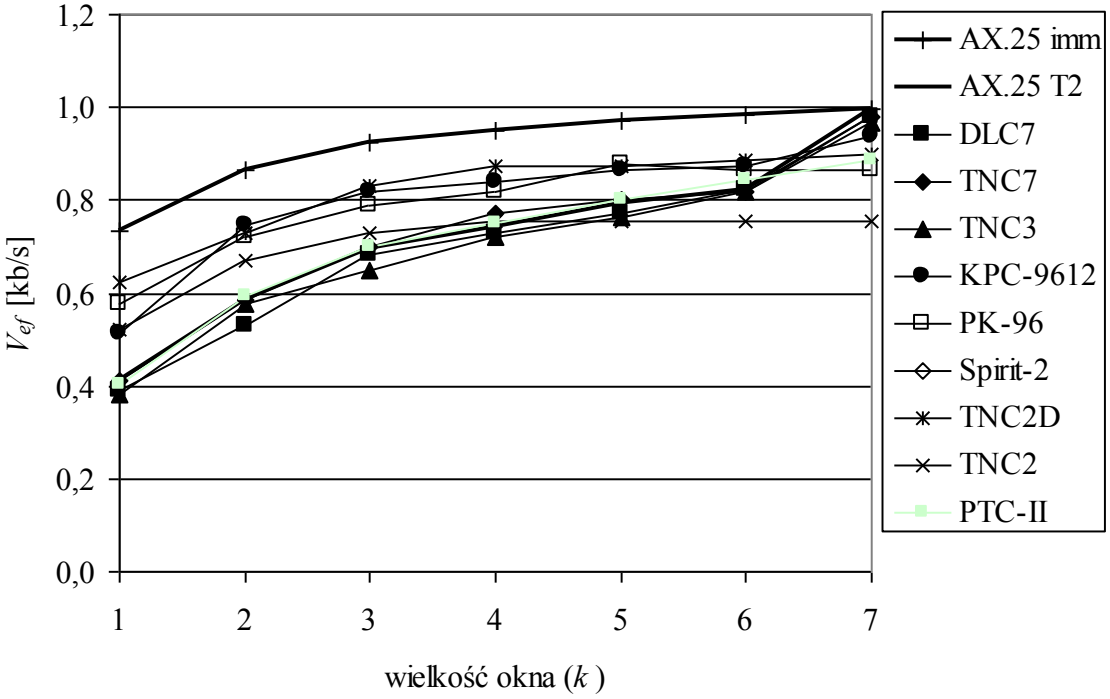
Rys. 2.21. Doświadczalna sieć Packet Radio
Fig. 2.21. Experimental Packet Radio network

2.5.1. Wpływ kontrolera na prędkość transmisji

Badania wykonywano, przesyłając plik o rozmiarze 8 lub 16 KB (zależnie od prędkości transmisji) przy różnych pojemnościach pola danych ramki protokołu AX.25 (N_1) i wielkościach okna (k). Mniejszy plik umożliwia porównanie z wynikami uzyskanymi we wcześniejszych badaniach [127, 141, 154], większy natomiast pozwala na zmniejszenie błędu pomiaru, wynikającego m. in. z konieczności transmisji danych między kontrolerami a komputerami PC oraz stosunkowo krótszego czasu transmisji przy użyciu nowocześniejszych kontrolerów TNC.

Na rys. 2.22 pokazano wyniki pomiarów efektywnej prędkości transmisji V_{ef} dla kilku wybranych kontrolerów TNC przy różnych wielkościach okna (k) i długości pola danych w ramce (N_1) równej 256 bajtów; prędkość łącza szeregowego wynosiła 19,2 kb/s, a radiowego – 1,2 kb/s [120, 151]. Na wykresie zamieszczono także, dla porównania, krzywą określającą teoretyczne możliwości protokołu AX.25. Z wykresu wynika, że osiągi kontrolerów nie

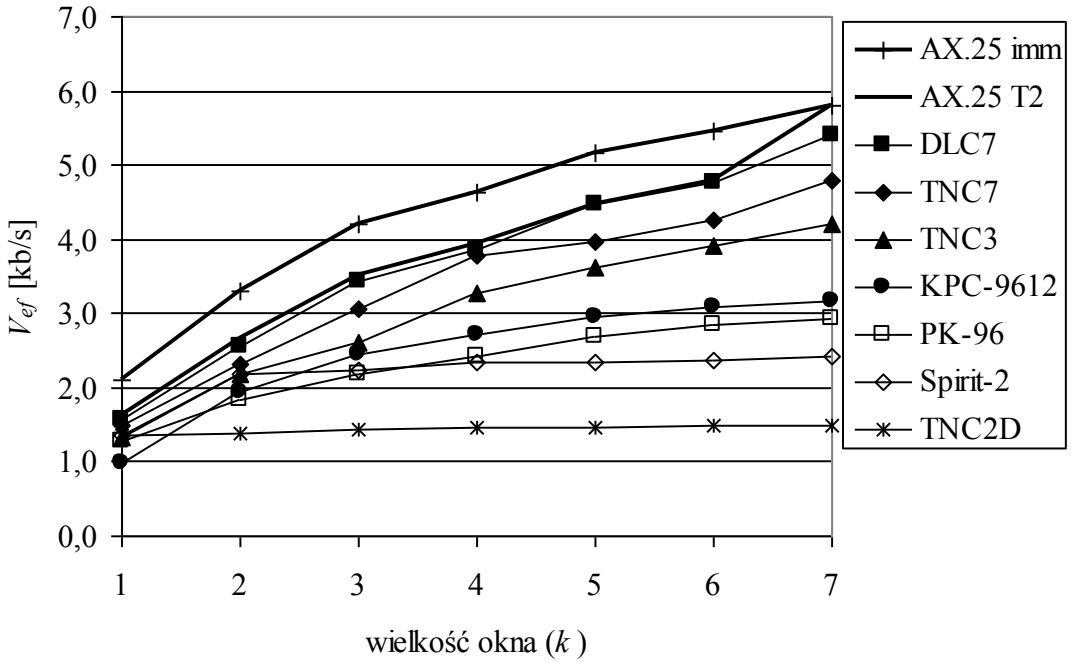
różnią się znacznie. Paradoksalnie, szybsze kontrolery TNC3 i TNC7 wypadają gorzej od pozostałych przy rozmiarze okna mniejszym niż 7. Dokładniejsza analiza wykonana w trybie monitorowania łącza wykazała, iż kontrolery te, w przeciwieństwie do pozostałych, nie wymuszają natychmiastowego przesłania potwierdzenia (bit P/F w polu sterującym ostatniej ramki w oknie). Dlatego też odbiorca czeka przez czas T_2 na nadejście kolejnej ramki i dopiero potem przesyła potwierdzenie. W przypadku maksymalnej wielkości okna zarówno TNC3, jak i TNC7 oraz DLC7 osiągają prędkości bliskie wartości teoretycznej, a jednocześnie wyższe niż pozostałe kontrolery. Jest to możliwe dlatego, że wówczas potwierdzenie jest przesyłane bez odliczania czasu T_2, gdyż zaimplementowana wersja protokołu nie dopuszcza przesłania większej liczby ramek w oknie.



Rys. 2.22. Wpływ użytego kontrolera TNC na efektywną prędkość transmisji (1,2 kb/s)
Fig. 2.22. TNC controller influence on effective throughput (1.2 kbps)

Na rys. 2.23 przedstawiono wyniki podobnych pomiarów, przeprowadzonych dla prędkości łącza radiowego 9,6 kb/s [120, 151]. W tym przypadku różnice są już znacznie bardziej wyraźne. W zależności od użytego kontrolera TNC największa efektywna prędkość transmisji waha się od około 1,5 kb/s (TNC2D) do prawie 5 kb/s (TNC7, DLC7). Nietrudno zauważyć – biorąc pod uwagę dane konstrukcyjne kontrolerów – że układy zawierające mikroprocesor Z80 cechują się najniższą wydajnością, aczkolwiek występują tu także pewne różnice. Przykładowo, kontroler Spirit-2 pozwala uzyskać niemal dwukrotnie wyższą przepustowość przy tych samych parametrach protokołu. Dla porównania, na wykresie przedstawiono także dwie krzywe, pokazujące teoretyczne możliwości protokołu. Jedna z nich (AX.25) odpowiada wersji z natychmiastowym potwierdzaniem, druga (AX.25 T2) – z odliczaniem czasu T_2. Jak

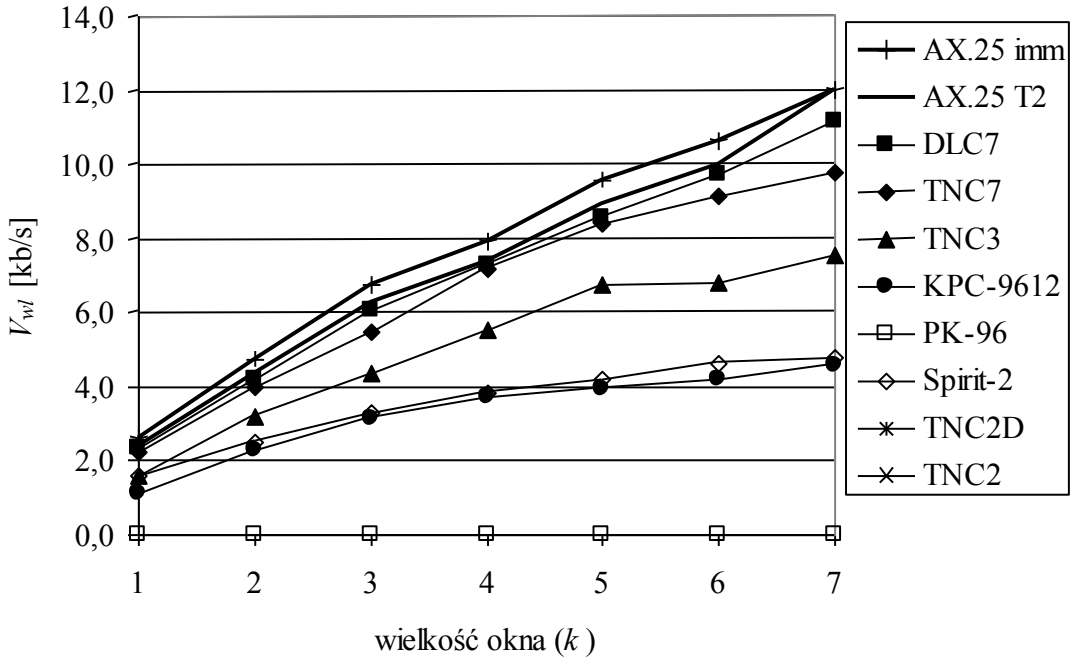
widać na wykresie, jedynie kontroler DLC7 umożliwia uzyskanie prędkości bliskiej wartościom teoretycznym, choć TNC7 i TNC3 mają osiągi niewiele gorsze. Wszystkie te kontrolery wypadają o wiele lepiej od pozostałych.



Rys. 2.23. Wpływ użytego kontrolera TNC na efektywną prędkość transmisji (9,6 kb/s)
Fig. 2.23. TNC controller influence on effective throughput (9.6 kbps)

Na rys. 2.24 przedstawiono wyniki podobnych pomiarów, wykonanych dla prędkości łącza radiowego 38,4 kb/s. Jest to najwyższa prędkość transmisji, która pozwala porównać większość dostępnych kontrolerów TNC. Jedynie konstrukcje oparte na mikroprocesorze Z80 (prócz kontrolerów Spirit-2) oraz zawierające modemy DSP nie zapewniają możliwości pracy z tą i wyższymi prędkościami. Z kolei kontroler PK-96, mimo iż może być skonfigurowany dla tej prędkości łącza radiowego, nie pracuje z nią poprawnie, chociaż kilka ramek udało się przesłać bezbłędnie. Prawdopodobną przyczyną jest ograniczenie szerokości pasma, uniemożliwiające stabilną pracę z prędkością 38,4 kb/s. Zniesienie tego ograniczenia wymaga modyfikacji filtrów analogowych modemu.

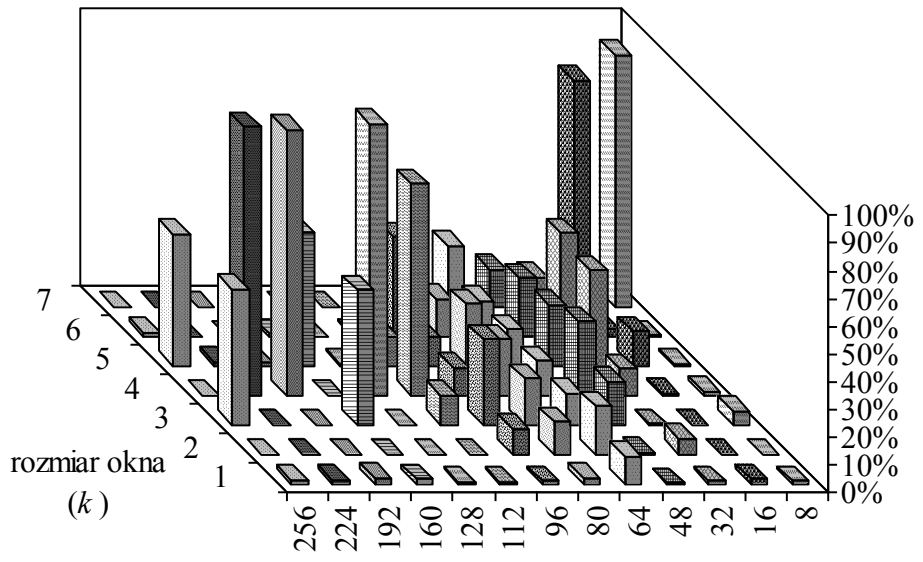
Warto zauważyć, iż różnica między kontrolerami Spirit-2 i KPC-9612+ nie jest duża. Wszystkie one jednak ograniczają efektywną prędkość transmisji do ok. 5 kb/s, podczas gdy TNC3 pozwala uzyskać prawie 8 kb/s, a TNC3 i DLC7 – odpowiednio 10 i 11 kb/s. Widać więc wyraźnie, iż moc obliczeniowa zastosowanego w kontrolerze mikroprocesora ma kluczowe znaczenie dla efektywnej prędkości transmisji.



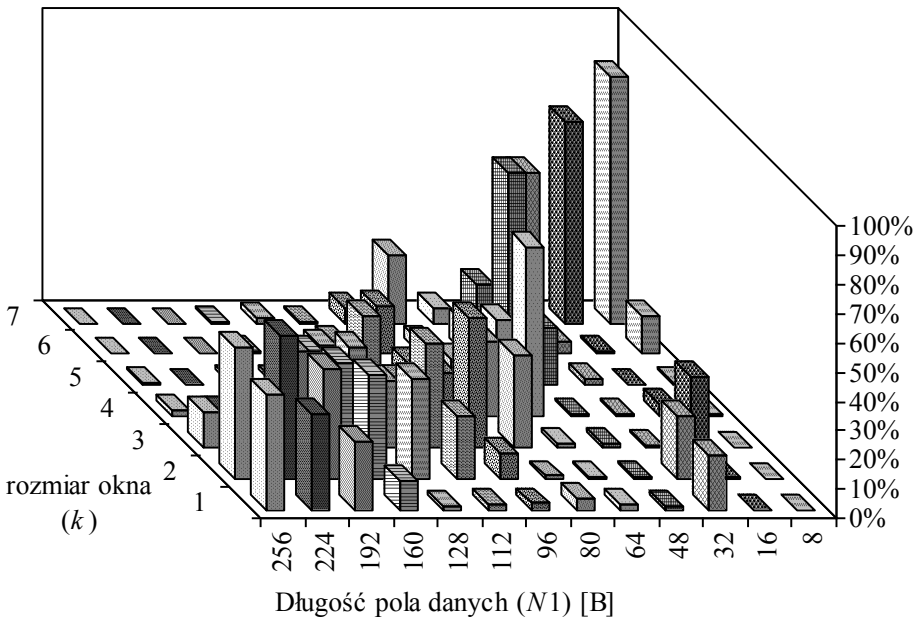
Rys. 2.24. Wpływ użytego kontrolera TNC na efektywną prędkość transmisji (38,4 kb/s)
Fig. 2.24. TNC controller influence on effective throughput (38.4 kbps)

W celu dokładniejszego określenia przyczyny niskiej wydajności kontrolerów zawierających procesor Z80 przeprowadzono dodatkowe badania. Polegały one na transmisji stosunkowo dużego pliku (64 KB) oraz rejestrowaniu przesyłanych ramek. Jest to możliwe w przypadku włączenia trybu monitorującego w TNC. Tryb ten pozwala na uzyskanie zdekodowanej informacji o przebiegu transmisji. W szczególności dla każdej ramki są podawane zawartości pól adresowych, typ i numer ramki oraz wartość bitu P/F. Na podstawie tak uzyskanego raportu określono faktyczne wielkości okna, występujące podczas transmisji. Histogramy, przedstawiające rozkład tych wielkości dla kilku wybranych kontrolerów i prędkości transmisji, pokazano na rys. 2.25÷2.29.

Z przedstawionych danych wynika, że kontrolery zawierające mikroprocesor Z80 nie są w stanie wykorzystać maksymalnej długości okna, i to praktycznie niezależnie od prędkości transmisji [120, 151]. Pomimo to wraz ze zmniejszającą się długością pola danych ramki, stosowane wielkości okna są coraz większe. Co ciekawe, dla ramek dłuższych (N_I>127 B) zaobserwowano, iż stosowane wielkości okna wynoszą na przemian 3 i 5 (jak na rys. 2.25), 2 i 6, a nawet 1 i 7, zależnie od typu kontrolera i wersji zastosowanego w nim oprogramowania. Można zatem przyjąć, że kontroler nie jest w stanie przetworzyć odpowiednio dużej ilości danych w odpowiednio krótkim czasie. Świadczyć to może o zbyt małej wydajności procesora, niskiej wydajności oprogramowania (brak odpowiedniej optymalizacji), zbyt małej pojemności pamięci użytej na bufory nadawczo-odbiorcze bądź celowych ograniczeniach wprowadzonych w oprogramowaniu.



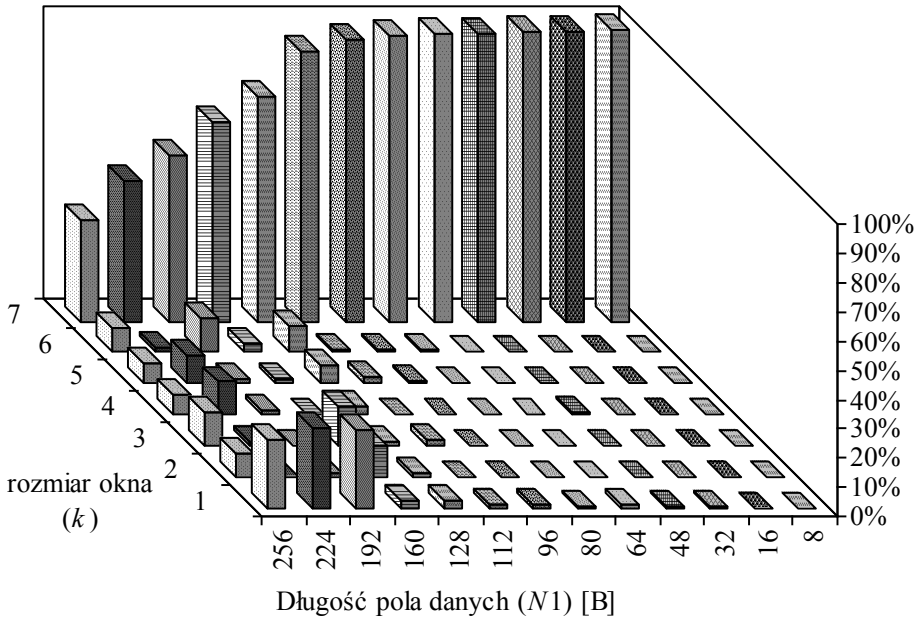
Rys. 2.25. Rozkład faktycznej wielkości okna dla kontrolera TNC2D przy prędkości 1,2 kb/s
Fig. 2.25. Real window size distribution for TNC2D controller at 1.2 kbps



Rys. 2.26. Rozkład faktycznej wielkości okna dla kontrolera TNC2H przy prędkości 9,6 kb/s
Fig. 2.26. Real window size distribution for TNC2H controller at 9.6 kbps

W kontrolerach zawierających procesor inny niż Z80, np. KPC-9612+, zdolność wykorzystania ustalonej wielkości okna jest znacznie większa (rys. 2.27). Dopiero dla ramek zawierających ponad 100 B danych można zauważyć spadek wydajności, jednak nawet przy maksymalnej długości ramki rozmiar okna 7 jest wielkością dominującą. Podobne zależności występują jednak także dla niższych prędkości transmisji łącza radiowego, np. 1,2 czy

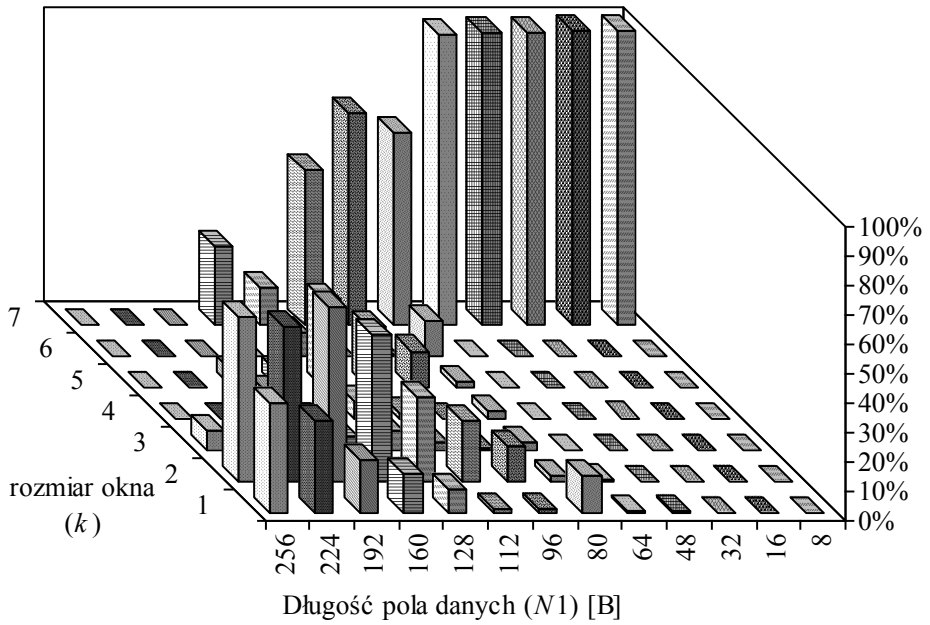
4,8 kb/s, a także dla wyższych. Zachowanie takie może świadczyć o celowym ograniczeniu wielkości okna, wprowadzonym w oprogramowaniu sterującym pracą kontrolera.



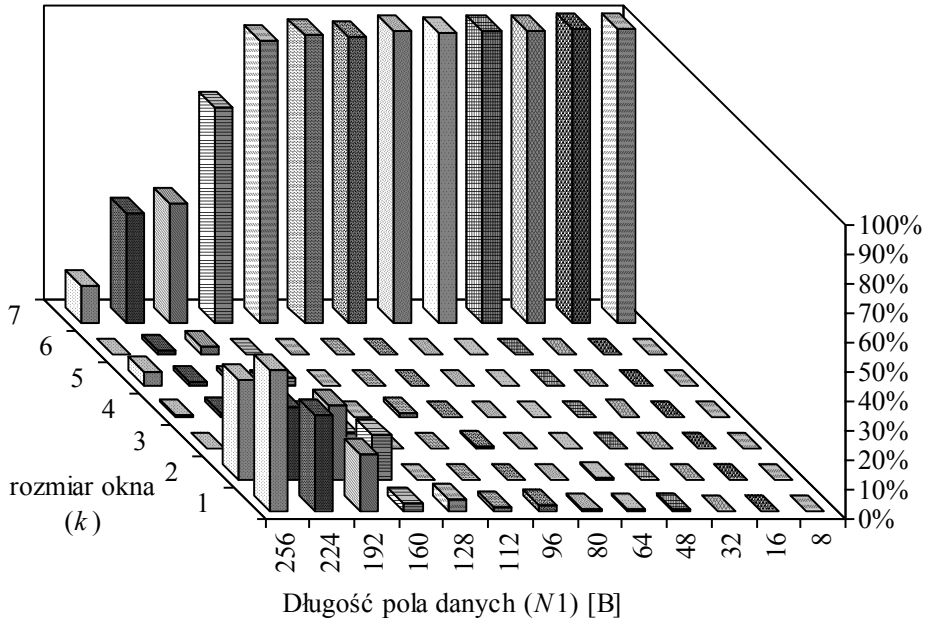
Rys. 2.27. Rozkład faktycznej wielkości okna dla kontrolera KPC-9612+ przy prędkości 9,6 kb/s
Fig. 2.27. Real window size distribution for KPC-9612+ controller at 9.6 kbps

Jeszcze lepsze osiągi uzyskują kontrolery TNC3, TNC7 i DLC7. W przypadku pierwszego z nich spadek wydajności spowodowany niepełnym wykorzystaniem ustawionej wielkości okna zauważa się dopiero przy prędkościach łącza radiowego przekraczających 150 kb/s. W dwóch pozostałych nie zauważono takiego efektu, nawet przy najwyższych możliwych prędkościach transmisji – odpowiednio 102,2 i 614,4 kb/s. Rozkład rzeczywistej wielkości okna dla kontrolera TNC3 przy prędkości łącza radiowego 307,2 kb/s pokazano na rys. 2.28, a przy prędkości 614,4 kb/s – na rys. 2.29. Porównując wyniki uzyskane dla dłuższych ramek, zawierających co najmniej 128 B danych, nietrudno zauważyć, iż – wbrew oczekiwaniom – wzrost prędkości transmisji nie zawsze powoduje zmniejszenie rzeczywistej wielkości okna.

Zdolność kontrolera do pełnego wykorzystania ustalonej wielkości okna jest ważną cechą. Brak jej bowiem powoduje zwiększenie narzutu protokołu, a co za tym idzie – zmniejszenie efektywnej prędkości transmisji. Wynika to po pierwsze z częstszych zmian kierunku transmisji oraz większej liczby potwierdzeń, po drugie natomiast – z występowania w cyklach transmisyjnych czasu T_2. Drugie z wymienionych zjawisk jest szczególnie dokuczliwe, gdy ustalona wielkość okna wynosi 7, gdyż w przypadku pełnego jej wykorzystania, jak pokazano na rys. 2.6, czas T_2 nie występuje.



Rys. 2.28. Rozkład faktycznej wielkości okna dla kontrolera TNC3 przy prędkości 307,2 kb/s
Fig. 2.28. Real window size distribution for TNC3 controller at 307.2 kbps



Rys. 2.29. Rozkład faktycznej wielkości okna dla kontrolera TNC3 przy prędkości 614,4 kb/s
Fig. 2.29. Real window size distribution for TNC3 controller at 614.4 kbps

2.5.2. Wpływ oprogramowania na wydajność transmisji

Podczas badań opisanych w poprzednim rozdziale zauważono występowanie różnic w implementacji protokołu AX.25. Dotyczą one szczególnie wykorzystania bitu P/F pola sterującego ramki informacyjnej. Niektóre wersje oprogramowania oznaczają ostatnią ramkę

okna przez ustawienie tego bitu, co – zgodnie ze specyfikacją protokołu – wymusza natychmiastową odpowiedź odbiorcy. Brak takiego oznaczenia powoduje, że odbiorca w wielu przypadkach czeka przez czas T_2 na ewentualną kolejną ramkę, co oczywiście zmniejsza efektywną prędkość transmisji. Można także zaobserwować różnice w czasie przygotowywania ramek do transmisji.

Dla kontrolerów TNC zawierających mikroprocesor Z-80 istnieje kilka rodzajów oprogramowania sterującego, dostępnych często w wielu wersjach. Niestety, wyboru takiego nie ma dla nowocześniejszych układów, dlatego też badając wpływ oprogramowania na wydajność transmisji, można było wykorzystać jedynie kontrolery rodziny TNC2 i Spirit. Przy takim podejściu – zawężając badania do pojedynczej platformy sprzętowej – unika się wpływu innych możliwych czynników na uzyskane wyniki.

Badania przeprowadzono, używając czterech typów kontrolerów TNC (TNC2, TNC2D, Spirit-2 Standard oraz Spirit-2 High Speed) [148, 152], pracujących przy różnych częstotliwościach taktowania mikroprocesora (tabela 2.5). Różnice między tymi układami występują także w zakresie prędkości transmisji łącza przewodowego i bezprzewodowego, a dostępność poszczególnych prędkości wynika z możliwości wbudowanych modemów (tabela 2.6). Wymienione powyżej kontrolery są ze sobą w pełni kompatybilne, chociaż układy Spirit-2 wymagają odpowiedniej konfiguracji. Oprogramowanie można zatem łatwo wymienić przez zmianę pamięci EPROM zawierającej program sterujący pracą kontrolera.

Podczas badań w kontrolerach wykorzystano następujące oprogramowanie sterujące:

- MFJ (od inicjałów autora, Martina F. Jue) w wersjach 1.1.4, 1.1.9 oraz 1.2.6, dostarczone z kontrolerami TNC2, TNC2D i podobnym TNC2H;
- TF (ang. *The Firmware*, *Turbo Firmware*) w wersjach 2.1d, 2.3b oraz 2.7b, dostarczone z kontrolerami TNC2D i podobnym TNC2H;
- Spirit-2 w wersji 5.0, dostarczone wraz z kontrolerami Spirit-2.

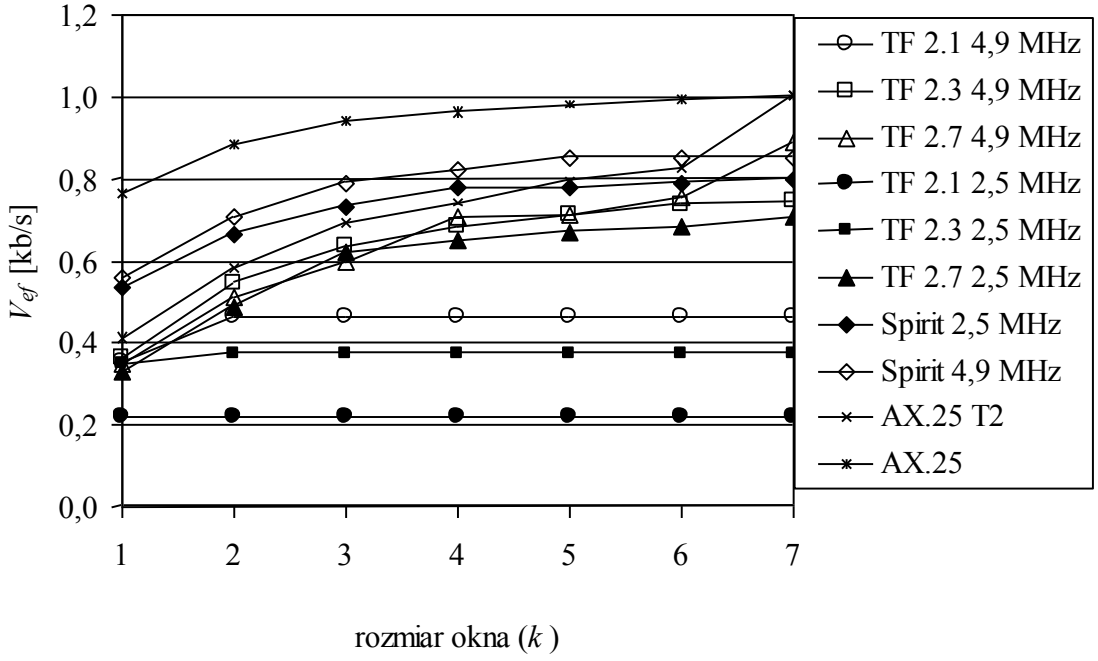
Badane kontrolery pełniły funkcję nadajników lub odbiorników. W obu przypadkach współpracowały one ze znacznie szybszymi (por. rozdział 2.5.1) kontrolerami TNC3 i TNC7. Ze względu na ich szybkość nie powinny one znacząco spowalniać transmisji przy połączeniu z testowanymi kontrolerami.

Podczas testów przesyłano plik o rozmiarze 8 KB. Wartość tę wybrano jako kompromis pomiędzy czasem transmisji a dokładnością pomiaru. Protokół AX.25 skonfigurowano tak, aby uzyskiwał najwyższą możliwą wydajność (k=7, N_1=256). Ze względu na parametry użytych kontrolerów – szczególnie prędkości transmisji łączy – testy przeprowadzono w dwóch konfiguracjach, „szybszej” i „wolniejszej”.

2.5.2.1. Konfiguracja „wolniejsza”

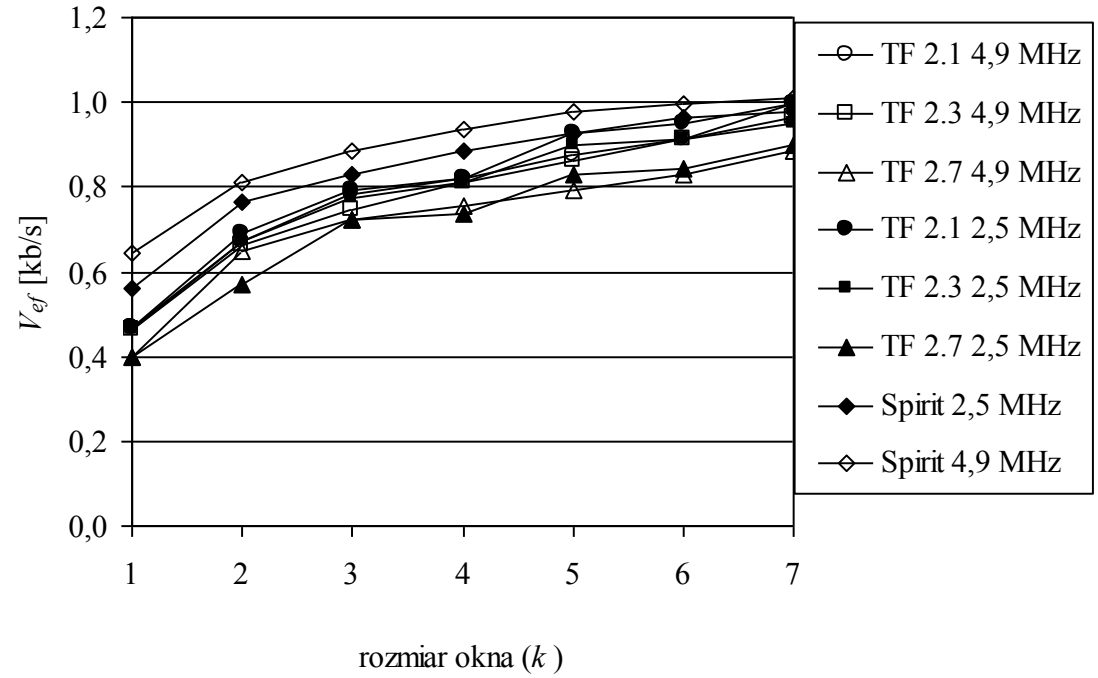
W konfiguracji „wolniejszej” użyto kontrolerów TNC2 oraz TNC2D z prędkościami transmisji R_w=9,6 kb/s oraz R_{wl}=1,2 kb/s [148, 152]. Wyniki pomiarów dla kontrolerów pracujących jako nadajniki przedstawiono na rys. 2.30, a jako odbiorniki – na rys. 2.31. Dla porównania, na wykresach przedstawiono także krzywe odpowiadające teoretycznej przepustowości protokołu AX.25, obliczonej według równania (2.9). Niezbędne wartości obliczono dla wersji protokołu z odliczaniem czasu T_2 (AX.25 T2) oraz z natychmiastowym przesyłaniem potwierdzeń (AX.25).

Jeśli badany kontroler TNC pracuje jako nadajnik, efektywna przepustowość sieci istotnie zależy od użytego oprogramowania. Wszystkie zbadane wersje oprogramowania Spirit-2 i MFJ wykazują zbliżoną wydajność (dla zwiększenia czytelności wykresu zawiera on wyniki tylko dla oprogramowania Spirit). Efektywna prędkość transmisji rośnie szybko przy wzroście wielkości okna (k) od 1 do 4. Dalsze zwiększanie tej wartości nie przynosi już znaczącego wzrostu prędkości transmisji. Może to świadczyć o celowych ograniczeniach w oprogramowaniu – np. rozmiaru bufora – które nie pozwalają przesłać w oknie więcej niż średnio 4 ramek o maksymalnej długości. Maksymalna prędkość transmisji zmienia się od około 0,75 do 0,8 kb/s dla wolniejszych kontrolerów (TNC2) i od 0,8 do 0,9 kb/s dla szybszych (TNC2D). Wyniki dla obu kontrolerów są jednak znacznie poniżej teoretycznej przepustowości protokołu AX.25, wynoszącej około 1 kb/s. Oprogramowanie TF zachowuje się odmiennie. Wersja 2.1 wykazuje najmniejszą efektywność w roli nadajnika, a możliwa do uzyskania prędkość transmisji nie zależy od wielkości okna. Podobna zależność jest widoczna dla wersji 2.3, jednak tylko w przypadku wolniejszego kontrolera TNC2. Na kontrolerze szybszym (TNC2D) zachowuje się podobnie do wersji 2.7. Można zatem przypuszczać, iż wersja 2.3 wymaga większej mocy obliczeniowej niż dostępna w kontrolerze TNC2. Z kolei wersja 2.7 wykazuje najwyższą wydajność niezależnie od kontrolera – prędkość efektywna rośnie wraz ze wzrostem wielkości okna. Tym niemniej, na kontrolerze TNC2D, jeśli k=7, uzyskuje się podobną prędkość efektywną jak w przypadku oprogramowania Spirit. Może to świadczyć o dłuższym przygotowywaniu ramek do transmisji, niż ma to miejsce w oprogramowaniu Spirit. Można to tłumaczyć następująco. Oprogramowanie Spirit, podobnie jak MFJ, uzyskuje zbliżoną wydajność dla $k \geq 4$. Pozwala to sądzić, iż rzeczywista wielkość okna nie przekracza średnio 4 nawet wówczas, gdy $k>4$. Z kolei wzrost efektywnej prędkości transmisji dla oprogramowania TF 2.7 pozwala sądzić, iż rzeczywista wielkość okna rośnie wraz ze wzrostem k. Powinno się to przełożyć na wyższą efektywną prędkość transmisji. Tak się jednak nie dzieje, co może świadczyć o dłuższym czasie przetwarzania informacji.



Rys. 2.30. Wpływ oprogramowania nadajnika na efektywną prędkość transmisji – konfiguracja „wolniejsza”

Fig. 2.30. Transmitter software influence upon effective throughput – “slower” configuration



Rys. 2.31. Wpływ oprogramowania odbiornika na efektywną prędkość transmisji – konfiguracja „wolniejsza”

Fig. 2.31. Receiver software influence upon effective throughput – “slower” configuration

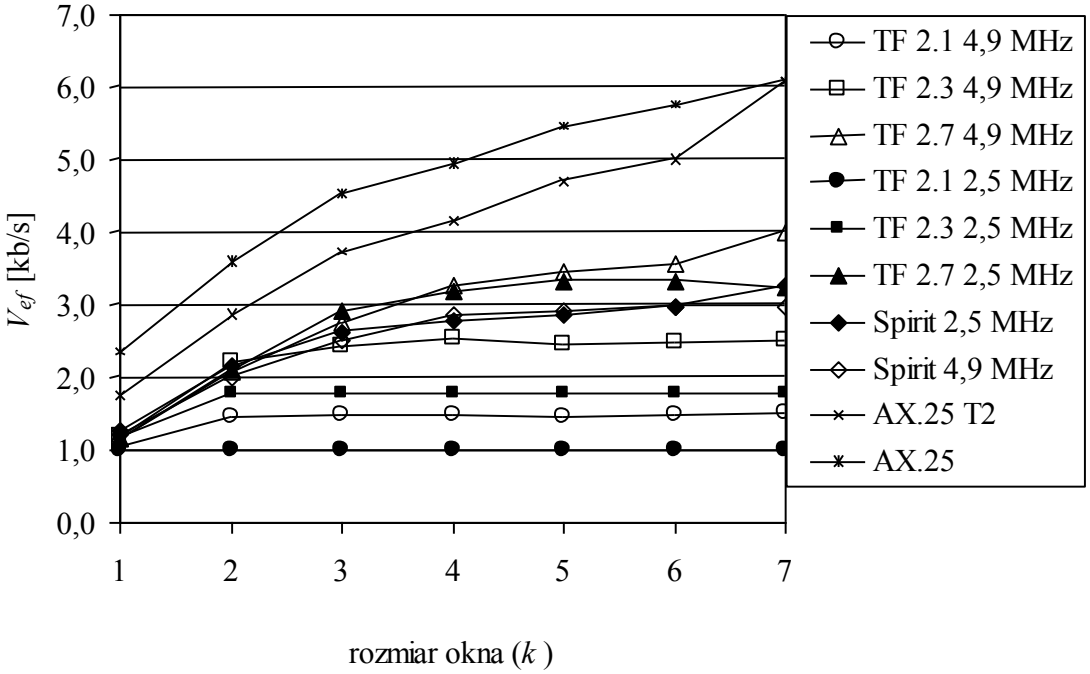
Gdy badany kontroler pełni funkcję odbiornika, różnica między najwolniejszym i najszybszym oprogramowaniem jest znacznie mniejsza. Można jednak zauważyć, że najsprawniejszy

odbiór jest możliwy w przypadku oprogramowania Spirit, szczególnie uruchomionego na szybszym kontrolerze TNC2D. Najwolniejszy odbiór występuje natomiast w przypadku oprogramowania TF 2.7, niezależnie od częstotliwości zegara taktującego mikroprocesor. Warto także zauważyć, iż wyniki te są znacznie bliższe wartościom teoretycznym i zmieniają się od około 0,9 do 1 kb/s, niezależnie od częstotliwości taktowania. Może to wynikać z faktu użycia znacznie szybszych kontrolerów (TNC3 lub TNC7) w roli nadajników. Pozwala to sądzić, iż moc obliczeniowa nadajnika jest znacznie bardziej istotna dla efektywnej prędkości transmisji niż moc obliczeniowa odbiornika. Tym niemniej, rodzaj oprogramowania sterującego pracą odbiornika nadal nie pozostaje bez znaczenia.

2.5.2.2. Konfiguracja „szybsza”

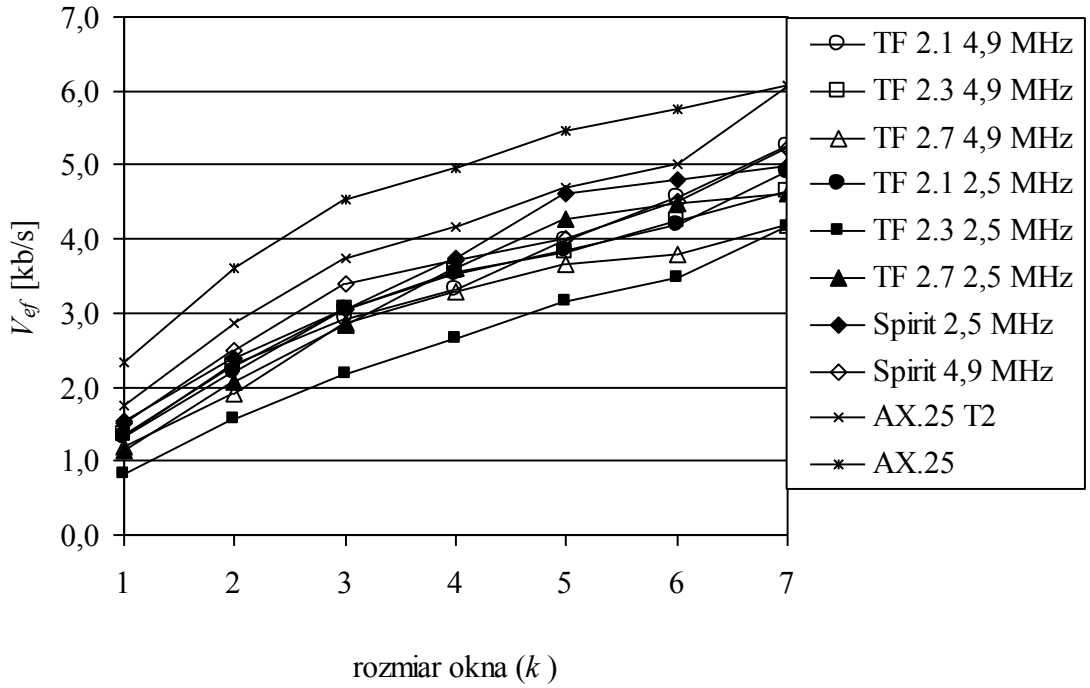
W konfiguracji „szybszej” użyto kontrolerów Spirit-2 w wersji Standard oraz High Speed z prędkościami transmisji R_{wl}=9,6 kb/s oraz R_w=57,6 kb/s [148, 152]. Wyniki pomiarów dla kontrolerów pełniących rolę nadajnika pokazano na rys. 2.32, odbiornika zaś – 2.33. Na wykresach zamieszczono także krzywe pokazujące teoretyczne możliwości protokołu AX.25 dla przyjętych w testach wartości parametrów.

Jeśli testowany kontroler pełni rolę nadajnika, wyniki są podobne do uzyskanych w wolniejszej konfiguracji. Ponownie najniższą prędkość transmisji uzyskuje się dla oprogramowania TF 2.1, niezależnie od częstotliwości taktowania mikroprocesora. Nieco lepsza jest wersja TF 2.3. Obie wersje pozwalają na osiągnięcie wyższej prędkości transmisji, gdy są uruchomione na szybszym mikroprocesorze. Tym niemniej wypadają one gorzej od jakiejkolwiek wersji oprogramowania Spirit lub MFJ, które uzyskują podobną wydajność niezależnie od częstotliwości taktowania mikroprocesora. Obserwacja ta może prowadzić do wniosku, że procedury transmisyjne są w tym przypadku dobrze zoptymalizowane, ale w oprogramowaniu istnieją pewne ograniczenia, które nie pozwalają uzyskać wyższej wydajności. Jest to szczególnie widoczne dla $k \geq 4$, kiedy to – podobnie jak w konfiguracji wolniejszej – zwiększanie rozmiaru okna nie przynosi zauważalnej poprawy efektywnej prędkości transmisji. Oprogramowanie TF 2.7 uzyskuje podobne prędkości jak MFJ oraz Spirit. W przypadku uruchomienia na szybszym mikroprocesorze (20 MHz) oprogramowanie TF 2.7 wykazuje najwyższą wydajność, choć tylko nieznacznie większą od wydajności innych programów. Najwyższa efektywna prędkość transmisji uzyskana w tym teście wynosi około 4 kb/s, podczas gdy według obliczeń może ona być nawet półtorakrotnie wyższa (około 6 kb/s). Obserwacja ta może prowadzić do wniosku, iż kontrolery TNC oparte na mikroprocesorze Z80 charakteryzują się zbyt niską mocą obliczeniową, aby efektywnie wykorzystać prędkość łącza radiowego 9,6 kb/s, bądź też oprogramowanie – szczególnie TF 2.7 – nie zostało odpowiednio zoptymalizowane.



Rys. 2.32. Wpływ oprogramowania nadajnika na efektywną prędkość transmisji – konfiguracja „szybsza”

Fig. 2.32. Transmitter software influence upon effective throughput – “faster” configuration



Rys. 2.33. Wpływ oprogramowania odbiornika na efektywną prędkość transmisji – konfiguracja „szybsza”

Fig. 2.33. Receiver software influence upon effective throughput – “faster” configuration

Jeśli testowany kontroler pełni funkcję odbiornika, wyniki także są podobne do uzyskanych w wolniejszej konfiguracji. Różnice między najszybszym i najwolniejszym odbiorni-

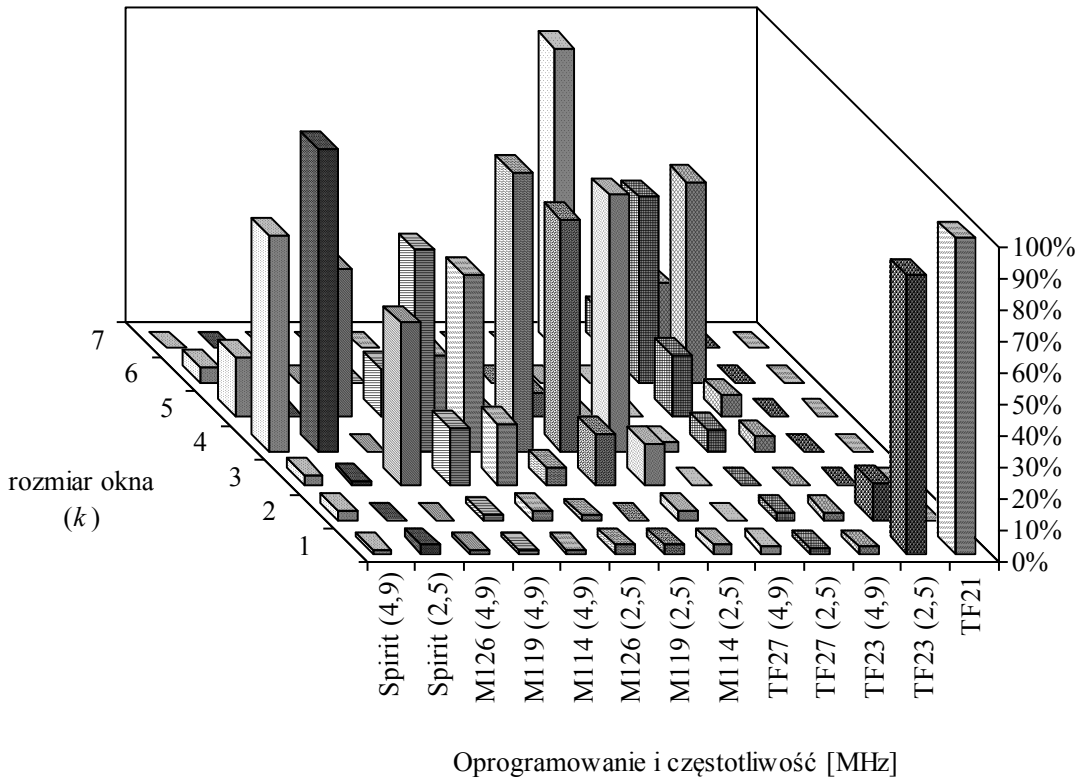
kiem nie są duże – efektywna przepustowość wynosi od około 4,1 kb/s do około 5,2 kb/s. Oba wyniki są poniżej oszacowania teoretycznego – inaczej niż w konfiguracji wolniejszej, gdzie wyniki doświadczalne są zbliżone do wyliczonych wartości teoretycznych. Najszybszym odbiornikami są programy TF 2.1 oraz Spirit, o ile uruchomiono je na szybszym mikroprocesorze (20 MHz). Zaskakująco zachowuje się natomiast oprogramowanie TF w wersji 2.7 – pozwala ono osiągnąć wyższą efektywną prędkość transmisji przy niższej częstotliwości taktowania mikroprocesora.

2.5.2.3. Rzeczywista wielkość okna

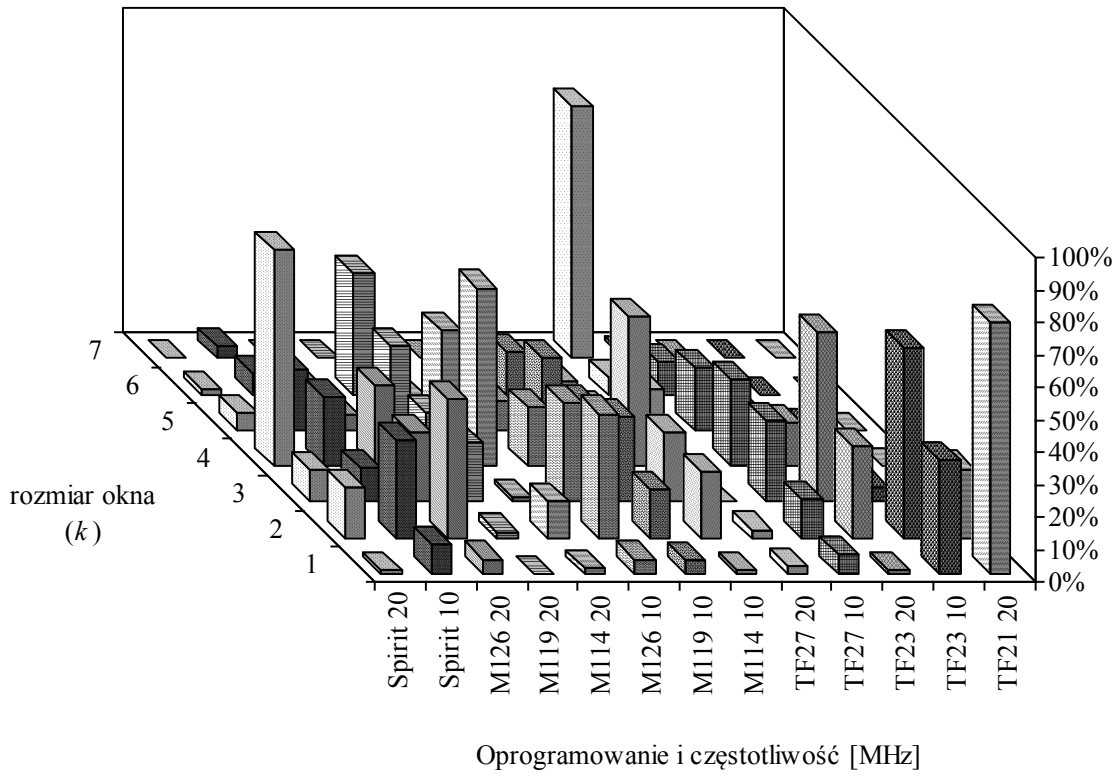
Powyższe badania wykazują, że jedną z przyczyn różnic w wydajności poszczególnych typów i wersji oprogramowania może być różny stopień wykorzystania ustawionej maksymalnej wielkości okna (parametr k). W celu dokładniejszego zbadania tego zagadnienia zapisywano raport transmisji w trybie monitorowania. Przeglądanie takiego raportu pozwala przeanalizować rzeczywisty proces wymiany ramek. Z tak zebranych danych można uzyskać rozkład rzeczywistej wielkości okna dla poszczególnych typów kontrolerów i programów sterujących ich pracą. Badania przeprowadzono dla najdłuższych możliwych ramek (N_1=256) oraz wielkości okna k=7 [148, 152]. Wyniki badań dla konfiguracji „wolniejszej" i „szybszej" pokazano na rys. 2.34 i 2.35.

Na przedstawionych histogramach wyraźnie widać, iż oprogramowanie TF zachowuje się zupełnie inaczej niż MFJ czy Spirit. W przypadku konfiguracji „wolniejszej" oprogramowanie MFJ oraz Spirit, niezależnie od wersji, osiąga rzeczywistą wielkość okna równą około 4, przy czym podczas transmisji występują okna o wielkości 3 i 5 (naprzemiennie), 2 i 6 (naprzemiennie) lub 4. Dla konfiguracji szybszej dominacja wartości k=4 nie jest tak oczywista – podczas transmisji częściej występują okna znacznie mniejsze.

Zachowanie oprogramowania TF jest bardziej uzależnione od częstotliwości taktowania mikroprocesora. Jeśli jest ona zbyt niska – powiedzmy 2,5 MHz – wersje 2.1 oraz 2.3 używają wielkości okna 1 lub 2. Sytuacja ta poprawia się w przypadku wersji 2.3, gdy częstotliwość wzrasta do co najmniej 4,9 MHz. Wersja 2.7 jest jeszcze bardziej wydajna i osiąga nawet wielkość okna równą 7, o ile tylko częstotliwość taktowania mikroprocesora jest wystarczająca dla danej prędkości transmisji. Przykładowo, dla prędkości 1,2 kb/s średnia wielkość okna wynosi około 6 przy częstotliwości taktowania 2,5 MHz i prawie 7 przy 4,9 MHz. Podobnie dla prędkości 9,6 kb/s średnia wielkość okna wynosi około 4 przy częstotliwości 10 MHz i prawie 7 przy 20 MHz. Niestety, nie przekłada się to na wzrost efektywnej prędkości transmisji (rys. 2.30 i 2.32), prawdopodobnie dlatego, że oprogramowanie TF 2.7 potrzebuje na przetworzenie wymaganej ilości informacji więcej czasu niż MFJ lub Spirit.



Rys. 2.34. Rozkład rzeczywistej wielkości okna dla konfiguracji „wolniejszej”
Fig. 2.34. Real window size distribution for “slower” configuration



Rys. 2.35. Rozkład rzeczywistej wielkości okna dla konfiguracji „szybszej”
Fig. 2.35. Real window size distribution for “faster” configuration

Gdy oprogramowanie TF w wersji 2.1 lub 2.3 jest uruchomione na mikroprocesorze taktowanym zegarem o częstotliwości co najmniej 10 MHz, wykorzystuje ono większe rozmiary okna pomimo wyższych prędkości transmisji. Oprogramowanie TF 2.7 przy częstotliwości 20 MHz zachowuje zdolność pełnego wykorzystania ustawionego rozmiaru okna. Tym niemniej nie uzyskuje ono znacząco większej wydajności niż oprogramowanie MFJ lub Spirit, które wykorzystuje wielkość okna średnio równą 4. Można zatem przypuszczać, iż możliwości oprogramowania TF w wersji 2.7 osiągnięto kosztem zmniejszonej prędkości przetwarzania ramek.

Przedstawione wyniki pokazują, że wydajność protokołu AX.25 zależy od rodzaju i wersji oprogramowania w podobnym stopniu jak od mocy obliczeniowej sprzętu transmisyjnego. Reakcja oprogramowania na zmianę mocy obliczeniowej sprzętu jest jednak różna. Niektóre rodzaje oprogramowania (np. MFJ czy Spirit) prawdopodobnie mają ograniczenia, które nie pozwalają im w pełni wykorzystać ustawionego rozmiaru okna przy ramkach o maksymalnej długości. Inne rodzaje oprogramowania mogą wykorzystać rozmiar okna w pełni. Nie przekłada się to jednak na wzrost efektywnej prędkości transmisji, ponieważ oprogramowanie to wymaga więcej czasu na przygotowanie do wysłania odpowiedniej liczby ramek. Można się zatem spodziewać wyższości tego typu oprogramowania w przypadku użycia procesorów o większej mocy obliczeniowej.

Podczas opisanych badań stwierdzono, że uzyskana w danej konfiguracji efektywna prędkość transmisji zależy nie tylko od sprzętu – a w szczególności od typu procesora i częstotliwości jego taktowania – lecz również od właściwości oprogramowania sterującego kontrolerem TNC. Wpływ na osiągi układu mogą mieć następujące czynniki, zależne wyłącznie od oprogramowania:

- pełne wykorzystanie ustawionej wielkości okna (k) przy każdej długości ramki (N_1),
- odpowiednio duża szybkość przetwarzania ramek protokołu AX.25,
- natychmiastowe (bez opóźnień) generowanie potwierdzeń poprawnego odebrania ramki informacyjnej,
- wymuszanie natychmiastowego potwierdzenia przez ustawienie bitu P/F w ostatniej ramce okna.

Brak możliwości pełnego wykorzystania ustawionej wielkości okna jest szczególnie dokuczliwy w kontrolerach zawierających mikroprocesor Z80, praktycznie niezależnie od prędkości jego taktowania i pojemności pamięci. Pewien wpływ na to ma rodzaj użytego oprogramowania. I tak, wersje obsługujące zbiór poleceń TAPR rzadko wykorzystują ustawioną wielkość okna w pełni – praktycznie nigdy nie następuje przesłanie w jednym cyklu więcej niż 5 ramek o maksymalnej długości. Nieco lepiej wypada tu oprogramowanie TF, które,

szczególnie w ostatniej wersji 2.7, potrafi wysłać nawet 7 ramek w jednym cyklu. Wydaje się jednak, iż możliwość ta jest okupiona dłuższym czasem przygotowania ramek do transmisji.

Kontrolery, obsługujące zbiór poleceń TAPR, ale zawierające inne mikroprocesory, znacznie lepiej potrafią wykorzystać ustawioną wielkość okna, i to nawet przy większych prędkościach. Ponieważ jednak brak dla nich innego oprogramowania, trudno określić, czy opisana możliwość wynika z większej mocy obliczeniowej mikroprocesora, czy też z optymalizacji kodu programu.

Dodatkowym czynnikiem wpływającym na efektywną prędkość transmisji jest sposób traktowania przez odbiorcę wielkości okna mniejszej niż 7. Jeśli nadawca nie oznacza ostatniej ramki cyklu transmisyjnego bitem P/F, oprogramowanie TF przesyła potwierdzenie dopiero po odliczeniu czasu T_2, natomiast TAPR – natychmiast. W niektórych wersjach oprogramowania TF czas ten można ustawić, w innych – np. w TNC3 – jest on obliczany automatycznie, bez możliwości modyfikacji. Można także ustawić ten parametr w niektórych wersjach oprogramowania TAPR.

Niektóre wersje oprogramowania, rozpoczynając transmisję, ograniczają początkowo wielkość okna, a następnie stopniowo ją powiększają aż do ustalonej wartości maksymalnej. Działanie to może być celowe, pozwala bowiem ustalić możliwości stacji odbierającej. Przy małym rozmiarze przesyłanej informacji powoduje to jednak spadek wydajności sieci.

2.6. Podsumowanie rozdziału

Nieniejszy rozdział poświęcono sieci Packet Radio oraz stosowanemu w niej protokołowi AX.25 i kontrolerom TNC, które są używane jako adaptery tej sieci. Na podstawie analizy działania protokołu AX.25 stworzono jego model analityczny, umożliwiający określenie jego wydajności, efektywnej prędkości transmisji oraz opóźnień występujących podczas przesyłu danych. W modelu uwzględniono najważniejsze parametry protokołu. Model pozwala określić zachowanie protokołu w warunkach idealnych, może zatem stanowić punkt odniesienia dla wyników osiągniętych w warunkach rzeczywistych. Przez porównanie takich wyników można wówczas ocenić wpływ implementacji protokołu oraz sprzętu i oprogramowania transmisyjnego na rzeczywiste osiągi sieci. Wykorzystując stworzony model, przeanalizowano wpływ poszczególnych parametrów protokołu na jego wydajność dla obu wariantów łącza radiowego, pracującego z różnymi prędkościami transmisji. Wykorzystując model protokołu AX.25, stworzono także model analityczny kontrolera TNC. Model ten pozwala oszacować teoretyczny wpływ kontrolera na efektywną prędkość oraz opóźnienia transmisji, pozwala także oszacować pojemność bufora w kontrolerze TNC, która gwarantuje ciągłość transmisji po stronie nadawczej.

Wykonano także liczne testy w doświadczalnej sieci Packet Radio, w której zapewniono warunki transmisji możliwie jak najbardziej zbliżone do idealnych. Dzięki temu wyniki doświadczalne mogły posłużyć do weryfikacji dokładności modelu analitycznego. Z drugiej strony, przez porównanie z wynikami uzyskanymi za pomocą modelu, można oszacować wpływ poszczególnych typów i wersji sprzętu i oprogramowania transmisyjnego na osiągi protokołu. Wyniki badań wykazały silną zależność parametrów użytkowych sieci zarówno od mocy obliczeniowej kontrolera TNC, jak i od oprogramowania sterującego jego pracą, a w szczególności od pewnych szczegółów implementacji protokołu AX.25.

Do najważniejszych, oryginalnych fragmentów rozdziału można zaliczyć:

- stworzenie modelu analitycznego, pozwalającego określić wydajność, efektywną prędkość transmisji oraz opóźnienia transmisji protokołu AX.25,
- stworzenie modelu analitycznego, pozwalającego określić wydajność, efektywną prędkość transmisji oraz opóźnienia transmisji przy stosowaniu w sieci kontrolerów TNC, a także określić minimalną pojemność jego bufora wymaganą w pewnych przypadkach,
- przeprowadzenie wielu testów w doświadczalnej sieci Packet Radio i określenie wpływu zarówno sprzętu, jak i oprogramowania transmisyjnego na efektywną prędkość transmisji w sieci przy różnych parametrach protokołu.

3. STANDARD IEEE 802.11

Standard IEEE 802.11 [52] można uznać obecnie za najistotniejsze rozwiązanie w zakresie bezprzewodowych sieci lokalnych. Może o tym świadczyć względnie duża liczba różnorodnych urządzeń, pozwalających na transmisję zgodnie z zasadami standardu, jak również stały proces modyfikacji i optymalizacji protokołu w celu uzyskania wyższych prędkości transmisji, większej efektywności czy możliwości wdrożenia nowych zastosowań, np. multimedialnych. Warto zauważyć, że w ciągu około 10 lat istnienia prędkość transmisji wzrosła ponad kilkudziesięciokrotnie (od 2 Mb/s do około 300 Mb/s), a przy tym ceny urządzeń również zanotowały kilkudziesięciokrotny spadek. Ponadto, wykorzystywanie pasma częstotliwości, należącego do zakresu ISM (ang. *Industrial, Scientific and Medical*), nie wymaga uzyskania zezwoleń czy licencji. Z tego względu trudno się dziwić, iż standard IEEE 802.11 jest używany do tworzenia nie tylko sieci lokalnych, ale także łączy dwupunktowych o większym zasięgu, co jest zrozumiałe wobec braku alternatywnych rozwiązań – sieci standardu WiMax (IEEE 802.16) czy GSM/GPRS/UMTS wymagają znacznie większych nakładów na infrastrukturę. Ponadto, sieci WiMax dopiero są wdrażane, zaś GSM/GPRS/UMTS mają ograniczoną prędkość transmisji. Z drugiej strony, standard IEEE 802.11 bywa wykorzystywany także w zastosowaniach, dla których stworzono bezprzewodowe sieci osobiste, a mianowicie do przesyłu informacji multimedialnej – wiele urządzeń o takim charakterze ma wbudowany interfejs bezprzewodowy IEEE 802.11, a nie np. 802.15.3 [55], który zaprojektowano z myślą o takich właśnie aplikacjach.

3.1. Rys historyczny

Pierwsze urządzenia do transmisji bezprzewodowej w sieciach lokalnych pojawiły się na rynku w początku lat 90. XX wieku [8]. Ze względu na brak jakichkolwiek standardów takiej transmisji można było zaobserwować dużą różnorodność oferowanych urządzeń. Z jednej strony, było to zjawisko korzystne, gdyż można było łatwiej dobrać urządzenie do danej aplikacji, ponadto umożliwiało poznanie zalet i wad poszczególnych rozwiązań, co w początkowym okresie rozwoju sieci było ważne. Z drugiej strony, urządzenia różnych producentów nie mogły ze sobą współpracować, nawet jeśli były stosowane podobne rozwiązania na poziomie

warstwy fizycznej i liniowej. Warto jednak zaznaczyć, że nieliczni wytwórcy sprzętu do transmisji bezprzewodowej opracowywali pewne elementy kart sieciowych wspólnie, tak więc ich urządzenia mogły ze sobą współpracować. Na podstawie tak uzyskanych doświadczeń opracowano podstawy standardu IEEE 802.11.

W chwili pojawienia się standardu (1999 rok) większość wytwórców rozpoczęła produkcję niemal wyłącznie kart sieciowych i punktów dostępu. Dopiero później zaczęły pojawiać się inne urządzenia, jak np. mosty, umożliwiające bezprzewodowe podłączenie dowolnego urządzenia wyposażonego jedynie w interfejs sieci Ethernet, bez konieczności modyfikacji jego struktury sprzętowej. Obecnie interfejs sieci bezprzewodowej standardu IEEE 802.11 jest praktycznie obowiązkowym wyposażeniem komputerów przenośnych (laptopów) i kieszonkowych (palmtopów). Spotyka się go także w niektórych cyfrowych tunerach satelitarnych, odtwarzaczach płyt DVD, aparatach fotograficznych, domowych centrach multimedialnych (ang. *media center*) czy kamerach sieciowych. Jako ciekawostkę można także wymienić możliwość użycia sieci standardu 802.11 w celu przenoszenia ramek standardu Bluetooth, wprowadzoną w najnowszej jego wersji (3.0) [96].

Tak wielka liczba urządzeń korzystających z sieci bezprzewodowej standardu IEEE 802.11 oraz nowe jego zastosowania wymusiły liczne modyfikacje standardu. Przede wszystkim warto wymienić nowe (w stosunku do pierwotnej wersji) warstwy fizyczne, umożliwiające transmisję ze znacznie wyższymi prędkościami (802.11 – 2 Mb/s, 802.11b – 11 Mb/s, 802.11a/g – 54 Mb/s). Na uwagę zasługuje także zwiększenie poziomu bezpieczeństwa sieci i zastąpienie przestarzałego już (i złamanego) protokołu WEP (ang. *Wired Equivalent Privacy*) protokołami z rodziny WPA (ang. *Wireless Protected Access*). Warto przy tym zauważyć, iż sieci 802.11 są praktycznie jedynymi w grupie lokalnych sieci bezprzewodowych, w których zagadnienie braku należytego bezpieczeństwa zostało dostrzeżone i w dużym stopniu rozwiązane, w przeciwieństwie np. do standardów Bluetooth [96] czy IrDA [97, 102, 111]. Ostatnie modyfikacje (802.11e) zapewniają wsparcie dla transmisji multimedialnych o różnych klasach ruchu, a także zwiększają wydajność protokołu poprzez wprowadzenie nowych strategii potwierdzeń. Prowadzone obecnie prace mają na celu stworzenie nowej warstwy fizycznej (802.11n), w której, ze względu na wymóg zachowania wysokiej wydajności sieci, wprowadza się zmiany prowadzące do znacznej redukcji narzutu protokołu nawet przy prędkościach transmisji rzędu kilkuset Mb/s. Jest to o tyle warte odnotowania, że „klasyczne" zasady wymiany ramek, wprowadzone w pierwotnej wersji standardu, już przy prędkościach rzędu kilkudziesięciu Mb/s obarczały transmisję narzutem rzędu kilkudziesięciu procent, co przekładało się na niewystarczającą efektywną prędkość transmisji.

Tak szeroka liczba zagadnień związanych ze standardem IEEE 802.11 powoduje, iż niektóre z nich – jak np. szczegóły realizacji kodowania i modulacji na poziomie warstwy fi-

zycznej czy też sposoby realizacji mechanizmów zabezpieczających – wykraczają poza ramy niniejszego opracowania.

3.2. Opis standardu IEEE 802.11

Opis standardu IEEE 802.11 można podzielić na kilka ważniejszych części:

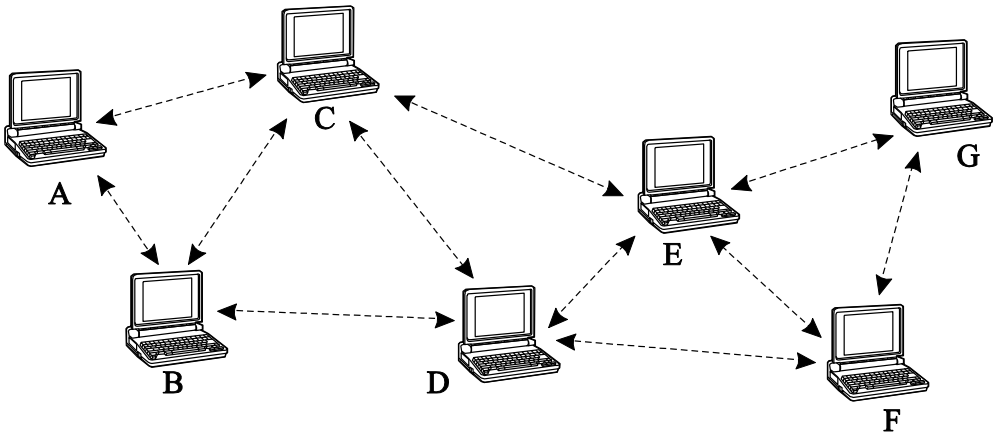
- opis topologii sieci,
- opis warstw fizycznych, stosowanych w standardzie 802.11,
- opis warstwy liniowej.

3.2.1. Topologie sieci

Standard IEEE 802.11 umożliwia tworzenie sieci bezprzewodowych o różnych strukturach. Sieci takie mogą zawierać elementy infrastruktury, umożliwiające m. in. połączenie sieci bezprzewodowej z siecią przewodową, ale można także tworzyć sieci pozbawione tych elementów.

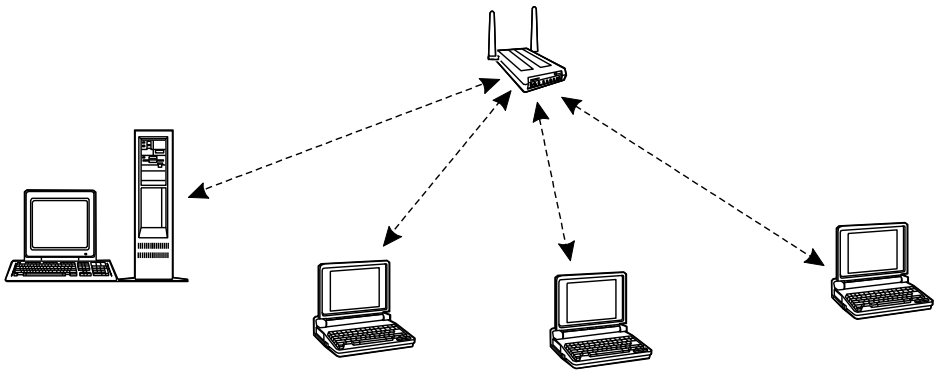
Podstawowe trzy konfiguracje sieci noszą nazwę zbiorów usług (ang. *service set*). Zbiór taki jest określany jako logiczne zgrupowanie urządzeń. Definicja ta wynika z faktu, iż stacja może znajdować się w zasięgu wielu sieci, które można rozróżnić za pomocą unikalnych identyfikatorów zbiorów usług (SSID, ang. *Service Set Identifier*). W przeciwieństwie do sieci przewodowych, w których wybór sieci wymaga podłączenia przewodu do odpowiedniego gniazdka, w sieci bezprzewodowej proces włączania stacji do sieci wymaga odpowiedniej wymiany informacji na poziomie warstwy liniowej.

Najprostsza konfiguracja sieci to tzw. niezależny, podstawowy zbiór usług (IBSS, ang. *Independent Basic Service Set*). W sieci takiej nie występują punkty dostępu, a wszystkie stacje są równoważne. Mogą one komunikować się ze sobą tylko bezpośrednio, ponieważ standard nie zapewnia mechanizmu przekazywania informacji do stacji znajdujących się poza zasięgiem nadawcy. Sieci takie często są tworzone naprędce i mają charakter tymczasowy, z tego też powodu nie wykonuje się wstępnych planów i pomiarów, choć oczywiście jest to możliwe. Nie ma także ograniczeń co do liczby urządzeń włączonych do sieci, należy jednak mieć na uwadze, iż w dużej sieci koordynacja i synchronizacja pracy stacji jest procesem czasochłonnym, a to ze względu na rozproszone wykonanie odpowiednich procedur. Sieci takie często są określane mianem sieci ad hoc. Przykład złożonej sieci typu IBSS, w której nie wszystkie stacje mogą komunikować się bezpośrednio, pokazano na rys. 3.1. Warto zauważyć, iż w sieci tej można prowadzić jednocześnie wiele transmisji, o ile tylko komunikujące się ze sobą grupy stacji znajdują się poza swoim zasięgiem. Przykładowo, w sieci pokazanej na rys. 3.1, stacje A i B mogą wymieniać informacje w tym samym czasie co F i G.



Rys. 3.1. Przykład złożonej sieci IBSS
Fig. 3.1. An example of a complex IBSS network

Podstawowa konfiguracja sieci zawierającej punkt dostępu to tzw. podstawowy zbiór usług (BSS, ang. *Basic Service Set*). Punkt dostępu pełni tu funkcje koordynatora pracy sieci i jest odpowiedzialny m. in. za uwierzytelnianie stacji, można więc powiedzieć, że prawidłowa konfiguracja tego urządzenia gwarantuje odpowiedni poziom bezpieczeństwa sieci. Komunikacja między stacjami może odbywać się wyłącznie za pośrednictwem punktu dostępu, choć niektóre elementy protokołu dopuszczają bezpośrednią transmisję między stacjami za zezwoleniem uzyskanym od punktu dostępu. Pewną wadą sieci typu BSS jest możliwość prowadzenia tylko jednej transmisji w danym czasie. Przykład sieci typu BSS pokazano na rys. 3.2.

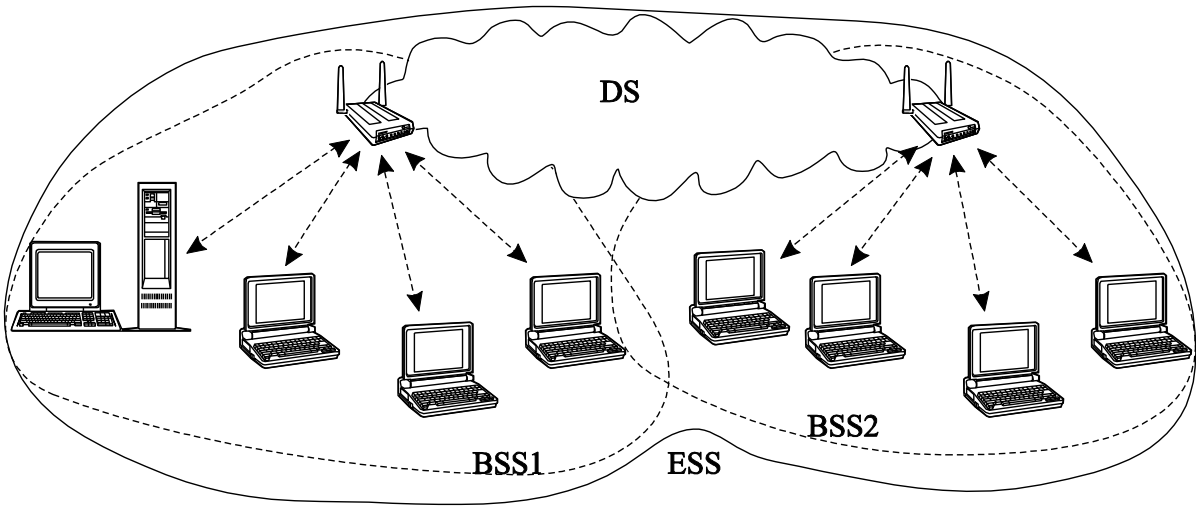


Rys. 3.2. Przykład sieci BSS
Fig. 3.2. An example of a BSS network

Bardziej rozbudowana konfiguracja, tzw. rozszerzony zbiór usług (ESS, ang. *Extended Service Set*) składa się z kilku jednostek BSS, połączonych za pomocą tzw. systemu dystrybucyjnego (DS, ang. *Distribution System*). Standard nie definiuje przy tym, czym dokładnie jest system dystrybucyjny. Wymaga się jedynie, aby umożliwiał on przekazywanie ramek łącza bezprzewodowego pomiędzy stacjami, które należą do różnych jednostek BSS. Wymóg taki może zostać spełniony przez wiele różnych rozwiązań sieciowych, jednak w praktyce jest

stosowana w tej roli niemal jedynie sieć Ethernet. Wynika to zapewne z filozofii standardu IEEE 802.11, który został pomyślany jako „bezprzewodowy Ethernet".

Sieć typu ESS ma bardzo zbliżoną charakterystykę do sieci BSS, z tą wszakże różnicą, iż występuje tu większa liczba punktów dostępu. Można zatem prowadzić jednocześnie większą liczbę transmisji, zwiększony jest także zasięg terytorialny sieci. Przykład sieci typu ESS pokazano na rys. 3.3.



Rys. 3.3. Przykład sieci ESS
Fig. 3.3. An example of an ESS network

3.2.2. Warstwa fizyczna

Warstwa fizyczna standardu IEEE 802.11 zawiera dwie podwarstwy:

- zależną od medium (PMD, ang. *Physical Medium Dependent*), której funkcje opisują charakterystykę użytego bezprzewodowego medium transmisyjnego oraz sposób nadawania i odbioru informacji przesyłanej między stacjami,
- niezależną od medium (PLCP, ang. *Physical Layer Convergence Protocol*), która dostosowuje możliwości podwarstwy PMD do wymagań warstwy liniowej, a to w celu zapewnienia jej odpowiednich usług; jeśli usługi te udostępnia podwarstwa PMD, to podwarstwa PLCP może nie realizować żadnych funkcji.

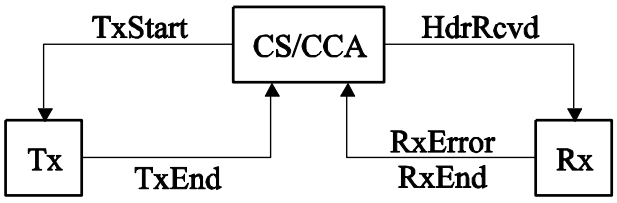
Opis mechanizmów stosowanych na poziomie warstwy fizycznej wykracza poza ramy niniejszego opracowania. Tym niemniej, pewne elementy tej warstwy – jak np. formaty ramek stosowanych na poziomie podwarstwy PLCP – mają istotny wpływ na wydajność protokołu warstwy liniowej i będą przydatne w dalszych rozważaniach.

3.2.2.1. Podwarstwa PLCP

Podwarstwa PLCP zapewnia przesył informacji pomiędzy warstwą liniową a podwarstwą PMD. Ponieważ rozwiązania przyjęte w poszczególnych odmianach podwarstwy PMD mogą być (i w rzeczywistości są) różne, na poziomie PLCP jest przesyłana informacja dodatkowa

w stosunku do jednostek występujących na poziomie warstwy liniowej. Zawiera ona preambułę i nagłówek. Preambuła pozwala na uzyskanie synchronizacji pomiędzy nadajnikiem a odbiornikiem, może być zatem potraktowana jako zapowiedź nadejścia ramki. Podczas odbioru preambuły stacje wyposażone w kilka anten mogą także dokonać wyboru anteny zapewniającej najlepszą jakość odbioru, jest także możliwe wprowadzenie niezbędnych pomiarów odbieranego sygnału w celu wprowadzenia odpowiednich korekcji [39].

Podwarstwa PLCP zawiera mechanizmy, pozwalające na informowanie o momencie rozpoczęcia lub zakończenia transmisji. W pewnym uproszczeniu można przyjąć, iż PLCP działa według schematu przedstawionego na rys. 3.4 [89].



Rys. 3.4. Graf stanów podwarstwy PLCP [89]
Fig. 3.4. PLCP sublayer state diagram

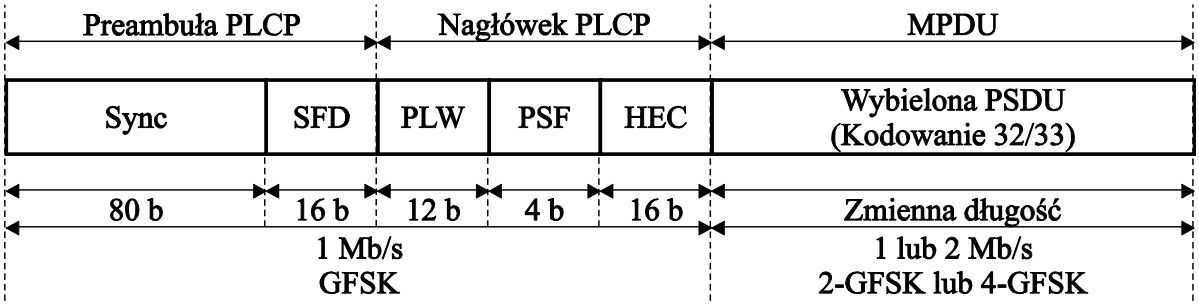
Większość czasu podwarstwa PLCP spędza w stanie CS/CCA (ang. *Carrier Sense/Clear Chanel Assessment*), w którym może ona wykryć rozpoczęcie transmisji przez inną stację oraz określić zajętość kanału. Wskutek otrzymania żądania rozpoczęcia nadawania (TxStart) przechodzi do stanu nadawania (Tx). W stanie tym podwarstwa PMD jest przełączona w tryb nadawania, co wiąże się z wysłaniem ramki. Po zakończeniu transmisji (TxEnd) powraca do stanu CS/CCA. Stan ten można także opuścić wskutek wykrycia rozpoczęcia transmisji przez inną stację (HdrRcvd). Przechodzi ona wówczas do stanu odbioru (Rx), który może zakończyć się prawidłowo (RxEnd) lub błędnie (RxError).

3.2.2.2. Warstwa fizyczna FHSS

Warstwa fizyczna z przeskokami częstotliwości (FHSS, ang. *Frequency Hopping Spread Spectrum*) została wprowadzona w pierwotnej wersji standardu [53]. Wykorzystuje ona fale radiowe z zakresu ISM (2,4÷2,4835 GHz). Pasmo to podzielono na 79 kanałów o szerokości 1 MHz[4]. Określono także 78 pseudolosowych sekwencji przeskoków, które podzielono na 3 grupy. Minimalna częstotliwość przeskoków nie jest określona i zależy od ustaleń obowiązujących w danym kraju. Prędkość modulacji jest stała i wynosi 1 Mbd, natomiast w zależności od użytej metody modulacji (2-GFSK czy 4-GFSK) można uzyskać prędkość transmisji, odpowiednio, 1 lub 2 Mb/s. Druga z wymienionych prędkości nie jest przy tym obowiązkowa.

[4] Taki podział obowiązuje w USA i większości krajów europejskich; dla niektórych innych krajów, jak np. Francja, Hiszpania czy Japonia, zdefiniowano osobne podziały.

Na poziomie podwarstwy PLCP informacje są przesyłane w ramkach, zawierających preambułę (ang. PLCP preamble), nagłówek (ang. PLCP header) oraz jednostkę PSDU. Format ramki na poziomie podwarstwy PLCP w warstwie fizycznej FHSS pokazano na rys. 3.5.



Rys. 3.5. Format ramki podwarstwy PLCP w warstwie fizycznej FHSS
Fig. 3.5. PLCP frame format in FHSS physical layer

Preambuła podwarstwy PLCP zawiera:

- 80-bitowy ciąg synchronizacyjny (Sync), składający się z wielokrotnie powtarzanego ciągu bitów o wartości 0 i 1,
- 16-bitowy znacznik początku ramki (SFD, ang. *Start Frame Delimiter*).

Z kolei nagłówek ramki zawiera:

- 12-bitowe pole PLW (ang. *PSDU Length Word*), określające długość ramki warstwy liniowej wyrażoną w bajtach, nie większą niż 4095 B;
- 4-bitowe pole PSF (ang. *PLCP Signalling Field*), określające prędkość transmisji tej ramki, wyrażoną w jednostkach 500 kb/s i mieszczącą się w zakresie 1,0÷4,5 Mb/s;
- 16-bitową sumę kontrolną nagłówka (HEC, ang. *Header Error Check*), obliczaną według wielomianu $G(x) = x^{16} + x^{12} + x^5 + 1$ (CRC-16).

Jednostka PSDU jest poddawana tzw. wybielaniu (ang. *whitening*) w celu upodobnienia przesyłanej informacji do losowego białego szumu. Proces ten polega na wstawieniu nadmiarowego bitu na każde 32 bity informacyjne. Preambuła i nagłówek są przesyłane zawsze z prędkością 1 Mb/s, natomiast PSDU można przesyłać z prędkością 1 lub 2 Mb/s, zależnie od możliwości stacji i warunków panujących w otoczeniu sieci.

Pewnym mankamentem warstwy FHSS jest stałość sekwencji przeskoków – ustalona sekwencja jest realizowana bez uwzględnienia jakości transmisji czy mocy sygnału w poszczególnych kanałach częstotliwościowych, nie zapewniając także koordynacji między punktami dostępu znajdującymi się w niedużej odległości [89]. Prowadzić to może do licznych kolizji, jeśli sekwencje przeskoków stosowane w sąsiednich sieciach nie są wzajemnie ortogonalne. Może wówczas się zdarzyć, że kilka sieci wybierze jednocześnie tę samą częstotliwość nośną.

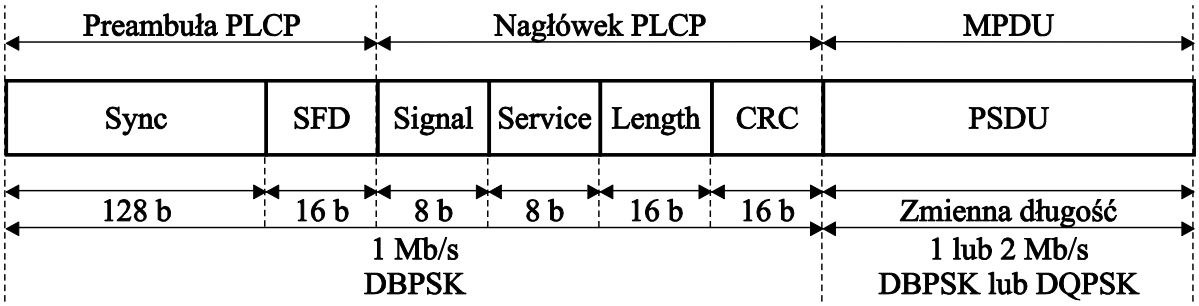
W chwili obecnej można już stwierdzić, iż warstwa fizyczna FHSS ma znaczenie wyłącznie historyczne. Nie była ona bowiem modyfikowana od chwili wprowadzenia standardu

(1999 rok), a użyte prędkości transmisji (1 lub 2 Mb/s) nie osiągnęły wartości zdefiniowanych w polu PSF (do 4,5 Mb/s). Ponadto, metodę tę wykorzystano w standardzie Bluetooth [96] – występuje tam ten sam podział na kanały częstotliwościowe i te same pseudolosowe sekwencje przeskoków. Mając na uwadze, iż oba standardy pracują w tym samym pasmie częstotliwości, można stwierdzić, iż rezygnacja z rozwoju tej metody w ramach standardu IEEE 802.11 podyktowana była także chęcią zmniejszenia ryzyka zakłócania jednej sieci przez drugą, tym bardziej, że w związku z mniejszą częstotliwością przeskoków sieć 802.11 i tak byłaby bardziej narażona na błędy transmisji [51].

3.2.2.3. Warstwa fizyczna DSSS

Warstwa fizyczna z kluczowaniem bezpośrednim (DSSS, ang. *Direct Sequence Spread Spectrum*) także została wprowadzona w pierwotnej wersji standardu [53]. Wykorzystuje ona to samo pasmo częstotliwości radiowych co warstwa FHSS, ale podzielono je na 14 kanałów o szerokości 22 MHz, przy czym odstęp pomiędzy sąsiednimi kanałami wynosi tylko 5 MHz (dostępność poszczególnych kanałów zależy od kraju, np. w Polsce są to kanały 1÷13). Jak nietrudno się domyślić, kanały częściowo nakładają się na siebie. Aby zmniejszyć poziom zakłóceń wynikających ze stosowania kilku sieci na wspólnym obszarze, należy wprowadzić odstęp o 5 kanałów (np. kanały 1, 6 i 11 lub 1, 7 i 13). W celu uzyskania sygnału o rozproszonym widmie stosuje się 11-bitowe sekwencje Barkera o prędkości transmisji 11 Mb/s. Uzyskany sygnał jest poddawany modulacji z prędkością 11 Mbd. W zależności od użytej metody modulacji (DBPSK lub DQPSK) uzyskuje się zatem prędkość transmisji, odpowiednio, 1 lub 2 Mb/s. Druga z prędkości nie jest przy tym obowiązkowa.

Na poziomie podwarstwy PLCP informacje są przesyłane w ramkach, zawierających preambułę (ang. PLCP preamble), nagłówek (ang. PLCP header) oraz jednostkę PSDU. Format ramki na poziomie podwarstwy PLCP w warstwie fizycznej DSSS pokazano na rys. 3.6.



Rys. 3.6. Format ramki podwarstwy PLCP w warstwie fizycznej DSSS
Fig. 3.6. PLCP frame format in DSSS physical layer

Preambuła podwarstwy PLCP zawiera:

- 128-bitowy ciąg synchronizacyjny (Sync), składający się z bitów o wartości 1,
- 16-bitowy znacznik początku ramki (SFD, ang. *Start Frame Delimiter*).

Z kolei nagłówek ramki zawiera:

- 8-bitowe pole Signal, określające prędkość transmisji jednostki PSDU, wyrażoną w jednostkach 100 kb/s;
- 8-bitowe pole Service, zarezerwowane dla przyszłych zastosowań,
- 16-bitowe pole Length, określające długość (a dokładniej czas trwania) jednostki PSDU w μs,
- 16-bitową sumę kontrolną nagłówka, obliczaną – podobnie jak w warstwie FHSS – według wielomianu CRC-16.

Wszystkie bity przesyłane na poziomie podwarstwy PLCP są poddawane działaniu skramblera, dzięki czemu informacja podczas transmisji ma postać zbliżoną do białego szumu. Preambuła i nagłówek są zawsze przesyłane z prędkością 1 Mb/s, natomiast PSDU można przesyłać z prędkością 1 lub 2 Mb/s, zależnie od możliwości stacji i warunków panujących w otoczeniu sieci.

Warstwa fizyczna DSSS jest jedyną spośród wprowadzonych w początkowej wersji standardu, która jest jeszcze wykorzystywana. Została ona bowiem rozwinięta w standardach IEEE 802.11b i 802.11g, które poprzez kompatybilność wsteczną zachowują zdolność komunikacji z urządzeniami typu DSSS.

3.2.2.4. Warstwa fizyczna Ir

Warstwa fizyczna wykorzystująca podczerwień rozproszoną również została wprowadzona w pierwotnej wersji standardu [53]. Wykorzystano tu fale elektromagnetyczne z zakresu bliskiej podczerwieni, tj. 850÷950 nm, podobnie jak w pilotach zdalnego sterowania czy w standardzie IrDA [102]. Przyjęto założenie, iż urządzenia używające tej warstwy fizycznej będą przeznaczone wyłącznie do pracy wewnątrz budynków, a zasięg transmisji może wynosić około 10÷20 m. Nie wprowadzono podziału na kanały częstotliwościowe, ponieważ sieci takie można bardzo łatwo odseparować fizycznie, np. za pomocą ścianek działowych. Praktycznie eliminuje to także ryzyko podsłuchu.

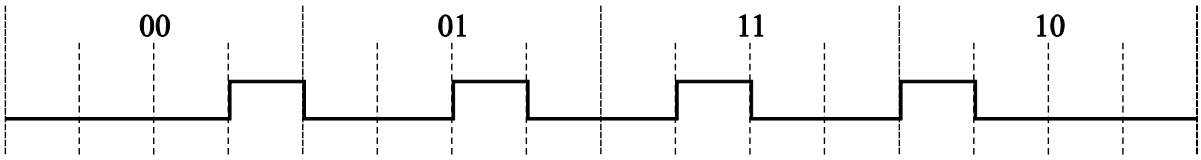
Do formowania sygnału optycznego użyto modulacji położenia impulsów (PPM, ang. *Pulse Position Modulation*). Dla obowiązkowej prędkości 1 Mb/s jest to modulacja 16-PPM, dla opcjonalnej zaś prędkości 2 Mb/s – 4-PPM. Każdy symbol modulacji składa się z odpowiedniej liczby szczelin (16 lub 4) o czasie trwania 250 ns. Dla zmniejszenia bitowej stopy błędu tworzenie symboli modulacji na podstawie wartości grup bitów wykorzystuje kod Graya – wartości grup bitów, odpowiadające symbolom różniącym się położeniem impulsu światła o jedną szczelinę, są sąsiednie logicznie. Dzięki temu przesunięcie impulsu światła w symbolu o jedną szczelinę – spowodowane np. interferencją międzysymbolową – powoduje błąd odbioru zawsze tylko jednego bitu. Przykładowe mapowanie wartości grup bitów na

symbole modulacji 4-PPM pokazano w tabeli 3.1, a odpowiadające im przebiegi sygnałów – na rys. 3.7.

Tabela 3.1

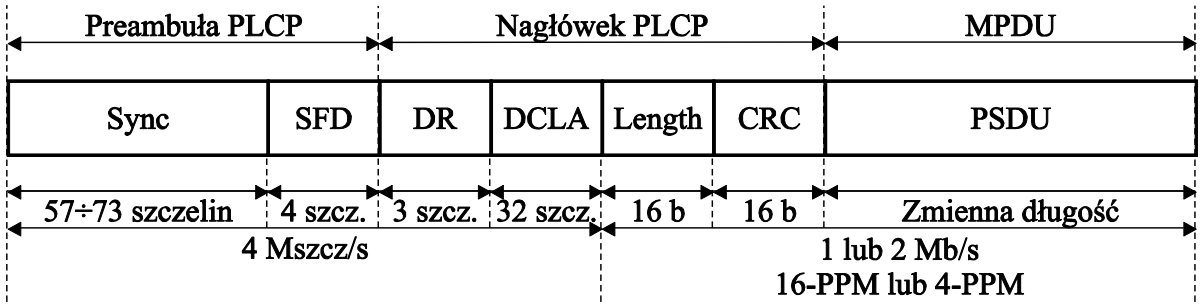
Mapowanie wartości bitów na symbole modulacji 4-PPM

Dane	Symbol
00	0001
01	0010
11	0100
10	1000



Rys. 3.7. Przebiegi czasowe w modulacji 4-PPM
Fig. 3.7. 4-PPM time diagram

Na poziomie podwarstwy PLCP informacje są przesyłane w ramkach, zawierających preambułę (ang. *PLCP preamble*), nagłówek (ang. *PLCP header*) oraz jednostkę PSDU. Format ramki na poziomie podwarstwy PLCP w warstwie fizycznej DSSS pokazano na rys. 3.8.



Rys. 3.8. Format ramki podwarstwy PLCP w warstwie fizycznej Ir
Fig. 3.8. PLCP frame format in Ir physical layer

Preambuła podwarstwy PLCP zawiera:

- ciąg synchronizacyjny (Sync) o długości 57÷73 szczelin (14,25÷18,25 μs), z których co druga zawiera impuls światła,
- znacznik początku ramki (SFD, ang. *Start Frame Delimiter*) o długości 4 szczelin (1 μs).

Z kolei nagłówek ramki zawiera:

- pole DR (ang. *Data Rate*) o długości 3 szczelin (0,75 μs), określające prędkość transmisji jednostki PSDU (1 lub 2 Mb/s),

- pole DCLA (ang. *DC Level Adjustment*) o długości 32 szczelin (8 μs), zawierające ciąg impulsów światła umożliwiający odbiornikowi stabilizację napięcia stałego po odebraniu początkowych pól ramki,
- 16-bitowe pole Length, określające długość jednostki PSDU w bajtach,
- 16-bitową sumę kontrolną nagłówka, obliczaną – podobnie jak w warstwie FHSS – według wielomianu CRC-16.

Preambuła i początkowe pola nagłówka (do DCLA włącznie), a więc te, których długość podano w szczelinach, są kodowane przy użyciu symboli niedozwolonych, czyli niewystępujących w ciągu danych. Pola te przesyła się z prędkością 4 Mszczelin na sekundę, co odpowiada prędkości 1 Mb/s. Pozostałe pola (Length, CRC oraz PSDU) są kodowane już zgodnie z zasadami przyjętej metody modulacji, a prędkość ich transmisji wynosi 1 lub 2 Mb/s, zależnie od możliwości stacji i warunków panujących w otoczeniu sieci.

Warstwa Ir standardu IEEE 802.11 nie znalazła zbyt wielu praktycznych zastosowań – nie są znane żadne szeroko dostępne urządzenia, które wykorzystywałyby to rozwiązanie. Może to być spowodowane większą elastycznością i wsparciem mobilności w sieciach radiowych. Wynikająca z tego wygoda używania sieci może przesłaniać zalety bezprzewodowych sieci optycznych, takie jak np. wyższa odporność na zakłócenia, wyższy poziom bezpieczeństwa sieci czy wreszcie brak jakichkolwiek ograniczeń prawnych dotyczących ich stosowania [39]. Z drugiej strony, w chwili opublikowania standardu 802.11 (1999 rok) na rynku były już dostępne stosunkowo liczne interfejsy standardu IrDA, które wprawdzie oferowały znacznie mniejszy zasięg (około 1 m), ale za to wyższą prędkość transmisji (4 Mb/s) [102]. Interfejsy takie charakteryzowały się ponadto niską ceną. Być może zatem ich popularność przyczyniła się do braku zainteresowania producentów warstwą Ir standardu 802.11.

3.2.2.5. Warstwa fizyczna HR-DSSS

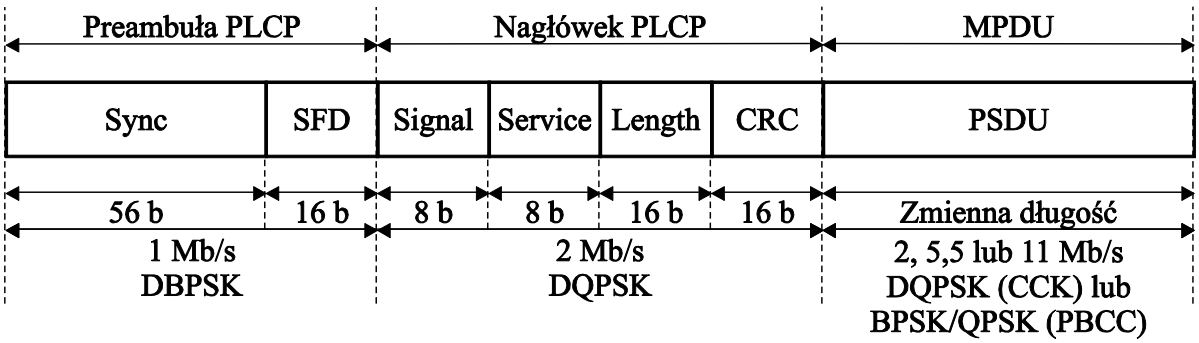
Warstwę fizyczną HR-DSSS (ang. *High Rate Direct Sequence Spread Spectrum*) wprowadzono w standardzie 802.11b. Przy zachowaniu wstecznej kompatybilności z warstwą DSSS zdefiniowano dwie nowe metody transmisji, pozwalające na uzyskanie prędkości 5,5 oraz 11 Mb/s – CCK i PBCC.

W modulacji CCK (ang. *Complementary Code Keying*) są wykorzystywane kody składające się nie z 11 (jak w DSSS), lecz z 8 części (ang. *chips*). Prędkość jego transmisji wynosi jednak nadal 11 Mb/s. Na jeden symbol modulacji przypada 4 (5,5 Mb/s) lub 8 (11 Mb/s) bitów. Dwóch z nich używa się do określenia zmiany fazy modulacji DQPSK, pozostałych natomiast – do wyboru jednej z 4 (5,5 Mb/s) lub 64 (11 Mb/s) sekwencji rozpraszających.

W modulacji PBCC (ang. *Packet Binary Convolutional Coding*) wykorzystuje się binarny koder splotowy (ang. *Binary Convolutional Coder*) o liczbie stanów równej 64 i sprawności ½ (tj. na każdy bit wejściowy przypadają dwa wyjściowe). Tak uzyskany ciąg bitów jest na-

stępnie mapowany na symbole modulacji. Przy prędkości 5,5 Mb/s dwa kolejne bity z wyjścia kodera tworzą dwa symbole modulacji BPSK, natomiast przy 11 Mb/s – pojedynczy symbol modulacji QPSK. W obu przypadkach sposób mapowania określa się za pomocą cyklicznie powtarzanego 256-bitowego, pseudolosowego ciągu pokrywającego (ang. *cover sequence*). Użycie kodera splotowego wymaga, aby po zakończeniu transmisji ramki wprowadzić dodatkowy bajt, pozostawiający koder w określonym stanie. Z tego też powodu szczegóły obliczania długości PSDU są różne dla CCK i PBCC.

Ze względu na konieczność zachowania zgodności wstecznej z warstwą DSSS, format ramki na poziomie podwarstwy PLCP pozostał niezmieniony, z wyjątkiem określenia znaczenia pola Service. Rozwiązanie to, zwane długą preambułą (ang. *Long Preamble*), może być użyte z dowolną prędkością transmisji. Dla podniesienia jednak wydajności protokołu przy wyższych prędkościach transmisji wprowadzono opcjonalną tzw. krótką preambułę (ang. *Short Preamble*), zmniejszającą dwukrotnie narzut protokołu na poziomie podwarstwy PLCP. Krótka preambuła może być użyta z prędkościami 2, 5,5 oraz 11 Mb/s. Znaczenie pól ramki PLCP nie zależy od użytej preambuły. W przypadku użycia długiej preambuły format ramki na poziomie podwarstwy PLCP jest identyczny jak w warstwie DSSS (rys. 3.6). Natomiast w przypadku użycia krótkiej preambuły format ramki jest taki, jak pokazano na rys. 3.9. Jak widać, w formacie tym skrócono preambułę (a dokładniej pole Sync) oraz zwiększono prędkość transmisji nagłówka do 2 Mb/s.



Rys. 3.9. Format ramki podwarstwy PLCP w warstwie fizycznej HR-DSSS
Fig. 3.9. PLCP frame format in HR-DSSS physical layer

Preambuła podwarstwy PLCP zawiera:

- 56- lub 128-bitowy ciąg synchronizacyjny (Sync), składający się z bitów o wartości 1 (preambuła długa) lub 0 (preambuła krótka), poddanych działaniu skramblera,
- 16-bitowy znacznik początku ramki (SFD, ang. *Start Frame Delimiter*).

Z kolei nagłówek ramki zawiera:

- 8-bitowe pole Signal, określające prędkość transmisji jednostki PSDU, wyrażoną w jednostkach 100 kb/s;

- 8-bitowe pole Service, zawierające m. in. 1-bitowe rozszerzenie pola Length w celu uniknięcia niejednoznaczności oznaczania długości (czasu trwania) ramki przy prędkościach transmisji powyżej 8 Mb/s, a także określające sposób kodowania (CCK lub PBCC);
- 16-bitowe pole Length, określające długość (a dokładniej czas trwania) jednostki PSDU w μs;
- 16-bitową sumę kontrolną nagłówka, obliczaną – podobnie jak w warstwie FHSS – według wielomianu CRC-16.

Wszystkie bity przesyłane na poziomie podwarstwy PLCP są poddawane działaniu skramblera, dzięki czemu informacja podczas transmisji ma postać zbliżoną do białego szumu. W przypadku długiej preambuły transmisja poszczególnych elementów ramki PLCP odbywa się tak, jak w warstwie DSSS. W przypadku natomiast preambuły krótkiej prędkość transmisji pola Sync i SFD wynosi 1 Mb/s, nagłówka PLCP – 2 Mb/s, zaś PSDU – 2, 5,5 lub 11 Mb/s. Nietrudno zauważyć, iż użycie krótkiej preambuły zmniejsza narzut podwarstwy PCLP o połowę.

3.2.2.6. Warstwa fizyczna OFDM

Warstwę fizyczną OFDM (ang. *Orthogonal Frequency Division Multiplexing*) wprowadzono w standardzie 802.11a. W przeciwieństwie do wcześniej opisanych warstw fizycznych, standard ten jest przeznaczony do pracy w pasmie 5 GHz. Dla Stanów Zjednoczonych i Japonii określono podział na kanały częstotliwościowe o szerokości 20, 10 i 5 MHz, w Europie są dostępne tylko kanały o szerokości 20 MHz. Znajdują się one w pasmie 5,15÷5,35 GHz (kanały 36÷64) oraz 5,47÷5,725 GHz (kanały 100÷140). Kanały te pokrywają się z zakresami częstotliwości przeznaczonymi dla europejskiej sieci HiPeRLAN [16], dlatego też można spotkać się z informacją, iż standard 802.11a nie jest dopuszczony do stosowania w Europie [39]. Najnowsze regulacje prawne w tym zakresie, również w Polsce [90], nie zastrzegają tych pasm dla sieci HiPeRLAN, wymaga się jedynie zgodności z normą ETSI EN 301 893 [17].

Modulacja OFDM pozwala na transmisję z prędkościami 6, 9, 12, 18, 24, 36, 48 lub 54 Mb/s, przy czym prędkości 6, 12 oraz 24 Mb/s są obowiązkowe. Dla kanałów o szerokości 10 lub 5 MHz zdefiniowano także prędkości odpowiednio 2 lub 4 razy mniejsze, nie mają one jednak zastosowania w Europie ze względu na brak odpowiednich kanałów.

Każdy kanał częstotliwościowy jest podzielony na 52 podkanały (o numerach -26÷26), z których 4 (-21, -7, 7, 21) są użyte jako nośne pilotujące (ang. *pilot carrier*), pozostałe natomiast służą do transmisji danych. Nośnej o numerze 0, odpowiadającej częstotliwości środkowej, nie używa się ze względu na przyjęte zasady przetwarzania sygnałów. Prędkość modulacji jest stała i wynosi V=250000 symboli na sekundę, natomiast zmiana prędkości transmisji

odbywa się poprzez zmianę metody modulacji i sprawności kodera splotowego, a w efekcie – zmianę liczby bitów zakodowanych w pojedynczym symbolu modulacji.

W standardzie 802.11a określono 8 schematów modulacji i kodowania (MCS, ang. *Modulation and Coding Scheme*). Znaczenie i wartości parametrów, opisujących właściwości tych schematów, zawarto w tabeli 3.2.

Wynikową prędkość transmisji można obliczyć według wzoru:

$$R_{wl} = N_{DBPS} \cdot V = \frac{R \cdot N_{SC} \cdot N_{BPSC}}{T_{FIT} + T_{GI}}. \tag{3.1}$$

Tabela 3.2

Zależności między parametrami modulacji OFDM

Parametr	Opis	Wartość
R	sprawność kodera konwolucyjnego	½, ⅔, ¾
N_{BPSC}	liczba zakodowanych bitów na podnośną	zależne od modulacji
N_{SC}	liczba podnośnych przeznaczonych do transmisji	48
N_{CBPS}	liczba zakodowanych bitów na symbol	$N_{SC} \cdot N_{BPSC}$
N_{DBPS}	liczba bitów danych na symbol	$R \cdot N_{CBPS}$
T_{FIT}	czas integracji FFT (ang. *FFT Integration Time*)	3,2 μs
T_{GI}	czas trwania okresu ochronnego (ang. *Guard Interval*)	0,8 μs
V	prędkość modulacji	$1/(T_{FIT} + T_{GI})$

Ważniejsze parametry modulacji i kodowania dla wszystkich określonych w standardzie prędkości transmisji zestawiono w tabeli 3.3 [52].

Tabela 3.3

Schematy modulacji i kodowania w standardzie 802.11a [52]

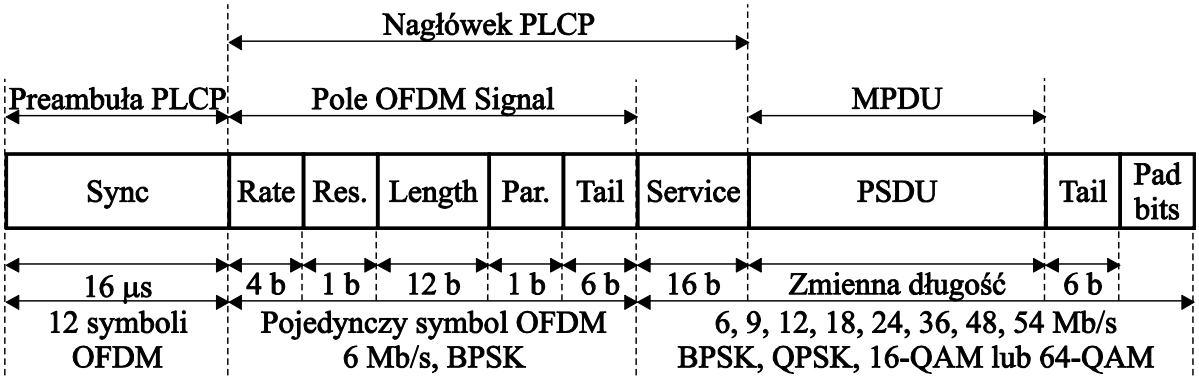
Modulacja	N_{BPSC}	N_{CBPS}	R	N_{DBPS}	R_{wl} [Mb/s]
BPSK	1	48	½	24	6
			¾	36	9
QPSK	2	96	½	48	12
			¾	72	18
16-QAM	4	192	½	96	24
			¾	144	36
64-QAM	6	288	⅔	192	48
			¾	216	54

Na poziomie podwarstwy PLCP informacje są przesyłane w ramkach, zawierających preambułę (ang. *PLCP preamble*), nagłówek (ang. *PLCP header*) oraz jednostkę PSDU. Format ramki na poziomie podwarstwy PLCP w warstwie fizycznej OFDM pokazano na rys. 3.10.

Preambuła ramki przesyłanej na poziomie podwarstwy PLCP, podobnie jak w innych warstwach, jest przeznaczona dla potrzeb synchronizacji. Składa się ona z 10 powtórzeń

krótkiego ciągu treningowego (ang. *training sequence*) oraz 2 powtórzeń ciągu długiego. Czas trwania preambuły wynosi w sumie 16 μs.

Nagłówek ramki na poziomie podwarstwy PLCP zawiera dwa pola: Signal (24 bity) oraz Service (16 bitów). Pole Signal określa prędkość transmisji jednostki PSDU (4-bitowe pole Rate) oraz jej długość (12-bitowe pole Length). Pola te są chronione pojedynczym bitem parzystości (Parity). Ostatnia część pola Signal to końcówka (Tail), potrzebna dla pozostawienia kodera splotowego w określonym stanie (6 bitów). Kolejne pole, Service, jest używane w celu inicjalizacji skramblera.



Rys. 3.10. Format ramki podwarstwy PLCP w warstwie fizycznej OFDM
Fig. 3.10. PLCP frame format in OFDM physical layer

Pole Signal zajmuje pojedynczy symbol modulacji OFDM, natomiast pole Service, wraz z jednostką PSDU oraz końcówką ramki (Tail) i bitami uzupełniającymi (ang. *Pad bits*), podlega kodowaniu i modulacji stosownie do wybranej prędkości transmisji (tabela 3.3).

3.2.2.7. Warstwa fizyczna ERP

Warstwa fizyczna ERP (ang. *Enhanced Rate Physical*) została wprowadzona w standardzie IEEE 802.11g. Łączy ona zalety warstw 802.11b i 802.11a, tj. umożliwia transmisję z prędkościami do 54 Mb/s w pasmie ISM 2,4÷2,4835 GHz przy zachowaniu kompatybilności wstecznej z sieciami 802.11b. Standard opisuje następujące sposoby transmisji:

- ERP-OFDM (obowiązkowy), przejęty z warstwy 802.11a z uwzględnieniem zmiany pasma częstotliwości oraz niektórych parametrów łącza,
- ERP-PBCC (opcjonalny), w którym prędkości transmisji wariantu PBCC standardu 802.11b wzbogacono o prędkości 22 oraz 33 Mb/s,
- DSSS-OFDM (opcjonalny), w którym preambuła i nagłówek warstwy PLCP są przesyłane zgodnie z wymogami standardu 802.11b, natomiast PSDU – w sposób zbliżony do 802.11a.

Wydaje się, iż metodę DSSS-OFDM stworzono dla sytuacji, gdy w sieci pracują urządzenia zarówno wykorzystujące technikę DSSS (warstwa fizyczna 802.11 DSSS lub 802.11b HR-DSSS), jak i nowsze, wykorzystujące technikę OFDM. Standard 802.11g wprowadzono

dopiero w roku 2003, toteż liczba używanych urządzeń 802.11(b) była już dość duża. Urządzenia takie, nawet jeśli nie potrafią „zrozumieć" transmisji prowadzonej przy użyciu OFDM, powinny umożliwić taką transmisję między pozostałymi urządzeniami w sieci. Jednocześnie te nowsze urządzenia powinny także móc komunikować się z urządzeniami starszymi, niewyposażonymi w technikę OFDM. Warto jednocześnie zauważyć, iż konieczność taka występuje jedynie w standardzie 802.11g, gdyż wcześniej wprowadzony standard 802.11a, mimo iż także wykorzystujący technikę OFDM, pracuje w pasmie nieużywanym wcześniej przez inne urządzenia standardu 802.11.

W standardzie 802.11g występuje aż 5 formatów ramek na poziomie podwarstwy PLCP. Jest to w dużej mierze wynikiem konieczności zapewnienia wstecznej kompatybilności ze standardem 802.11(b), a jednocześnie wprowadzenia nowych mechanizmów wspierających modulację OFDM. Oprócz całkowicie nowego formatu ERP-OFDM występują formaty „zwykły" (802.11b) oraz DSSS-OFDM, oba w wersji z długą lub krótką preambułą.

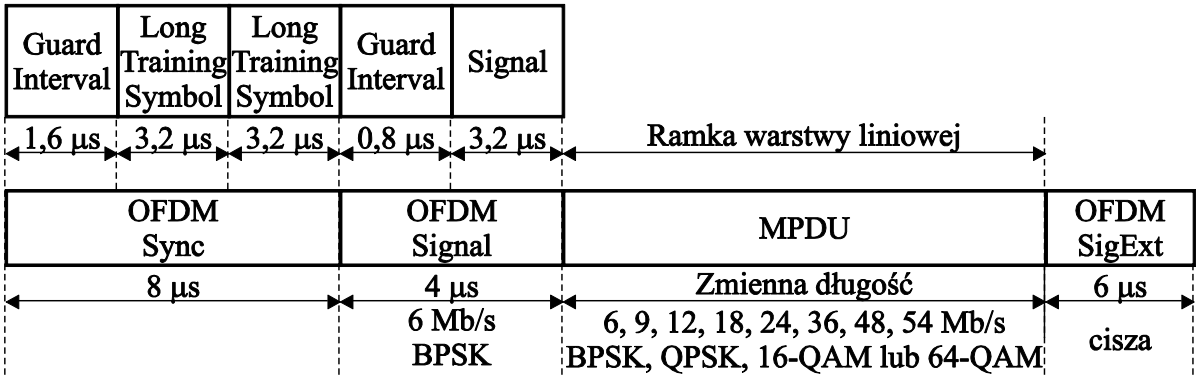
Format ramki ERP-OFDM jest identyczny jak w standardzie 802.11a (rys. 3.10). Jedynie w zakończeniu ramki występuje dodatkowy okres ciszy o czasie trwania 6 μs, którego zadaniem jest wyrównanie odstępów międzyramkowych (czas trwania SIFS w warstwach DSSS i HR-DSSS jest o 6 μs krótszy niż w OFDM). Format ten umożliwia osiągnięcie prędkości transmisji z zakresu 6÷54 Mb/s oraz maksymalną wydajność protokołu, ale może być stosowany tylko wówczas, gdy wszystkie stacje sieci obsługują ten rodzaj transmisji. W przypadku pojawienia się w sieci stacji standardu 802.11(b) konieczna jest zmiana formatu, np. na DSSS-OFDM z długą lub krótką preambułą.

Format ramki DSSS-OFDM jest hybrydą, która umożliwia stosowanie techniki OFDM nawet wówczas, gdy w sieci występują stacje standardu 802.11(b). Preambuła i nagłówek podwarstwy PLCP są wówczas zgodne z formatem preambuły krótkiej (rys. 3.9) lub długiej (rys. 3.6) z wykorzystaniem techniki DSSS. Umożliwia to „zrozumienie" zawartej tam informacji przez stacje standardu zarówno 802.11(b), jak i 802.11g. Jednostkę PSDU natomiast przesyła się już z wykorzystaniem metody OFDM. PSDU zawiera wówczas:

- pole OFDM Sync, zawierające dwa długie symbole treningowe (ang. *Long Training Symbol*), poprzedzone odstępem ochronnym (ang. *Guard Interval*),
- pole OFDM Signal, zawierające pole Signal, określające prędkość transmisji jednostki PSDU, poprzedzone odstępem ochronnym,
- pole MPDU, w którym jest umieszczana ramka warstwy liniowej,
- pole OFDM Signal Extension, zawierające 6 μs ciszy.

Format tak utworzonej jednostki PSDU pokazano na rys. 3.11. Jednostka ta zastępuje pole PSDU w ramce z preambułą długą (rys. 3.6) lub krótką (rys. 3.9). Aby stacje nieobsługujące OFDM nie próbowały dekodować PSDU, w polu Signal nagłówka DSSS podaje się pręd-

kość transmisji 3 Mb/s, nieobsługiwaną przez żadną stację standardu 802.11(b). Prawidłowa prędkość transmisji ramki warstwy liniowej jest natomiast podana w polu Signal części OFDM. Także pole Service nagłówka DSSS zostało zmodyfikowane, m. in. przez dodanie kolejnych bitów rozszerzających pole Length w celu precyzyjnego określenia długości (czasu trwania) jednostki PSDU przy stosowaniu różnych metod modulacji. Dzięki temu stacje mogą powstrzymać się od nadawania przez odpowiedni czas.



Rys. 3.11. Format pola PSDU w wariancie DSSS-OFDM standardu 802.11g
Fig. 3.11. PSDU field format in DSSS-OFDM variant of 802.11g standard

W metodzie ERP-PBCC format ramki na poziomie podwarstwy PLCP jest identyczny jak w standardzie 802.11b. Wyższe prędkości transmisji uzyskano przez zastosowanie modulacji 8-PSK, a w przypadku 33 Mb/s – dodatkowo przez podniesienie prędkości modulacji z 11 do 16,5 Mbd. W tym wariancie, niezależnie od prędkości transmisji, można stosować zarówno preambułę długą (rys. 3.6), jak i krótką (rys. 3.9).

3.2.3. Warstwa liniowa – formaty ramek

Informacje na poziomie warstwy liniowej sieci standardu IEEE 802.11 – podobnie jak w innych sieciach lokalnych – są wymieniane za pomocą ramek. Od razu warto jednak zauważyć, iż warstwa liniowa 802.11 realizuje o wiele więcej zadań niż analogiczna sieć przewodowa (np. Ethernet). Konieczne jest bowiem wprowadzenie odpowiednich typów ramek, niezbędnych dla realizacji mechanizmów dostępu do łącza. Ponadto, proces uwierzytelniania stacji i użytkowników oraz ich łączenia z siecią bezprzewodową także jest realizowany na poziomie warstwy liniowej poprzez wymianę odpowiednich typów ramek. Analogiczne operacje w przypadku sieci przewodowej są realizowane albo na poziomie warstwy fizycznej (podłączanie do sieci), albo w ogóle poza siecią (uwierzytelnianie). Różnice te wynikają m. in. z charakterystyki medium przewodowego i bezprzewodowego. O ile bowiem ograniczenie dostępności sieci przewodowej tylko do wybranych pomieszczeń nie stanowi problemu, o tyle w przypadku sieci bezprzewodowej – a szczególnie radiowej – jest często fizycznie niemożliwe. Nietrudno zatem zauważyć, że uwierzytelnianie użytkowników sieci przewodowej można wykonać przez sprawdzenie ich tożsamości, np. przy pobieraniu klucza do

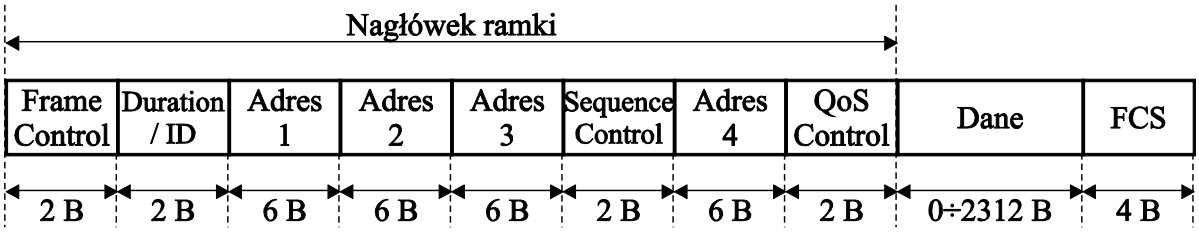
pomieszczenia, w którym zainstalowano gniazdka sieciowe. W sieci bezprzewodowej, ze względu na bardziej swobodną propagację fal elektromagnetycznych, jest to utrudnione, gdyż bardzo często moc sygnału sieci bezprzewodowej jest wystarczająca także w miejscach swobodnie dostępnych. Dlatego też warstwa liniowa standardu 802.11 zawiera mechanizmy podwyższające poziom bezpieczeństwa sieci, dokładny ich opis wykracza jednak poza ramy niniejszego opracowania.

3.2.3.1. Ogólny format ramki

Każda ramka w standardzie IEEE 802.11 zawiera następujące składniki:

- nagłówek (ang. *MAC header*), m. in. z polem sterującym i adresowym,
- rdzeń ramki (ang. *frame body*) o zmiennej długości, przenoszący informacje zależne od typu ramki,
- sumę kontrolną (FCS, ang. *Frame Check Sequence*).

Ogólny format ramki w sieci standardu 802.11 przedstawiono na rys. 3.12.



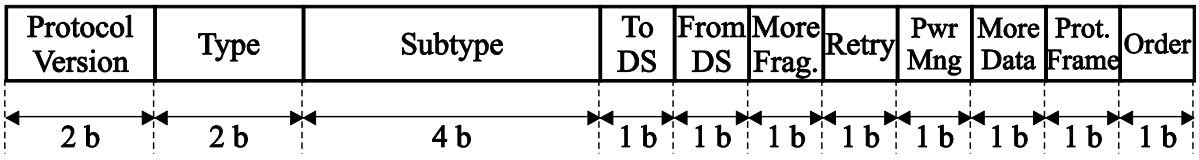
Rys. 3.12. Ogólny format ramki w standardzie 802.11
Fig. 3.12. General frame format in 802.11 standard

Pole sterujące (ang. *Frame Control*) zawiera ogóle informacje o ramce. Dotyczą one typu ramki oraz opcji jej przesyłu. Informacje te są zakodowane w kilkunastu polach, których znaczenie opisano w tabeli 3.4. Format pola sterującego przedstawiono na rys. 3.13.

Tabela 3.4

Składniki pola sterującego ramki [52]

Nazwa ang.	Znaczenie	Opis
Protocol Version	wersja protokołu	obowiązuje wersja 0
Type	typ ramki	sterująca, zarządzająca lub danych
Subtype	podtyp ramki	znaczenie ramki w kontekście jej typu
To DS	do DS	ramka kierowana do systemu dystrybucyjnego
From DS	z DS	ramka pochodzi z systemu dystrybucyjnego
More Fragments	więcej fragmentów	ramka jest podzielona na fragmenty
Retry	retransmisja	ramka jest retransmitowana
Power management	zarządzanie energią	stacja jest w trybie energooszczędnym
More data	więcej danych	punkt dostępu buforuje dane dla adresata
Protected Frame	szyfrowanie	ramkę zaszyfrowano
Order	porządek	przesył ramek dokładnie w ustalonej kolejności



Rys. 3.13. Format pola sterującego ramki w standardzie 802.11
Fig. 3.13. Frame control field format in 802.11 standard

Pole **Duration/ID** może mieć jedno z dwóch znaczeń, zależnie od typu ramki. Na ogół określa ono czas trwania wymiany ramek (ang. *duration*) wyrażony w μs, na podstawie którego stacje ustawiają i zerują wektor zajętości sieci NAV. W ramkach protokołu PCF wartość ta wynosi 32768 μs. Czas ten powinien być obliczony tak, aby obejmować całą wymianę ramek wraz z potwierdzeniami i odstępami międzyramkowymi. Natomiast w ramkach PS-Poll pole określa identyfikator skojarzenia (AID, ang. *Association Identifier*).

Pola adresowe zawierają po 48 bitów, a ich format jest zgodny z formatem adresów używanych w sieciach typu Ethernet. Nie każda ramka zawiera wszystkie pola adresowe. Przykładowo, ramki sterujące zawierają tylko 1 lub 2 pola, natomiast w ramkach danych liczba pól wynosi 3 lub 4 i zależy od konfiguracji sieci. Ze względu na to, że ramka taka może być przesyłana między segmentem przewodowym i bezprzewodowym, a w danej lokalizacji może pracować wiele sieci bezprzewodowych, wyróżnia się kilka typów adresów:

- źródłowy (SA, ang. *Source Address*) – adres stacji, która wysłała ramkę (w segmencie przewodowym lub bezprzewodowym),
- docelowy (DA, ang. *Destination Address*) – adres stacji, do której wysłano ramkę (w segmencie przewodowym lub bezprzewodowym),
- nadajnika (TA, ang. *Transmitter Address*) – adres stacji, nadającej ramkę w segmencie bezprzewodowym,
- odbiornika (RA, ang. *Receiver Address*) – adres stacji, odbierającej ramkę w segmencie bezprzewodowym,
- identyfikator sieci (BSSID, ang. *Basic Service Set Identifier*), używany m. in. w celu rozróżnienia sieci pracujących w jednym kanale; w sieciach typu BSS jest to adres punktu dostępowego.

Wykorzystanie poszczególnych pól adresowych wskazuje się za pomocą bitów To DS i From DS pola sterującego ramki. Zależność tę pokazano w tabeli 3.5.

Pole **Sequence Control** określa kolejny numer ramki (12 bitów) oraz fragmentu ramki (4 bity). Pole to stosuje się we wszystkich ramkach z wyjątkiem ramek sterujących.

Pole **QoS Control** określa strumień lub klasę ruchu (ang. *traffic stream*, *traffic class*), do którego należy dana ramka, a także dodatkowe informacje zależne od typu ramki. Pole to występuje we wszystkich ramkach danych, przeznaczonych do obsługi ruchu z gwarancją jakości usług. Zawiera ono między innymi identyfikator ruchu (TID, ang. *traffic identifier*),

określa strategię potwierdzeń (ang. *Ack policy*) i umożliwia sterowanie prawem transmisji (TXOP, ang. *Transmission Opportunity*). W ramkach ze wskazaniem QoS CF-Poll wysłanych przez koordynatora pole zawiera limit TXOP, natomiast w ramkach typu QoS Data, QoS Null i QoS Data+CF-Ack – określa stan bufora punktu dostępowego. Ramki QoS Data (z ewentualnym wskazaniem CF-Ack), wysłane przez pozostałe stacje, określają zapotrzebowanie na TXOP (ang. *TXOP Duration Requested*) lub rozmiar kolejki (ang. *Queue Size*). Limit i zapotrzebowanie TXOP podaje się jako wielokrotność 32 μs; mogą one przyjmować wartości z zakresu 32÷8160 μs, wliczając w to narzut warstwy fizycznej i odstępy międzyramkowe. Rozmiar kolejki określa ilość informacji, czekającej na wysłanie, dla danej kategorii lub strumienia ruchu. Wielkość ta jest podawana jako wielokrotność 256 B.

Tabela 3.5

Wykorzystanie pól adresowych ramki [52]

To DS	From DS	Adres 1	Adres 2	Adres 3	Adres 4
0	0	RA = DA	TA = SA	BSSID	–
0	1	RA = DA	TA = BSSID	SA	–
1	0	RA = BSSID	TA = SA	DA	–
1	1	RA	TA	DA	SA

Pole danych (ang. *frame body field*) ma zmienną długość, a jego zawartość jest ściśle uzależniona od typu i podtypu ramki. Długość pola waha się od 0 do 2312 B, przy czym największy rozmiar MSDU wynosi 2304 B, pozostałe 8 B wykorzystuje się do realizacji zabezpieczeń zgodnie z wymogami protokołu WEP.

Suma kontrolna (FCS, ang. *Frame Check Sequence*) występuje we wszystkich typach ramek i służy do sprawdzenia poprawności ich odbioru. W celu obliczenia sumy stosuje się następujący wielomian:

$$G(x) = x^{32}+x^{26}+x^{23}+x^{22}+x^{16}+x^{12}+x^{11}+x^{10}+x^{8}+x^{7}+x^{5}+x^{4}+x^{2}+x+1. \tag{3.2}$$

3.2.3.2. Ramki sterujące

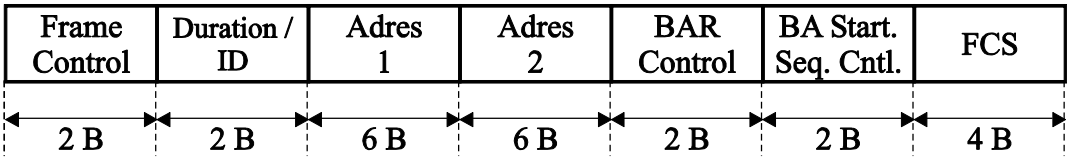
Ramki sterujące służą do przesyłania wewnętrznych informacji sterujących sieci 802.11. Z tego powodu nie są one przesyłane do systemu dystrybucyjnego, nie są one także z niego odbierane. Ponadto, ramki sterujące nie podlegają fragmentacji ani retransmisji, nie są także szyfrowane.

Do typu ramek sterujących należą następujące ramki:

- **RTS** (ang. *Request To Send*) – zgłoszenie zamiaru transmisji danych,
- **CTS** (ang. *Clear To Send*) – odpowiedź na ramkę RTS,
- **Ack** (ang. *Acknowledge*) – potwierdzenie prawidłowego odbioru ramki danych w protokole DCF,

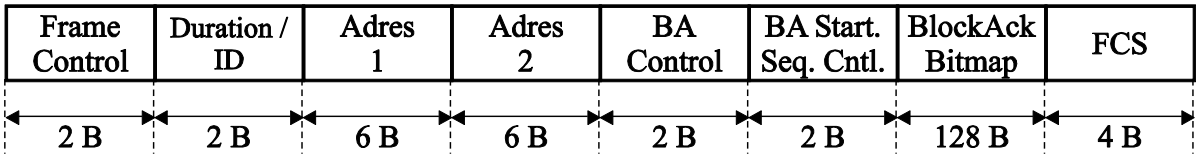
- **PS-Poll** (ang. *Power Save Poll*) – żądanie przesłania ramek zgromadzonych w punkcie dostępowym dla stacji w trybie energooszczędnym,
- **CF-End** (ang. *Contention Free End*) – zakończenie okresu transmisji bezkolizyjnej zgodnie z zasadami protokołu PCF,
- **CF-End+CF-Ack** – jak w ramce CF-End, ale z potwierdzeniem prawidłowego odbioru ramki danych bezpośrednio poprzedzającej tę ramkę w protokole PCF,
- **BlockAckReq** (BAR, ang. *Block Acknowledge Request*) – żądanie przesłania potwierdzenia blokowego,
- **BlockAck** (BA, ang. *Block Acknowledge*) – potwierdzenie blokowe.

Większość ramek sterujących ma niewielką długość, zazwyczaj 14 B (CTS, Ack) lub 20 B (RTS, PS-Poll, CF-End, CF-End+CF-Ack). Długość ramki BlockAckReq wynosi 24 B, natomiast BlockAck – 152 B, a to ze względu na możliwość indywidualnego potwierdzenia każdego przesłanego fragmentu ramki. Formaty tych ramek wykraczają poza ramy typowej ramki sterującej; pokazano je na rys. 3.14 i 3.15.



Rys. 3.14. Format ramki sterującej BlockAckReq
Fig. 3.14. A format of BlockAckReq control frame

Pola BA Control i BAR Control określają identyfikator ruchu, dla którego zestawiono wymianę z użyciem potwierdzenia blokowego. Pole BA Starting Sequence Control zawiera numer kolejny ramki, która występuje jako pierwsza w danym bloku ramek. Z kolei pole BlockAck Bitmap określa, które ramki i fragmenty bloku odebrano poprawnie.



Rys. 3.15. Format ramki sterującej BlockAck
Fig. 3.15. A format of BlockAck control frame

3.2.3.3. Ramki danych

Ramki danych służą do przesyłania informacji użytkownika. W ramkach tego typu rdzeń ramki przenosi informację, pochodzącą z wyższych warstw sieci. Zawartość tych ramek może pochodzić z systemu dystrybucyjnego, może także być do niego kierowana. Ramki takie mogą podlegać fragmentacji, retransmisji oraz szyfrowaniu. Ze względu na szczególne właściwości standardu 802.11 ramki danych mogą także pełnić dodatkowe funkcje, podnoszące wydajność sieci podczas używania protokołu PCF.

Do typu ramek danych należą następujące ramki:

- **Data** – „zwykła” ramka danych, używana do przesyłu informacji w protokole DCF,
- **Data+CF-Ack** – ramka danych z potwierdzeniem prawidłowego odbioru ramki bezpośrednio poprzedzającej tę ramkę w protokole PCF,
- **Data+CF-Poll** – ramka danych z wywołaniem następnej stacji w protokole PCF,
- **Data+CF-Ack+CF-Poll** – ramka danych z potwierdzeniem i wywołaniem, stosowana w protokole PCF,
- **Null** – „ramka danych bez danych”, służąca np. do poinformowania o braku danych do przesłania w protokole PCF, zaś w DCF używana do informowania o zmianach trybu oszczędzania energii,
- **CF-Ack** – potwierdzenie prawidłowego odbioru ramki bezpośrednio poprzedzającej tę ramkę w protokole PCF,
- **CF-Poll** – wywołanie następnej stacji w protokole PCF,
- **CF-Ack+CF-Poll** – potwierdzenie i wywołanie, stosowane w protokole PCF,
- **QoS Data** – „zwykła” ramka danych, używana do przesyłu informacji w protokole HCF (EDCA i HCCA),
- **QoS Data+CF-Ack** – ramka danych z potwierdzeniem prawidłowego odbioru ramki bezpośrednio poprzedzającej tę ramkę w protokole HCF (HCCA),
- **QoS Data+CF-Poll** – ramka danych z wywołaniem następnej stacji w protokole HCF (HCCA),
- **QoS Data+CF-Ack+CF-Poll** – ramka danych z potwierdzeniem i wywołaniem, stosowana w protokole HCF (HCCA),
- **QoS Null** – „ramka danych bez danych”, służąca do poinformowania o braku danych do przesłania w protokole HCF (HCCA),
- **QoS CF-Poll** – wywołanie następnej stacji w protokole HCF (HCCA),
- **QoS CF-Ack+CF-Poll** – potwierdzenie i wywołanie, stosowane w protokole HCF (HCCA).

W ramkach danych przyjęto interesujące kodowanie pola podtypu ramki. Jako wartość wyjściową przyjęto 0, oznaczające podstawowy podtyp ramki danych. Każdy spośród czterech bitów pola podtypu ramki określa sposób modyfikacji właściwości ramki. I tak, bit 0 oznacza wskazanie (ang. *indication*) CF-Ack, bit 1 – CF-End, bit 2 – brak danych, natomiast bit 3 – ramki QoS. Obecność wskazania CF-Ack lub CF-End nie zmienia istotnie formatu ramki, modyfikuje tylko jej znaczenie. Wskazanie braku danych informuje o braku pola danych w ramce. Wskazanie QoS informuje o obecności pola QoS Control w nagłówku ramki.

Ramki ze wskazaniem QoS są używane do wymiany informacji zawsze, o ile tylko nadawca i odbiorca obsługują mechanizmy gwarancji jakości usług. Jeśli co najmniej jedna

z komunikujących się stacji nie obsługuje tych mechanizmów, stosuje się zawsze ramki „tradycyjne". Ramki ze wskazaniem QoS można także wykorzystać do transmisji wieloadresowych lub rozgłoszeniowych, jeśli tylko nadawca ma pewność, że wszyscy adresaci obsługują mechanizmy gwarancji jakości usług.

3.2.3.4. Ramki zarządzające

Ramki zarządzające pełnią wiele ważnych funkcji w sieciach standardu 802.11. Wynika to z przyjętych założeń co do zadań realizowanych przez warstwę liniową – odpowiada ona nie tylko za bezbłędny przesył informacji pochodzącej z wyższych warstw, ale także za bezpieczeństwo sieci. Z tego też powodu występują tutaj ramki używane m. in. podczas odnajdywania sieci, uwierzytelniania klientów i podłączania ich do sieci. W ramkach zarządzających niezbędne informacje są przenoszone w rdzeniu ramki w postaci tzw. elementów informacyjnych (ang. *information elements*), których rodzaj i liczba mogą zależeć od podtypu ramki i konkretnej sytuacji. Ramki zarządzające mogą być dzielone na fragmenty.

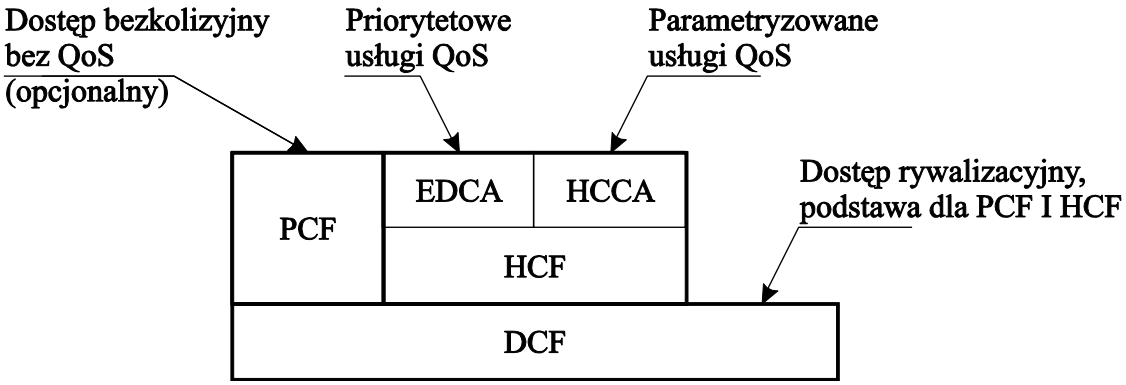
Do typu ramek zarządzających należą następujące ramki:

- „radiolatarnia" (ang. *beacon*) – informująca o istnieniu sieci i jej parametrach,
- sondowanie (ang. *Probe Request*) – używana podczas aktywnego przeglądania dostępnych sieci,
- odpowiedź na sondowanie (ang. *Probe Response*) – odpowiedź na ramkę Probe Request,
- ATIM (ang. *Announcement Traffic Indication Message*) – używana w sieciach IBSS w celu poinformowania odbiorcy o tym, iż nadawca ma dlań ramki zbuforowane w czasie, gdy był on w trybie energooszczędnym,
- uwierzytelnienie (ang. *Authentication*) – używana podczas uwierzytelniania stacji przed podłączeniem jej do sieci,
- anulowanie uwierzytelnienia (ang. *Deauthentication*) – używana do anulowania uwierzytelnienia danej stacji wraz z podaniem jego przyczyny,
- skojarzenie (ang. *Association Request*) – używana w celu przyłączenia do sieci stacji, która pomyślnie przeszła proces uwierzytelniania,
- ponowne skojarzenie (ang. *Reassociation Request*) – używana w celu ponownego przyłączenia do sieci,
- odpowiedź na skojarzenie (ang. *Association Response*) i ponowne skojarzenie (ang. *Reassociation Response*) – używana do poinformowania stacji próbującej połączyć się z siecią o wyniku tej próby,
- anulowanie skojarzenia (ang. *Deassociation*) – używana w celu odłączenia stacji od sieci, wraz z podaniem jej przyczyny odłączenia,
- akcja (ang. *Action*) – używana w celu przeprowadzenia określonych dodatkowych działań związanych z zarządzaniem siecią, a odnoszących się m. in. do:

- zarządzania widmem częstotliwości (ang. *spectrum management*),
- gwarancji jakości usług,
- rozpoczęcia i zakończenia wymiany z użyciem potwierdzenia blokowego.

3.2.4. Warstwa liniowa – dostęp do łącza

Mechanizm dostępu do łącza, zdefiniowany w standardzie IEEE 802.11, jest niezależny od zastosowanej w danej sieci warstwy fizycznej. Protokół ten początkowo określał dwa sposoby dostępu do łącza, z których jeden był oparty na rywalizacji, drugi natomiast wykorzystywał mechanizmy rezerwacyjne. W chwili obecnej są wprowadzone kolejne dwa sposoby dostępu do łącza, których celem jest zapewnienie wsparcia dla jakości usług (QoS, ang. *Quality of Service*). Ogólną architekturę protokołu dostępu do łącza pokazano na rys. 3.16 [52].



Rys. 3.16. Architektura protokołu dostępu do łącza standardu 802.11
Fig. 3.16. An architecture of MAC protocol in 802.11 standard

Protokół dostępu do łącza zawiera następujące elementy:

- DCF (ang. *Distributed Coordination Function*),
- PCF (ang. *Point Coordination Function*),
- HCF (ang. *Hybrid Coordination Function*), w skład którego wchodzą:
  - EDCA (ang. *Enhanced Distributed Chanel Access*),
  - HCCA (ang. *HCF Controlled Chanel Access*).

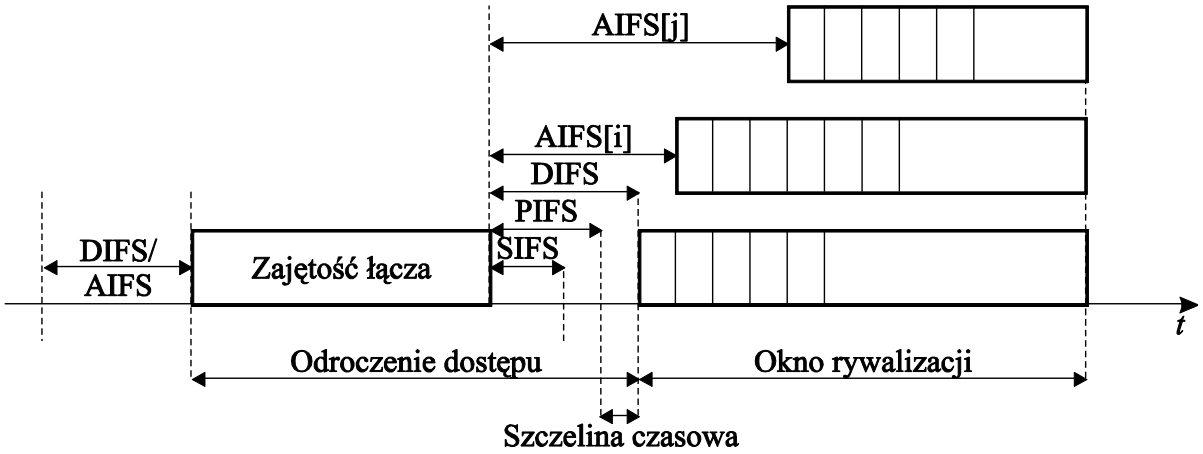
Protokół DCF jest obowiązkowy i musi być zaimplementowany w każdej stacji zgodnej ze standardem IEEE 802.11. Protokół HCF, zawierający elementy EDCA i HCCA, jest obowiązkowy w stacjach z obługą mechanizmów gwarancji jakości usług (ang. *QoS stations*) i nie występuje w stacjach, nieposiadających tej funkcjonalności. Protokół PCF jest natomiast opcjonalny we wszystkich stacjach.

3.2.4.1. Zależności czasowe

Przy realizacji wymienionych elementów protokołu są używane odpowiednie odstępy międzyramkowe (IFS, ang. *Inter-Frame Space*). W obecnej wersji standardu określono pięć różnych odstępów:

- SIFS (ang. *Short Inter-Frame Space*), rozdzielający ramki, tworzące jeden cykl wymiany (np. RTS-CTS-Data-Ack),
- PIFS (ang. *PCF Inter-Frame Space*), stosowany przez protokół PCF w celu uzyskania dostępu do łącza przez punkt dostępowy,
- DIFS (ang. *DCF Inter-Frame Space*), stosowany przez protokół DCF w celu uzyskania dostępu do łącza przez dowolną stację lub punkt dostępu,
- AIFS (ang. *Arbitration Inter-Frame Space*), stosowany przez protokół EDCA w celu określenia pierwszeństwa stacji przy dostępie do łącza,
- EIFS (ang. *Extended Inter-Frame Space*), stosowany dla rozwiązania sytuacji powstałej m. in. w wyniku odebrania błędnej ramki.

Niektóre zależności pomiędzy wymienionymi odstępami pokazano na rys. 3.17. Dokładne czasy ich trwania są zależne od wariantu warstwy fizycznej.

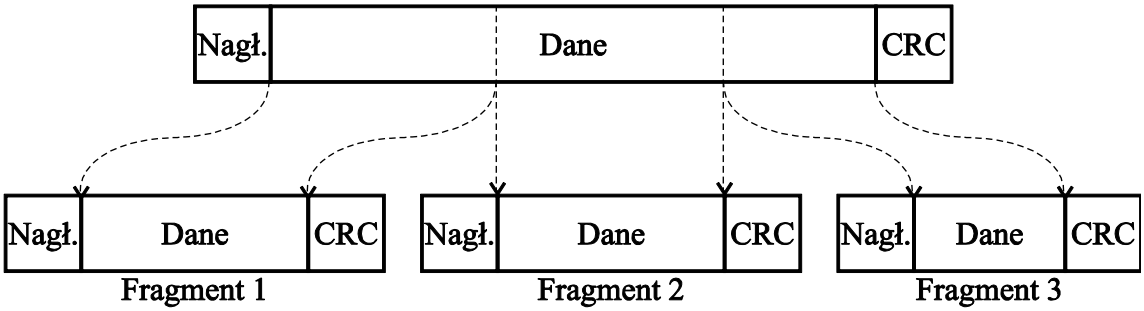


Rys. 3.17. Niektóre zależności między odstępami międzyramkowymi [52]
Fig. 3.17. Some inter-frame space relationships

3.2.4.2. Fragmentacja i defragmentacja ramek

Warstwa liniowa standardu 802.11 umożliwia przesyłanie ramek, podzielonych na mniejsze fragmenty w celu zmniejszenia ramkowej stopy błędu. W tym przypadku jednostka MSDU (ang. *MAC Service Data Unit*), która bez fragmentacji byłaby przesłana w jednej ramce, jest dzielona na mniejsze części. Każda z nich jest prawidłową, indywidualnie potwierdzaną, ramką danych standardu 802.11. Rozwiązanie takie pomaga zwiększyć przepustowość łącza, w przypadku gdy często występujące błędy transmisji wymagają ponawiania przesyłu ramek. Wprawdzie fragmentacja zwiększa narzut protokołu, gdyż przy podziale ramki na n fragmentów nagłówek i suma kontrolna ramki są przesyłane n-krotnie, ale retransmisja pojedynczego fragmentu jest mniej czasochłonna niż ponowny przesył całej ramki. Zazwyczaj przyjmuje się, że długość wszystkich fragmentów – z wyjątkiem ostatniego – jest identyczna. Idea fragmentacji jest przedstawiona na rys. 3.18.

Poszczególne fragmenty można w pewnym sensie traktować jako niezależne od siebie, indywidualnie potwierdzane ramki. Z punktu widzenia protokołu dostępu do łącza stanowią one jednak nierozerwalną całość, o czym świadczy możliwość przesyłu całego ciągu fragmentów w pojedynczej próbie dostępu do łącza, niezależnie od stosowanego w danej sieci wariantu unikania kolizji.



Rys. 3.18. Zasada działania fragmentacji ramek [52]
Fig. 3.18. Frame fragmentation rules

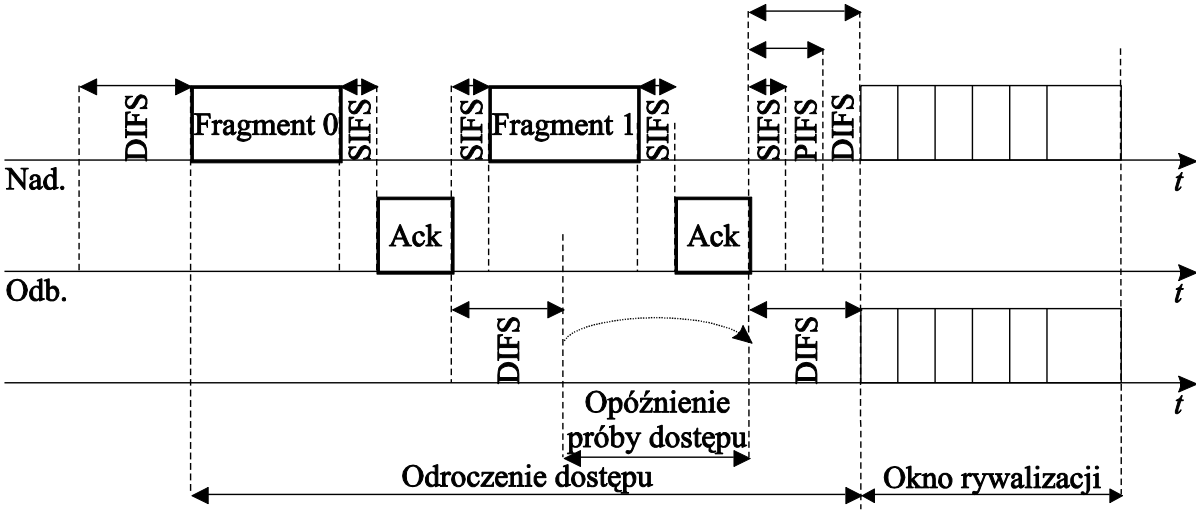
Adresat rozpoznaje użycie fragmentacji na podstawie bitu More Fragments w polu sterującym ramki. Fragmenty zostają użyte do odtworzenia całości w kolejności wskazanej w polu Sequence Control. Proces ten jest kontynuowany aż do otrzymania ostatniej ramki, w której bit More Fragments przyjmuje wartość 0. Ewentualne odebrane duplikaty są odrzucane.

3.2.4.3. Protokół DCF

Protokół DCF jest podstawowym i wymaganym elementem protokołu dostępu do łącza standardu IEEE 802.11. Wykorzystuje on mechanizmy rywalizacyjne i stanowi podstawę dla protokołów PCF i HCF. Protokół ten może być użyty w sieciach typu IBSS, BSS oraz ESS, ponieważ – jak sama nazwa wskazuje – koordynacja pracy stacji odbywa się w sposób rozproszony, zatem obecność stacji sterującej nie jest konieczna. Celem użycia protokołu jest zmniejszenie ryzyka kolizji między ramkami pochodzącymi od różnych stacji. Cel ten jest osiągnięty przez nasłuch łącza przed rozpoczęciem transmisji, z wykorzystaniem mechanizmu wykrywania nośnej. Ponieważ jednak w sieciach bezprzewodowych użycie tego mechanizmu może prowadzić do błędnej oceny stanu łącza (por. rozdz. 1.2), wprowadzono także opcjonalną metodę wykorzystującą wymianę ramek sterujących. Dzięki temu stacje znajdujące się poza zasięgiem nadajnika można powiadomić o prowadzonej transmisji. Mechanizm ten jest nazwany wirtualnym wykrywaniem nośnej (ang. *virtual carrier sense*). Nazwa taka wydaje się uzasadniona, ponieważ uzyskany efekt jest podobny jak przy „fizycznym" wykrywaniu nośnej, a jednocześnie tej „fizycznej" nośnej brak. Mechanizm „fizycznego" wykrywania nośnej jest realizowany przez warstwę fizyczną, a wirtualnego – przez protokół dostępu do łącza.

Przy dostępie do łącza z użyciem „fizycznego" wykrywania nośnej jest stosowany protokół, który można uznać za odmianę *p*-trwałej wersji protokołu CSMA (por. rozdz. 1.4.2.3).

Przed rozpoczęciem transmisji stacja prowadzi nasłuch łącza przez czas równy co najmniej DIFS. Jeśli w tym czasie łącze jest wolne, stacja rozpoczyna transmisję natychmiast po zakończeniu nasłuchu. Jeżeli natomiast łącze jest zajęte, stacja czeka aż do jego zwolnienia i także prowadzi nasłuch przez czas DIFS. Po zakończeniu nasłuchu nie rozpoczyna jednak transmisji natychmiast, tylko dzieli czas na szczeliny o stałej długości i pseudolosowo wybiera szczelinę, w której ponawia badanie stanu łącza. Jeśli jest ono wolne, stacja rozpoczyna transmisję. W przeciwnym przypadku stacja przyjmuje, iż inna stacja rozpoczęła transmisję wcześniej, wygrywając tym samym rywalizację. W takim przypadku stacja, która przegrała rywalizację, podwaja wielkość okna rywalizacji, tj. liczbę szczelin, spośród których losuje moment rozpoczęcia transmisji. Odpowiada to stosowanej także w sieci Ethernet metodzie BEB (ang. *Binary Exponential Backoff*). Dla każdej warstwy fizycznej są zdefiniowane wartości CWmin oraz CWmax, określające najmniejszą i największą liczbę szczelin okna rywalizacji. W przypadku gdy rozmiar okna rywalizacji osiągnie wartość CWmax, nie będzie ono już dalej powiększane. Jeśli natomiast stacja uzyska dostęp do łącza, okno rywalizacji zostaje skrócone do wartości CWmin. Zasadę dostępu do łącza z wykorzystaniem fizycznego wykrywania nośnej zilustrowano na rys. 3.19.



Rys. 3.19. Dostęp do łącza z wykorzystaniem fizycznego wykrywania nośnej
Fig. 3.19. Medium access using physical carrier sense

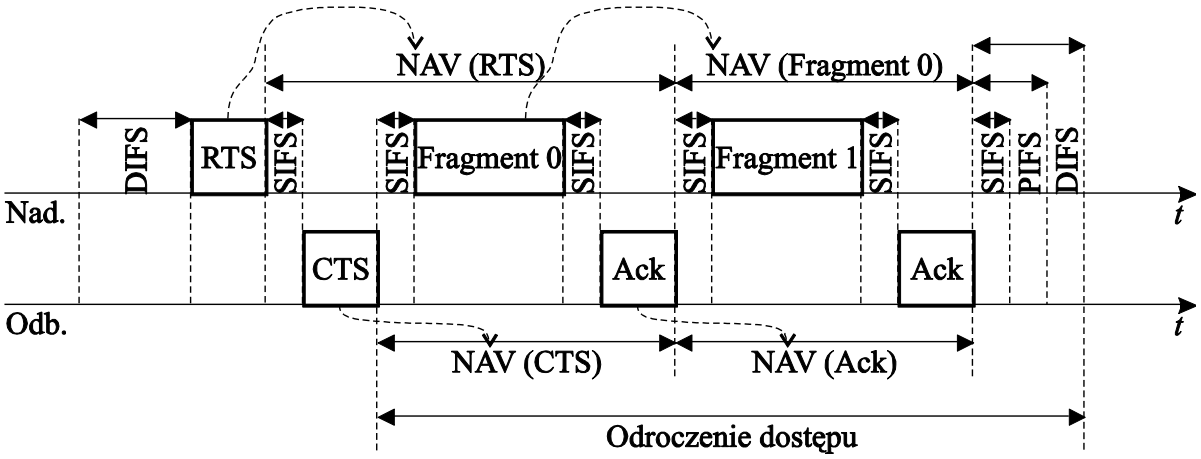
Przy dostępie do łącza z użyciem „wirtualnego" wykrywania nośnej jest stosowany protokół, który można uznać za zbliżony do protokołów MACAW czy FAMA (por. rozdz. 1.4.5 i 1.4.6). W celu eliminacji negatywnych zjawisk wynikających z obecności w sieci stacji ukrytych stosuje się wymianę ramek sterujących RTS (ang. *Request To Send*) oraz CTS (ang. *Clear To Send*). Podobnie jak w protokole MACAW stosuje się także potwierdzenia na poziomie warstwy liniowej, jednak nie używa się pozostałych wprowadzonych w tym protokole ramek sterujących. Podobnie jak w przypadku „fizycznego" wykrywania nośnej przed rozpoczęciem transmisji stacja prowadzi nasłuch łącza przez czas równy co najmniej DIFS. Jeśli

w tym czasie łącze jest wolne, stacja rozpoczyna transmisję ramki RTS natychmiast po zakończeniu nasłuchu. Następnie po upływie czasu SIFS odbiorca odpowiada ramką CTS. Z kolei nadawca, także po upływie czasu SIFS, wysyła ramkę danych (Data), której prawidłowy odbiór jest sygnalizowany przez odbiorcę ramką potwierdzenia Ack. Jeśli ramka przesyłana jest w całości, wymiana ramek kończy się, a po czasie DIFS są możliwe kolejne próby dostępu do łącza. Natomiast w przypadku użycia fragmentacji ramek występują kolejne wymiany ramek Data i Ack, oddzielone czasem SIFS.

Mechanizm „wirtualnego" wykrywania nośnej opiera się na parametrach czasowych, przekazywanych m. in. w ramkach RTS, CTS, Data i Ack. Na podstawie tej informacji stacje ustawiają tzw. wektor zajętości sieci NAV (ang. *Network Allocation Vector*). Jeśli w danej stacji wektor ten jest ustawiony, to – nawet gdy nie wykrywa ona nośnej w sposób „fizyczny" – powinna ona powstrzymać się przed rozpoczęciem nadawania. Po zakończeniu bieżącej transmisji wektor ten jest zerowany za pomocą kończącej wymianę ramki Ack.

Działanie mechanizmu NAV można wytłumaczyć następująco. Każda z przesyłanych ramek (RTS, CTS, Data, Ack) zawiera pole (Duration), określające przewidywany czas zajętości łącza. Czas ten powinien obejmować przesłanie nie tylko ramki danych, ale także towarzyszących jej ramek sterujących (np. CTS czy Ack) oraz wynikających z zasad wymiany ramek odstępów międzyramkowych. Każda stacja, odbierająca określoną ramkę, kopiuje przesłaną w ramce wartość do licznika, którego zawartość ulega stopniowemu zmniejszeniu. Póki zawartość licznika jest niezerowa, stacja przyjmuje, iż łącze jest zajęte i powstrzymuje się od transmisji. Wyzerowanie licznika może nastąpić zarówno wskutek naturalnego upływu czasu, jak i w wyniku odebrania ramki Ack z zerowym czasem zajętości łącza.

Wymiana ramek sterujących umożliwia także przesył ciągu fragmentów w jednym cyklu dostępu. Przykładową wymianę tego typu pokazano na rys. 3.20 [52].

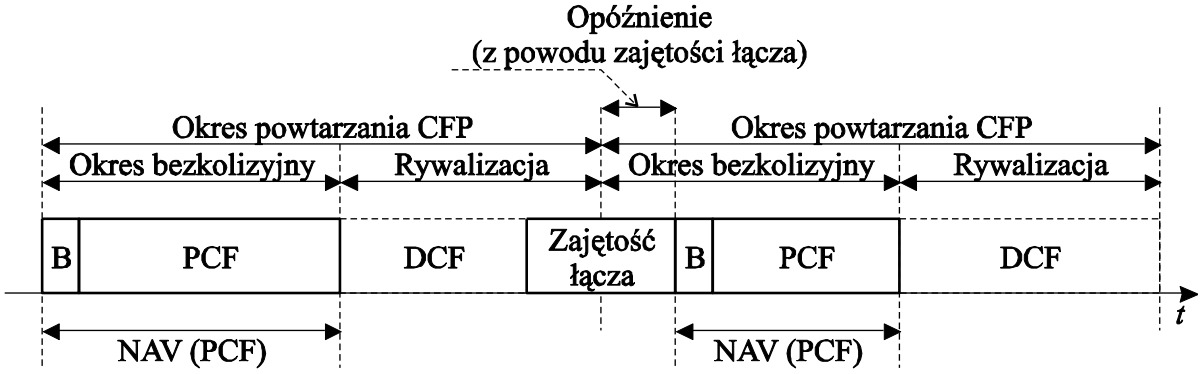


Rys. 3.20. Wymiana informacji z użyciem ramek sterujących
Fig. 3.20. Information exchange with the use of control frames

3.2.4.4. Protokół PCF

Protokół PCF wprowadzono w celu zapewnienia bezkolizyjnej transmisji ramek. Protokół ten wymaga obecności w sieci punktu dostępowego, który pełni wówczas funkcję koordynatora punktowego (ang. *Point Coordinator*). Z tego powodu PCF nie może być stosowany w sieciach typu IBSS. Protokół ten nie jest zresztą obowiązkowy i może działać nawet wówczas, gdy w danej sieci występują stacje nieobsługujące PCF.

Protokół PCF obowiązuje podczas wymiany ramek w okresie bezkolizyjnym (CFP, ang. *Contention-Free Period*), który występuje naprzemiennie z okresem rywalizacyjnym (CP, ang. *Contention Period*), obsługiwanym przez DCF. Każdy okres CFP rozpoczyna się od przesłania ramki „radiolatarni" (ang. *beacon*). Ramka ta powinna być przesyłana regularnie, w stałych odstępach czasu. Ze względu jednak na zajętość łącza rozpoczęcie okresu CFP może się opóźnić o czas nieprzekraczający najdłuższego czasu wymiany ramek w protokole DCF (tj. RTS–CTS–Data–Ack). Okres CFP ulega wówczas skróceniu, tak że mimo opóźnionego przesłania ramki „radiolatarni" częstość powtarzania CFP nie jest zakłócona. Długość okresu CFP może się zmieniać także w zależności od natężenia ruchu. Podczas okresu CFP – niezależnie od jego długości – wszystkie stacje ustawiają wektor NAV, a nadawać mogą jedynie stacje jawnie wywoływane przez punkt dostępowy. Ogólną organizację czasową łącza z działającymi protokołami DCF i PCF pokazano na rys. 3.21.



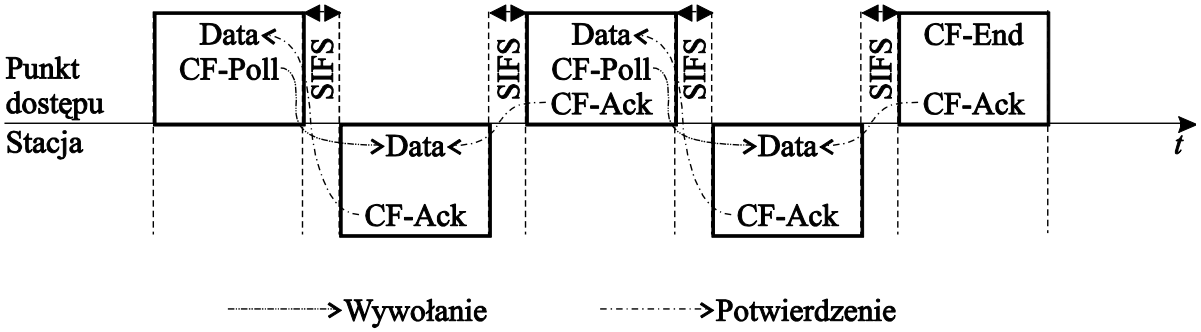
Rys. 3.21. Organizacja czasowa łącza z protokołami DCF i PCF
Fig. 3.21. Link time organization with DCF and PCF protocols

Punkt dostępowy może rozpocząć działanie zgodnie z protokołem PCF po uzyskaniu dostępu do łącza. Dostęp taki jest możliwy, jeśli łącze jest wolne przez czas równy co najmniej PIFS. Jak widać na rys. 3.17, rozwiązanie takie nadaje protokołowi PCF wyższy priorytet niż protokołowi DCF, ponieważ w chwili, gdy zgodnie z DCF można by rozpocząć transmisję, łącze jest już zajęte przez PCF. Z drugiej strony, protokół PCF nie może przerwać pojedynczej wymiany ramek, gdyż oddziela je tylko czas SIFS, który jest krótszy od PIFS.

Po uzyskaniu dostępu do łącza punkt dostępowy wysyła ramkę „radiolatarni" , zawierającej element informacyjny DTIM (ang. *Delivery Traffic Indication Message*). Następnie – po upływie czasu SIFS – punkt dostępowy może:

- przesłać ramkę danych (Data),
- wywołać stację (ramka CF-Poll),
- przesłać ramkę danych wraz z wywołaniem stacji (ramka Data+CF-Poll),
- przesłać ramkę zarządzającą,
- zakończyć okres PCF (ramką CF-End), o ile nie ma żadnej informacji do przesłania w trybie bezkolizyjnym.

Podczas kolejnych wymian punkt dostępowy może przesyłać do stacji ramki danych (Data) z opcjonalnym wskazaniem CF-Poll, gdy zachodzi konieczność wywołania adresata ramki, lub CF-Ack, gdy występuje konieczność potwierdzenia poprzedniej ramki. Mogą także występować oba wskazania. Jeśli punkt dostępowy nie ma danych do wysłania, może przesłać ramkę odpowiednio CF-Poll, CF-Ack lub CF-Ack+CF-Poll bez danych. Ramki Data+CF-Ack można także wykorzystać w celu przesłania danych do stacji niewywoływalnej lub do grupy stacji (transmisja wieloadresowa). Łączenie przesyłu danych z potwierdzeniami i wywołaniami stacji pozwala na redukcję narzutu protokołu. Przykładowo, użycie ramki Data+CF-Ack+CF-Poll zamiast sekwencji ramek CF-Ack, Data i CF-Poll pozwala zmniejszyć narzut protokołu około trzykrotnie. Użycie wskazań CF-Ack i CF-Poll ilustruje rys. 3.22.



Rys. 3.22. Ilustracja mechanizmów CF-Poll i CF-Ack protokołu PCF
Fig. 3.22. An illustration of CF-Poll and CF-Ack mechanisms of PCF protocol

W przypadku użycia ramek Data+CF-Ack lub Data+CF-Ack+CF-Poll dane i wywołanie są kierowane do innej stacji niż potwierdzenie. Z tego powodu stacja oczekująca potwierdzenia powinna dekodować typ ramki wysłanej przez punkt dostępowy bezpośrednio po przesłaniu ramki przez tę stację. Jeśli ramka wymagająca potwierdzenia zostanie odebrana przez stację niewywoływalną, stacja ta powinna wysłać potwierdzenie Ack po upływie czasu SIFS (podobnie jak w protokole DCF).

Stacja wywołana ramką CF-Poll lub inną ramką zawierającą takie wskazanie tymczasowo ignoruje – lecz nie zeruje – wektor NAV. Może ona zatem:

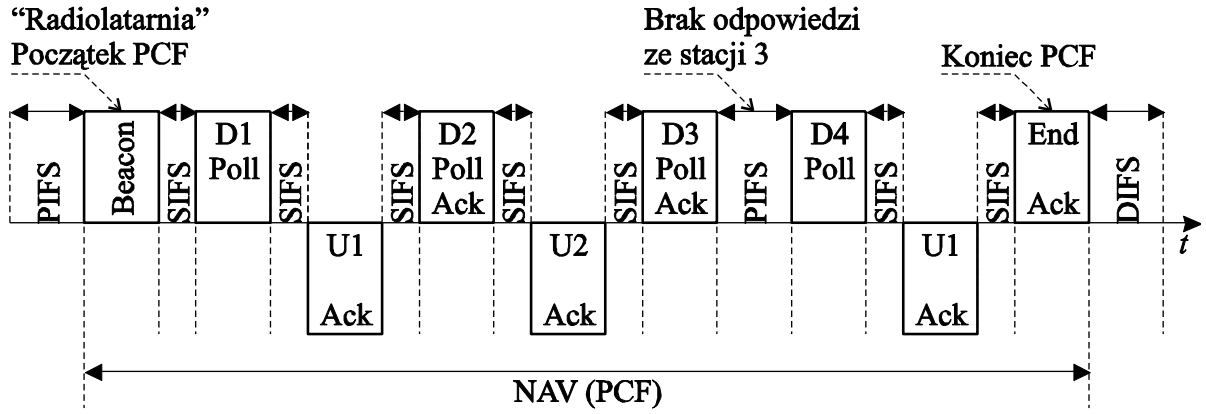
a) przesłać pojedynczą ramkę danych (Data) do punktu dostępowego lub innej stacji sieci, w tym do stacji nieobsługującej protokołu PCF,

b) poinformować punkt dostępowy, iż nie ma informacji do wysłania (ramką Null lub CF-Ack),
c) poinformować punkt dostępowy, iż transmisja nie zmieści się przed zakończeniem okresu CFP (ramką Null lub CF-Ack z opcjonalnie ustawionym bitem More Data w celu umożliwienia punktowi dostępowemu rozróżnienia sytuacji b) i c)).

W czasie realizacji protokołu PCF wszystkie ramki powinny być rozdzielone odstępem SIFS. Wyjątkiem od tej reguły jest sytuacja, gdy wywoływana stacja nie odpowiada. Wówczas, po upływie czasu PIFS, punkt dostępowy przechodzi do kolejnej transmisji. Dzięki tym zasadom stacja, która nie ustawiła wektora NAV np. wskutek błędnego odebrania ramki „radiolatarni", nie przerwie cyklu wymian PCF.

Okres bezkolizyjny kończy się ramką CF-End, przesłaną przez punkt dostępowy. Ramka taka może zawierać wskazanie CF-Ack, o ile występuje potrzeba potwierdzenia poprzedniej ramki, np. zawierającej dane. Ramkę CF-End można przesłać bezpośrednio po ramce „radiolatarni", jeśli punkt dostępowy nie ma żadnych danych do przesłania do stacji, a lista wywoływania (ang. *polling list*) jest pusta.

Przykładowy cykl wymian protokołu PCF pokazano na rys. 3.23.



Rys. 3.23. Przykładowa wymiana ramek w protokole PCF
Fig. 3.23. An example of PCF protocol frame transfer

Stosując protokół PCF, można wykorzystać listę wywoływania. Jest ona konieczna, jeśli protokół jest wykorzystany nie tylko dla dostarczania ramek stacjom, lecz także dla obsługi ruchu przychodzącego. Lista wywoływania wymusza wywoływanie określonych stacji niezależnie od tego, czy punkt dostępowy ma dla nich jakieś dane do przesłania. Aby stacja została umieszczona na liście, musi ona zgłosić swoją wywoływalność (ang. *CF-pollability*) podczas przyłączania się do sieci. Zmiana tego stanu wymaga ponownego przyłączenia się do sieci.

Pewną wadą protokołu PCF jest brak gwarancji rozpoczęcia transmisji w ściśle określonym czasie, co wynika z organizacji czasowej łącza (rys. 3.21). Po pierwsze, może wystąpić opóźnienie wysłania ramki radiolatarni, rozpoczynającej okres bezkolizyjny. Stacje pracujące zgodnie z protokołem DCF mogą bowiem rozpocząć transmisję nawet wówczas, gdy nie mo-

że ona się zakończyć przed pożądanym momentem przesłania ramki radiolatarni. Opóźnienie takie w przypadku standardu 802.11a może sięgnąć około 5 ms [71]. Ponadto, nie jest z góry znany czas transmisji ramek przez odpytywane stacje. Ramki te mogą mieć różną długość, mogą podlegać fragmentacji, mogą wreszcie być przesyłane z różnymi prędkościami transmisji. Żaden z wymienionych czynników nie podlega nadzorowi punktu dostępu. Ostatni problem może wynikać z istnienia stacji ukrytych, znajdujących się poza zasięgiem punktu dostępu. Z tego powodu nie są one powiadamiane o rozpoczęciu okresu transmisji bezkolizyjnej i kontynuują rywalizacyjną wymianę danych. Standard nie określa także algorytmu wyznaczania kolejności wywoływania stacji, co może rodzić obawy, że przydział łącza dla danej stacji nastąpi zbyt późno. Biorąc pod uwagę wymienione problemy, nietrudno zauważyć, iż wsparcie dla aplikacji uwarunkowanych czasowo jest ograniczone. Pomimo to można oszacować opóźnienie dla najgorszego przypadku, tak więc determinizm protokołu jest w dużej mierze zachowany. Protokół PCF jest implementowany tylko w niektórych urządzeniach.

3.2.4.5. Protokół HCF

Protokół HCF, wraz z jego dwiema składowymi – EDCA oraz HCCA – wprowadzono w standardzie 802.11e w celu zapewnienia lepszego wsparcia dla jakości usług (ang. *QoS support*) niż w przypadku protokołu PCF.

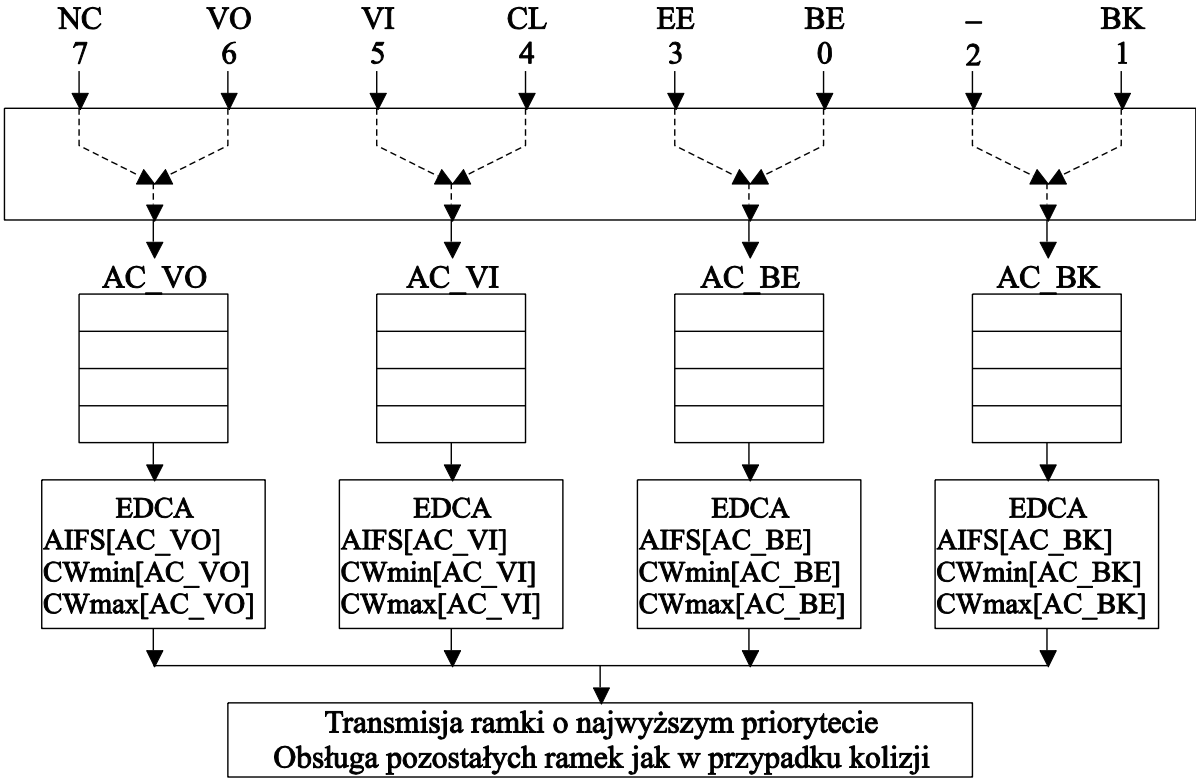
Podstawową zmianą w stosunku do wcześniejszych rozwiązań (DCF i PCF) jest wprowadzenie prawa transmisji TXOP (ang. *Transmission Opportunity*). Prawo to określa moment rozpoczęcia transmisji oraz maksymalny czas jej trwania, nie oznacza jednak, że transmisja faktycznie będzie miała miejsce. Prawo transmisji można uzyskać przez wygranie rywalizacji w protokole EDCA w okresie dostępu rywalizacyjnego bądź przez odbiór wywołania QoS CF-Poll w okresie dostępu rywalizacyjnego lub bezkolizyjnego. Pierwszy przypadek określa się jako EDCA TXOP, drugi – HCCA TXOP. Gwarantuje się przy tym, że HCCA TXOP ulegnie zakończeniu przed wyznaczonym momentem przesłania ramki radiolatarni (TBTT, ang. *Target Beacon Transmission Time*); jest to istotna zmiana w stosunku do protokołu PCF.

Wybór sposobu uzyskania dostępu odbywa się na podstawie identyfikatora ruchu, przypisanego ramce na podstawie wymagań odnośnie do gwarancji jakości usług. Identyfikator ten może przyjąć wartości z zakresu 0÷15. Wartości z zakresu 0÷7 odpowiadają priorytetom użytkownika i skutkują obsługą ramki, zgodnie z protokołem EDCA. Pozostałe wartości skutkują przypisaniem ramki do jednej z ośmiu kolejek dla poszczególnych strumieni ruchu i obsługą zgodnie z protokołem HCCA.

3.2.4.6. Protokół EDCA

Protokół EDCA umożliwia rywalizacyjny dostęp do łącza, z uwzględnieniem zróżnicowanych kategorii dostępu (ang. *Access Category*). Wyróżnia się cztery kategorie, wyprowadzone z priorytetów użytkownika (ang. *User Priority*), zdefiniowanych w standardzie 802.1D

[5]. Kategorie te są przypisane do czterech kolejek, w których funkcje dostępu do łącza są realizowane niezależnie, z wykorzystaniem indywidualnych parametrów dla każdej kolejki. Zasadę działania protokołu EDCA wyjaśniono na rys. 3.24. Zależności między kategoriami dostępu i priorytetami użytkownika pokazano w tabeli 3.6. Warto zauważyć przy tym, iż niektóre źródła, pochodzące sprzed momentu opublikowania standardu (np. [71]), podają inne sposoby przypisania kategorii dostępu poszczególnym priorytetom. Może to świadczyć o ewolucji proponowanych rozwiązań przed ich standaryzacją.



Rys. 3.24. Zasada działania protokołu EDCA [52]
Fig. 3.24. Operating rules of EDCA protocol

Ze względu na niezależną realizację funkcji dostępu do łącza w poszczególnych kolejkach, może wystąpić „wewnętrzna kolizja" ramek, rywalizujących o dostęp do łącza w ramach jednej stacji. W tej sytuacji pierwszeństwo uzyskuje ramka o priorytecie najwyższym spośród rywalizujących ramek. W odniesieniu do pozostałych ramek obowiązują wówczas takie same zasady jak w przypadku kolizji w łączu, z tą wszakże różnicą, że w polach sterujących tych ramek nie ustawia się bitu retransmisji (*Retry*). Stacja, która wygrała rywalizację i otrzymała prawo transmisji EDCA, może przesłać więcej ramek, o ile tylko należą one do jednej kategorii dostępu. Przydzielony danej stacji czas powinien umożliwiać także przesłanie ramek, będących natychmiastową odpowiedzią na ramkę wysłaną przez posiadacza prawa transmisji. Przykładem takiej odpowiedzi jest ramka potwierdzenia Ack. Przekroczenie przy-

dzielonego czasu transmisji może nastąpić jedynie w przypadku retransmisji ramki z obniżoną prędkością; w takim jednak przypadku można przesłać tylko jedną ramkę.

Tabela 3.6

Kategorie dostępu w protokole EDCA standardu IEEE 802.11

Priorytet	Standard 802.1D [5]			Standard 802.11 [52]	
	Priorytet użytkownika	Ozn.	Przeznaczenie	Kategoria dostępu	Przeznaczenie
najwyższy	7	NC	*Network Control*	AC_VO	*Voice*
↑	6	VO	*Voice*		
	5	VI	*Video*	AC_VI	*Video*
	4	CL	*Controlled Load*		
	3	EE	*Excellent Effort*	AC_BE	*Best Effort*
	0	BE	*Best Effort*		
	2	–	*Spare*	AC_BK	*Background*
najniższy	1	BK	*Background*		

O ile wyznaczenie ramki o najwyższym priorytecie w ramach jednej stacji jest względnie proste, o tyle rywalizacja między stacjami, gwarantująca pierwszeństwo takiej ramce, wymaga modyfikacji mechanizmu stosowanego w protokole DCF. W miejsce odstępu międzyramkowego DIFS, używanego przez wszystkie stacje, wprowadzono zróżnicowane odstępy AIFS o długości zależnej od kategorii dostępu i przypisanego jej priorytetu. W ten sposób ramki należące do kategorii o priorytecie wyższym uzyskują pierwszeństwo przed pozostałymi, które w chwili rozpoczęcia okna rywalizacji stwierdzają zajętość łącza. Jest to zatem mechanizm zbliżony do stosowanego w celu nadania protokołowi PCF pierwszeństwa w stosunku do DCF. Dodatkowo, dla każdej kategorii dostępu są używane indywidualne okna rywalizacji. Również minimalna i maksymalna wielkość tego okna jest określona dla każdej kategorii z osobna. Wartości te wyprowadza się z parametrów warstwy fizycznej.

W tabeli 3.7 zestawiono parametry protokołu EDCA, obowiązujące dla wszystkich kategorii dostępu. Są one podawane przez punkt dostępu w elemencie informacyjnym, występującym m. in. w ramkach radiolatarni. W przypadku braku takiej informacji stacje przyjmują domniemane wartości parametrów.

Analizując dane zawarte w tabeli 3.7, można stwierdzić, iż:

- okna rywalizacji kategorii dostępu AC_VO, AC_VI i AC_BE następują bezpośrednio po sobie, dzięki czemu ramki należące do kategorii wyższej nie rywalizują z ramkami należącymi do kategorii niższej,
- okna rywalizacji kategorii AC_BE i AC_BK mają identyczną długość, a pierwszeństwo kategorii AC_BE wynika z użycia krótszego odstępu międzyramkowego, a co za tym idzie, wcześniejszego rozpoczęcia okna rywalizacji,

- dwie najwyższe kategorie dostępu umożliwiają przesłanie wielu ramek w ramach jednego prawa dostępu, o ile tylko całkowity czas ich transmisji wraz z odpowiedziami nie przekroczy limitu TXOP; możliwość taka występuje jedynie w warstwach fizycznych DSSS, HR-DSSS, OFDM i ERP,
- dwie najniższe kategorie dostępu oraz pozostałe warstwy fizyczne (FHSS i Ir) umożliwiają przesłanie co najwyżej jednej ramki wraz z odpowiedzią w ramach jednego prawa dostępu (limit TXOP wynosi 0),
- moment rozpoczęcia okna rywalizacji zależy od kategorii dostępu i określonej dla niej wartości parametru AIFSN, wpływającej na długość odstępu międzyramkowego AIFS; przykładowo, jeżeli AIFSN=2, to AIFS=DIFS,
- dwie najwyższe kategorie dostępu rozpoczynają rywalizację jednocześnie z DCF, ale, biorąc pod uwagę mniejszą długość ich okien rywalizacji, stacje posługujące się protokołem DCF mają mniejsze szanse na uzyskanie dostępu do łącza w danej próbie,
- dwie najniższe kategorie rozpoczynają rywalizację później niż DCF, zatem, biorąc pod uwagę taką samą długość okien rywalizacji, stacja posługująca się protokołem DCF ma większe szanse na uzyskanie dostępu do łącza w danej próbie,
- zastosowane mechanizmy zmniejszają ryzyko kolizji między ramkami o różnych kategoriach dostępu, pochodzącymi z różnych stacji,
- możliwe są kolizje między ramkami, należącymi do tej samej kategorii dostępu, a pochodzącymi z różnych stacji; wydaje się to szczególnie prawdopodobne w przypadku kategorii AC_VO, jeśli okno rywalizacji ma minimalną wielkość (3 szczeliny).

Powyższe zależności wyjaśniono na rys. 3.25.

Tabela 3.7

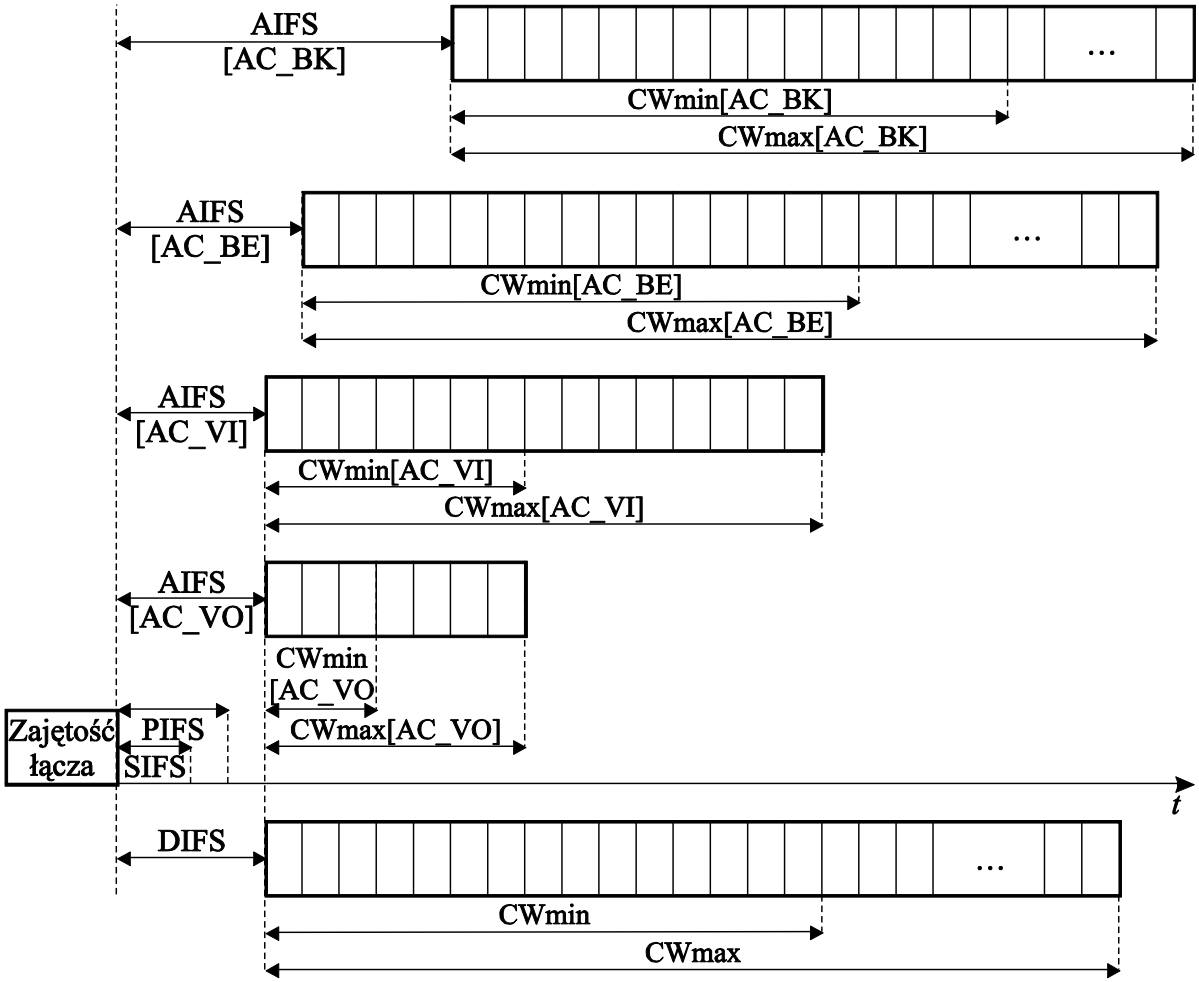
Parametry protokołu EDCA w standardzie 802.11 [52]

Kategoria dostępu	CWmin	CWmax	AIFSN	Limit TXOP dla warstwy		
				DSSS	OFDM	pozostałe
AC_VO	(aCWmin+1)/4-1	(aCWmin+1)/2-1	2	3,264 ms	1,504 ms	0
AC_VI	(aCWmin+1)/2-1	aCWmin	2	6,016 ms	3,008 ms	0
AC_BE	aCWmin	aCWmax	3	0	0	0
AC_BK	aCWmin	aCWmax	7	0	0	0

W niektórych typach ramek stosowanych w standardzie 802.11 kategoria ruchu nie jest jawnie określona. W takim przypadku kategorie przydziela się następująco:

- ramki zarządzające przesyła się przy użyciu kategorii AC_VO bez ograniczeń wynikających z kontroli przyjmowania (ang. *admission control*),

- stacje obsługujące mechanizmy QoS przesyłają ramki zarządzające przed skojarzeniem z siecią także przy użyciu kategorii AC_VO nawet wówczas, gdy sieć nie obsługuje tych mechanizmów,
- ramki potwierdzenia blokowego (ang. *Block Acknowledge*) oraz żądania potwierdzenia blokowego (ang. *Block Acknowledge Request*) przesyła się z wykorzystaniem tych samych parametrów (w tym kategorii dostępu) co potwierdzane w ten sposób ramki danych,
- ramkom PS-Poll przypisuje się kategorię AC_BE w celu zmniejszenia ryzyka kolizji po ramce radiolatarni,
- ramkom RTS przypisuje się tę samą kategorię dostępu, jaka została określona dla pozostałych ramek danych lub sterujących przesyłanych w tym samym cyklu wymiany.



Rys. 3.25. Odstępy międzyramkowe i długość okna rywalizacji w protokołach DCF i EDCA [52]
Fig. 3.25. Interframe spaces and contention window length in DCF and EDCA protocols

3.2.4.7. Protokół HCCA

Protokół HCCA łączy zalety protokołów PCF i DCF. Występuje tu organizacja czasowa oparta na strukturze superramki oraz centralna stacja sterująca – koordynator hybrydowy (HC, ang. *Hybrid Coordinator*). Stacje uzyskują dostęp do łącza, gdy zostaną wywołane

przez koordynatora, przy czym wywołanie takie może nastąpić w dowolnym momencie superramki – w szczególności w okresie rywalizacyjnym CP. Okres bezkolizyjny CFP nie jest zatem wymagany dla wywoływania stacji, a jego wystąpienie w superramce jest opcjonalne. Aby koordynator mógł prawidłowo realizować swoje funkcje, ma on wyższy priorytet przy dostępie do łącza, zatem nawet w okresie rywalizacyjnym może uzyskać pierwszeństwo względem pozostałych stacji. Warto także zaznaczyć, że ograniczona i z góry znana długość prawa transmisji uniemożliwia przekroczenie czasu trwania superramki, co sprzyja prawidłowej obsłudze ruchu uwarunkowanego czasowo.

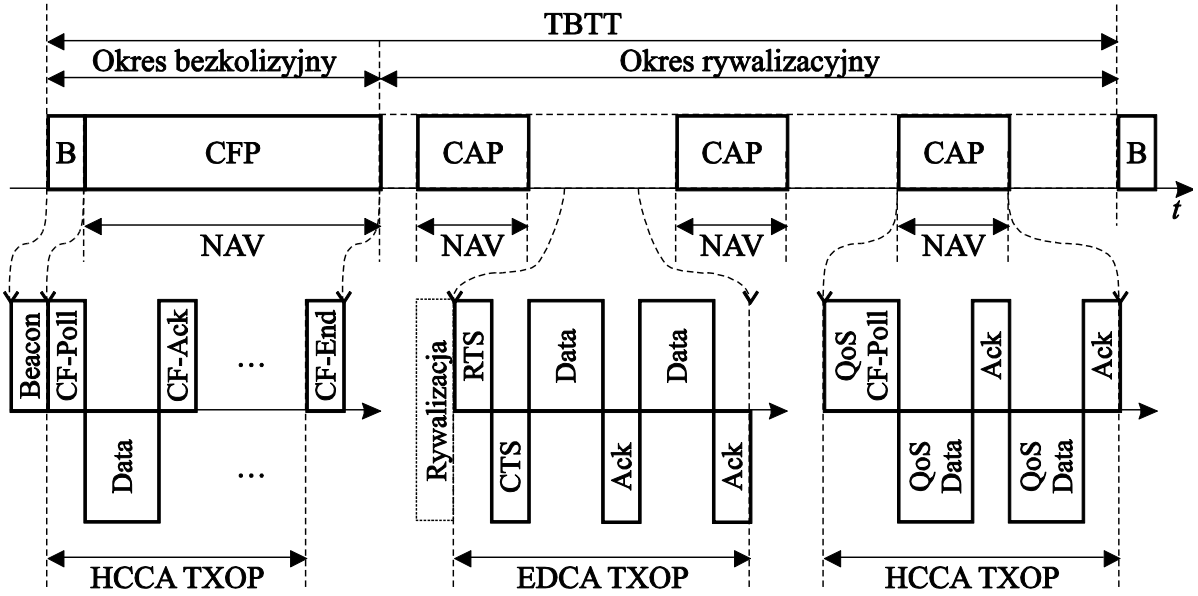
Superramka rozpoczyna się w chwili, gdy upływa czas TBTT. Wówczas koordynator przesyła ramkę radiolatarni, po czym może rozpocząć się okres bezkolizyjny. Podlega on takim samym regułom jak w przypadku protokołu PCF, z tą różnicą, że do jego zakończenia można użyć jedynie ramki CF-End. Z drugiej strony, okres bezkolizyjny można także zakończyć ramką CF-End+CF-Ack, ale tylko wtedy, gdy punkt dostępu, pełniący funkcję koordynatora, komunikuje się ze stacjami, nieobsługującymi mechanizmów gwarancji jakości usług, zgodnie z protokołem PCF. Wywoływanie stacji obsługujących te mechanizmy przy użyciu ramek ze wskazaniem CF-Poll, ale bez wskazania QoS, nie jest zalecane. Można natomiast wywoływać je przy użyciu ramek ze wskazaniami CF-Poll i QoS. Użycie okresu bezkolizyjnego w tym celu nie jest jednak obowiązkowe.

Po zakończeniu okresu bezkolizyjnego rozpoczyna się okres rywalizacyjny. W tym czasie stacje mogą uzyskać prawo transmisji, zgodnie z protokołem EDCA. Pomimo to koordynator może uzyskać prawo transmisji w celu przesłania ramek lub wywołania innej stacji w ramach fazy dostępu sterowanego (CAP, ang. *Controlled Access Phase*), jeśli stwierdzi brak zajętości łącza przez czas PIFS. Dzięki temu charakteryzuje się on priorytetem wyższym od pozostałych stacji, które przed rozpoczęciem rywalizacji muszą stwierdzić brak zajętości łącza przez dłuższy o jedną szczelinę czas DIFS. Jest to podobna zależność jak w przypadku protokołów PCF i DCF, ale HCCA pozwala koordynatorowi uzyskać dostęp wielokrotnie w czasie trwania superramki. Po uzyskaniu dostępu do łącza koordynator może rozpocząć dowolną wymianę ramek, z tym że jeśli upłynął czas TBTT, pierwszą ramką musi być ramka radiolatarni. Wymiana ramek jest zbliżona do wymiany w protokole PCF, ale używa się ramek ze wskazaniem QoS (m. in. QoS CF-Poll, QoS Data itp.). Wszystkie stacje – prócz jawnie wywołanej ramką ze wskazaniem CF-Poll – mają ustawiony wektor NAV, nie rywalizują zatem o dostęp do łącza. Po zakończeniu wymiany ramek, jeśli nie wygasło jeszcze prawo transmisji, jego posiadacz ma prawo rozpocząć kolejną wymianę po odczekaniu czasu SIFS. Jeśli łącze pozostaje wolne przez czas PIFS po wygaśnięciu prawa transmisji, koordynator może ponownie uzyskać dostęp i zainicjować kolejną wymianę ramek. W przypadku braku takiej akcji faza dostępu sterowanego ulega zakończeniu, a stacje mogą ubiegać się o prawo transmisji, zgod-

nie z protokołem EDCA. Wszystkie transmisje muszą zakończyć się przed upływem czasu TBTT, aby nie naruszyć wyznaczonego momentu transmisji ramki radiolatarni.

Przykładową superramkę z wyróżnionym okresem bezkolizyjnym CFP i fazami dostępu sterowanego CAP pokazano na rys. 3.26.

Każda ramka ze wskazaniami QoS i CF-Poll zawiera limit TXOP w polu QoS Control. Zawartość pola Duration tej ramki jest ustawiona tak, że transmisja wszystkich ramek w czasie przyznanego prawa transmisji jest chroniona przez mechanizm wirtualnego wykrywania nośnej (w tym czasie stacje mają ustawiony wektor NAV). Jeśli podany limit wynosi 0, stacja powinna przesłać pojedynczą ramkę lub – w przypadku braku danych do przesłania – ramkę QoS Null. Po udzieleniu stacji prawa transmisji koordynator także ustawia wektor NAV, może jednak odzyskać to prawo, jeśli stacja go nie używa lub zakończyła transmisję przed upływem jego ważności.



Rys. 3.26. Przykładowa superramka protokołu HCCA [52]
Fig. 3.26. An example of HCCA protocol superframe

Podsumowując, protokół HCCA wprowadza wiele zmian w stosunku do PCF, a najważniejsze z nich są następujące:

- transmisje w protokole HCCA są możliwe zarówno w okresie dostępu rywalizacyjnego, jak i bezkolizyjnego, podczas gdy protokół PCF obsługuje tylko dostęp bezkolizyjny,
- prawo transmisji udzielone przez koordynator hybrydowy ma ściśle określoną długość, podczas gdy w protokole PCF długość ta nie jest określona,
- stacja może przesłać wiele ramek w ramach przydzielonego prawa transmisji, podczas gdy w protokole PCF może przesłać tylko jedną ramkę,
- w protokole HCCA stacje mogą być wywoływane wielokrotnie i w dowolnym momencie superramki, natomiast w PCF – jedynie w okresie bezkolizyjnym; możliwość wywołania

stacji w dowolnym superramki momencie pozwala istotnie zmniejszyć opóźnienia przy dostępie do łącza, ułatwia także obsługę strumieni ruchu wymagających częstego dostępu do łącza i zwiększa elastyczność protokołu,

- w protokole HCCA moment rozpoczęcia superramki jest ściśle określony i nienaruszalny, podczas gdy w PCF może się opóźnić wskutek zajętości łącza (opóźnienie to może sięgnąć około 5 ms).

3.2.4.8. Potwierdzenie blokowe

Wprowadzone – podobnie jak obsługa gwarancji jakości usług – wraz ze standardem 802.11e potwierdzenie blokowe (ang. *Block Acknowledge*) pozwala na podniesienie wydajności transmisji przez połączenie pewnej liczby potwierdzeń w pojedynczą ramkę. Potwierdzenie blokowe może być natychmiastowe (ang. *immediate*) lub opóźnione (ang. *delayed*). Pierwsze stosuje się podczas przesyłu informacji z wysoką prędkością i małym opóźnieniem, drugie natomiast może być przydatne, gdy aplikacja może zaakceptować zwiększone opóźnienia.

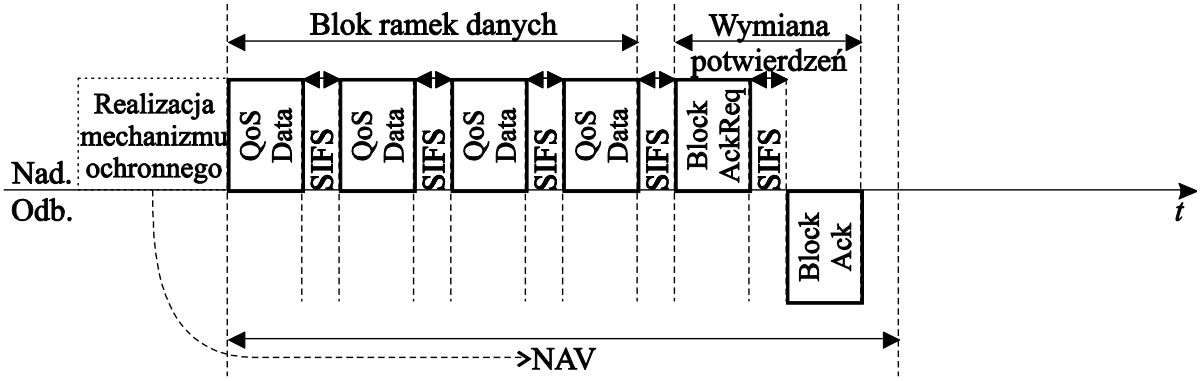
Transmisja z użyciem potwierdzenia blokowego musi być uzgodniona przez nadawcę (ang. *originator*) i odbiorcę (ang. *recipient*). Uzgodnienie to jest wykonywane za pomocą ramek akcji ADDBA Request i ADDBA Response. Są to ramki zarządzające, zawierające następujące informacje:

- wybór strategii potwierdzenia (natychmiastowe lub opóźnione),
- identyfikator klasy lub strumienia ruchu, dla którego uzgadnia się potwierdzenie blokowe,
- liczbę buforów przeznaczonych dla obsługi transmisji,
- czas, po którym transmisja z potwierdzeniem blokowym jest automatycznie zrywana, jeśli nie wystąpił przesył danych,
- numer kolejny pierwszej ramki danych przesyłanej w ramach potwierdzenia blokowego.

Wybór strategii potwierdzenia i liczba buforów, podane w ramce ADDBA Request, mają wyłącznie charakter informacyjny i mogą być zmienione przez odbiorcę. Jeśli potwierdzenie blokowe dotyczy określonego strumienia danych, jego uzgodnienie powinno zostać poprzedzone ustaleniem wymaganych parametrów transmisji tego strumienia.

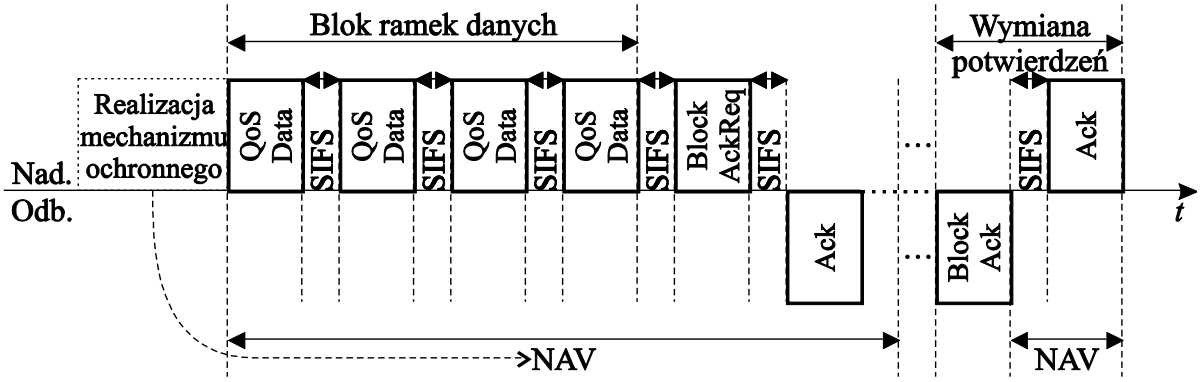
Po uzgodnieniu potwierdzenia blokowego nadawca może przesłać blok ramek typu QoS Data, rozdzielonych odstępem SIFS. Pole QoS Control tych ramek powinno wskazywać strategię potwierdzenia blokowego. Liczba ramek nie może przekroczyć rozmiaru bufora, podanego w ramce ADDBA Response. Nadawca może zażądać potwierdzenia odbioru przesłanych ramek za pomocą ramki BlockAckReq. Jeśli uzgodniono strategię potwierdzenia natychmiastowego, odbiorca przesyła ramkę BlockAck po upływie czasu SIFS. Jeśli natomiast uzgodniono strategię potwierdzenia opóźnionego, ramka BlockAckReq jest potwierdzana „zwykłą” ramką potwierdzenia Ack, zaś potwierdzenie blokowe BlockAck zostanie przesłane

później, w ramach osobnego prawa transmisji udzielonego odbiorcy. Ramka taka powinna zostać przesłana z użyciem najwyższej kategorii dostępu tak szybko, jak tylko jest to możliwe. Nadawca potwierdza odbiór ramki BlockAck „zwykłą" ramką potwierdzenia Ack. Jeśli nadawcą lub odbiorcą jest koordynator, w miejsce ramki Ack może on użyć ramki ze wskazaniem CF-Ack, o ile potwierdzana w ten sposób ramka (odpowiednio BlockAck lub BlockAckReq) była ostatnią przesłaną w ramach udzielonego prawa transmisji. Typową transmisję z użyciem potwierdzenia blokowego natychmiastowego pokazano na rys. 3.27, a z użyciem potwierdzenia blokowego opóźnionego – na rys. 3.28.



Rys. 3.27. Transmisja z użyciem potwierdzenia blokowego natychmiastowego [52]
Fig. 3.27. Transmission with immediate block acknowledge

Ramka BlockAck może potwierdzić odbiór co najwyżej 64 ramek. Ponieważ użycie potwierdzenia blokowego nie wyklucza fragmentacji ramek, konieczne jest potwierdzenie każdego fragmentu z osobna. W tym celu ramka BlockAck zawiera 128-bajtowe pole bitowe, w którym każdy bit określa stan odbioru poszczególnych ramek i fragmentów. Numer kolejny pierwszej potwierdzanej ramki jest zawarty w polu Block Ack Starting Sequence Number.



Rys. 3.28. Transmisja z użyciem potwierdzenia blokowego opóźnionego [52]
Fig. 3.28. Transmission with delayed block acknowledge

W celu realizacji transmisji blokowej nadawca może:

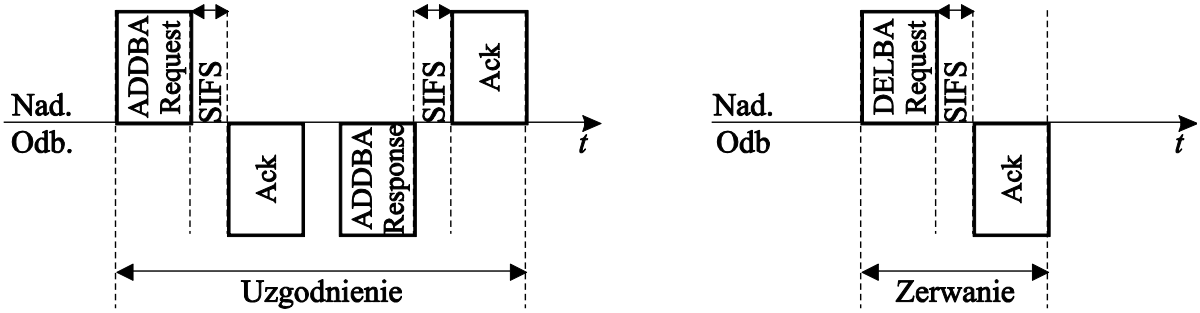
- przesyłać blok ramek danych i ramkę BlockAckReq w ramach osobnych praw transmisji,
- rozdzielić blok ramek danych lub ramkę bloku danych pomiędzy osobne prawa transmisji,

- przesłać w ramach jednego prawa transmisji ramki dla różnych identyfikatorów ruchu i o różnych adresach docelowych (RA).

Przed rozpoczęciem transmisji bloku ramek danych nadawca może użyć dowolnego mechanizmu ochronnego, jak np. HCCA, RTS/CTS w celu zmniejszenia ryzyka kolizji. Jeśli nie używa żadnego takiego mechanizmu, pierwsza ramka bloku powinna wywołać odpowiedź odbiorcy. Pole Duration tej ramki powinno zawierać odpowiednią wartość, tak aby pozostałe stacje mogły ustawić wektor NAV na cały czas przesyłu bloku lub przynajmniej na czas ważności uzyskanego przez nadawcę prawa nadawania.

W przypadku stosowania potwierdzenia opóźnionego dopuszcza się przesłanie kolejnego bloku ramek po wysłaniu ramki BlockAckReq, ale przed otrzymaniem odpowiadającej jej ramki BlockAck.

Gdy nadawca nie ma już danych do przesłania i zakończono ostatnią wymianę blokową, może zasygnalizować koniec przesyłu blokowego ramką akcji DELBA. Ramka ta jest potwierdzana przez odbiorcę jedynie „zwykłą" ramką potwierdzenia Ack. Zerwanie (ang. *teardown*) przesyłu blokowego jest możliwe, gdy nie ma już do przesłania ramek BlockAck, BlockAckReq i QoS Data dla danego identyfikatora ruchu. Etapy uzgodnienia i zerwania transmisji z użyciem potwierdzenia blokowego pokazano na rys. 3.29.



Rys. 3.29. Uzgodnienie i zerwanie transmisji z użyciem potwierdzenia blokowego
Fig. 3.29. Block Acknowledge transmission setup and teardown

3.2.4.9. Transmisja bez potwierdzeń

Oprócz transmisji z wykorzystaniem potwierdzenia blokowego jest możliwa także transmisja bez potwierdzenia, wprowadzona w standardzie 802.11e. W przypadku jej użycia warstwa liniowa nie realizuje żadnego mechanizmu odzyskania ramek danych, utraconych w wyniku kolizji lub błędów transmisji. Aby zmniejszyć ryzyko kolizji, zaleca się użycie odpowiednich mechanizmów, jak np. transmisja w trybie HCCA lub wymiana ramek sterujących RTS i CTS.

Transmisja bez potwierdzenia w środowisku bezprzewodowym jest raczej mało wiarygodna. Pomimo to może być stosowana, gdy mechanizm odzyskiwania utraconych fragmentów informacji zaimplementowano w wyższych warstwach sieci lub gdy retransmisja na po-

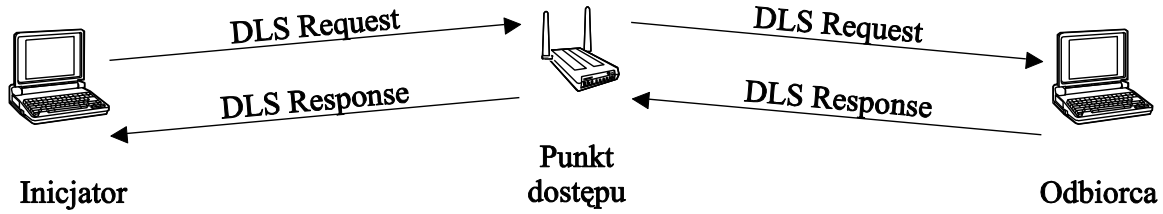
ziomie warstwy liniowej spowodowałaby zbyt duże opóźnienia. Ostatnia z wymienionych sytuacji może wystąpić np. podczas transmisji głosu.

3.2.4.10. Transmisja bezpośrednia

W sieciach bezprzewodowych typu BSS lub ESS wymiana ramek między stacjami musi odbywać się za pośrednictwem punktu dostępu. Rozwiązanie takie ma pewne zalety, umożliwia bowiem transmisję między stacjami, które nie mają bezpośredniej łączności pomimo przynależności do jednej sieci. Z drugiej jednak strony każda ramka musi być przesłana dwukrotnie, co zmniejsza efektywną pojemność sieci o połowę. Wprawdzie protokół PCF dopuszcza przesłanie ramki bezpośrednio do stacji docelowej, w chwili wyznaczonej przez punkt dostępu, lecz z jego pominięciem, ale protokół ten, jak już wspomniano, jest implementowany w stosunkowo niewielkiej liczbie urządzeń.

Standard 802.11e, prócz opisanych powyżej rozwiązań, umożliwia także bezpośrednią wymianę ramek pomiędzy stacjami za pomocą tzw. łącza bezpośredniego (ang. *Direct Link*, *Station-to-Station link*). Wymiana taka, podobnie jak potwierdzenie blokowe, jest inicjowana za pomocą ramek akcji DLS (ang. *Direct Link Setup*). Inicjator połączenia bezpośredniego przesyła ramkę DLS Request, która jest retransmitowana przez punkt dostępu. Powodów takiego działania jest kilka. Odbiorca ramki DLS Request może być w stanie energooszczędnym, z którego może być wybudzony tylko przez punkt dostępu. Ponadto, można uzgodnić pewne parametry połączenia, jak np. prędkości transmisji. Wydaje się także, że poinformowanie punktu dostępowego o nawiązywaniu połączenia bezpośredniego może być celowe.

Jeżeli adresat ramki DLS Request jest stacją obsługującą mechanizmy gwarancji jakości usług, a w sieci są dopuszczalne połączenia bezpośrednie, punkt dostępu przesyła ramkę do odbiorcy. Gdy zaakceptuje on nawiązanie połączenia, odsyła ramkę DLS Response do punktu dostępu, który retransmituje ją do inicjatora. Sposób nawiązywania połączenia bezpośredniego przedstawiono na rys. 3.30.



Rys. 3.30. Nawiązanie połączenia bezpośredniego
Fig. 3.30. Direct Link setup

Po nawiązaniu połączenia stacje mogą komunikować się bezpośrednio, podobnie jak w sieci IBSS, przy użyciu dowolnego mechanizmu dostępu do łącza. Mogą także użyć potwierdzenia blokowego i utworzyć strumień ruchu w celu zapewnienia wystarczającej przepustowości. Zaleca się użycie mechanizmu zmniejszającego ryzyko kolizji, jak np. dostęp sterowany HCCA czy wymiana ramek sterujących RTS i CTS.

Inicjator może zerwać połączenie bezpośrednie, przesyłając ramkę DLS Teardown do punktu dostępowego. Ramka ta jest następnie przesyłana do odbiorcy. Również punkt dostępu może zainicjować zerwanie połączenia bezpośredniego, jeśli np. jedna ze stacji zaangażowanych w to połączenie opuści sieć.

Wszystkie ramki akcji związane z połączeniem bezpośrednim zawierają w polu informacyjnym źródłowy i docelowy adres MAC. Prócz tego ramki DLS Request i DLS Response zawierają informacje o prędkościach transmisji obsługiwanych przez obie stacje i możliwościach tych stacji.

3.2.5. Elementy zarządzania siecią

Każda stacja zgodna ze standardem 802.11 może znajdować się w jednym z trzech stanów, różniących się uprawnieniami stacji:

- początkowym, gdy stacja nie jest jeszcze ani uwierzytelniona, ani skojarzona z siecią,
- dopuszczenia, gdy stacja przeszła pomyślnie proces uwierzytelniania, ale jeszcze nie połączyła się z siecią,
- połączenia, gdy zarówno proces uwierzytelniania, jak i skojarzenia z siecią przebiegły pomyślnie.

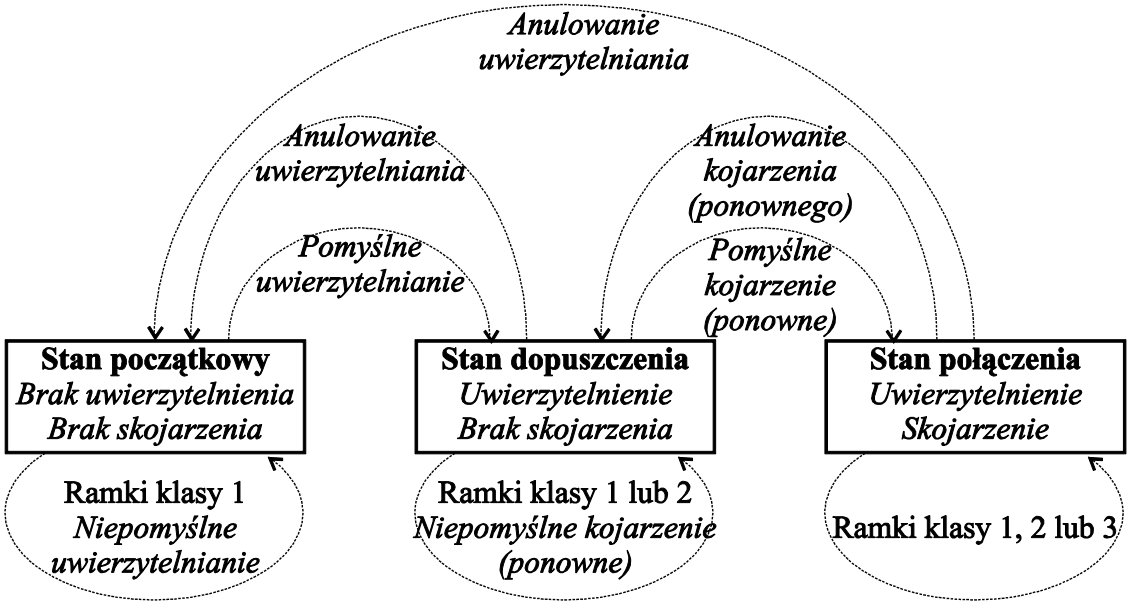
W stanie początkowym stacja może przesyłać jedynie ramki klasy 1. Są to m. in. ramki sterujące RTS, CTS, Ack, CF-End i CF-End+CF-Ack, zarządzające związane z poszukiwaniem sieci i uwierzytelnianiem stacji oraz ramki danych niepochodzące z systemu dystrybucyjnego ani do niego niekierowane.

W stanie uwierzytelnienia stacja może przesyłać ramki klasy 1 i 2. Do tej ostatniej należą ramki związane z podłączaniem stacji do sieci.

W stanie połączenia stacja może przesyłać ramki klasy 1, 2 lub 3. Do tej ostatniej należą ramki sterujące PS-Poll oraz ramki danych pochodzące z systemu dystrybucyjnego lub do niego kierowane. Ramki typu 3 można przesyłać tylko w sieciach zawierających punkty dostępowe.

Może się zdarzyć, że punkt dostępu odbierze ramkę innej klasy niż dopuszczone do przesyłania w danym stanie stacji. Wówczas odpowiada on ramką anulowania uwierzytelnienia, jeśli stacja nie była uwierzytelniona, lub ramką anulowania skojarzenia, jeśli stacja była uwierzytelniona, ale nie połączona z siecią. Opisana sytuacja może wystąpić wskutek błędów transmisji podczas przyłączania stacji do sieci lub wskutek celowego działania, mającego na celu uzyskanie informacji o sieci. Z tego powodu, jeśli stacja nie jest dopuszczona do uczestnictwa w danej sieci na podstawie adresu MAC karty, ramka pochodząca z tej stacji zostanie po prostu zignorowana.

Diagram stanów stacji sieci 802.11 przedstawia rys. 3.31.



Rys. 3.31. Ogólny diagram stanów stacji w standardzie 802.11 [52]
Fig. 3.31. General station state diagram in 802.11 standard

3.2.5.1. Przeglądanie sieci

Przed rozpoczęciem korzystania z sieci 802.11 należy przeprowadzić wyszukiwanie sieci, które mogą działać w pobliżu stacji. Mechanizm ten w pewnym sensie zastępuje szukanie przewodu gniazdka sieciowego, umożliwiającego połączenie przewodowe.

Podczas przeglądania sieci można ograniczyć poprzez określenie parametrów, między innymi takich jak:

- rodzaj sieci – IBSS, BSS lub oba rodzaje,
- rodzaj identyfikatora – indywidualny (sieć zamknięta) lub rozgłaszany (sieć dla otwartego dostępu),
- identyfikator – pozwala znaleźć konkretną sieć,
- przeglądanie – aktywne lub pasywne,
- lista kanałów do przeglądania.

Przeglądanie pasywne polega na nasłuchiwaniu określonych kanałów i dekodowaniu odebranych ramek „radiolatarni". Ramki te są zapamiętywane, gdyż zawierają parametry istotne dla dalszej części procesu łączenia z siecią. Przeglądanie aktywne polega natomiast na wysyłaniu przez stację ramek Probe Request kolejno w każdym kanale. Ramki te są wysyłane zgodnie z zasadami protokołu DCF i mogą być kierowane do wszystkich sieci bądź tylko do pewnych konkretnych sieci, zależnie od zawartości pól adresowych ramki. Następnie stacja przez określony czas zbiera wszystkie ramki Probe Response, wysłane w danym kanale. Prawidłowy odbiór każdej ramki Probe Response powinien zostać potwierdzony ramką Ack.

Niezależnie od sposobu przeglądania sieci wynikiem jego jest lista sieci dostępnych w danej lokalizacji wraz z niezbędnym parametrami, określającymi m. in. odstęp czasowy

między ramkami „radiolatarni”, parametry warstwy fizycznej oraz listę obsługiwanych przez sieć prędkości transmisji. Lista dostępnych sieci jest przedstawiana użytkownikowi celem dokonania wyboru sieci. Można jednak skonfigurować adapter sieciowy, tak że połączenie z określoną siecią następuje automatycznie, bez ingerencji użytkownika.

3.2.5.2. Uwierzytelnianie stacji

Uwierzytelnianie ma na celu dopuszczenie do pracy w sieci tylko uprawnionych stacji bądź użytkowników. Konieczność wprowadzenia takiego mechanizmu jest podyktowana propagacją fal radiowych. W przeciwieństwie do sieci przewodowych możliwość podłączenia do sieci bezprzewodowej występuje także w miejscach niezamierzonych, np. ogólnodostępnych. Stwarza to ryzyko podłączenia do sieci nieupoważnionych urządzeń lub użytkowników.

W początkowej wersji standardu określono dwie metody uwierzytelniania: otwarte (ang. *open*) oraz ze współdzielonym kluczem (ang. *shared key*). Pierwsza z metod jest przeznaczona dla sieci „otwartych”, tj. takich, z których może korzystać każdy. Druga natomiast umożliwia ograniczenie dostępu do wybranej grupy urządzeń.

W przypadku uwierzytelniania otwartego klient wysyła prośbę o uwierzytelnienie (ramka Authentication), która jest przyjmowana przez punkt dostępowy bez żadnych zastrzeżeń. Zatem, każda próbująca stacja do sieci uzyskuje zgodę na przyłączenie się do sieci.

Metoda uwierzytelniania ze współdzielonym kluczem jest bardziej złożona. Wymagane jest mianowicie, aby wszystkie stacje korzystały z szyfrowania WEP i posługiwały się tym samym kluczem. Podczas uwierzytelniania sprawdza się zgodność kluczy. Proces uwierzytelniania odbywa się w czterech etapach:

1. Klient wysyła do punktu dostępowego prośbę o uwierzytelnienie.
2. Punkt dostępowy przesyła ramkę zawierającą tekst wezwania (ang. *challenge text*) w postaci jawnej.
3. Klient odsyła tekst wezwania zaszyfrowany swoim kluczem WEP.
4. Jeśli odszyfrowanie tekstu wezwania kluczem WEP punktu dostępowego powiedzie się, klient zostaje uwierzytelniony.

Jak nietrudno zauważyć, przechwycenie ramek wymienianych między klientem a punktem dostępowym w drugim i trzecim etapie uwierzytelniania z kluczem współdzielonym umożliwia poznanie klucza przez osoby nieupoważnione. Klucz ten nie tylko może posłużyć do odszyfrowania informacji przesyłanej w sieci, lecz także do uwierzytelnienia stacji nieupoważnionych użytkowników. Większość dostępnych punktów dostępowych używa jednak dodatkowej metody uwierzytelniania, opartej na adresie MAC adaptera sieciowego. Pozwala to na utworzenie zarówno „czarnej listy”, zawierającej adresy kart niedopuszczonych do pracy w danej sieci, jak i „białej listy”, zawierającej adresy jedynych kart dopuszczonych do pracy w danej sieci. Ponieważ ta metoda uwierzytelniania nie jest określona przez standard

802.11, punkty dostępowe mogą różnie reagować na próbę uwierzytelnienia stacji o niedozwolonym w danej sieci adresie MAC [91]. Przykładowo, punkt dostępu może odmówić uwierzytelnienia takiej stacji lub całkowicie zignorować ramki z niej pochodzące.

Ze względu na możliwość łatwego złamania klucza WEP obecnie zaleca się uwierzytelnianie oparte na protokole WPA. W przypadku małych sieci, składających się z kilku lub kilkunastu stacji klienckich, wystarcza wersja protokołu z kluczem współdzielonym (WPA-PSK, ang. *Pre-Shared Key*). Podobnie jak w przypadku WEP wymaga ona ręcznej dystrybucji kluczy w sieci, co w przypadku większej liczby klientów może być uciążliwe. Warto jednak zauważyć, że zapewnia ona znacznie wyższy poziom bezpieczeństwa niż WEP, ponieważ klucze podlegają okresowym wymianom. Ponadto, w miejsce algorytmu szyfrującego RC4, użytego w WEP, w WPA zastosowano znacznie skuteczniejszy i bezpieczniejszy algorytm AES. W rezultacie jedynym sposobem poznania klucza WPA jest atak słownikowy.

W przypadku większych sieci jest wskazane stosowanie uwierzytelniania z wykorzystaniem serwera RADIUS (ang. *Remote Authentication Dial In User Service*). Serwer ten stanowi centralną bazę danych nazw użytkowników i ich haseł. Dzięki temu wiele punktów dostępowych może uwierzytelniać klientów bez konieczności ręcznego wprowadzania informacji o nich do każdego punktu dostępowego. Do serwera RADIUS można kierować zapytania także z punktów dostępowych znajdujących się w dużej odległości od niego. Wydaje się, że to spostrzeżenie legło u podstaw projektu Eduroam [31]. Projekt ten zakłada, że użytkownik sieci bezprzewodowej danej organizacji może korzystać z sieci innej organizacji będącej członkiem tego projektu przy użyciu tej samej nazwy użytkownika i hasła co w jednostce macierzystej.

3.2.5.3. Kojarzenie stacji z siecią

Kojarzenie (ang. *association*) stacji z siecią polega na przypisaniu tej stacji identyfikatora skojarzenia (AID, ang. *Association ID*) przez punkt dostępu. Przypisanie to odbywa się na prośbę stacji, która przesyła ramkę Association Request. Ramka ta zawiera elementy informacyjne, określające między innymi:

- prędkości transmisji, obsługiwane przez stację,
- identyfikator sieci (SSID),
- długość okresu nasłuchiwania (ang. *listen interval*), używanego przez stacje w celu oszczędzania energii.

Podczas pracy w środowisku zawierającym wiele sieci bezprzewodowych pojedynczy kanał transmisyjny może być używany przez więcej niż jedną sieć. Aby uniknąć kojarzenia stacji z niewłaściwą siecią, punkt dostępu zwykle ignoruje żądania skojarzenia, zawierające identyfikator sieci inny niż ustalony w punkcie dostępu.

Jeśli żądanie skojarzenia może zostać spełnione, punkt dostępu odpowiada ramką Associattion Response. Ramka ta zawiera elementy informacyjne, określające:

- prędkości transmisji, obsługiwane przez punkt dostępu,
- identyfikator skojarzenia (AID),
- kod wyniku operacji (ang. *status code*) wraz z podaniem ewentualnej przyczyny odmowy skojarzenia (ang. *reason code*).

3.2.6. Powstające rozszerzenia standardu IEEE 802.11

Standard 802.11 podlega ciągłym modyfikacjom. Za najważniejsze spośród nowych rozwiązań można uważać rozszerzenie standardu – 802.11n [54]. Proponuje się w nim nową warstwę fizyczną, opartą na transmisji MIMO-OFDM (ang. *Multiple Input Multiple Output OFDM*), pozwalającą osiągnąć prędkości transmisji sięgające 600 Mb/s, a więc około 10-krotnie wyższe niż osiągane obecnie. Aby efektywnie wykorzystać tę prędkość i uzyskać odpowiednio wysoką efektywną przepustowość sieci, wprowadza się także szereg modyfikacji na poziomie warstwy liniowej. Podstawową metodą, służącą podniesieniu sprawności transmisji, jest agregacja (ang. *aggregation*), czyli łączenie ramek w ciąg poprzedzony pojedynczym, wspólnym nagłówkiem warstwy fizycznej. Agregację można zatem, w pewnym uproszczeniu, uznać za rozwinięcie koncepcji potwierdzenia blokowego.

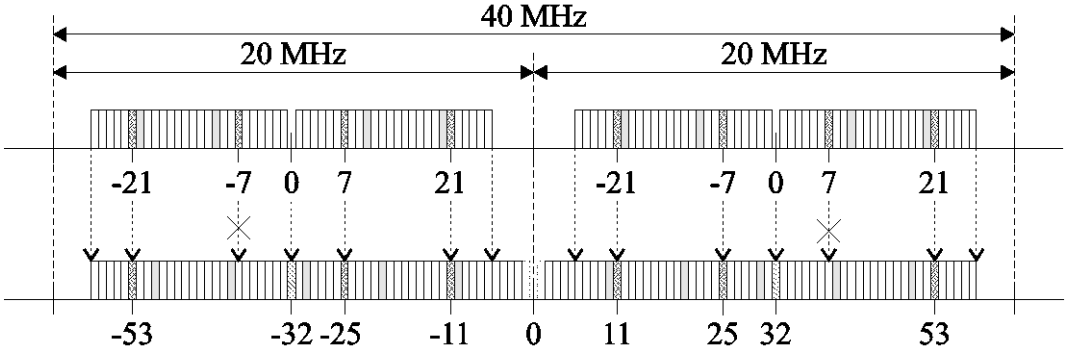
Prace nad rozszerzeniem 802.11n trwały wiele lat i jeszcze przed jego ostatecznym zatwierdzeniem na rynku było dostępnych wiele urządzeń zgodnych z jego projektem. Nie było jednak pewności, czy przyjęte w projekcie rozwiązania będą wystarczająco zgodne z wersją końcową standardu, aby wystarczyła aktualizacja sterowników i oprogramowania sterującego pracą urządzeń [40].

3.2.6.1. Warstwa fizyczna MIMO-OFDM (HT)

Proponowana w standardzie 802.11n warstwa HT (ang. *High Throughput*) wykorzystuje technikę transmisji MIMO-OFDM. Polega ona na równoległym użyciu niezależnych strumieni fal radiowych, tzw. strumieni przestrzennych (ang. *Spatial Stream*), przesyłanych za pomocą osobnych anten. Pojedynczą ramkę można wówczas rozbić na części, z których każda jest przesyłana innym strumieniem. Pozwala to na znaczne podniesienie prędkości transmisji. Dodatkowym czynnikiem, wpływającym na przyspieszenie transmisji, jest dwukrotne zwiększenie szerokości kanału w stosunku do używanej w warstwie fizycznej OFDM. Przykładowo, użycie kanału o szerokości 40 MHz i czterech strumieni fal radiowych pozwala na osiągnięcie prędkości transmisji do 600 Mb/s. Pomimo wprowadzonych modyfikacji warstwa fizyczna zachowuje wsteczną kompatybilność z warstwami OFDM, HR-DSSS i ERP.

Warstwa HT jest w dużej części rozwinięciem warstwy OFDM, wprowadzonej w standardzie 802.11a. W zależności od potrzeb i możliwości można użyć jednej z czterech metod mo-

dulacji (BPSK, QPSK, 16-QAM lub 64-QAM) oraz sprawności kodera konwolucyjnego (½, ⅔, ¾ lub $^5/_6$). Sieć może działać zarówno w pasmie 2,4 GHz, jak i 5 GHz, a szerokość kanału może wynosić 20 lub 40 MHz. Dla zapewnienia wstecznej zgodności z sieciami 802.11a/g warstwa HT obsługuje podział kanału 20 MHz na 52 podnośne, w tym 4 pilotujące (ang. *pilot carrier*) i 48 przeznaczonych dla transmisji danych. Przez „sklejenie" dwóch sąsiednich kanałów uzyskuje się kanał o szerokości 40 MHz. Kanał taki jest podzielony na 128 podnośnych, w tym 6 pilotujących i 108 dla transmisji danych. Efektywna pojemność szerszego kanału jest zatem 2,25-krotnie większa niż kanału węższego. Wynika to z możliwości użycia kilku dodatkowych podnośnych w pobliżu granicy kanałów [40]. Jeśli łączone kanały nie są sąsiednie, dostępnych dla transmisji jest tylko 96 podnośnych. Organizację 20 MHz i 40 MHz kanału w projekcie standardu ilustruje rys. 3.32.



Rys. 3.32. Organizacja kanału 20 MHz i 40 MHz w standardzie 802.11n [54]
Fig. 3.32. Organisation of 20 MHz and 40 MHz channel in 802.11n standard

Czas integracji FFT, odpowiadający czasowi trwania symbolu, wynosi 3,2 μs i jest taki sam jak w standardach 802.11a i 802.11g. Przy użyciu przedziału ochronnego (GI, ang. *Guard Intereval*) o czasie trwania 800 ns uzyskuje się prędkość modulacji 250000 bd. Opcjonalne skrócenie przedziału ochronnego o połowę zwiększa tę prędkość do 312500 bd.

W projekcie standardu 802.11n określono kilkadziesiąt schematów modulacji i kodowania (MCS, ang. *Modulation and Coding Scheme*). Znaczenie i wartości parametrów, opisujących właściwości tych schematów, zawarto w tabeli 3.8.

Wynikową prędkość transmisji można obliczyć według wzoru:

$$R_{wl} = V \cdot N_{DBPS} = \frac{R \cdot N_{SS} \cdot N_{SC} \cdot N_{BPSC}}{T_{DFT} + T_{GI}}. \tag{3.3}$$

Schematy o numerach 0÷7, określone dla kanałów o szerokości 20 MHz i pojedynczego strumienia fal radiowych, są obowiązkowe. W przypadku punktów dostępu jest konieczna także obsługa schematów o numerach 8÷15. Pozwalają one na dwukrotnie większą prędkość transmisji niż odpowiadające im schematy 0÷7. Uzyskano ją dzięki obsłudze dwóch strumieni fal radiowych. Parametry schematów 0÷15 dla kanałów 20 MHz oraz kilku przykładowych innych schematów zebrano w tabeli 3.9.

Tabela 3.8

Zależności między parametrami modulacji OFDM

Parametr	Opis	Wartość
R	sprawność kodera konwolucyjnego	½, ⅔, ¾, $^5/_6$
N_{SS}	liczba strumieni przestrzennych	1÷4
N_{BPSC}	liczba zakodowanych bitów na podnośną (dla 1 strumienia przestrzennego)	zależne od modulacji
N_{SC}	liczba podnośnych przeznaczonych do transmisji	48, 52, 96 lub 108
N_{CBPS}	liczba zakodowanych bitów na symbol	$N_{SS} \cdot N_{SC} \cdot N_{BPSC}$
N_{DBPS}	liczba bitów danych na symbol	$R \cdot N_{CBPS}$
T_{FIT}	czas integracji FFT (ang. *FFT Integration Time*)	3,2 μs
T_{GI}	czas trwania okresu ochronnego (ang. *Guard Interval*)	0,4 μs lub 0,8 μs
V	prędkość modulacji	$1/(T_{FIT}+T_{GI})$

Tabela 3.9

Wybrane schematy modulacji i kodowania w standardzie 802.11n [54]

Schemat	Modulacja	N_{BPSC}	N_{SS}	N_{SC}	N_{CBPS}	R	N_{DBPS}	R_{wl} [Mb/s]	
								T_{GI}=800 ns	T_{GI}=400 ns
0	BPSK	1	1	52	52	½	26	6,5	7,2
1	QPSK	2			104	½	52	13,0	14,4
2						¾	78	19,5	21,7
3	16-QAM	4			208	½	104	26,0	28,9
4						¾	156	39,0	43,3
5	64-QAM	6			312	⅔	208	52,0	57,8
6						¾	234	58,5	65,0
7						5/6	260	65,0	72,2
8	BPSK	1	2	52	104	½	52	13,0	14,4
9	QPSK	2			208	½	104	26,0	28,9
10						¾	156	39,0	43,3
11	16-QAM	4			416	½	208	52,0	57,8
12						¾	312	78,0	86,7
13	64-QAM	6			624	⅔	416	104,0	115,6
14						¾	468	117,0	130,0
15						5/6	520	130,0	144,4
7	64-QAM	6	1	108	648	5/6	540	135,0	150,0
15			2		1296		1080	270,0	300,0
23			3		1944		1620	405,0	450,0
31			4		2592		2160	540,0	600,0
32	BPSK	1	1	48	48	½	24	6,0	6,7

Warto zauważyć, że niektóre schematy określone dla kanałów 40 MHz (N_{SC}=108) i 20 MHz (N_{SC}=52) mają taki sam numer. Podstawowe schematy MCS charakteryzują się

jednakowymi parametrami dla poszczególnych strumieni przestrzennych i należą do grupy EQM (ang. *Equal Modulation*). Określono jednak także schematy, umożliwiające transmisję z różnymi parametrami w każdym strumieniu. Należą one do grupy UEQM (ang. *Unequal Modulation*). Wydaje się, że przyjęcie różnych parametrów w poszczególnych strumieniach pozwala w pewnych sytuacjach lepiej uwzględnić charakterystykę kanału transmisyjnego.

Na poziomie podwarstwy PLCP są zdefiniowane trzy formaty ramek. Pierwszy z nich wywodzi się z warstw fizycznych OFDM (802.11a) i ERP (802.11g) i ma format przedstawiony odpowiednio na rys. 3.10 oraz 3.11. W drugim – mieszanym (ang. *HT-mixed*) – ramka dzieli się na dwie części. Format preambuły umożliwia zdekodowanie jej przez stacje, nieobsługujące podwyższonych prędkości transmisji. Pozostała część ramki może zostać odczytana tylko przez stacje, które obsługują te prędkości. W trzecim formacie (ang. *HT-greenfield*) ramka zawiera wyłącznie elementy przeznaczone dla stacji obsługujących podwyższone prędkości transmisji. Dla zachowania kompatybilności wstecznej z istniejącymi sieciami 802.11a/g obsługa pierwszych dwóch formatów jest obowiązkowa, trzeciego natomiast – opcjonalna. Wymienione formaty ramek PLCP pokazano na rys. 3.33.

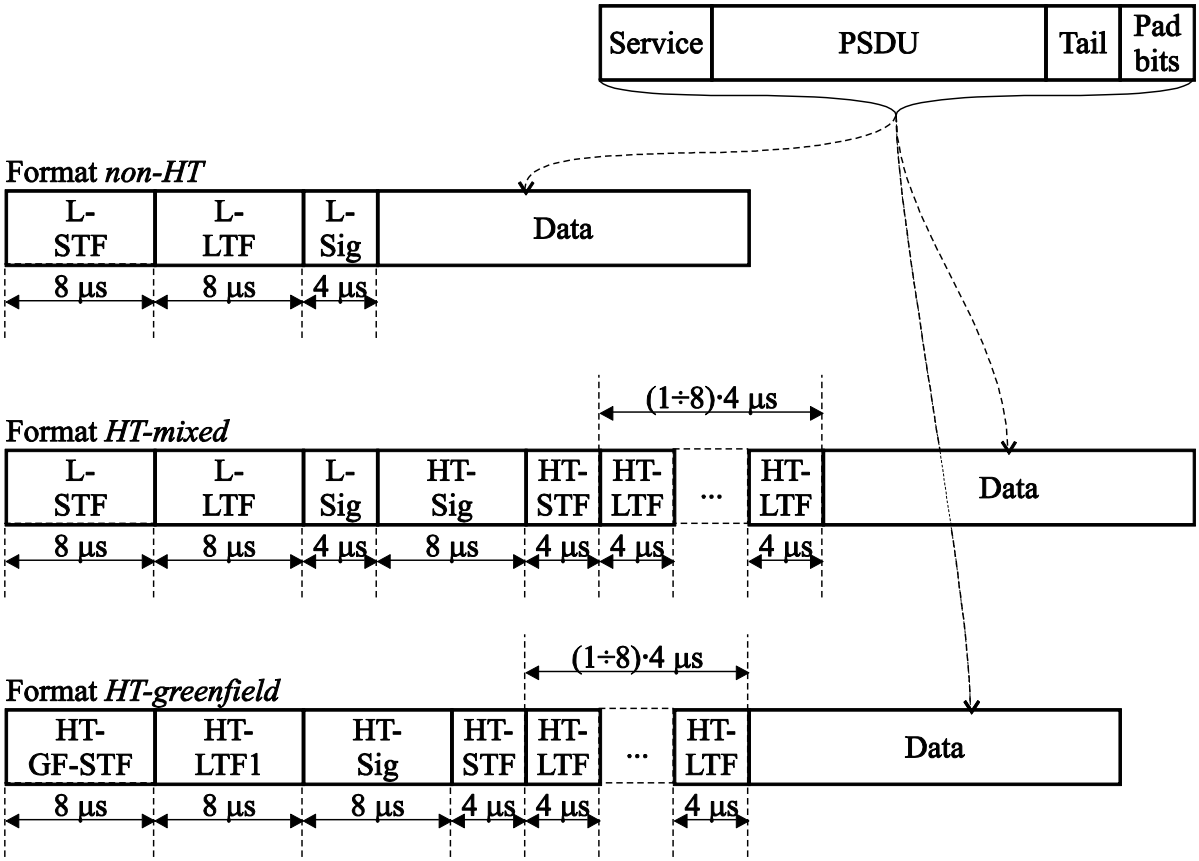
Ramka formatu mieszanego rozpoczyna się preambułą, zawierającą krótkie i długie pola treningowe (L-STF, ang. *Legacy Short Training Field* oraz L-LTF, ang. *Legacy Long Training Field*) oraz pole sygnału (L-Sig, ang. *Legacy Signal*). Format tych pól jest taki sam jak w warstwie OFDM (por. rozdział 3.2.2.6 i rys. 3.10). Z kolei występuje pole HT-Sig (ang. *High Throughput Signal*) oraz krótkie i długie pola treningowe (HT-STF, ang. *High Throughput Short Training Field* oraz HT-LTF, ang. *High Throughput Long Training Field*). Pola te nie mogą być odczytane przez stacje, nieobsługujące podwyższonych prędkości transmisji. Pole HT-Sig składa się z dwóch części po 24 bity każda i określa m. in.:

- numer schematu modulacji i kodowania (MCS), użytego do przesłania części ramki zawierającej dane,
- szerokość kanału (ang. *Channel Bandwidth*) – 20 lub 40 MHz,
- długość jednostki PSDU, z zakresu 0÷65535 B,
- użycie agregacji (łączenia) ramek,
- liczbę strumieni przestrzennych (1÷4),
- rodzaj użytego kodera FEC,
- długość przedziału ochronnego (długi lub krótki).

Wszystkie elementy pola HT-Sig są chronione 8-bitową sumą kontrolną CRC, wytwarzaną za pomocą wielomianu $G(x) = x^8 + x^2 + x + 1$. Pole HT-Sig kończy się bitami wypełniającymi, koniecznymi dla pozostawienia kodera splotowego w określonym stanie.

Krótkie pole treningowe (HT-STF) umożliwia optymalizację parametrów automatycznej regulacji wzmocnienia (AGC, ang. *Automatic Gain Control*) systemu MIMO. Z kolei długie

pola treningowe (HT-LTF) można podzielić na dwie grupy. Pierwsza z nich – obowiązkowa – zawiera 1÷4 pola, umożliwiające demodulację danych, przesyłanych z podwyższonymi prędkościami. Druga natomiast – opcjonalna – zawiera 0÷4 pola i służy do próbkowania właściwości kanału fizycznego.



Rys. 3.33. Formaty ramki podwarstwy PLCP w warstwie fizycznej HT [54]
Fig. 3.33. PLCP frame formats in HT physical layer

Ramka formatu *greenfield* rozpoczyna się od dwóch pól treningowych – krótkiego (HT-GF-STF, ang. *High Throughput Greenfield Short Training Field*) oraz długiego (HT-LTF1, ang. *High Throughput First Long Training Field*). Pozostałe składniki ramki PLCP są analogiczne jak w ramce formatu mieszanego. Jak nietrudno zauważyć, ramka formatu mieszanego jest dłuższa tylko o 4 µs od ramki formatu *greenfield*, a umożliwia pracę w sieci, zawierającej stacje starszego typu.

Pole danych jest kodowane przy użyciu kodera konwolucyjnego (BCC, ang. *Binary Convolutional Code*), stosowanego także w warstwach fizycznych OFDM i ERP, bądź kodera LDPC (ang. *Low Density Parity Check*). W przypadku prędkości powyżej 300 Mb/s stosuje się podwójny koder BCC, natomiast koder LDPC jest zawsze pojedynczy. W przypadku kodowania BCC pole danych zawiera liczbę symboli modulacji równą

$$N_{SYM} = m_{STBC}\left\lceil \frac{8L + 16 + 6N_{ES}}{m_{STBC}N_{DBPS}} \right\rceil, \tag{3.4}$$

gdzie L_D – długość PSDU, N_{ES} – liczba strumieni kodera BCC (1 lub 2), N_{DBPS} – liczba bitów na symbol modulacji, natomiast m_{STBC} przyjmuje wartość 2, jeśli użyto kodowania STBC (ang. *Space Time Block Code*) i 1 w przeciwnym przypadku. Jeśli natomiast użyto kodowania LDPC, można przyjąć N_{ES}=0.

3.2.6.2. Agregacja ramek

Agregacja (ang. *aggregation*), czyli łączenie ramek ma na celu redukcję narzutu wprowadzanego przez warstwę fizyczną [116]. Ponieważ format ramki na poziomie podwarstwy PLCP jest ustalony, jedyna możliwość zmniejszenia narzutu to użycie pojedynczego nagłówka warstwy fizycznej (wraz z preambułą) dla większej liczby ramek warstwy liniowej. Jest to istotne o tyle, że narzut warstwy fizycznej rośnie wraz ze wzrostem prędkości transmisji, gdyż początkowe elementy ramki podwarstwy PLCP są przesyłane zawsze z najniższą prędkością określoną dla danej warstwy. Można zatem powiedzieć, że czas transmisji preambuły i nagłówka jest stały, natomiast czas transmisji jednostki PSDU zmniejsza się wraz ze wzrostem prędkości transmisji. W związku z tym zwiększa się narzut protokołu, a jego efektywność maleje. Aby tego uniknąć, w projekcie standardu 802.11n proponuje się dwie metody agregacji ramek: A-MSDU oraz A-MPDU.

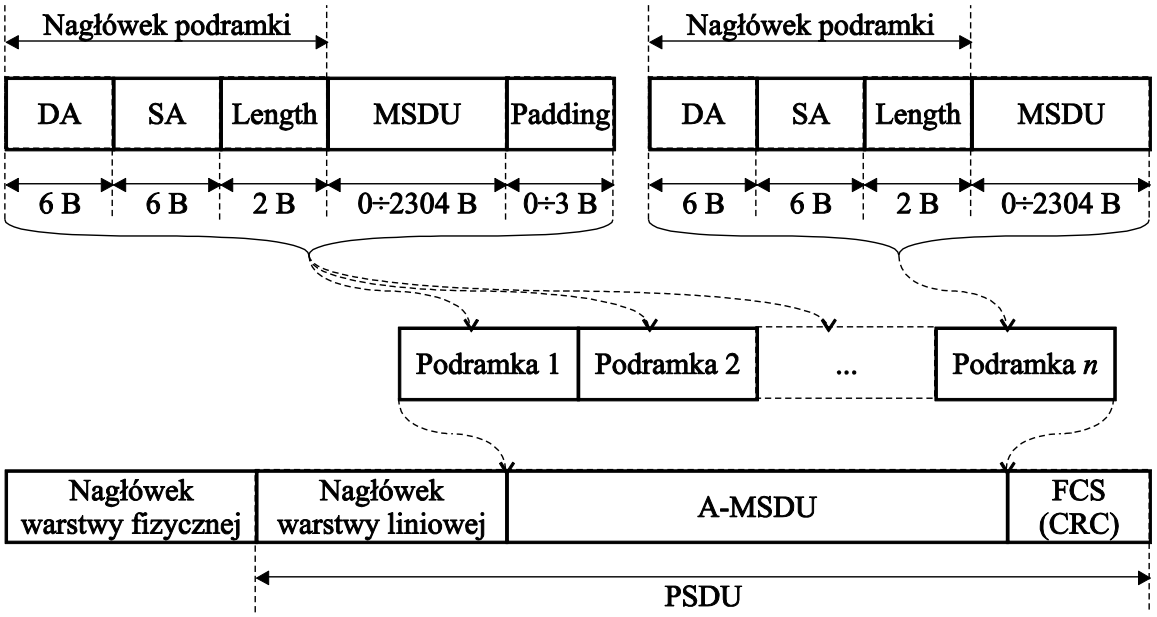
Agregacja A-MSDU polega na połączeniu wielu jednostek MSDU, kierowanych do jednego odbiorcy, w pojedynczą jednostkę MPDU. Operacja taka może znacznie zwiększyć wydajność sieci, szczególnie gdy przesyłane jednostki MSDU są niewielkich rozmiarów. Pakiety przychodzące z wyższej warstwy są zbierane i umieszczane w buforze aż do osiągnięcia maksymalnego rozmiaru łączonej ramki lub upłynięcia określonego czasu oczekiwania. Zależnie od możliwości stacji maksymalny rozmiar rdzenia łączonej ramki wynosi 3839 B lub 7935 B, tj. o 256 B mniej niż limit wielkości PSDU (odpowiednio 4095 lub 8191 B). Limit czasu oczekiwania zwykle wynosi 1 μs, ale może być ustawiony niezależnie dla każdej z czterech kategorii dostępu [95, 117].

Ramka A-MSDU zawiera ciąg podramek (ang. *subframe*), składających się z nagłówka, MSDU oraz bajtów rozciągających (ang. *padding*). Nagłówek podramki określa jej adres docelowy (DA) i źródłowy (SA) oraz długość (ang. *Length*) MSDU. Długość podramki – z wyjątkiem ostatniej – powinna być całkowitą wielokrotnością 4 B; w tym celu stosuje się bajty rozciągające. Ciąg podramek poprzedza się standardowym nagłówkiem warstwy liniowej. Strukturę ramki A-MSDU oraz sposób jej tworzenia pokazano na rys. 3.34.

Ramkę A-MSDU przesyła się, używając ramek typu QoS Data z opcjonalnymi wskazaniami CF-Ack i CF-Poll.

Agregacja A-MSDU wprowadza pewne ograniczenia:

- wszystkie podramki muszą mieć ten sam identyfikator ruchu (TID),
- adresy docelowy (DA) i źródłowy (SA) w podramkach muszą odpowiadać adresom nadawcy (TA) i odbiorcy (RA) w nagłówku warstwy liniowej,
- transmisje wieloadresowe i rozgłoszeniowe nie są możliwe.



Rys. 3.34. Struktura ramki A-MSDU w projekcie standardu 802.11n [54]
Fig. 3.34. A-MSDU frame structure in 802.11 draft standard

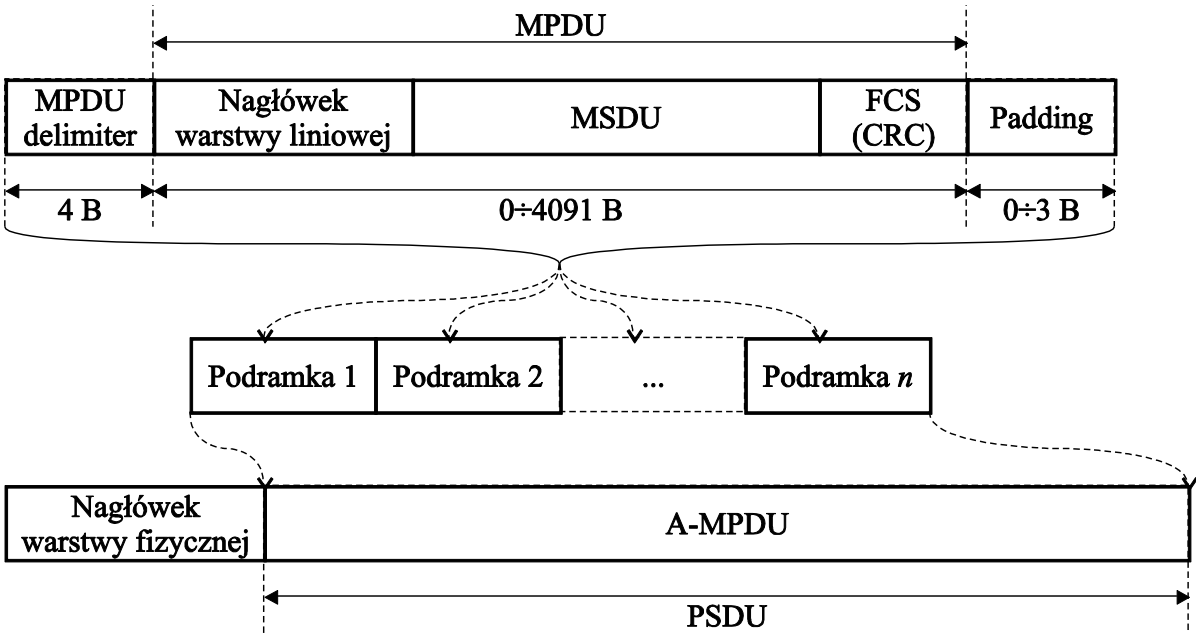
Poważną wadą agregacji A-MSDU jest nieefektywna reakcja na błędy transmisji [117]. Ponieważ wszystkie podramki są połączone w pojedynczą ramkę warstwy liniowej o jednym numerze sekwencyjnym, przekłamanie tylko jednej podramki powoduje konieczność retransmisji całej jednostki MPDU, zawierającej wszystkie podramki.

Agregacja A-MPDU polega na połączeniu wielu jednostek MPDU i poprzedzeniu tak uzyskanego ciągu pojedynczym nagłówkiem warstwy fizycznej. Ponieważ każda jednostka MPDU jest już przetworzona przez warstwę fizyczną, poszczególne ramki mogą należeć do różnych strumieni ruchu. Pomimo to wymaga się, aby wszystkie połączone ramki były kierowane do jednego odbiorcy. Nie określono także czasu oczekiwania na sformowanie połączonej ramki, tak więc liczba jednostek MPDU podlegających łączeniu zależy tylko od liczby ramek znajdujących się w kolejce [117]. Liczba ta nie może jednak przekroczyć 64 ze względu na możliwości procedury potwierdzenia blokowego. Maksymalna długość jednostki A-MPDU wynosi 65535 B – tyle co limit długości PSDU w warstwie fizycznej. Długość ta może jednak być dodatkowo ograniczona przez możliwości stacji, w szczególności zaś przez pojemność bufora odbiorcy. Pojemność ta może wynosić 2^i-1 B, gdzie i=13÷16. Z kolei, rozmiar pojedynczej jednostki MPDU nie może przekroczyć 4095 B [117], nawet gdy nadawca i odbiorca mogą przesyłać ramki o długości do 7935 B.

Każda składowa jednostka MPDU jest poprzedzona ogranicznikiem (ang. *MPDU delimiter*), określającym długość MPDU, 8-bitową sumę kontrolną i sygnaturę. Zadaniem sygnatury jest ułatwienie odnalezienia kolejnego ogranicznika, jeśli poprzedni został przekłamany. Obecność sumy kontrolnej zmniejsza ryzyko potraktowania przypadkowego fragmentu ramki jako ogranicznika, tym bardziej że długość MPDU musi być wielokrotnością 4 B. Aby spełnić ten warunek, za każdą jednostką MPDU – prócz ostatniej – występują 0÷3 bajty wypełniające (ang. *padding*). Strukturę ramki A-MPDU oraz sposób jej tworzenia wyjaśniono na rys. 3.35.

W przypadku przesyłania krótkich jednostek MPDU z dużymi prędkościami czas transmisji każdej z nich może być krótszy od minimalnego czasu pomiędzy kolejnymi jednostkami (ang. *Minimum MPDU Start Spacing*). W takim przypadku w połączonej ramce mogą wystąpić ograniczniki, w których długość MPDU wynosi 0. Oczywiście, po takim ograniczniku nie występuje jednostka MPDU, tylko kolejny ogranicznik.

Ramka A-MPDU jest potwierdzana ramką potwierdzenia blokowego BlockAck. Mechanizm ten jest jednak realizowany w sposób nieco zmodyfikowany. Mianowicie, nie jest konieczne uzgodnienie potwierdzenia blokowego ani jawne wywołanie potwierdzenia ramką BlockAckReq. Dodatkowo, ponieważ agregacja A-MPDU wyklucza użycie fragmentacji, mapa bitowa stanu odebranych ramek w ramce BlockAck może zostać zmniejszona do 8 B.



Rys. 3.35. Struktura ramki A-MPDU w projekcie standardu 802.11n [54]
Fig. 3.35. A-MPDU frame structure in 802.11 draft standard

3.2.6.3. Pozostałe zmiany w warstwie liniowej

W projekcie standardu wprowadzono wiele modyfikacji na poziomie warstwy liniowej. Są one konieczne dla efektywnego wykorzystania wysokich prędkości transmisji. Do najważniejszych zmian można zaliczyć:

- wprowadzenie zredukowanego odstępu międzyramkowego RIFS (ang. *Reduced Inter-Frame Space*),
- wprowadzenie protokołu transmisji w przeciwnym kierunku (ang. *Reverse Direction*),
- modyfikację ogólnego formatu ramki (por. rys. 3.12):
  - wprowadzenie 4-bajtowego pola HT Control,
  - zwiększenie długości pola danych ramki do 7955 B.

Odstęp międzyramkowy RIFS można zastosować zamiast SIFS w celu rozdzielenia ramek wysyłanych przez jednego nadawcę, gdy nie oczekuje się odpowiedzi, która powinna nastąpić po czasie SIFS. Ramki powinny też być kierowane do jednego odbiorcy, chyba że zastosowano mechanizm PSMP (ang. *Power Save Multi-Poll*), którego zadaniem jest sterowanie trybem uśpienia w stacjach klienckich. Odstęp RIFS można zastosować tylko podczas transmisji z podwyższonymi prędkościami transmisji. W standardzie 802.11n wartość RIFS wynosi 2 μs, podczas gdy SIFS – 16 μs.

3.3. Analiza wydajności standardu IEEE 802.11

Od samego początku swojego istnienia standard IEEE 802.11 jest przedmiotem intensywnych badań, mających na celu ocenę jego wydajności. Spośród wielu publikacji, dotyczących analizy wydajności standardu IEEE 802.11, za jedne z ważniejszych można uważać prace [12, 13]. Przeprowadzono w nich analizę pod kątem przepustowości w warunkach nasycenia sieci, tzn. dla największego obciążenia, przy którym może ona jeszcze pracować stabilnie. W tym celu zaproponowano uproszczony model oparty na łańcuchach Markowa, a umożliwiający ocenę wydajności protokołu DCF. Pomimo przyjęcia upraszczającego założenia, że prawdopodobieństwo kolizji jest równe i niezmienne dla wszystkich stacji, model wykazuje dużą dokładność, szczególnie dla większych sieci. Model ten był dalej rozwijany, przykładowo w pracy [156] zaproponowano model dokładniejszy – bez upraszczających założeń odnośnie do prawdopodobieństwa kolizji. Niestety, charakteryzuje się on dużą złożonością. Wymienione modele opisują raczej działanie pojedynczych stacji, wchodzących w skład sieci. Natomiast w pracy [20] zaproponowano inne podejście, polegające na stworzeniu modelu całej sieci, składającej się z pewnej liczby stacji.

Wydajność protokołu można ocenić też w inny sposób [85, 115]. W pracach tych wyprowadzono zależności analityczne, pozwalające obliczyć m. in. efektywną prędkość transmisji. Rozpoczynając analizę od podobnych założeń, prace te podążają jednak w różnych kierun-

kach. I tak, w [85] oszacowano efektywną prędkość transmisji z uwzględnieniem możliwości wystąpienia błędów transmisji (ang. *goodput*), ale bez kolizji między nadającymi stacjami. Natomiast w [115] wykazano, że dla standardu IEEE 802.11 istnieje górna granica przepustowości (TUL, ang. *Throughput Upper Limit*) oraz dolna granica opóźnienia (DLL, ang. *Delay Lower Limit*). Ponieważ granice te wynikają z zasady wymiany ramek, zwiększanie prędkości transmisji w pewnym momencie nie powoduje już podniesienia przepustowości i zmniejszenia opóźnień. Dlatego też w celu uzyskania poprawy parametrów użytkowych sieci konieczna jest modyfikacja zasad wymiany ramek. Ostatnią z wymienionych metod zastosowano w dalszych rozważaniach dla porównania wydajności standardu IEEE 802.11 na poziomie warstwy liniowej dla różnych warstw fizycznych i różnych sposobów wymiany informacji, w tym potwierdzenia blokowego i agregacji ramek.

3.3.1. Uwagi wstępne

Jedną z ważniejszych cech wszystkich warstw fizycznych standardu 802.11 jest możliwość prowadzenia transmisji z różnymi prędkościami. Pozwala to użyć prędkości optymalnej dla danych warunków, co jest szczególnie istotne dla sieci bezprzewodowych ze względu na ich zmienność w czasie. Prędkość transmisji warstwy fizycznej nie może jednak być widoczna dla warstw wyższych, konieczne jest zatem wprowadzenie odpowiedniej sygnalizacji na poziomie warstwy fizycznej. Niestety, zwiększa to narzut protokołu. Dodatkowy narzut powodują także pola synchronizujące, niezbędne dla transmisji radiowej.

Warto zauważyć, że długość preambuły i nagłówka warstwy fizycznej oraz czasy DIFS i SIFS zależą od wariantu tej warstwy. Tym niemniej preambuła i nagłówek są przesyłane zwykle z najniższą prędkością transmisji określoną dla danej warstwy fizycznej (zwykle 1, 2 lub 6 Mb/s), aby jak najwięcej stacji mogło prawidłowo odebrać zawarte tam informacje. Elementy pochodzące z warstwy liniowej przesyła się natomiast z najwyższą prędkością możliwą do uzyskania w danych warunkach. Dla potrzeb transmisji potwierdzeń wykorzystuje się tzw. prędkości podstawowe, nieprzekraczające prędkości, z jaką odebrano potwierdzane właśnie dane (np. ramka Data – 54 Mb/s, Ack – 24 Mb/s).

W celu zmniejszenia narzutu warstwy fizycznej przy użyciu wyższych prędkości transmisji, wprowadzono tzw. „krótką preambułę”. W tym przypadku sama preambuła jest rzeczywiście dwukrotnie krótsza od „długiej”, a dodatkowo nagłówek jest przesyłany dwukrotnie szybciej. W sumie narzut warstwy fizycznej maleje o połowę.

W tabeli 3.10 zebrano parametry zależne od warstwy fizycznej, niezbędne dla dalszych rozważań i oszacowania wydajności protokołu.

W celu wyprowadzenia potrzebnych zależności użyto sposobu podobnego do opisanego w pracy [85]. Zakłada on transmisję w warunkach idealnych w celu wyznaczenia górnej granicy przepustowości (TUL, ang. *Throughput Upper Limit*). Przyjęto następujące założenia:

- sieć składa się z dwóch stacji, a w czasie transmisji nie występują kolizje ani inne błędy transmisji – nie ma więc retransmisji,
- wymiana danych odbywa się na poziomie warstwy liniowej, zatem nie uwzględnia się narzutu warstw wyższych, jak np. TCP/IP,
- nie używa się mechanizmów zabezpieczających,
- czas przetwarzania ramek można zaniedbać.

Ponadto załóżmy, iż:

- prędkość transmisji łącza radiowego podczas transmisji ramek danych wynosi R_{wl}, zaś podczas transmisji potwierdzeń – R'_{wl};
- długość pola danych w ramce jest stała i wynosi L;
- długość ramki potwierdzenia (Ack) na poziomie warstwy liniowej wynosi 14 bajtów;
- długość ramki danych (Data) na poziomie warstwy liniowej wynosi 28+L bajtów.

Tabela 3.10

Parametry protokołu 802.11 zależne od warstwy fizycznej

Warstwa fizyczna	CWmin	CWmax	Czas [μs]				Dodatkowy narzut
			SIFS	Slot	Preambuła	Nagłówek	
DSSS	31	1023	10	20	144	48	
FHSS	15	1023	28	50	96	32	kodowanie 32/33
Ir (1 Mb/s)	63	1023	10	8	16	41	
Ir (2 Mb/s)			10	8	20	25	
HR-DSSS sp	31	1023	10	20	72	24	
OFDM	15	1023	16	9	20	4	≥22 bity
ERP-DSSS lp	15 / 31	1023	10	9 / 20	144	48	18 μs
ERP-DSSS sp	15 / 31	1023	10	9 / 20	72	24	18 μs
HT	15	1023	16	9	16	4	
HT-mixed					16	16÷40	
HT-GF					16	12÷36	

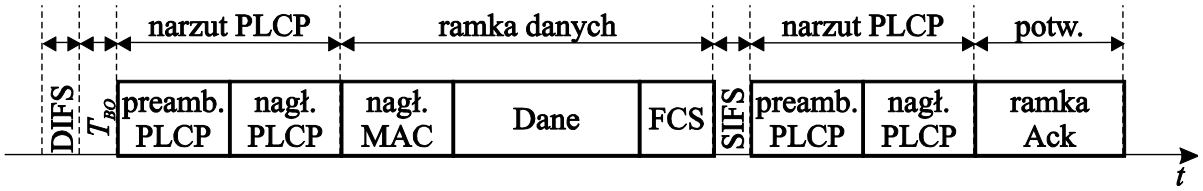
Przy każdej metodzie wymiany danych dla wszystkich wariantów warstwy fizycznej wydajność protokołu można wyrazić następująco:

$$\eta = \frac{L}{R_{wl} T_p}, \tag{3.5}$$

gdzie: L – objętość informacji przesyłana w ramach cyklu transmisyjnego [B], R_{wl} – prędkość transmisji łącza [b/s], a T_p – czas trwania cyklu transmisyjnego [s]. Cykl transmisyjny jest to niezmienny i powtarzający się fragment wymiany ramek, wystarczający dla potrzeb analizy wydajności protokołu.

3.3.2. Podstawowa wymiana informacji

W podstawowej metodzie wymiany informacji z użyciem protokołu DCF ramki danych (Data) i potwierdzeń (Ack) są przesyłane naprzemiennie. Każda ramka musi być poprzedzona preambułą i nagłówkiem warstwy fizycznej. Wymiana ramek może zatem przebiegać tak, jak pokazano na rys. 3.36.



Rys. 3.36. Przebieg transmisji przy użyciu metody podstawowej [135]
Fig. 3.36. Transmission course using basic method

Biorąc pod uwagę elementy wymiany ramek pokazane na rysunku 3.36, czas trwania cyklu transmisyjnego można określić jako [135, 138, 155]:

$$T_p = T_{DIFS} + T_{BO} + T_{SIFS} + 2 \cdot T_{PLCP} + T_{Data} + T_{Ack}, \tag{3.6}$$

gdzie: T_{DIFS} oraz T_{SIFS} określają czas trwania okresów DIFS i SIFS, natomiast T_{PLCP} – czas trwania preambuły i nagłówka warstwy fizycznej. Wartości te są zdefiniowane w specyfikacji poszczególnych warstw fizycznych i zebrane w tabeli 3.10. T_{Data} i T_{Ack} oznaczają czas transmisji odpowiednio ramek danych i potwierdzenia i są określone zależnościami (3.8) i (3.9). Z kolei T_{BO} określa czas trwania okresu wycofywania (ang. *backoff period*), który przy założeniu idealnych warunków transmisji i zgodnie z wyjaśnieniami w pracy [85] można uprościć do postaci:

$$T_{BO} = \frac{CW_{min}}{2} \cdot T_{slot}. \tag{3.7}$$

T_{slot} określa czas trwania szczeliny [s] (tabela 3.10), natomiast CW_{min} (ang. *Contention Window*) określa minimalną liczbę szczelin okresu rywalizacji. Z kolei

$$T_{Data} = \frac{8 \cdot (28 + L)}{R_{wl}} \tag{3.8}$$

oraz

$$T_{Ack} = \frac{8 \cdot 14}{R'_{wl}}, \tag{3.9}$$

gdzie: L – pojemność pola danych ramki [B], R_{wl} – prędkość transmisji ramki danych [b/s], natomiast R'_{wl} – prędkość transmisji ramki potwierdzenia [b/s].

Podczas obliczania T_{Data} i T_{Ack} należy zwrócić uwagę na dodatkowy narzut wynikający z przyjętej metody modulacji, np. kodowanie 32/33 w warstwie 802.11 FHSS. Podobnie mo-

dulacja OFDM wprowadza dodatkowy narzut o wielkości co najmniej 22 bitów na każdą przesyłaną ramkę.

W czasie pojedynczego cyklu transmisyjnego przesłanych zostanie dokładnie *L* bajtów.

Przykładowe obliczenia wykonano dla następujących pojemności pola danych:

- 2304 B (największa dla standardu 802.11),
- 1500 B (największa dla sieci Ethernet, z którą 802.11 często współpracuje),
- 256 B (największa dla protokołu AX.25 używanego w sieci Packet Radio [6]),
- 48 B (zbliżona do rozmiaru komórki ATM oraz najmniejszej ramki sieci Ethernet [103]).

Dla warstw fizycznych DSSS, FHSS, Ir oraz HR-DSSS przyjęto, że ramki danych i potwierdzeń są przesyłane z tą samą prędkością. Natomiast dla warstw OFDM i ERP przyjęto, że potwierdzenia przesyła się z prędkością podstawową (6, 12 lub 24 Mb/s), nieprzekraczającą prędkości, z jaką odebrano potwierdzane dane. Takie zachowanie sieci potwierdzono testami z wykorzystaniem monitora łącza bezprzewodowego.

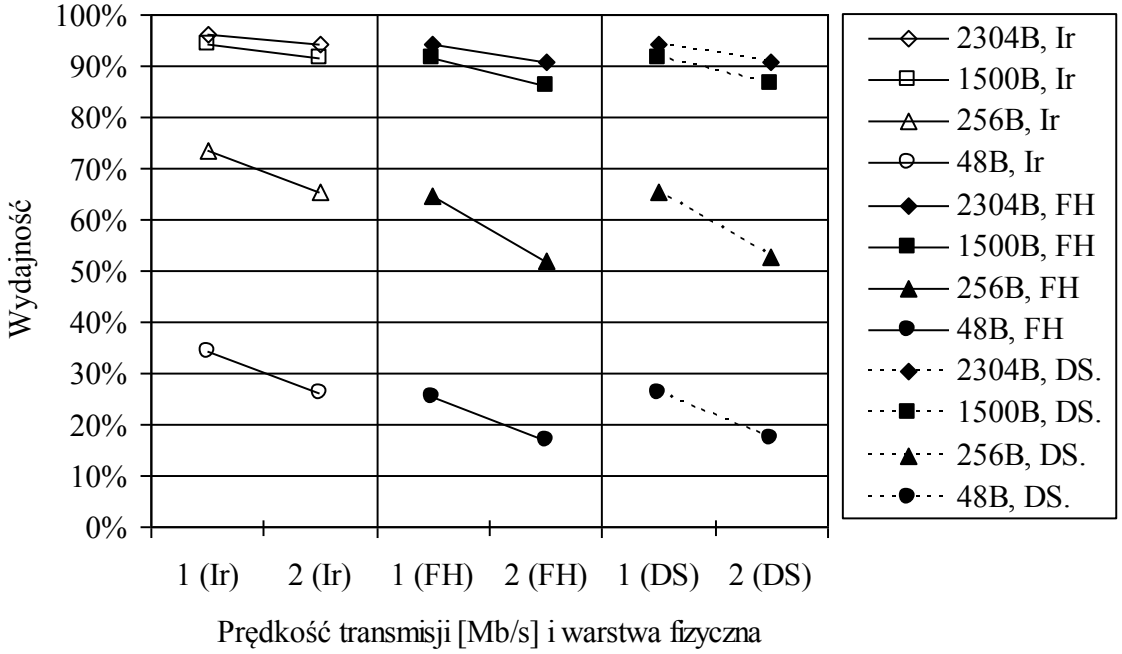
3.3.2.1. Warstwy DSSS, FHSS i Ir

Wyniki otrzymane dla wszystkich wariantów „podstawowego" standardu 802.11 – podczerwieni (Ir), przeskoków częstotliwości (FH) i kluczowania bezpośredniego (DS) – przedstawiono razem na rys. 3.37 [138, 155]. Najbardziej wydajną warstwę tworzy podczerwień, najmniej – DS. W każdej odmianie transmisja z prędkością 2 Mb/s jest mniej efektywna (ale szybsza) niż 1 Mb/s. Różnica między tymi prędkościami jest najmniej widoczna dla podczerwieni, a dla dłuższych ramek – 1500 lub 2304 B – jest pomijalna. W każdej odmianie zmniejszanie rozmiaru pola danych zmniejsza zarówno wydajność protokołu, jak i przepustowość efektywną.

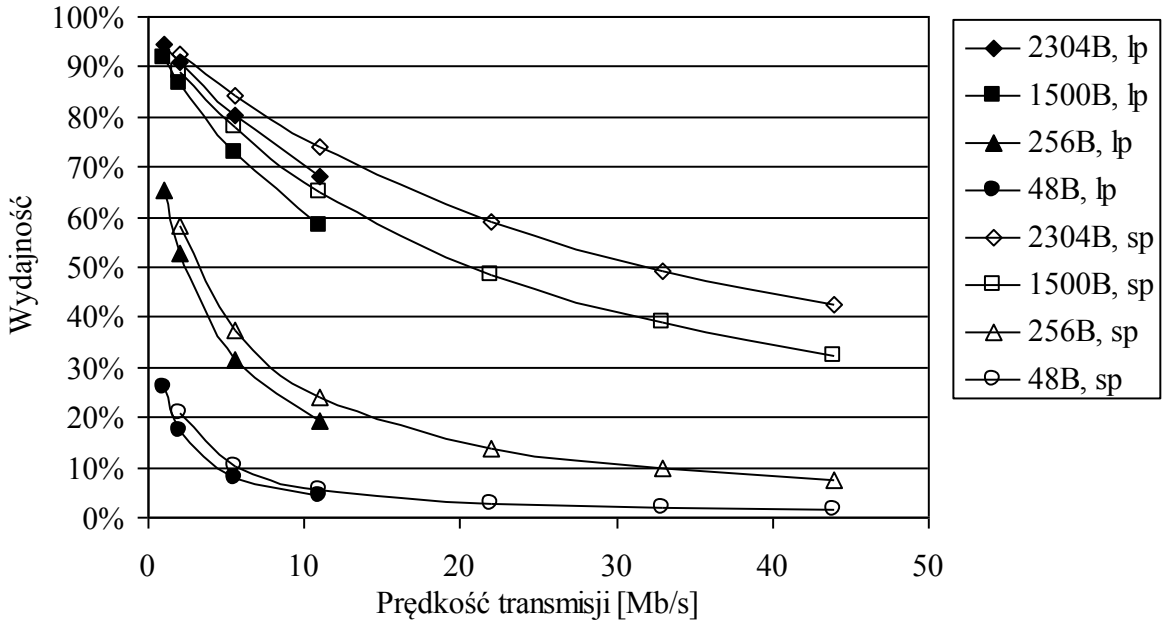
3.3.2.2. Warstwa HR-DSSS

Wyniki otrzymane dla warstwy fizycznej 802.11b, z użyciem długiej (lp) lub krótkiej (sp) preambuły, pokazano na rys. 3.38 [138, 155]. Nietrudno zauważyć, że skrócenie preambuły przynosi pewne korzyści. Przykładowo, wydajność protokołu z najdłuższymi ramkami danych (2304 B) i długą preambułą jest zbliżona do wydajności z ramkami 1500 B i krótką preambułą. Różnica między preambułami jest najbardziej widoczna w przypadku ramek zawierających 256 B danych. Można także zauważyć, iż wraz ze wzrostem prędkości transmisji spada wydajność protokołu, lecz rośnie efektywna przepustowość.

Pomimo że standard określa prędkości transmisji nie większe niż 11 Mb/s, przeprowadzono obliczenia także dla prędkości wyższych, mianowicie 22, 33 i 44 Mb/s. Prędkości te są zdefiniowane jako opcjonalne w standardzie 802.11g, który można rozpatrywać jako rozszerzenie 802.11b. Ponadto, są dostępne urządzenia, w których użycie tych prędkości transmisji jest możliwe. Obliczenia dla tych prędkości przeprowadzono przy założeniu występowania wyłącznie krótkiej preambuły.



Rys. 3.37. Wydajność protokołu 802.11 dla warstw fizycznych DS, FH i Ir
Fig. 3.37. 802.11 protocol performance for DS, FH and Ir physical layers

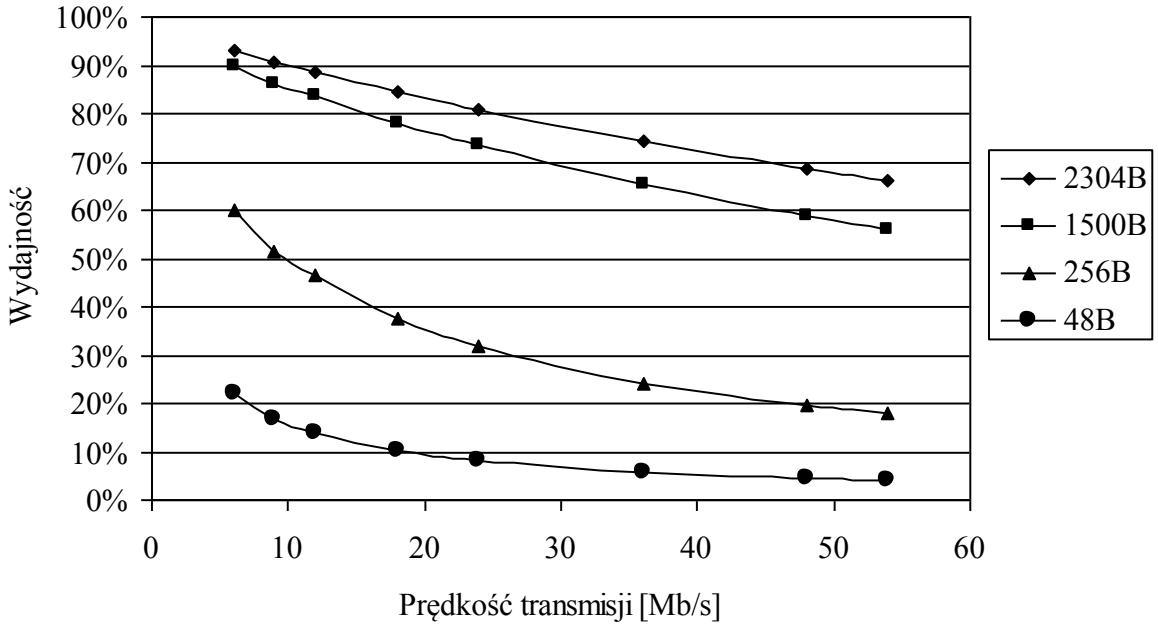


Rys. 3.38. Wydajność protokołu 802.11 dla warstwy fizycznej HR-DSSS – preambuła długa (lp) i krótka (sp)
Fig. 3.38. 802.11 protocol performance for HR-DSSS physical layer – long (lp) and short (sp) preamble

3.3.2.3. Warstwa OFDM

Wyniki obliczeń dla warstwy fizycznej 802.11a, uzyskane dla różnych długości pola danych, pokazano na rys. 3.39 [138, 155]. Warto zauważyć, że wraz ze wzrostem prędkości

transmisji wydajność protokołu szybko maleje – nawet dla ramek o maksymalnej długości – od około 92% dla 6 Mb/s do około 66% dla 54 Mb/s. Spadek ten – podobnie jak dla 802.11b – jest liniowy dla ramek dłuższych i hiperboliczny dla krótszych.

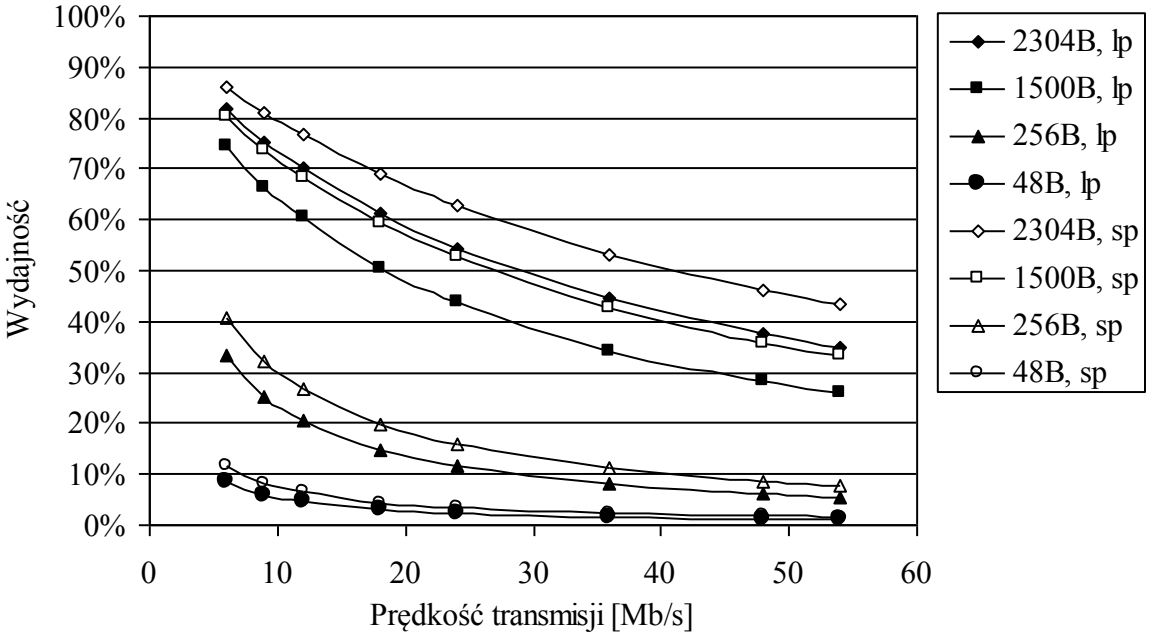


Rys. 3.39. Wydajność protokołu 802.11 dla warstwy fizycznej OFDM
Fig. 3.39. 802.11 protocol performance for OFDM physical layer

3.3.2.4. Warstwa ERP

Wyniki uzyskane dla warstwy fizycznej 802.11g pokazano na rys. 3.40 [138, 155]. Obliczono je dla wariantu DSSS-OFDM z użyciem krótkiej lub długiej preambuły. Wyniki dla wariantu ERP-OFDM byłyby zbliżone do osiągów warstwy OFDM (rys. 3.39), ale tylko wówczas, gdy żadne urządzenia starszego typu (802.11b lub 802.11 DSSS) nie są podłączone do sieci. W tym przypadku bowiem ERP-OFDM zastępuje się przez DSSS-OFDM. Z kolei wyniki dla ERP-CCK i ERP-PBCC są zbliżone do osiągów warstwy HR-DSSS (rys. 3.38).

Nietrudno zauważyć, że w porównaniu z warstwą OFDM (rys. 3.39) spadek wydajności następuje znacznie szybciej. Jest to spowodowane użyciem preambuł zapewniających zgodność wsteczną, które są znacznie dłuższe od używanych w modulacji OFDM. Przykładowo, dla prędkości 54 Mb/s wydajność spada poniżej 50% nawet dla najdłuższych ramek. Podobnie jak w przypadku warstwy DSSS, różnice efektywności między ramkami zawierającymi 1500 B i 2304 B danych nie są duże. Natomiast różnica między krótką i długą preambułą jest najbardziej widoczna dla ramek 1500 B.



Rys. 3.40. Wydajność protokołu 802.11 dla warstwy fizycznej DSSS-OFDM – preambuła długa (lp) i krótka (sp)
Fig. 3.40. 802.11 protocol performance for DSSS-OFDM physical layer – long (lp) and short (sp) preamble

3.3.2.5. Porównanie warstw fizycznych

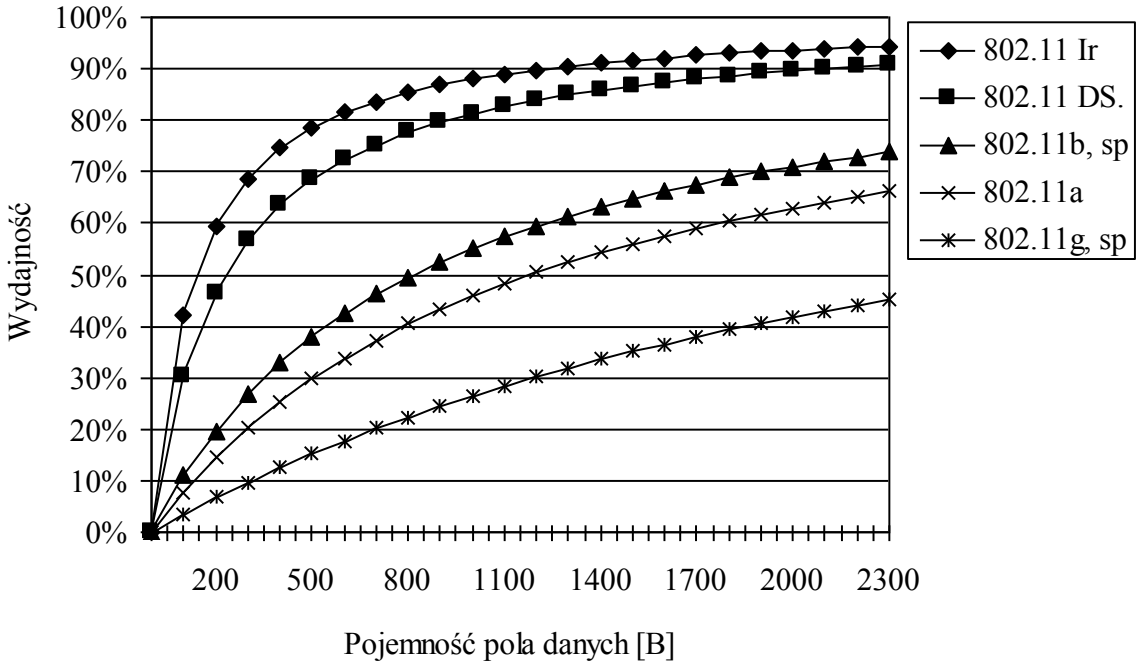
Opierając się na poprzednich wyliczeniach, wybrano kilka odmian warstw fizycznych dla dalszych porównań, których celem jest określenie wpływu pojemności pola danych ramki na wydajność protokołu. Wybrano następujące warstwy:

- 802.11 (podczerwień) z prędkością transmisji 2 Mbps,
- 802.11 (DSSS) z prędkością 2 Mbps,
- 802.11b (krótka preambuła) z prędkością 11 Mbps,
- 802.11a z prędkością 54 Mbps,
- 802.11g (krótka preambuła) z prędkością 54 Mbps.

Dzięki takiemu wyborowi można porównać najwyższe prędkości transmisji określone dla danej warstwy fizycznej. Podejście takie jest uzasadnione, ponieważ niższe prędkości używane są rzadko – gdy moc sygnału jest zbyt niska, aby zapewnić prawidłowy odbiór z prędkościami wyższymi. Ponadto, różnice między osiągami poszczególnych warstw przy małych prędkościach transmisji nie są duże – wydajność przekracza wówczas 85%.

Wyniki porównania są pokazane na rys. 3.41 [138, 155]. Nietrudno zauważyć, że wraz ze wzrostem prędkości transmisji spada wydajność protokołu. Jest to spowodowane narzutem warstwy fizycznej, wynikającym z użycia preambuły i nagłówka, przesyłanych z najniższą prędkością dla danej warstwy. Istotne jest przy tym, że najważniejsze i najpopularniejsze od-

miany warstw fizycznych –DSSS, HR-DSSS i ERP (z wyjątkiem ERP-OFDM) mają taki sam format preambuły i nagłówka. Biorąc pod uwagę zasady wymiany ramek (rys. 3.36) i prędkości transmisji elementów warstwy fizycznej, nietrudno zauważyć, iż, wraz ze zwiększaniem prędkości transmisji, czas transmisji ramki warstwy liniowej maleje, co oczywiście przynosi wzrost efektywnej przepustowości. Jednakże elementy warstwy fizycznej przesyła się nadal z tymi samymi (niskimi) prędkościami, zatem narzut protokołu wynikający z ich użycia rośnie, a jednocześnie wydajność protokołu maleje. Wyjaśnienie to nie dotyczy przesyłów realizowanych wyłącznie z użyciem modulacji OFDM, np. w sieci 802.11a czy 802.11g ERP-OFDM. Tym niemniej, narzut warstwy fizycznej jest tam i tak wystarczająco duży, by wydajność tych protokołów przy najwyższych prędkościach transmisji była niższa niż dla 802.11b.



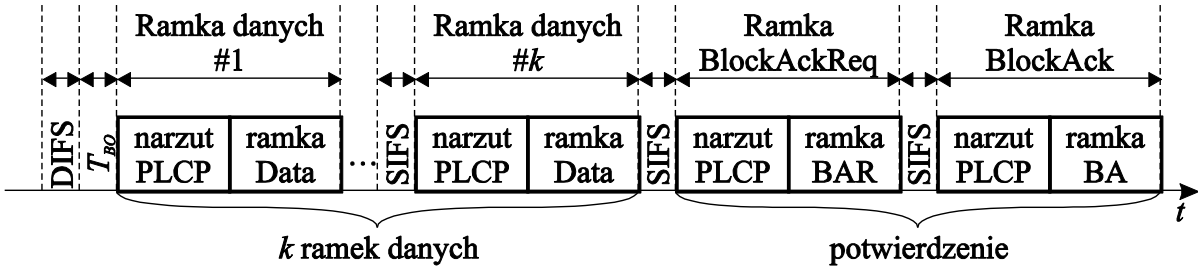
Rys. 3.41. Wpływ długości pola danych na wydajność protokołu 802.11 dla różnych warstw fizycznych
Fig. 3.41. Data field length influence on 802.11 protocol performance for various physical layers

Opisane powyżej rozważania dotyczą wydajności podstawowej metody wymiany informacji w protokole IEEE 802.11. Standard ten podlega jednak ciągłym modyfikacjom nie tylko w zakresie warstwy fizycznej, ale także liniowej. Modyfikacje te obejmują wprowadzenie mechanizmów podnoszących wydajność protokołu przy wysokich prędkościach transmisji. Można do nich zaliczyć potwierdzenie blokowe, wprowadzone w standardzie 802.11e, a także agregację ramek, proponowaną w standardzie 802.11n.

3.3.3. Potwierdzenie blokowe

Mechanizm potwierdzenia blokowego (por. rozdz. 3.2.4.8) umożliwia transmisję ciągu składającego się z wielu ramek danych, które są następnie potwierdzane wspólnie. Mechanizm ten umożliwia zastosowanie strategii potwierdzenia natychmiastowego lub opóźnionego. Przyjmuje się, iż pierwsza z nich umożliwia osiągnięcie wyższej wydajności transmisji i dlatego powinna być stosowana przy przesyle z dużą prędkością [52]. Przesył z wykorzystaniem potwierdzenia blokowego wymaga uprzedniego uzgodnienia między nadawcą i odbiorcą. Dla potrzeb oszacowania wydajności protokołu wykorzystującego potwierdzenie blokowe przyjmijmy jednak, iż objętość przesyłanej informacji jest wystarczająco duża, aby narzut wynikający z czynności administracyjnych był pomijalny.

W przypadku potwierdzenia blokowego natychmiastowego cykl transmisyjny składa się z ciągu ramek danych (Data). Po ostatniej ramce danych pojawia się ramka BlockAckReq, po której następuje ramka BlockAck. Wszystkie ramki są oddzielone od siebie odstępem SIFS, a przed każdą z nich występuje preambuła i nagłówek warstwy fizycznej. Proces wymiany informacji z użyciem potwierdzenia blokowego pokazano na rys. 3.42.



Rys. 3.42. Przebieg transmisji z użyciem potwierdzenia blokowego [135]
Fig. 3.42. Transmission course using block acknowledge

Biorąc pod uwagę przebieg transmisji, czas trwania cyklu można określić jako [134, 135]:

$$T_p = T_{DIFS} + T_{BO} + (k+1)\cdot T_{SIFS} + (k+2)\cdot T_{PLCP} + kT_{Data} + T_{BAR} + T_{BA}, \tag{3.10}$$

gdzie: k – rozmiar bloku danych, natomiast T_{BAR} i T_{BA} określają czas transmisji ramek odpowiednio BlockAckReq i BlockAck. Biorąc pod uwagę formaty tych ramek,

$$T_{BAR} = \frac{8\cdot 24}{R'_{wl}} \tag{3.11}$$

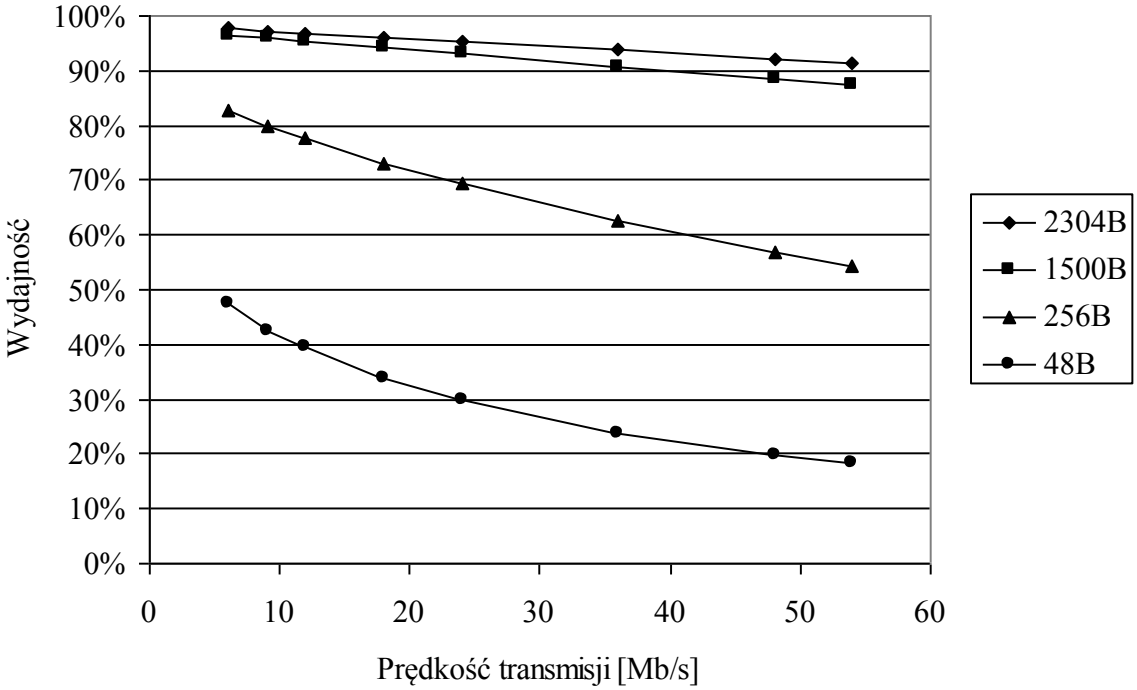
oraz

$$T_{BA} = \frac{8\cdot(24+128)}{R'_{wl}}. \tag{3.12}$$

Również i tu należy uwzględnić dodatkowy narzut, wprowadzany przez niektóre warstwy fizyczne, np. FHSS, OFDM i ERP-OFDM. Należy także pamiętać o ograniczeniu liczby ramek w bloku do 64.

Przy założeniu stałej długości ramek danych w czasie pojedynczego cyklu transmisyjnego przesłanych zostanie $L_p = k \cdot L$ bajtów.

Przykładowe obliczenia wykonano dla tych samych pojemności pola danych ramki co poprzednio (2304, 1500, 256 i 48 B), przyjmując maksymalną wielkość bloku (k=64). Wyniki uzyskane dla warstwy OFDM (802.11a, 802.11g) pokazano na rys. 3.43, natomiast dla HT (802.11n) – na rys. 3.44.



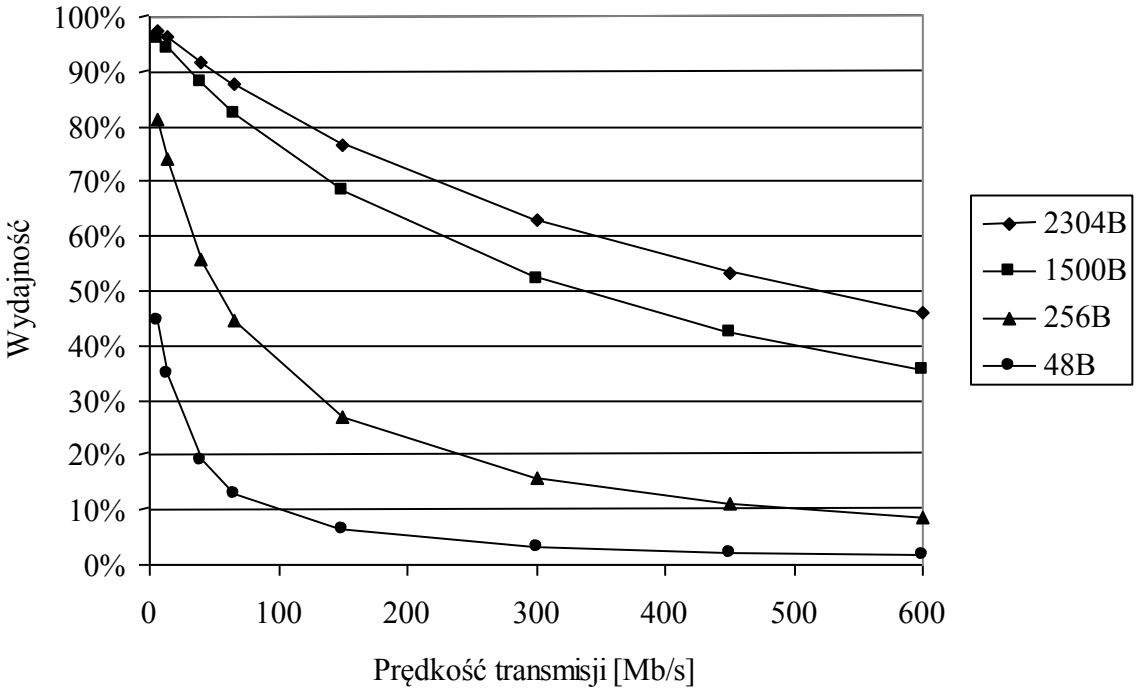
Rys. 3.43. Wydajność protokołu 802.11 z warstwą fizyczną OFDM przy użyciu potwierdzenia blokowego

Fig. 3.43. 802.11 protocol performance for OFDM physical layer using block acknowledge

Jak nietrudno zauważyć, potwierdzenie blokowe znacznie podnosi wydajność protokołu 802.11 z warstwą fizyczną OFDM. Przy użyciu najdłuższych ramek (2304 B) uzyskuje się wydajność przekraczającą 90%, nawet dla najwyższej prędkości transmisji; przy użyciu ramek 1500 B wydajność spada tylko nieznacznie. Krótsze ramki powodują widoczny spadek wydajności protokołu, jednak nawet dla ramek 256 B nie spada ona poniżej 50%. Można zatem powiedzieć, że potwierdzenie blokowe jest mechanizmem wystarczającym dla uzyskania wystarczającej wydajności obecnie stosowanych sieci 802.11a i 802.11g.

W przypadku warstwy fizycznej HT (802.11n) sytuacja nie wygląda już tak optymistycznie. Już przy prędkościach transmisji rzędu 100 Mb/s wydajność spada poniżej 90%, nawet przy użyciu najdłuższych ramek; przy prędkości 600 Mb/s jest ona nawet niższa niż 50%. Można zatem powiedzieć, że potwierdzenie blokowe, jakkolwiek podnosi wydajność sieci z warstwą fizyczną HT, nie jest mechanizmem wystarczającym dla efektywnego działania

sieci o wysokiej prędkości transmisji, nawet przy maksymalnej wielkości bloku. Sieć taka charakteryzuje się bowiem podobną wydajnością jak sieć z warstwą OFDM bez żadnych mechanizmów podnoszących wydajność.



Rys. 3.44. Wydajność protokołu 802.11 z warstwą fizyczną HT przy użyciu potwierdzenia blokowego
Fig. 3.44. 802.11 protocol performance for HT physical layer using block acknowledge

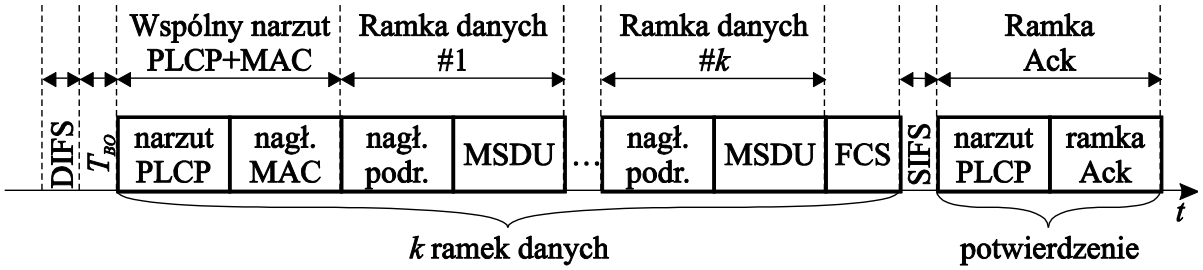
W przypadku gdy bloki zawierają niewielką liczbę ramek danych, potwierdzenie blokowe może nie przynieść wzrostu wydajności. Nawet jeśli – zgodnie z przyjętymi założeniami – pominąć narzut związany z uzgodnieniem i zerwaniem takiego typu transmisji, narzut związany z przesłaniem ramki potwierdzenia blokowego jest większy niż w przypadku zwykłego potwierdzenia. Przeprowadzono zatem dodatkowe obliczenia w celu określenia minimalnej liczby ramek, dla której transmisja z użyciem potwierdzenia blokowego przynosi wzrost wydajności w porównaniu z podstawową metodą wymiany danych. Wynika z nich, że przy dużych prędkościach transmisji, np. 54 Mb/s, potwierdzenie blokowe jest wydajniejsze od zwykłego już przy rozmiarze bloku k=2, natomiast dla mniejszych prędkości transmisji, np. 6 Mb/s – przy k=3. Właściwość ta nie zależy od rozmiarów przesyłanych ramek.

3.3.4. Agregacja A-MSDU

Agregacja A-MSDU, podobnie jak i potwierdzenie blokowe, umożliwia transmisję ciągu ramek, które są potwierdzane wspólnie. O ile jednak potwierdzenie blokowe wymaga, aby każda ramka była w pewnym sensie niezależną jednostką, zawierającą preambułę i nagłówek podwarstwy PLCP, o tyle agregacja ramek umożliwia poprzedzenie całego ciągu pojedynczą

preambułą i nagłówkiem, które są wspólne dla wszystkich ramek danych. Również wspólny jest nagłówek warstwy liniowej. Każda podramka ciągu jest natomiast uzupełniana kilkunastobajtowym nagłówkiem indywidualnym.

W przypadku agregacji A-MSDU cykl wymiany składa się z ciągu podramek, z których każda jest poprzedzona indywidualnym nagłówkiem. Podramki te poprzedza preambuła i nagłówek warstwy fizycznej oraz typowy nagłówek ramki warstwy liniowej. Za ciągiem podramek występuje wspólna suma kontrolna. Ramka jest potwierdzana za pomocą ramki Ack, która jest poprzedzona odstępem SIFS oraz preambułą i nagłówkiem warstwy fizycznej. Proces wymiany z użyciem agregacji A-MSDU pokazano na rys. 3.45.



Rys. 3.45. Przebieg transmisji z użyciem agregacji A-MSDU [135]
Fig. 3.45. Transmission course using A-MSDU aggregation

Biorąc pod uwagę przebieg transmisji, czas trwania cyklu można określić jako [135]:

$$T_p = T_{DIFS} + T_{BO} + T_{SIFS} + 2T_{PLCP} + T_{MAC} + kT_{SubFr} + T_{Ack}, \tag{3.13}$$

gdzie: k – liczba agregowanych ramek, natomiast T_{MAC} i T_{SubFr} określają czas transmisji odpowiednio nagłówka warstwy liniowej (MAC) oraz podramki wraz z jej nagłówkiem. Biorąc pod uwagę formaty tych elementów,

$$T_{MAC} = \frac{8 \cdot 28}{R_{wl}} \tag{3.14}$$

oraz

$$T_{SubFr} = \frac{8 \cdot 4 \cdot \left\lceil \frac{14 + L}{4} \right\rceil}{R_{wl}}. \tag{3.15}$$

Również i tu należy uwzględnić dodatkowy narzut, wprowadzany przez niektóre warstwy fizyczne, np. FHSS, OFDM i ERP-OFDM. Należy także pamiętać o ograniczeniu sumarycznej długości wynikowej ramki do $L_{\max}$=3839 lub 7935 B, zależnie od możliwości komunikujących się stacji. Ograniczenie to może mieć dwojaki wpływ na liczbę i długość podramek, zależnie od przyjętych zasad agregacji.

W pierwszym wariancie nadawca zbiera jednostki MSDU o ustalonej długości do momentu, gdy pozostała ilość miejsca nie pozwala już na dodanie kolejnej jednostki. Wówczas liczba scalanych jednostek MSDU wynosi

$$k = \left\lfloor \frac{L_{\max}}{4\left\lceil \frac{14+L}{4} \right\rceil} \right\rfloor, \tag{3.16}$$

zatem całkowita objętość informacji przesłanej w cyklu transmisyjnym jest równa

$$L_D = k \cdot L = \left\lfloor \frac{L_{\max}}{4\left\lceil \frac{14+L}{4} \right\rceil} \right\rfloor \cdot L. \tag{3.17}$$

W drugim wariancie nadawca zbiera jednostki MSDU, a gdy pozostała ilość miejsca nie pozwala na dodanie jednostki o założonej stałej długości, jest dodawana jednostka krótsza o długości tak dobranej, aby całkowicie wykorzystać ustalony limit długości wynikowej ramki. Jest to wariant mniej realny ze względu na trudności w jego praktycznej implementacji, ale ze względu na efektywniejsze wykorzystanie ustalonych limitów długości ramki powinien pozwalać na uzyskanie wyższej wydajności. W wariancie tym liczba scalanych jednostek MSDU wynosi

$$k = \left\lceil \frac{L_{\max}}{4\left\lceil \frac{14+L}{4} \right\rceil} \right\rceil, \tag{3.18}$$

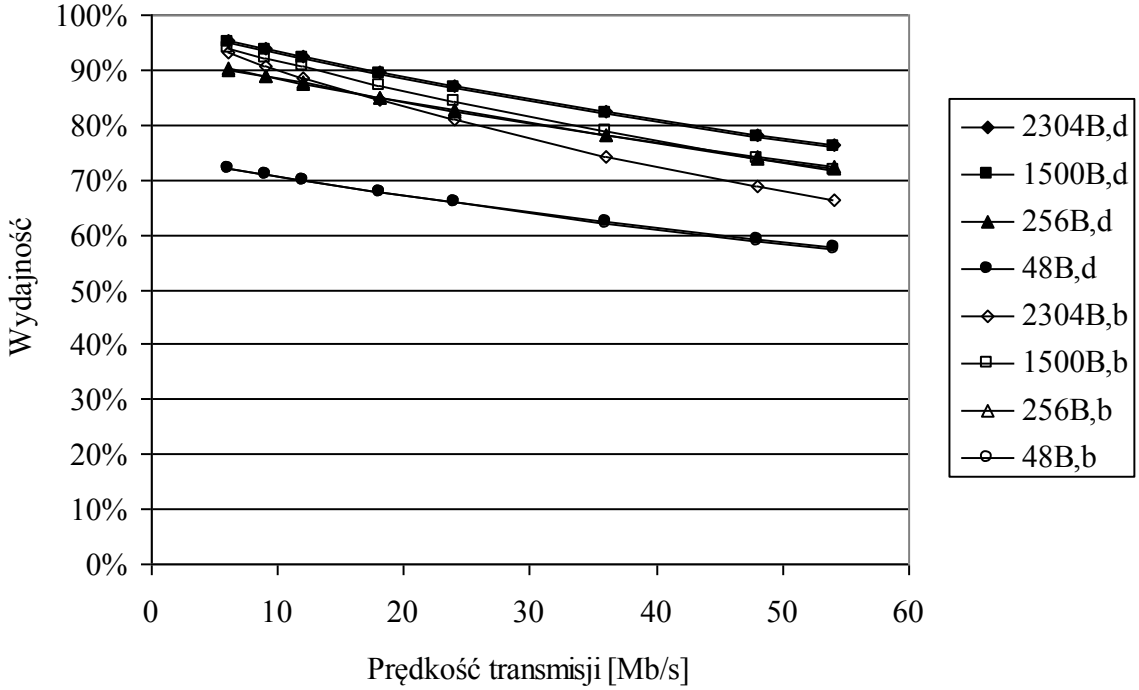
a całkowita objętość informacji przesłanej w cyklu transmisyjnym jest równa wartości $L_{\max}$ pomniejszonej o liczbę bajtów, zajętych przez informacje organizacyjne (nagłówki podramek i bajty rozciągające). W rezultacie

$$L_D = L_{\max} - \left\lfloor \frac{L_{\max}}{4\left\lceil \frac{14+L}{4} \right\rceil} \right\rfloor \cdot \left(4\left\lceil \frac{14+L}{4} \right\rceil - L \right) - 14. \tag{3.19}$$

Niezależnie od ustalonej pojemności pola danych jednostki A-MSDU sumaryczna długość wynikowej ramki w drugim wariancie wynosi zawsze $L_{\max}$.

Na rys. 3.46 zilustrowano wyniki obliczeń wydajności protokołu 802.11 z agregacją A-MSDU w obu rozważanych powyżej wariantach i z ograniczeniem długości wynikowej ramki do $L_{\max}$=3839 B. Okazuje się, że wariant z dopełnianiem do tej wartości przynosi poprawę wydajności tylko dla ramek długich (1500 i 2304 B), natomiast dla ramek krótkich (48 i 256 B) wyniki dla obu wariantów są zbliżone, a krzywe na wykresie pokrywają się. Różnica między wariantem z dopełnianiem i bez dopełniania jest najbardziej widoczna przy najwięk-

szej prędkości transmisji (54 Mb/s). Dla ramek o długości 2304 B różnica ta wynosi 10%, natomiast dla 1500 B – 5%.

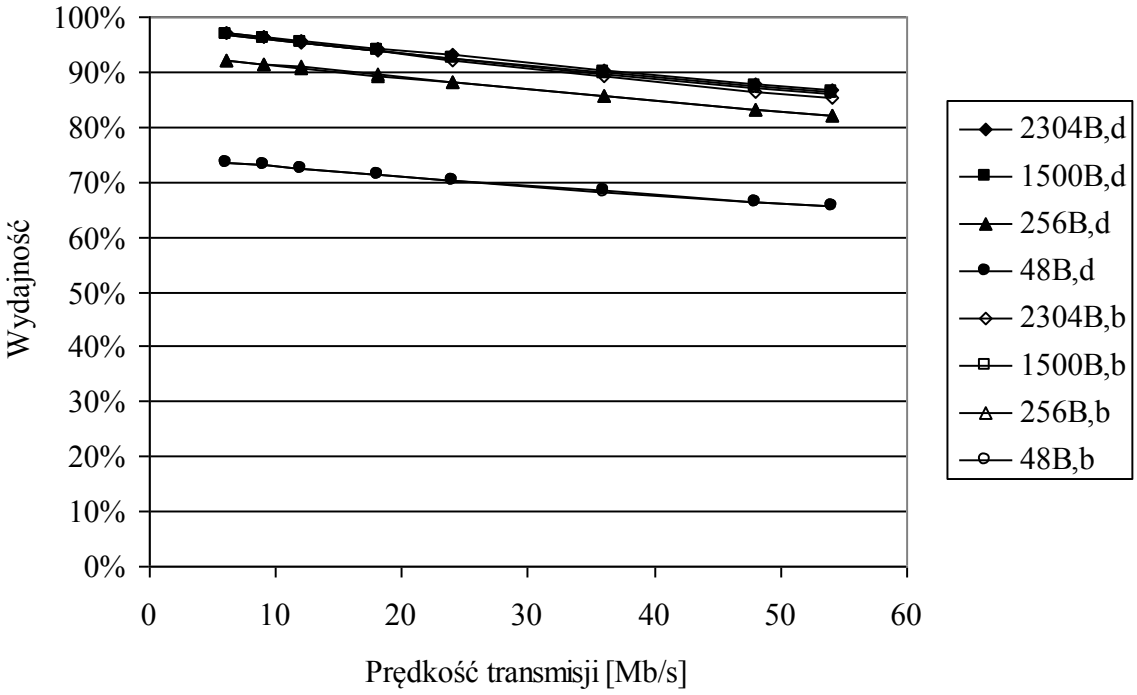


Rys. 3.46. Wydajność protokołu 802.11 z warstwą fizyczną OFDM przy użyciu agregacji A-MSDU z dopełnianiem (d) i bez dopełniania (b) dla L_{max}=3839 B

Fig. 3.46. 802.11 protocol performance for OFDM physical layer using A-MSDU aggregation with padding (d) and without padding (b) for L_{max}=3839 B

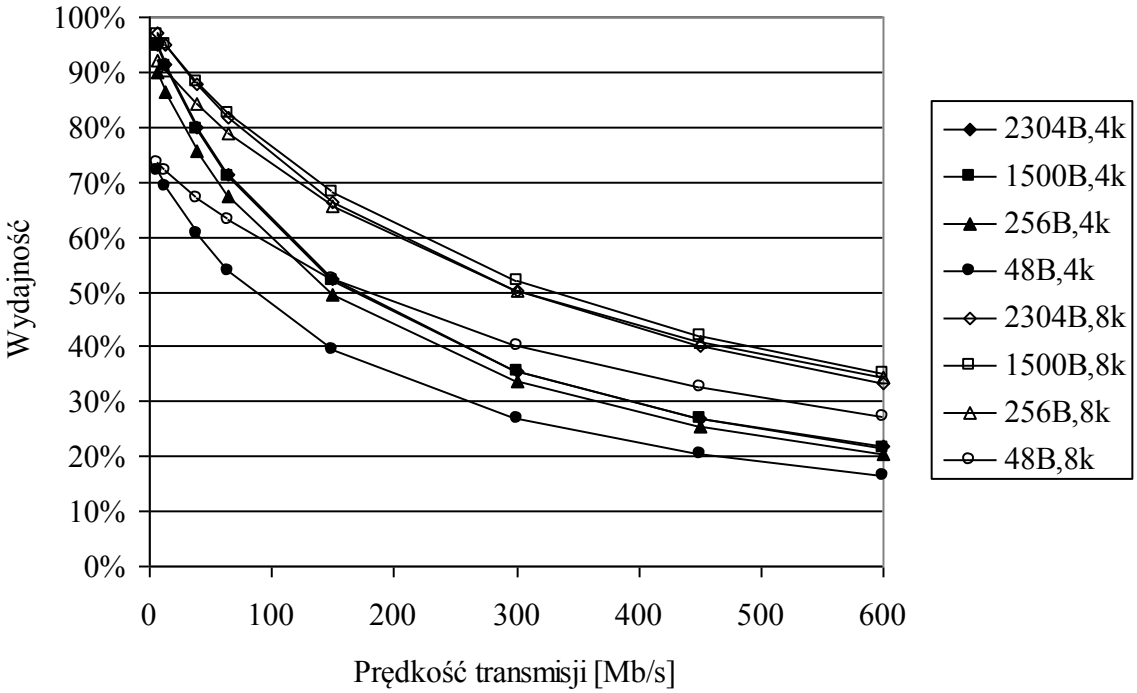
Na rys. 3.47 pokazano wyniki analogicznych obliczeń, ale wykonanych dla wartości L_{max}=7935 B. W tym przypadku różnice między wariantem z dopełnianiem i bez dopełniania są praktycznie niewidoczne – nie przekraczają one 1%. Biorąc pod uwagę wyniki uzyskane dla obu limitów długości ramki wynikowej, można stwierdzić, że rozważany mechanizm dopełniania, poza nielicznymi przypadkami, nie przynosi oczekiwanych korzyści. Mając zatem na uwadze trudności w jego realizacji praktycznej, stosowanie go należy uznać za niecelowe.

Na rys. 3.48 przedstawiono wyniki obliczeń, wykonanych dla warstwy fizycznej HT, proponowanej w standardzie 802.11n. Wyniki te uzyskano dla obu limitów długości ramki wynikowej, tj. L_{max}=3839 B oraz L_{max}=7935 B. Jak widać na przedstawionym wykresie, wyższy limit długości ramki sprzyja uzyskaniu wyższej wydajności protokołu, ale dla wysokich prędkości transmisji jest ona i tak stosunkowo niska. Przykładowo, dla najwyższej prędkości (600 Mb/s) nawet użycie najdłuższych możliwych ramek (2304 B danych) zapewnia wydajność zaledwie około 35%. Warto zauważyć, że zmniejszanie długości ramek nie ma dużego wpływu na wydajność – nawet przy ramkach zawierających 256 B danych wydajność wynosi około 33% i spada do 28% dopiero dla ramek 48 B. W przypadku krótszego limitu długości ramki wynikowej wydajność przy najwyższej prędkości transmisji waha się między 16% a 22%.



Rys. 3.47. Wydajność protokołu 802.11 z warstwą fizyczną OFDM przy użyciu agregacji A-MSDU z dopełnianiem (d) i bez dopełniania (b) dla L_{max}=7935 B

Fig. 3.47. 802.11 protocol performance for OFDM physical layer using A-MSDU aggregation with padding (d) and without padding (b) for L_{max}=7935 B



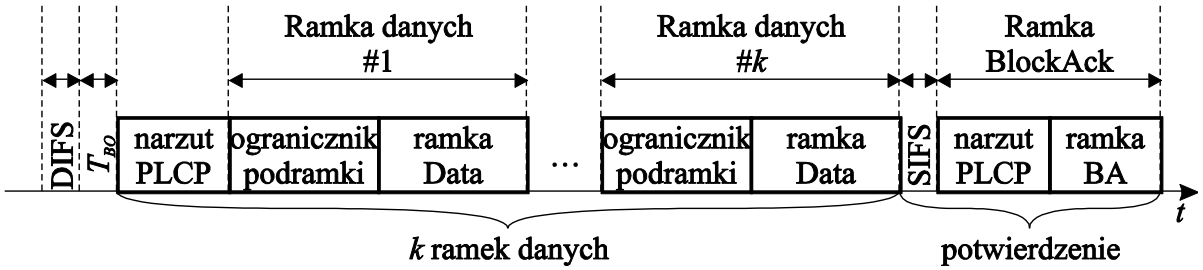
Rys. 3.48. Wydajność protokołu 802.11 z warstwą fizyczną HT przy użyciu agregacji A-MSDU bez dopełniania dla L_{max}=3839 B (4k) i L_{max}=7935 B (8k)

Fig. 3.48. 802.11 protocol performance for HT physical layer using A-MSDU aggregation without padding for L_{max}=3839 B (4k) and L_{max}=7935 B (8k)

Agregacja A-MSDU przy dłuższych ramkach wypada gorzej niż potwierdzenie blokowe, natomiast przy krótszych ramkach wypada lepiej. Ponieważ mechanizm agregacji A-MSDU wprowadzono w celu podniesienia wydajności podczas transmisji wielu krótkich ramek, można stwierdzić, że cel ten został osiągnięty. Warto jednak zauważyć, że potwierdzenie blokowe pozwala przesłać w cyklu transmisyjnym do 64 ramek o długości do 2304 B każda, co daje w sumie ponad 140 kB. Przy prędkości 600 Mb/s czas trwania cyklu wynosi około 4300 μs. Agregacja A-MSDU pozwala natomiast przesłać w jednym cyklu nie więcej niż około 7900 B, a przy prędkości 600 Mb/s długość cyklu transmisyjnego wynosi około 290 μs. Potwierdzenie blokowe pozwala zatem przesłać ponad 18-krotnie więcej informacji w 14-krotnie dłuższym czasie, zatem nie należy się dziwić, iż jego wydajność dla długich ramek jest wyższa. Przy krótkich ramkach wydajność potwierdzenia blokowego jest ograniczona maksymalną liczbą ramek w bloku. Przy długich ramkach natomiast większą rolę odgrywa ograniczenie długości wynikowej ramki agregacji A-MSDU.

3.3.5. Agregacja A-MPDU

W przypadku agregacji A-MPDU cykl wymiany składa się z ciągu ramek danych (Data), poprzedzonych preambułą i nagłówkiem warstwy fizycznej. Ramki te są przesyłane bezpośrednio jedna po drugiej, bez odstępu SIFS. Cykl transmisyjny kończy się wraz z przesłaniem nieco zmodyfikowanego potwierdzenia blokowego. Przy agregacji A-MPDU nie występuje bowiem ramka BlockAckReq, co jest uzasadnione, ponieważ agregacja wymusza niejako wspólne potwierdzanie ramek. Ponadto, ramka potwierdzenia jest krótsza od typowej ramki BlockAck, gdyż agregacja nie zezwala na fragmentację ramek danych. Ramka BlockAck jest poprzedzona odstępem SIFS oraz preambułą i nagłówkiem warstwy fizycznej. Proces wymiany z użyciem agregacji A-MPDU pokazano na rys. 3.49.



Rys. 3.49. Przebieg transmisji z użyciem agregacji A-MPDU [135]
Fig. 3.49. Transmission course using A-MPDU aggregation

Biorąc pod uwagę przebieg transmisji, czas trwania cyklu można określić jako [135]:

$$T_p = T_{DIFS} + T_{BO} + T_{SIFS} + 2 \cdot T_{PLCP} + kT'_{Data} + T'_{BA}, \tag{3.20}$$

gdzie: k – liczba agregowanych ramek, natomiast T'_{Data} – czas transmisji ramki danych z uwzględnieniem narzutu wnoszonego przez agregację. T'_{BA} określa czas transmisji skróconej ramki BlockAck. Biorąc pod uwagę formaty tych ramek,

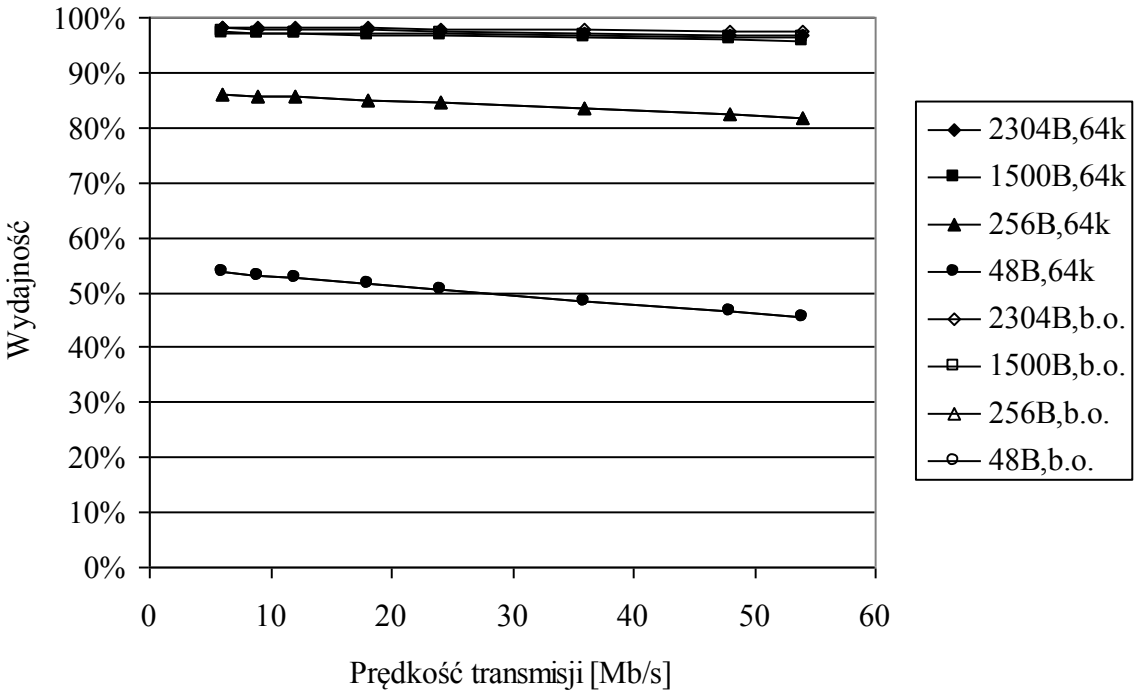
$$T'_{Data} = \frac{8 \cdot \left(4 + 28 + 4 \left\lceil \frac{L}{4} \right\rceil \right)}{R_{wl}} \tag{3.21}$$

oraz

$$T'_{BA} = \frac{8 \cdot (24 + 8)}{R'_{wl}}. \tag{3.22}$$

Również i tu należy uwzględnić dodatkowy narzut, wprowadzany przez niektóre warstwy fizyczne, np. FHSS, OFDM i ERP-OFDM. Należy także pamiętać o tym, że liczba łączonych ramek nie może przekroczyć 64, a ich sumaryczna długość wraz z narzutem wnoszonym przez agregację – 65535 B.

Na rys. 3.50 przedstawiono wyniki obliczeń wydajności protokołu 802.11 z warstwą fizyczną OFDM i agregacją A-MPDU. Dla porównania obliczono także wydajność z pominięciem ograniczenia długości wynikowej ramki. Jak widać na przedstawionym wykresie, zastosowanie agregacji umożliwia osiągnięcie wydajności protokołu rzędu 95÷97% dla wszystkich prędkości transmisji, o ile tylko zostaną użyte długie ramki. Różnica między ramkami zawierającymi 1500 B i 2304 B danych jest niewielka i niewidoczna na wykresie. W przypadku ramek zawierających 256 B danych wydajność jest także wysoka – powyżej 80% – a jej spadek wraz ze wzrostem prędkości transmisji – niewielki. W przypadku najkrótszych ramek wydajność protokołu waha się w granicach 45÷55%, zależnie od prędkości transmisji. Zniesienie limitu długości wynikowej ramki przynosi niewielkie korzyści w przypadku dłuższych ramek. Jest to zgodne z oczekiwaniami, gdyż zniesienie tego ograniczenia pozwala przesłać maksymalną liczbę (64) ramek o dowolnej długości w granicach ustalonych przez format ramki (tj. zawierających nie więcej niż 2304 B danych). W przypadku natomiast, gdy limit obowiązuje, liczba faktycznie przesłanych ramek zależy od ich długości. Efekt ten jest odczuwalny, gdy długość ramki przekracza około 1 KB; wraz ze zwiększaniem się długości ramki maleje liczba ramek podlegających agregacji.

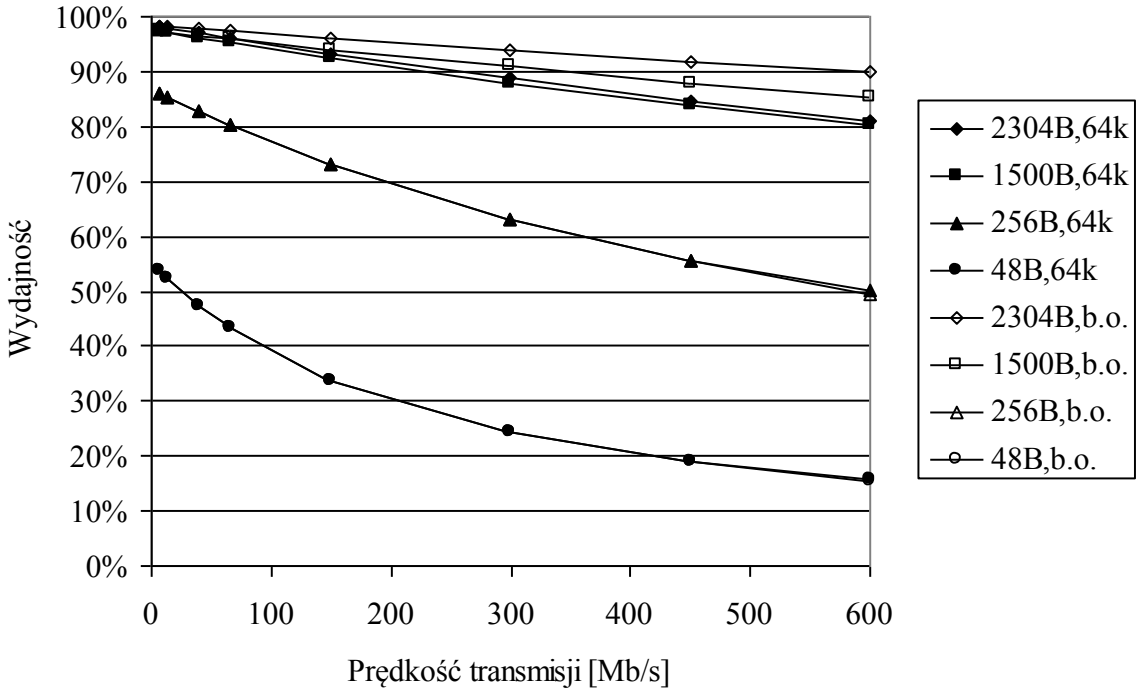


Rys. 3.50. Wydajność protokołu 802.11 z warstwą fizyczną OFDM przy użyciu agregacji A-MPDU z ograniczeniem długości ramki (64k) i bez ograniczenia (b.o.)
Fig. 3.50. 802.11 protocol performance for OFDM physical layer using A-MPDU aggregation with frame length limit (64k) and without limit (b.o.)

Na rys. 3.51 przedstawiono wyniki obliczeń wydajności protokołu 802.11 z warstwą fizyczną HT i agregacją A-MPDU. Dla porównania obliczono także wydajność z pominięciem ograniczenia długości wynikowej ramki. Jak widać na przedstawionym wykresie, zastosowanie agregacji umożliwia osiągnięcie wydajności protokołu przekraczającej 80%, niezależnie od prędkości transmisji, choć spadek wydajności wraz ze wzrostem prędkości jest widoczny. Przykładowo, dla ramek zawierających 1500 B lub 2304 B wydajność spada od około 96% dla prędkości 6 Mb/s do około 80% dla prędkości 600 Mb/s. Zniesienie ograniczenia długości wynikowej ramki pozwala przy najwyższej prędkości transmisji osiągnąć wydajność około 85% dla ramek 1500 B oraz 90% dla ramek 2304 B. W przypadku ramek krótszych wydajność jest oczywiście niższa i – zgodnie z oczekiwaniami – nie zależy ona od tego, czy obowiązuje limit długości ramki wynikowej. Dla ramek zawierających 256 B wydajność zmniejsza się od około 86% dla prędkości 6 Mb/s do około 50% dla 600 Mb/s; analogiczne wartości dla najkrótszych rozpatrywanych ramek wynoszą odpowiednio 55% i 16%.

W przypadku warstwy HT zniesienie limitu długości ramki wynikowej wydaje się o wiele bardziej atrakcyjne. Należy jednak pamiętać, że powoduje ono ponaddwukrotny wzrost zapotrzebowania na pamięć adaptera sieciowego. Przykładowo, dla przesłania ramki składającej się z 64 ramek o maksymalnej długości (2304 B danych), wymagana pojemność bufora wynosi około 150 KB. Wydaje się zatem, że koszt wynikający ze zniesienia ograniczenia długo-

ści ramki wynikowej jest niewspółmiernie wysoki w porównaniu z możliwymi do uzyskania korzyściami. W pewnych przypadkach koszt ten może jednak być uzasadniony.



Rys. 3.51. Wydajność protokołu 802.11 z warstwą fizyczną HT przy użyciu agregacji A-MPDU z ograniczeniem długości ramki (64k) i bez ograniczenia (b.o.)
Fig. 3.51. 802.11 protocol performance for HT physical layer using A-MPDU aggregation with frame length limit (64k) and without limit (b.o.)

3.4. Górna granica przepustowości

Można wykazać [115], że dla sieci standardu 802.11 z podstawową metodą wymiany informacji istnieje górna granica przepustowości (TUL, ang. *Throughput Upper Limit*). Oblicza się ją przy założeniu idealnych warunków pracy sieci i nieskończenie wysokiej prędkości transmisji. Wówczas czasy transmisji wszystkich ramek danych i potwierdzeń wynoszą 0, zatem przy obliczaniu czasu trwania cyklu transmisyjnego bierze się pod uwagę jedynie narzut warstwy liniowej i fizycznej, czyli czasy SIFS i DIFS oraz preambułę i nagłówek warstwy fizycznej. Nagłówek warstwy liniowej pomija się, ponieważ czas jego transmisji jest zależny od przyjętej prędkości i w tym przypadku także wynosi 0. Wynika z tego, że czas trwania cyklu transmisyjnego nie zależy od długości ramki danych, a dokładniej – od liczby bajtów zawartych w polu danych ramki. Górna granica przepustowości jest jednak od niej zależna. Od tego zależy bowiem liczba bajtów, efektywnie przesłanych w czasie cyklu transmisyjnego.

W tabeli 3.11 zebrano obliczone wartości górnej granicy przepustowości dla warstwy fizycznej OFDM [135, 153]. W obu przypadkach wykonano obliczenia dla tych samych długości ramek co w poprzednich rozważaniach – 2304, 1500, 256 i 48 B. W obliczeniach uwzględniono następujące metody wymiany danych:

- podstawową,
- potwierdzenie blokowe (długość bloku k=64),
- agregację A-MSDU z ograniczeniem długości ramki do 3839 B (bez dopełniania),
- agregację A-MSDU z ograniczeniem długości ramki do 7935 B (bez dopełniania),
- agregację A-MPDU z ograniczeniem długości ramki do 65535 B i długości bloku do 64.

Tabela 3.11

Górna granica przepustowości dla warstwy fizycznej OFDM

Rozmiar pola danych [B]	Metoda wymiany danych				
	Podst.	BlAck	A-MSDU (4k)	A-MSDU (8k)	A-MPDU
2304	117,78	434,25	111,37	184,12	3119,12
1500	76,68	282,72	145,02	183,35	3093,61
256	13,09	48,25	173,24	174,07	791,98
48	2,45	9,05	136,89	139,26	148,50

Jak wynika z danych przedstawionych w tabeli 3.11, podstawowa metoda wymiany informacji w standardzie 802.11 ogranicza prędkość efektywną do ok. 118 Mb/s przy założeniu nieskończenie wysokiej prędkości transmisji, i to jedynie przy użyciu najdłuższych możliwych ramek. Zmniejszenie długości ramki do 1500 B danych powoduje spadek efektywnej prędkości transmisji do około 75 Mb/s. Nietrudno więc zauważyć, iż metoda ta nie zapewnia efektywnego wykorzystania możliwości, jakie daje warstwa fizyczna HT, a nawet – choć w nieco mniejszym zakresie – OFDM.

Zastosowanie potwierdzenia blokowego powoduje niemal czterokrotne podwyższenie górnej granicy przepustowości. Przy najdłuższych ramkach pozwala ono osiągnąć 434 Mb/s, natomiast skrócenie pola danych do 1500 B zmniejsza tę wartość do ok. 283 Mb/s. Można zatem założyć, iż potwierdzenie blokowe umożliwia efektywne wykorzystanie warstwy fizycznej OFDM, nie pozwala natomiast efektywnie wykorzystać najwyższych prędkości transmisji określonych dla warstwy HT.

Paradoksalnie, agregacja A-MSDU dla ramek najdłuższych nie tylko nie przynosi poprawy, ale nawet może pogarszać osiągi sieci – przy ograniczeniu długości ramki wynikowej do 3839 B górna granica przepustowości jest nawet nieco niższa niż w przypadku podstawowej metody wymiany danych. Jest to spowodowane zwiększonym narzutem w przypadku agregacji, natomiast przesłać można i tak tylko jedną ramkę zawierającą 2304 B. Jednak już przy użyciu ramek krótszych (1500 B) górna granica prędkości jest około dwukrotnie wyższa niż dla podstawowej metody wymiany danych. Zwiększenie limitu długości ramki wynikowej do

7935 B znacznie poprawia osiągi, są one jednak poniżej możliwości potwierdzenia blokowego. Są one także praktycznie niezależne od długości pola danych łączonych ramek.

Agregacja A-MSDU wykazuje jednak dużą skuteczność przy ramkach krótkich – górna granica przepustowości jest rzędu 135÷140 Mb/s dla ramek zawierających 48 B danych i 173÷174 Mb/s dla ramek zawierających 256 B. Zarówno podstawowa metoda wymiany informacji, jak i potwierdzenie blokowe wypada znacznie poniżej tych osiągów. Można zatem powiedzieć, iż agregacja A-MSDU dobrze spełnia swoje zadanie, polegające na podniesieniu wydajności podczas transmisji krótkich ramek. Tym niemniej metoda ta nie zapewnia efektywnego wykorzystania prędkości transmisji warstwy fizycznej HT, pomimo że jest z nią w pewnym sensie skojarzona – oba rozwiązania zdefiniowano w standardzie 802.11n.

Agregacja A-MPDU wykazuje zdecydowanie najwyższą efektywność – górna granica przepustowości już przy ramkach zawierających 256 B danych przekracza 700 Mb/s, a przy najdłuższych – 3 Gb/s. Nawet przy najkrótszych ramkach wypada najlepiej ze wszystkich rozpatrywanych metod. Można zatem powiedzieć, iż agregacja A-MPDU pozwala nie tylko efektywnie wykorzystać prędkości transmisji określone dla warstwy fizycznej HT, ale nawet dysponuje pewną rezerwą dla ewentualnych przyszłych rozwiązań o jeszcze wyższych prędkościach transmisji.

W tabeli 3.12 zebrano wyniki analogicznych obliczeń, wykonanych dla warstwy fizycznej HT. Są one nieco gorsze niż dla warstwy OFDM. Wpływ na to ma większa długość preambuły i nagłówka podwarstwy PLCP, co nieznacznie zwiększa narzut protokołu, zmniejszając jednocześnie wydajność i efektywną prędkość transmisji.

Tabela 3.12

Górna granica przepustowości dla warstwy fizycznej HT

Rozmiar pola danych [B]	Metoda wymiany danych				
	Podst.	BlAck	A-MSDU (4k)	A-MSDU (8k)	A-MPDU
2304	106,85	363,58	101,55	167,89	2844,16
1500	69,57	236,71	132,23	167,18	2820,89
256	11,87	40,40	157,97	158,72	722,16
48	2,23	7,57	124,83	126,99	135,40

3.5. Standard IEEE 802.11 w praktyce

Wprowadzenie standardu IEEE 802.11 miało zapewnić możliwość współpracy różnego typu urządzeń sieciowych, stosujących tę samą warstwę fizyczną, także wówczas, gdy pochodzą one od różnych producentów. Niestety, nie zawsze wygląda to tak różowo, jak można by się spodziewać, a kilka przypadków opisanych poniżej to zapewne tylko niektóre problemy, z jakimi może spotkać się użytkownik.

Przykładowo, adapter sieciowy w komputerze przenośnym, oparty na układzie Intel(R) PRO/Wireless 2200BG, nie łączył się z punktem dostępowym D-Link DWL-2000AP+, mimo że bezproblemowo współpracował z innymi urządzeniami tego samego producenta. Problem udało się rozwiązać przez aktualizację oprogramowania w punkcie dostępu.

Podczas transmisji między urządzeniami pracującymi zgodnie ze standardem IEEE 802.11b/g – kartą sieciową D-Link DWL-510 (układ Realtek RTL8180 [30]) i punktem dostępu Linksys WRT54G (układ Broadcom BCM4712P [67]) – zaobserwowano, że przy prędkości 1 Mb/s efektywna przepustowość jest w przybliżeniu zgodna z oczekiwaniami, natomiast przy prędkościach wyższych – jest znacznie niższa. Również i w tym przypadku rozwiązano problem przez aktualizację oprogramowania w punkcie dostępu oraz sterowników karty sieciowej. Niestety, nie udało się wyjaśnić tego zachowania karty sieciowej, ponieważ nie dysponowano wówczas oprogramowaniem do monitorowania transmisji w standardzie 802.11. Możliwe zresztą, że nawet analiza przebiegu transmisji za pomocą takiego oprogramowania nie pozwoliłaby na wyjaśnienie tego przypadku.

Podobne problemy występują także w najnowszym sprzęcie. Przykładowo, układ Intel® Ultimate N WiFi Link 5300, według informacji producenta [58], zapewnia możliwość osiągnięcia prędkości transmisji 450 Mb/s. Podczas współpracy z routerem D-Link DIR-655, zawierającym układ firmy Atheros, prędkości transmisji – nawet w dobrych warunkach – nie przekraczają jednak 130 Mb/s, przy czym z wcześniejszymi wersjami sterownika, jeżeli wierzyć informacjom podawanym przez system operacyjny Windows XP, nie udało się uzyskać prędkości transmisji powyżej 53 Mb/s (*notabene* standard nie określa takiej prędkości!). W tych samych warunkach adaptery sieciowe firmy D-Link, również z układami firmy Atheros, osiągają deklarowaną przez producenta prędkość 300 Mb/s. W przypadku standardu 802.11n istnieje kilkadziesiąt schematów modulacji i kodowania (MCS, por. rozdz. 3.2.6.1), przy czym niektóre prędkości transmisji można uzyskać, korzystając z kilku takich schematów. Prawdopodobnie zatem w opisywanych urządzeniach zaimplementowano różne schematy MCS, a wspomniane 130 Mb/s to najwyższa prędkość, którą można uzyskać przy użyciu tego samego schematu w obu urządzeniach. Są to jednak jedynie przypuszczenia, a dokładna analiza opisanego zagadnienia wymaga dokładnych informacji o możliwościach zastosowanego w danym urządzeniu układu scalonego. Niestety, informacje takie na ogół nie są podawane do wiadomości publicznej.

Opisane przypadki nie są na szczęście zbyt częste, ale mogą świadczyć o pewnego rodzaju niedojrzałości technologii sieci bezprzewodowych w porównaniu z sieciami przewodowymi. Należy jednak zaznaczyć, że w przypadku sieci lokalnej Ethernet podobna sytuacja zazwyczaj świadczy o poważnej awarii sprzętu sieciowego lub okablowania.

3.6. Podsumowanie rozdziału

Niniejszy rozdział poświęcono bezprzewodowym sieciom lokalnym zgodnym ze standardem IEEE 802.11. W rozdziale największy nacisk położono na elementy warstwy liniowej, które mają wpływ na osiągi protokołu, a także na wybrane aspekty warstwy fizycznej, istotne z punktu widzenia oszacowania wydajności protokołu. Nie pominięto przy tym najnowszego rozwiązania, które pojawiło się w czasie pisania rozdziału, a mianowicie standardu IEEE 802.11n. Standard ten wnosi nie tylko nową warstwę fizyczną, umożliwiającą transmisję z prędkościami sięgającymi 600 b/s, lecz także nowe mechanizmy wymiany ramek, znacznie podnoszące wydajność transmisji.

Opisane powyżej mechanizmy poddano analizie pod kątem ich wpływu na wydajność protokołu w różnych warunkach, a co za tym idzie, możliwej do uzyskania maksymalnej teoretycznej prędkości transmisji. Na podstawie opisu zasad wymiany ramek na poziomie warstwy liniowej oraz formatów stosowanych na poziomie podwarstwy PLCP wyznaczono zależności analityczne, które pozwalają na oszacowanie wydajności protokołu dla różnych wariantów warstwy fizycznej, prędkości transmisji, pojemności pola danych ramki oraz przyjętej zasady wymiany ramek. Wyprowadzone zależności pozwalają oszacować wydajność protokołu oraz efektywną prędkość transmisji na poziomie warstwy liniowej. Uzyskane wyniki mogą więc być nieco lepsze niż uzyskane w rzeczywistej sieci, w której występuje bardziej rozbudowany stos protokołów, a w szczególności protokół TCP/IP obecny nad warstwą liniową. Dla celów obliczeń przyjęto, że transmisja odbywa się w cyklicznie powtarzających się, identycznych fragmentach, które nazwano cyklami transmisyjnymi. Liczba tych cykli zależy oczywiście od całkowitej objętości przesyłanej informacji, ale w ramach jednego cyklu przesyła się ściśle określoną ilość informacji pochodzącej z wyższych warstw sieci, zależną tylko od zasad wymiany ramek oraz parametrów protokołu warstwy liniowej. Dzięki temu, analizując transmisję na poziomie pojedynczego cyklu transmisyjnego, można uzyskać wyniki niezależne od całkowitej objętości przesyłanej informacji.

Wykorzystując wyprowadzone zależności, oszacowano efektywność protokołu na poziomie warstwy liniowej dla wszystkich opisanych warstw fizycznych i metod wymiany ramek. Uzyskane wyniki są w przybliżeniu zgodne z wynikami, jakie można uzyskać w rzeczywistej sieci bezprzewodowej. Drobne odstępstwa spowodowane są tym, że w obliczeniach uwzględniono osiągi warstwy liniowej, podczas gdy w sieci rzeczywistej stosowane są także wyższe warstwy sieci. Może to świadczyć o wystarczającej dokładności przyjętego modelu sieci. Uzyskane wyniki pokazują, że już przy obecnie stosowanych warstwach fizycznych wydajność protokołu jest niewystarczająca, a to przez zbyt duży narzut, wprowadzany głównie przez warstwę fizyczną. Natomiast przy zastosowaniu potwierdzenia blokowego czy

agregacji ramek wydajność sieci jest znacznie wyższa i pozwala na efektywne wykorzystanie prędkości transmisji określonych w obecnie używanych warstwach fizycznych.

Do najważniejszych, oryginalnych fragmentów rozdziału można zaliczyć:

- przeprowadzenie analizy teoretycznej wpływu warstwy fizycznej na efektywność protokołu na poziomie warstwy liniowej,
- przeprowadzenie teoretycznej analizy wydajności mechanizmów podnoszących wydajność protokołu, takich jak potwierdzenie blokowe czy agregacja ramek, dla kilku możliwych wariantów ich działania,
- oszacowanie górnej granicy przepustowości dla wszystkich rozpatrywanych zasad wymiany ramek.

PODSUMOWANIE

Praca podsumowuje wyniki badań autora nad efektywnością protokołów warstwy liniowej w bezprzewodowych sieciach komputerowych. W trakcie badań posługiwano się zarówno metodami analitycznymi, jak i pomiarami uzyskanymi w doświadczalnych sieciach bezprzewodowych. Dzięki takiemu podejściu udało się nie tylko zweryfikować zastosowane modele analityczne, lecz także określić wpływ sprzętu i oprogramowania transmisyjnego na parametry użytkowe sieci. Należy jednocześnie podkreślić, iż badania takie nie zawsze są możliwe, gdyż są uzależnione od dostępności wyposażenia sieciowego z zaimplementowanymi odpowiednimi mechanizmami. O ile jeszcze w przypadku sieci Packet Radio czy – w ograniczonym zakresie – protokołów dostępu do łącza można próbować własnoręcznej implementacji protokołów, o tyle w przypadku sieci zgodnych ze standardem IEEE 802.11 nie jest to praktycznie możliwe ze względu na duży stopień scalenia układów realizujących ten protokół. Ponadto, specyfikacja zarówno tych układów, jak i gotowych urządzeń transmisyjnych nie mówi wprost, czy interesujące mechanizmy – szczególnie potwierdzenie blokowe i agregacja ramek – są zaimplementowane. Pewnym utrudnieniem jest także stosowana terminologia – określenia takie jak *Extended Range* czy *Frame Burst* nie występują w opisie standardu, można zatem jedynie domyślać się ich znaczenia. Zazwyczaj można znaleźć informację o zgodności danego urządzenia ze standardem np. IEEE 802.11e czy 802.11n. Może to wprawdzie sugerować wsparcie dla mechanizmów opisanych w tych standardach, jednak należy do tych deklaracji podchodzić z pewną ostrożnością – przykładowo, urządzenia zgodne ze standardem 802.11 zazwyczaj nie wspierają protokołu PCF. Przed przystąpieniem do badań należy zatem potwierdzić możliwość zastosowania wspomnianych mechanizmów przy użyciu np. programu do monitorowania sieci. Warto jednak zauważyć, że funkcjonalność programów, takich jak np. AirMagnet Laptop czy CommView for WiFi zależy od rodzaju układu scalonego zastosowanego w użytej karcie sieciowej (tzn. nie z każdą kartą program działa tak samo dobrze). Nie oznacza to, że badania efektywności mechanizmów podnoszących wydajność transmisji nie są możliwe, jednak aby były one wiarygodne, należy odpowiednio przygotować bazę sprzętowo-programową. Można zresztą spodziewać się dużego wpływu zastosowanego sprzętu sieciowego na uzyskane wyniki, a wobec braku szczegółowej

dokumentacji urządzeń i stosowanych w nich układów trudno wyjaśnić przyczyny takiego czy innego zachowania sieci. Może też okazać się, że wyniki w dużym stopniu zależą od wersji sterownika urządzenia.

Przedstawiona praca oczywiście nie wyczerpuje wszystkich zagadnień związanych z bezprzewodowymi sieciami komputerowymi. Pomija ona nawet – ze względu na ogrom materiału – wiele interesujących zagadnień związanych z warstwą liniową, jak np. problematykę bezprzewodowych sieci ATM (ang. *Asynchronous Transfer Mode*) oraz protokołów dostępu do łącza dla tych sieci. Pominięto także wiele istniejących i używanych – dawniej i obecnie – systemów i standardów sieci bezprzewodowych, jak np. IrDA, BlueTooth, HiPeRLAN czy HomeRF, a także sieci cyfrowej telefonii bezsznurowej i komórkowej, jak np. DECT i GSM. Niniejsza praca nie ma jednak charakteru pracy encyklopedycznej.

Wpływ bezprzewodowego łącza transmisyjnego na sposób działania protokołów sieciowych nie kończy się na warstwie liniowej. W przypadku gdy sieć zawiera stacje ruchome, konieczne jest stworzenie protokołu warstwy sieciowej, zapewniającego odpowiedni wybór trasy w warunkach stale zmieniającej się topologii sieci. Jest to szczególnie istotne, gdy przemieszczanie się stacji wpływa na bezpośrednią łączność między nimi. Charakterystyka łącza bezprzewodowego może także wpływać na warstwę transportową – wyższa stopa błędów łącza bezprzewodowego sprzyja częstszej, niż w przypadku łącza przewodowego, utracie pakietów. Do niedawna – np. w protokole TCP/IP – utrata pakietu była traktowana jako oznaka przepełnienia bufora routera. Ponieważ w sieci bezprzewodowej przyczyną utraty pakietu może być błąd transmisji, konieczne okazało się zmodyfikowanie protokołów warstwy transportowej.

Ponieważ znaczny odsetek urządzeń komunikujących się bezprzewodowo stanowią urządzenia przenośne o często niewygórowanych zasobach, stworzono specjalny stos protokołów wyższych warstw sieci. Rozwiązanie to, znane pod wspólną nazwą WAP (ang. *Wireless Application Protocol*), można także uznać za przykład modyfikacji istniejących protokołów w celu zapewnienia lepszego – bardziej efektywnego – ich działania w sieci bezprzewodowej. Dotyczy to szczególnie sieci o niskiej prędkości transmisji, do których można zaliczyć m. in. sieci telefonii komórkowej. Obecnie prędkości te są wprawdzie znacznie wyższe, ale i tak znacznie ustępują osiągom współczesnych sieci lokalnych.

Jak widać, praca omawia zaledwie niewielką część problematyki bezprzewodowych sieci komputerowych. Tym niemniej może ona być interesującą pozycją dla osób zajmujących się tworzeniem sieci bezprzewodowych i protokołów dla nich przeznaczonych.

BIBLIOGRAFIA

1. 73M223 1200 Baud FSK Modem. TDK Semiconductor Corp., Apr. 2000.
2. Abramson N.: Sieć Aloha. W: Abramson N., Kuo F. K.: Sieci telekomunikacyjne komputerów. WNT, Warszawa 1978, rozdz. 14, s. 581÷599.
3. Abramson N.: The Aloha System – Another alternative for computer communications. Proceedings of AFIPS Joint Computer Conferences, Houston, Texas, 17-19.11.1970, s. 281÷285.
4. Am7910/11 WORLD-CHIP(R) FSK Modem. Advanced Micro Devices, Jun. 1989.
5. ANSI/IEEE Std 802.1D. IEEE Standard for Information technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – Common specifications. Part 3: Media Access Control (MAC) Bridges. IEEE, New York 1998.
6. Beech W. A., Nielsen D. E., Taylor J.: AX.25 Link Access Protocol for Amateur Packet Radio. Tucson Amateur Packet Radio Corporation, Tucson 1997.
7. Bem D. J.: Anteny i rozchodzenie się fal radiowych. WNT, Warszawa 1973.
8. Berline G., Perratore E.: Portable, Affordable, Secure: Wireless LANs. PC Magazine, 11.02.1992, s. 291÷314.
9. Bharghavan V.: A New Protocol for Medium Access in Wireless Packet Networks. Technical Report, University. of Illinois, Urbana-Champaign, 1996.
10. Bharghavan V.: Performance Evaluation of Algorithms for Wireless Medium Access. Proceedings of IEEE International Computer Performance and Dependability Symposium IPDS '98, 7-9.9.1998, s. 86÷95.
11. Bharghavan V., Demers A., Shenker S., Zhang L.: MACAW: A Media Access Protocol for Wireless LAN's. ACM SIGCOMM Computer Communication Review, Vol. 24, No. 4, Oct. 1994, s. 212÷225.
12. Bianchi G.: IEEE 802.11 – Saturation Throughput Analysis. IEEE Communications Letters, Vol. 2, No. 12, Dec. 1998, s. 318÷320.

13. Bianchi G.: Performance Analysis of the IEEE 802.11 Distributed Coordination Function. IEEE Journal on Selected Areas in Communications, Vol. 18, No. 3, March 2000, s. 535÷547.
14. Bing B.: High-Speed Wireless ATM and LANs. Artech House, 2000.
15. Black U.: Data link protocols. Prentice Hall, Englewood Cliffs, 1993.
16. Broadband Radio Access Networks (BRAN); HIgh PErformance Radio Local Area Network (HIPERLAN) Type 1; Functional specification. European Standard (Telecommunications series) EN 300 652 V1.2.1. ETSI, Sophia Antipolis 1998.
17. Broadband Radio Access Networks (BRAN); 5 GHz high performance RLAN; Harmonized EN covering the essential requirements of article 3.2 of the R&TTE Directive. Harmonized European Standard (Telecommunications series) ETSI EN 301 893 V1.5.1. ETSI, Sophia Antipolis 2008.
18. Chlamtac I., Franta W. R., Levin K. D.: BRAM: The Broadcast Recognizing Access Method. IEEE Transactions on Communications, Vol. COM-27, No. 8, Aug. 1979, s. 1183÷1190.
19. Chlamtac I., Conti M., Liu J. J.-N.: Mobile ad hoc networking: imperatives and challenges. Ad Hoc Networks, Vol. 1, No. 1, Jul. 2003, s. 13÷64.
20. Choi J., Yoo J., Kim C-K.: A Novel Performance Analysis Model for an IEEE 802.11 Wireless LAN. IEEE Communication Letters, Vol. 10, No. 5, May 2006, s. 335÷337.
21. CMX469A 1200/2400/4800 Baud FFSK/MSK Modem. Consumer Microcircuits Limited, 2001.
22. CMX589A GMSK Modem. Consumer Microsystems Plc, 2002.
23. Data Communication over the Telephone Network. 300 Bits Per Second Modem Standarized for Use in the General Switched Telephone Network. ITU-T Recommendation V.23. International Telecommunication Union, Genewa 1993.
24. Data Communication over the Telephone Network. 600/1200 Bits Per Second Duplex Modem Standarized for Use in the General Switched Telephone Network. ITU-T Recommendation V.21. International Telecommunication Union, Genewa 1993.
25. Dąbrowski A.: Amatorska komunikacja cyfrowa. PWN, Warszawa 1994.
26. Dąbrowski A.: Nie tylko fonia i CW. Bogmar, Olsztyn 1994. Wyd. 2 z roku 1998 dostępne w Internecie pod różnymi (i zmieniającymi się) adresami.
27. Deng J., Haas Z. J.: Dual Busy Tone Multiple Access (DBTMA): A New Medium Access Control for Packet Radio Networks. Proceedings of Internationa Conference on Universal Personal Communications ICUPC'98, Florence, Italy, 5-9.10.1998, vol. 2, s. 973÷977.

28. Deng J., Haas Z. J.: Dual Busy Tone Multiple Access (DBTMA) – A Multiple Access Control Scheme for Ad Hoc Networks. IEEE Transactions on Communications, Vol. 50, No. 6, Jun 2002, s. 975÷985.
29. DS1216 SmartWatch RAM (DS1216B/C/D/H); SmartWatch ROM (DS1216E/F). Maxim Integrated Products, Sunnyvale 2004.
30. D-Link DWL-510 11b Wireless PCI Card. [@:] http://www.modem-help.co.uk/D-Link/DWL-510-11b-Wireless-PCI-Card.html. Dostęp: 16.10.2011.
31. Eduroam – about. [@:]http://www.eduroam.org/index.php?p=about. Dostęp: 16.10.2011.
32. Fullmer C. L., Garcia-Luna-Aceves J. J.: Complete single-channel solutions to hidden terminal problems in wireless LANs. Proceedings of IEEE ICC'97, June 1997, s. 575÷579.
33. Fullmer C. L., Garcia-Luna-Aceves J. J.: FAMA-PJ: A Channel Access Protocol For Wireless LANs. Proceedings of ACM International Conference on Mobile Computing and Networking MOBICOM ’95, s. 76÷85.
34. Fullmer C. L., Garcia-Luna-Aceves J. J.: Floor Acquisition Multiple Access (FAMA) for Packet Radio Networks. ACM SIGCOMM Computer Communication Review, Vol. 25, No. 4, Oct. 1995, s. 262÷273.
35. Fullmer C. L., Garcia-Luna-Aceves J. J.: Solutions to Hidden Terminal Problems in Wireless Networks. ACM SIGCOMM Computer Communication Review, Vol. 27, No. 4, Oct. 1997, s. 39÷49.
36. FX604 V.23 Compatible Modem. Consumer Microcircuits Limited, 1996.
37. FX614 Bell 202 Compatible Modem. Consumer Microcircuits Limited, 1997.
38. Garcia-Luna-Aceves J.J, Tzamaloukas A.: The Effect of Exerting Adequate Persistence in Collision Avoidance Protocols. IEEE International Workshop on Mobile Multimedia Communications (MoMuC '99), San Diego, 15-17.11.1999, s. 328÷337.
39. Gast M. S.: 802.11. Sieci bezprzewodowe. Przewodnik encyklopedyczny. Helion, Gliwice 2003.
40. Gast M.: 802.11 Wireless Networks: The Definitive Guide, 2nd ed. O’Reilly, Bejing 2005.
41. Gummalla A., Limb J.: Design of an Access Mechanism for a High Speed Distributed Wireless LAN. IEEE Journal On Selected Areas In Communications, Vol. 18, No. 9, Sept. 2000, s. 1740÷1750.
42. Gummalla A., Limb J.: Wireless Collision Detect (WCD): Multiple Access with Receiver Initiated Feedback and Carrier Detect Signal. Proceedings of IEEE ICC’00, vol. 1, 2000, s. 397÷401.
43. Gummalla A. C. V., Limb J. O.: Wireless Medium Access Control Protocols. IEEE Communications Surveys & Tutorials, Vol. 3, No. 2, s. 2÷15, 2000.

44. Haas Z. J., Deng J., Tabrizi S.: Collision-Free Medium Access Control Scheme for Ad-Hoc Networks. Proceedings of IEEE MILCOM'99, Nov. 1999, s. 276÷280.
45. Haas Z. J., Deng J.: Dual Busy Tone Multiple Access (DBTMA) – Performance Evaluation. Proceedings of VTC'99, May 1999, Vol. 1, s. 314÷319.
46. Haas Z. J., Deng J.: Dual Busy Tone Multiple Access (DBTMA) – Performance Results. Proceedings of Wireless Communications and Networking Conference WCNC 1999, Sept. 1999, Vol. 3, s. 1328÷1332.
47. Harald Baumgart – Entwicklung elektronischer Komponenten. [@:] http://www.hbtron.de/. Dostęp: 16.10.2011.
48. Hołubowicz W., Płóciennik P.: GSM - cyfrowy system telefonii komórkowej. Wyd. EFP, Poznań 1995.
49. Hołubowicz W., Płóciennik P.: Cyfrowe systemy telefonii komórkowej GSM 900, GSM 1800, UMTS. Wyd. EFP, Poznań 1998.
50. Hołubowicz W., Płóciennik P., Różański A.: Systemy łączności bezprzewodowej. Wydawnictwa EFP, Poznań 1996.
51. IEEE Recommended Practice for Information technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – Specific requirements. Part 15.2: Coexistence of Wireless Personal Area Networks with Other Wireless Devices Operating in Unlicensed Frequency Bands. IEEE Std 802.15.2™-2003. New York 2003.
52. IEEE Standard for Information technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE Std 802.11™-2007. IEEE, New York 2007.
53. IEEE Standard for Information technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. ANSI/IEEE Std 802.11, 1999 Edition. IEEE, New York 1999.
54. IEEE Standard for Information technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput. IEEE P802.11n™-2009. IEEE, New York 2009.

55. IEEE Standard for Information technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – Specific requirements. Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs). IEEE Std 802.15.3™-2003, IEEE, New York 2003.
56. Information technology – Telecommunications and information exchange between systems – High-level data link control (HDLC) procedures. ISO/IEC 13239:2002, ISO, Genewa 22.08.2002.
57. Information technology – Telecommunications and information exchange between systems – High-level data link control procedures – Description of the X.25 LAPB-compatible DTE data link procedures. ISO/IEC 7776:1995, ISO, Genewa 22.06.1995.
58. Intel® Ultimate N WiFi Link 5300 and Intel® WiFi Link 5100 Products. [@:] http://www.intel.com/network/connectivity/products/wireless/adapters/5000/index.htm. Dostęp: 16.10.2011.
59. Joong s. Ma.: On the Impact of HDLC Zero Insertion and Deletion on Link Utilization and Reliability. IEEE Transactions on Communications, Vol. COM-30, No. 2, Feb. 1982, s. 375÷381.
60. Kahn J. M., Barry J. R.: Wireless Infrared Communications. Proceedings of the IEEE, Vol. 85, No. 2, Feb. 1997, s. 265÷298.
61. Kantronics Radio Modems/TNC's. [@:] http://www.kantronics.com/modems.html.
62. Karn P.: MACA – A New Channel Access Method for Packet Radio. Proceedings of ARRL/CRRL Amateur Radio 9th Computer Networking Conference, 1990, s. 134÷140.
63. Karn P. R., Price H. E., Diersing J.: Packet Radio in the Amateur Service. IEEE Journal on Selected Areas of Communications, vol. 3, no. 3, May 1985, s. 431÷439.
64. Kleinrock L., Scholl M. O.: Packet Switching in Radio Channels: New Conflict-Free Multiple Access Schemes. IEEE Transactions on Communications, Vol. COM-28, No. 7, Jul. 1980, s. 1015÷1029.
65. Kleinrock L., Tobagi F. A.: Packet Switching in Radio Channels: Part I – Carrier Sense Multiple-Access Modes and Their Throughput-Delay Analysis. IEEE Transactions on Communications, Vol. COM-23, No. 12, Dec. 1975, s. 1400÷1416.
66. Контроллер пакетной радиосвязи TNC3W. [@:] http://ur7iwz.qrz.ru/hw/tnc3w.php.
67. Linksys WRT54G v3.1 Wireless-G Broadband Router. [@:] http://www.modem-help.co.uk/Linksys/WRT54G-v3-1-Wireless-G-Broadband-Router.html
68. Lo W. F., Mouftah H. T.: Carrier Sense Multiple Access with Collision Detection for Radio Channels. Proceedings of 13th International Communications and Energy Conference, 1984, s. 244÷247.

69. Lo W. F., Mouftah H. T.: Collision Detection and Multitone Tree Search for Multiple-Access Protocols on Radio Channels. IEEE Journal on Selected Areas of Communications, Vol. 5, No. 6, July 1987, s. 1035÷1040.
70. Mangold S., Choi S., May P., Klein O., Hiertz G., Stibor L.: IEEE 802.11e Wireless LAN for Quality of Service. Proceedings of European Wireless 2002, Florence, Italy, s. 32÷39.
71. Mangold S., Choi S., Hiertz G. R., Klein O., Walke B.: Analysis of IEEE 802.11e for QoS Support in Wireless LANs. IEEE Wireless Communications, Vol. 10, No. 6, Dec. 2003, s. 40÷50.
72. Metzner J.: On Improving Utilization in Aloha Networks. IEEE Transactions on Communications, Vol. COM-24, No. 4, Apr. 1976, s. 447÷448.
73. MUEL. [@:] http://www.muel.pl/. Dostęp: 16.10.2011.
74. MX604 V.23 Compatible Modem. MX●COM Inc., Winston-Salem, 1998.
75. MX469 1200/2400/4800bps MSK Modem. MX●COM Inc., Winston-Salem, 1998.
76. Myers A., Basagni S.: Wireless Media Access Control. W: Stojmenović I. (red.): Handbook of Wireless Networks and Mobile Computing. Wiley, 2002, rozdz. 6, s. 119÷143.
77. Nachrichtentechnik Marten Güttner. [@:] http://www.nt-g.de/. Dostęp: 16.10.2011
78. Nelson R., Kleinrock L.: Spatial TDMA: A Collision-Free Multihop Channel Access Protocol. IEEE Transaction on Communications, Vol. COM-33, No. 9, Sep. 1985, s. 934÷944.
79. Ni Q., Romdhani L., Turletti T.: A Survey of QoS Enhancements for IEEE 802.11 Wireless LAN. Journal of Wireless Communications and Mobile Computing, Vol. 4, No. 5, 2004, s. 547÷566.
80. Nowicki K., Woźniak J.: Przewodowe i bezprzewodowe sieci LAN. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2002.
81. Nowicki K., Woźniak J.: Sieci LAN, MAN i WAN – protokoły komunikacyjne. Wydawnictwo Fundacji Postępu Telekomunikacji, Kraków 1998.
82. PacComm Packet Radio Systems. [@:] http://www.paccomm.com/. Dostęp: 16.10.2011
83. Pahlavan K., Levesque A. H.: Wireless Data Communications. Proceedings of the IEEE, Vol. 82, No. 9, Sept. 1994, s. 1398÷1430.
84. Prasad N., Prasad A. (eds.): WLAN Systems and Wireless IP for Next Generation Communications. Artech House, 2002.
85. Qiao D., Choi S., Shin K. G.: Goodput Analysis and Link Adaptation for IEEE 802.11a Wireless LANs. IEEE Transactions on Mobile Computing, Vol. 1, No. 4, Oct-Dec 2002, s. 278÷292.
86. Roberts L. G.: Aloha packet system with and without slots and capture. ACM SIGCOMM Computer Communication Review, Vol. 5 , No. 2, Apr. 1975, s. 28÷42.

87. Rom R.: Collision Detection in Radio Channels. W: Pickholtz R. L. (red.): Local Area & Multiple Access Networks. Computer Science Press, 1986, rozdz. 12, s. 235÷249.
88. Rom R., Sidi M.: Multiple Access Protocols. Performance and Analysis. Springer Verlag, New York 1990.
89. Roshan P., Leary J.: Bezprzewodowe sieci LAN 802.11. Podstawy. Mikom, Warszawa 2004.
90. Rozporządzenie Ministra Transportu z dnia 3 lipca 2007 r. w sprawie urządzeń radiowych nadawczych lub nadawczo-odbiorczych, które mogą być używane bez pozwolenia radiowego. Dziennik Ustaw Nr 138, Poz. 972, s. 10051÷10071.
91. Sankar K., Sundaralingam S., Balinsky A., Miller D.: Bezpieczeństwo sieci bezprzewodowych. Ochrona sieci 802.11. Porady eksperta. Mikom, Warszawa 2004.
92. SCS. [@:]http://www.scs-ptc.com/. Dostęp: 16.10.2011.
93. Shenker S.: Some Conjectures on the Behavior of Acknowledgement-Based Transmission Control of Random Access Communication Channels. ACM SIGMETRICS Performance Evaluation Review, Vol. 15 , No. 1, May 1987, s. 245÷255.
94. Simoens, S., Pellati, P., Gosteau, G., Gosse, K., Ware, C.: The Evolution of 5 GHz WLAN toward Higher Throughputs. IEEE Wireless Communications, Vol. 10, No. 6, Dec. 2003, s. 6÷13.
95. Skordoulis D., Ni Q., Chen H., Stephens A. P., Liu C., Jamalipour A.: IEEE 802.11n MAC Frame Aggregation Mechanisms for Next-Generation High-Throughput WLANs. IEEE Wireless Communications, Vol. 15, No. 1, Feb. 2008, s. 40÷47.
96. Specification of the Bluetooth System. Core Version 3.0 + HS. Bluetooth SIG, 2009.
97. Seaborne A., Williams S., Novak F., Suvak D., Pennington R., Williams T.: Infrared Data Association Link Management Protocol Version 1.1. Infrared Data Association, Walnut Creek, 23.01.1996.
98. Symek Packet-Radio. [@:] http://www.symek.de/. Dostęp: 16.10.2011.
99. Szóstka J.: Fale i anteny. WKŁ, Warszawa 2000.
100. Takagi H., Kleinrock L.: Output Processes in Contention Packet Broadcasting Systems. IEEE Transactions on Communications, Vol. COM-33, No. 11, Nov. 1985, s. 1191÷1199.
101. Talucci F., Gerla M., Fratta L.: MACA-BI (MACA By Invitation): A Receiver Oriented Access Protocol For Wireless Multihop Networks. Proceedings of IEEE PIMRC '97, Vol. 2, s. 435÷439.
102. Tan W.-S., Petrilla J., Hirt W., Yuuki Y., Quek R., Tajnai J. (red.): Infrared Data Association Serial Infrared Physical Layer Specification Version 1.4. Infrared Data Association, Walnut Creek, 30.05.2001.
103. Tanenbaum A. S.: Computer Networks (4th ed.). Prentice Hall, 2003.

104. Tannenbaum A. S.: Sieci komputerowe. WNT, Warszawa 1988.
105. TCM3105DWL, TCM3105JE, TCM3105JL, TCM3105NE, TCM3105NL FSK Modem. Texas Instruments Inc., May 1994.
106. TimeWave Main Page. [@:] http://www.timewave.com/. Dostęp: 16.10.2011.
107. Tobagi F. A., Kleinrock L.: Packet Switching in Radio Channels: Part II – The Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy-Tone Solution. IEEE Transactions of Communications, Vol. COM-23, No. 12, Dec. 1975, s. 1417÷1433.
108. Tobagi F. A., Kleinrock L.: Packet Switching in Radio Channels: Part III – Polling and (Dynamic) Split-Channel Reservation Multiple Access. IEEE Transactions on Communications, Vol. COM-24, No. 8, Aug. 1976, s. 832÷845.
109. Wade I. (red.): Automatic Position Reporting System. APRS Protocol Reference. Protocol Version 1.0. Tucson Amateur Packet Radio Corporation, Tucson 2000.
110. Wesołowski K.: Systemy radiokomunikacji ruchomej. WKŁ, Warszawa 1999.
111. Williams T., Smith K., Suvak D., Cremer M., Sell D.: Infrared Data Association Serial Infrared Link Access Protocol (IrLAP) Version 1.1. Infrared Data Association, Walnut Creek, 16.06.1996.
112. Wojnar A.: Systemy radiokomunikacji ruchomej lądowej. Podstawy analizy i syntezy. WKiŁ, Warszawa 1989.
113. Wolisz A.: Podstawy lokalnych sieci komputerowych. Tom 1: Sprzęt sieciowy. WNT, Warszawa 1992.
114. Wu C., Li. V.: Receiver-Initiated Busy Tone Multiple Access in Packet Radio Networks. ACM SIGCOMM Computer Communication Review, Vol. 17, No. 5, Oct/Nov 1987, s. 336÷342.
115. Xiao Y., Rosdahl J.: Throughput and Delay Limits of IEEE 802.11. IEEE Communications Letters, Vol. 6, No. 8, Aug. 2002, s. 355÷357.
116. Xiao Y.: Packing Mechanisms for the IEEE 802.11n Wireless LANs. Proc. IEEE GLOBECOM'04, Vol. 5, 2004, s. 3275÷3279.
117. Xiao Y.: IEEE 802.11n: Enhancements for Higher Throughput in Wireless LANs. IEEE Wireless Communications, Vol. 12, No. 6, Dec. 2005, 82÷91.
118. Zhu C., Corson M. S.: A Five-Phase Reservation Protocol (FPRP) for Mobile Ad Hoc Networks. Wireless Networks, Vol. 7, No. 4, Jul. 2001, s. 371÷384.
119. Zieliński B.: A Comparison of Collision Avoidance Methods Effectiveness in a Mobile Ad-hoc Network. Proceedings of PhysCon 2007, [@:] http://lib.physcon.ru/?item=1186 (dostęp 16.10.2011). Rozszerzone streszczenie w: Kurths J., Fradkov A., Chen G. (red.): The 3rd International Conference on Physics and Control (PhysCon 2007). Abstract Collection. Universitätsverlag Potsdam, Potsdam, Germany, 2007, s. 298.

120. Zieliński B.: A Comparison of Various Types of TNC Controllers. Theoretical and Applied Informatics, Vol. 22 (2010), No. 3, s. 187÷202.
121. Zieliński B.: A comparison of various versions of TNC controllers. W: Slanina Z., Srovnal V. (red.): Preprints of IFAC Workshop on Programmable Devices and Embedded Systems PDeS 2009, Rožnov pod Radhoštěm, 10-12.02.2009, s. 80÷85.
122. Zieliński B.: An analytical model of TNC controller. Theoretical and Applied Informatics, Vol. 21 (2009), No. 1, s. 7÷22.
123. Zieliński B.: Analiza opóźnień w sieci Packet Radio zawierającej kontrolery TNC. W: Zieliński Z. (red.): Systemy czasu rzeczywistego. Postępy badań i zastosowania. WKŁ, Warszawa 2009, chapter 29, s. 333÷342.
124. Zieliński B.: Bezprzewodowe sieci ATM – protokoły dostępu do łącza z podziałem czasu (TDD). W: Kwiecień A, Grzywak A. (red.): Współczesne problemy sieci komputerowych. Zastosowanie i bezpieczeństwo. WNT, Warszawa 2004, rozdz. 19, s. 161÷168.
125. Zieliński B.: Bezprzewodowe sieci ATM – protokoły dostępu do łącza z podziałem częstotliwości (FDD). W: Kwiecień A, Grzywak A. (red.): Współczesne problemy sieci komputerowych. Zastosowanie i bezpieczeństwo. WNT, Warszawa 2004, rozdz. 18, s. 153÷160.
126. Zieliński B.: Bezprzewodowe sieci komputerowe. Helion, Gliwice 2000.
127. Zieliński B.: Bezprzewodowe sieci komputerowe wykorzystujące konwersję protokołów. Rozprawa doktorska, Instytut Informatyki Politechniki Śląskiej, Gliwice 1997.
128. Zieliński B.: Buffer Capacity Adjustment for TNC Controller. W: Kwiecień A., Gaj P., Stera P. (red.): Computer Networks CN'09 Proceedings. Communications in Computer and Information Science Series, Vol. 39, s. 119÷126.
129. Zieliński B.: Channel utilisation for contention MAC protocols using control frames exchange. Theoretical and Applied Informatics, Vol. 19 (2007), No. 1, s. 71÷85.
130. Zieliński B.: Contention MAC Protocols Efficiency Testing in a Small Wireless Network. DOI: 10.1109/ICDT.2006.27. Proceedings of International Conference of Digital Telecommunications ICDT 2006, Cap Esterel, Côte d'Azur, 2006, s. 58.
131. Zieliński B.: Contention Wireless MAC Protocols Using Carrier Detection. Theoretical and Applied Informatics, Vol. 18 (2006), No. 1, s. 75÷87.
132. Zieliński B.: Contention wireless medium access protocols using control frames exchange. Archive of Theoretical and Applied Computer Science, Vol. 16 (2004), No. 2, s. 121÷133.
133. Zieliński B.: Effective Transmission Speed in AX.25 Protocol. Proceedings of IEEE EUROCON 2009, Sankt Petersburg, 18-23.05.2009, s. 1763÷1768. DOI: 10.1109/EURCON.2009.5167882.

134. Zieliński B.: Efficiency Analysis of IEEE 802.11 Protocol with Block Acknowledge. W: Grzech A., Borzemski L., Świątek J., Wilimowska Z.: Information Systems Architecture and Technology. Networks and Networks' Services. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2010, chapter 10, s. 125÷135.
135. Zieliński B.: Efficiency analysis of IEEE 802.11 protocol with block acknowledge and frame aggregation. Bulletin of Polish Academy of Sciences: Technical Sciences, Vol. 59, No. 2, Mar 2011, s. 235÷243.
136. Zieliński B.: Efficiency Comparison of Classical Contention MAC Protocols in a Wireless Network. Archive of Theoretical and Applied Computer Science, Vol. 17 (2005), No. 3, s. 157÷172.
137. Zieliński B.: Efficiency estimation of AX.25 protocol. Theoretical and Applied Informatics, Vol. 20 (2008), No. 3, s. 199÷214.
138. Zieliński B.: Efficiency estimation of IEEE 802.11 protocol. W: Grzech A., Borzemski L., Świątek J., Wilimowska Z. (red.): Information Systems Architecture and Technology. Information Systems and Computer Communication Networks. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2008, chapter 10, s. 125÷136.
139. Zieliński B.: Factors Having Influence upon Efficiency of an Integrated Wired-Wireless Network. W: Cyran K. A., Kozielski S., Peters J. F., Stańczyk U., Wakulicz-Deja A. (red.): Man-Machine Interactions. Advances in Intelligent and Soft Computing, Vol. 59, s. 647÷654, Springer-Verlag, Berlin Heidelberg 2009. International Conference on Man-Machine Interactions, Kocierz, 25-27.09.2009.
140. Zieliński B.: Further considerations on AX.25 protocol performance. W: Grzech A., Borzemski L., Świątek J., Wilimowska Z. (red.): Information Systems Architecture and Technology. Service Oriented Distributed Systems: Concepts and Infrastructure. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2009, chapter 28, s. 327÷336.
141. Zieliński B.: Influence of protocol converter processing power upon network efficiency. W: Bradac Z., Zezulka F., Polansky M., Jirsik V. (red.): Proceedings of IFAC Workshop on Programmable Devices and Embedded Systems PDeS 2006, Brno 2006, s. 38÷43.
142. Zieliński B.: Medium access protocols for wireless ATM networks with frequency division duplex link. W: Grzech A., Wilimowska Z. (red.): Proceedings of Information Systems Architecture and Technology ISAT 2005 Conference, Szklarska Poręba 2005, s. 258÷264.
143. Zieliński B.: Medium access protocols for wireless ATM networks with time division duplex link. W: Grzech A., Wilimowska Z. (red.): Proceedings of Information Systems Architecture and Technology ISAT 2005 Conference, Szklarska Poręba 2005, s. 250÷257.

144. Zieliński B.: Medium access protocols for wireless networks – problem description. Archive of Theoretical and Applied Computer Science, Vol. 15 (2003), No. 4, s. 371÷383.
145. Zieliński B.: Metody unikania i wykrywania kolizji dla sieci ad-hoc. W: Kwiecień A., Grzywak A. (red.): Współczesne problemy sieci komputerowych. Zastosowanie i bezpieczeństwo. WNT, Warszawa 2004, rozdz. 20, s. 169÷176.
146. Zieliński B.: Model analityczny kontrolera TNC. W: Mazur Z., Huzar Z. (red.): Modele i zastosowania systemów czasu rzeczywistego. WKŁ, Warszawa 2008, rozdz. 11, s. 127÷136.
147. Zieliński B.: Ocena efektywności protokołu AX.25. W: Gaj P., Pochopień B., Kozielski S. (red.): Współczesne aspekty sieci komputerowych. Tom 1. WKŁ, Warszawa 2008, rozdz. 12, s. 127-136.
148. Zieliński B.: Performance Evaluation of Various Implementations of AX.25 Protocol. Theoretical and Applied Informatics, Vol. 21 (2009), No. 3-4, s. 225-237.
149. Zieliński B.: Physical properties and medium access control for wireless ad-hoc networks. DOI: 10.1109/PHYCON.2005.1513959. W: Fradkov A. L., Churilov A. N. (red.): PhysCon 2005 International Conference Physics and Control Proceedings, Sankt Petersburg 2005, s. 106-111.
150. Zieliński B.: Porównanie metod unikania kolizji w sieciach bezprzewodowych zawierających stacje ruchome. W: Węgrzyn S., Znamirowski L., Czachórski T., Kozielski S. (red.): Nowe technologie sieci komputerowych. Tom 1. WKŁ, Warszawa 2006, rozdz. 13, s. 145-152.
151. Zieliński B.: Porównanie wydajności różnych wersji kontrolerów TNC. W: Gaj P., Pochopień B., Kozielski S. (red.): Współczesne aspekty sieci komputerowych. Tom 1. WKŁ, Warszawa 2008, rozdz. 13, s. 137-146.
152. Zieliński B.: Software Influence upon AX.25 Protocol Performance. W: Kwiecień A., Gaj P., Stera P. (red.): Computer Networks CN'09 Proceedings. Communications in Computer and Information Science Series, Vol. 39, s. 111-118. DOI: 10.1007/978-3-642-02671-3_13.
153. Zieliński B.: Throughput Upper Limit for IEEE 802.11 Networks with Block Acknowledge and Frame Aggregation. W: Kwiecień A., Gaj P., Stera P. (red.): Computer Networks CN'10 Proceedings. Communications in Computer and Information Science Series, Vol. 79/2010, s. 67-75.
154. Zieliński B.: Using TNC controller as a model of protocol converter. W: Pułka A., Hrynkiewicz E., Kłosowski P. (red.): Proceedings of IFAC Workshop on Programmable Devices and Systems PDS2004, Kraków 2004, s. 269-273.

155. Zieliński B.: Wpływ warstwy fizycznej na wydajność protokołu IEEE 802.11. W: Kwiecień A., Gaj P., Jestratjew A. (red.): Techniczne i teoretyczne aspekty współczesnych sieci komputerowych. WKŁ, Warszawa 2009, chapter 9, s. 89-98.
156. Ziouva E., Antonakopoulos T.: CSMA/CA performance under high traffic conditions: throughput and delay analysis. Computer Communications, Vol. 25, No. 3, Feb. 2002, s. 313-321.

PROTOKOŁY WARSTWY LINIOWEJ W BEZPRZEWODOWYCH SIECIACH KOMPUTEROWYCH

STRESZCZENIE

Monografia jest poświęcona budowie i analizie właściwości protokołów warstwy liniowej w bezprzewodowych sieciach komputerowych.

W początkowej częci rozprawy określono zadania stawiane przed warstwą liniową w sięciach komputerowych, ze szczególnym uwzględnieniem sieci bezprzewodowych. Następnie sformułowano cel rozprawy i dokonano przeglądu zagadnień w niej poruszanych.

Zasadnicza część rozprawy jest podzielona na trzy rozdziały.

Pierwszy rozdział jest poświęcony zagadnieniu protokołów dostępu do łącza w sieciach bezprzewodowych, ze szczególnym uwzględnieniem sieci ad hoc. W części tej opisano zjawiska, występujące w sieciach bezprzewodowych, a mające istotny wpływ na działanie protokołu dostępu do łącza. Następnie przedstawiono metody unikania i wykrywania kolizji, możliwe do zrealizowania w sieciach bezprzewodowych. Na podstawie tych opisów określono warunki, w których wykrywanie kolizji – pomimo występowania efektu przechwytywania – jest możliwe w sieciach wykorzystujących promieniowanie podczerwone z wiązką rozproszoną. Porównano także zachowanie dwóch metod unikania kolizji w sieci ad hoc, zawierającej stacje ruchome i określono kryterium skuteczności unikania kolizji metodą wymiany ramek sterujących. Kolejny fragment rozdziału przedstawia rywalizacyjne protokoły dostępu do łącza, zaprojektowane dla sieci bezprzewodowych. W protokołach tych są stosowane opisane wcześniej metody unikania i wykrywania kolizji. Dla wybranych protokołów przeprowadzono analizę wydajności w różnych warunkach pracy sieci, włączając warunki typowe dla kilku przypadków istniejących sieci bezprzewodowych. Wybrane protokoły zostały także zaimplementowane w małej, doświadczalnej sieci bezprzewodowej, w której dokonano pomiaru ich wydajności dla kilku wybranych konfiguracji. Uzyskane wyniki doświadczalne odbiegają nieco od wyników analitycznych, co może świadczyć o występowaniu w sieci zjawisk, które nie zostały uwzględnione w modelu, np. efektu przechwytywania.

Drugi rozdział poświęcono sieci Packet Radio oraz stosowanemu w niej protokołowi AX.25 i kontrolerom TNC, które są używane jako adaptery tej sieci. W tej części opisano zasady działania protokołu AX.25, z uwzględnieniem najnowszej wersji (2.2) oraz ważniejszych różnic w stosunku do wersji wcześniejszych (szczególnie najczęściej stosowanej wersji 2.0). Opisano także ogólną zasadę pracy kontrolerów TNC, dokonano przeglądu ich konstrukcji i głównych funkcji oprogramowania. Na podstawie opisu protokołu AX.25 stworzono jego model analityczny, umożliwiający określenie jego wydajności, efektywnej prędkości transmisji oraz opóźnień występujących podczas przesyłu danych. Wykorzystując stworzony model, przeanalizowano wpływ poszczególnych parametrów protokołu na jego wydajność dla obu wariantów łącza radiowego, pracującego z różnymi prędkościami transmisji. Wykorzystując model protokołu AX.25, stworzono także model analityczny kontrolera TNC. Model ten pozwala oszacować teoretyczny wpływ kontrolera na efektywną prędkość oraz opóźnienia transmisji, pozwala także oszacować pojemność bufora w kontrolerze TNC, która gwarantuje ciągłość transmisji po stronie nadawczej. Wykonano także liczne badania w doświadczalnej sieci Packet Radio. Wyniki badań wykazały silną zależność parametrów użytkowych sieci zarówno od mocy obliczeniowej kontrolera TNC, jak i od oprogramowania sterującego jego pracą, a w szczególności od szczegółów implementacji protokołu AX.25.

Trzeci rozdział monografii poświęcono bezprzewodowym sieciom lokalnym zgodnym ze standardem IEEE 802.11. W monografii opisano topologie sieci, określone w standardzie, a także wybrane elementy warstwy fizycznej i liniowej. Opisano także formaty ramek podwarstwy PLCP, formaty i typy ramek warstwy liniowej, zasady ich wymiany w różnych wariantach dostępu do łącza (DCF, PFC, EDCA, HCCA) oraz wybrane elementy zarządzania siecią. Opisano także mechanizm potwierdzenia blokowego. Osobny fragment poświęcono najnowszemu rozwiązaniu – dodatkowi IEEE 802.11n, który został ukończony w trakcie pisania monografii (listopad 2009). W tym fragmencie opisano wybrane aspekty warstwy fizycznej – szczególnie formaty ramek na poziomie podwarstwy PLCP – oraz rozszerzenia warstwy liniowej – agregację ramek A-MSDU i A-MPDU. Mechanizmy te poddano analizie pod kątem ich wpływu na wydajność protokołu w różnych warunkach, a co za tym idzie, możliwej do uzyskania maksymalnej teoretycznej prędkości transmisji. Na podstawie opisu zasad wymiany ramek na poziomie warstwy liniowej oraz formatów stosowanych na poziomie podwarstwy PLCP wyznaczono zależności analityczne, które pozwalają na oszacowanie wydajności protokołu dla różnych wariantów warstwy fizycznej, prędkości transmisji, pojemności pola danych ramki oraz przyjętej zasady wymiany ramek. Wyprowadzone zależności pozwalają oszacować wydajność protokołu oraz efektywną prędkość transmisji na poziomie warstwy liniowej. Dla celów obliczeń przyjęto, że transmisja odbywa się w cyklicznie powtarzających się, identycznych fragmentach, które nazwano cyklami transmisyjnymi. Dzięki te-

mu, analizując transmisję na poziomie pojedynczego cyklu transmisyjnego, można uzyskać wyniki niezależne od całkowitej objętości przesyłanej informacji.

W podsumowaniu wskazano na nieomówione w monografii, a również ważne dla działania sieci zagadnienia wynikające z faktu stosowania transmisji bezprzewodowej.

DATA LINK LAYER PROTOCOLS IN WIRELESS COMPUTER NETWORKS

ABSTRACT

The monograph is devoted to the construction and analysis of the properties of the data link layer protocols in wireless networks

In the initial part, data link layer tasks, especially in wireless computer networks, are characterized. Next, the aim of the monograph is formulated and a review of presented problems is given.

The main part of the monograph is divided into three chapters.

The first chapter is devoted to medium access protocols in wireless networks, especially ad-hoc networks. In this part, phenomena present in wireless network that have influence on MAC protocols operation, have been characterized. Next, collion avoidance and detection methods for wireless networks have been presented. On the basis of their operation, the conditions that allow to use collision detection in wireless diffuse infrared networks have been determined. Two collision avoidance methods behaviour in a mobile ad-hoc network is also compared and the criterion of control frames exchange-based collision avoidance method effectiveness is given. Next, contention MAC protocols for wireless networks are presented. In these protocols, the aforementioned collision avoidance and detection methods are used. For selected protocols, performance comparison in various operating conditions, including those typical for wireless networks, is presented. Selected protocols have been also implemented is a small, experimental wireless networks, in which their properties have been measured. The experimental results are different from the theoretical ones, which may show that in the network there are some phenomena not taken into account in the theoretical model.

The second chapter is devoted to the amateur Packet Radio network together with AX.25 protocol and TNC controllers used as network adapters in this network. In this part, AX.25 protocol operation is presented according to its 2.2 version with emphasize to the differences to the earlier version (especially widely used 2.0). The chapter presents also TNC operation rules, their hardware structure and software functions. On the basis of AX.25 protocol de-

scriprtion an analytical model has been created; it allows to estimate protocol efficiency, effective throughput and transmission delays. Using this model, protocol parameter influence on its performance has been analysed for both radio link types, opearating at different transmission rates. A similar model of TNC controller is also presented; it allows to estimate TNC influence on throughput and delays, and to estimate TNC buffer capacity that guarantees sender-side transmission continuity. Numerous tests performed in an axperimetal Packet Radio networks are also presented. Results show strong dependency of effective network parameters on TNC processing power as well as software-dependent AX.25 protocol implementation details.

The third chapter is devoted to IEEE 802.11-comatible wireless local area networks. In this chapter, network topologies and selected elements of physical and data link layer are presented, such as: PLCP frame formats, data link layer frame formats, types and exchange rules in various medium access modes (DCF, PCF, EDCA, HCCA) and selected network management elements. Block acknowledge mechanism is also described. A separate part of the chapter describes 802.11n amendment that was finished during writing of this monograph (November 2009). In this fragment, PLCP frame formats as well as frame aggregation (A-MSDU and A-MPDU) methods are discussed. These mechanisms are then analysed for their influence on protocol performance under various conditions. On the basis of protocol description, an analytical model known from literature has been extended. It allows estimate protocol performance, effective throughput and throughput upper limit for all considered data transmission methods.

In the summary, interesting issues resulting from wireless transmission, that are not discussed in the monograph, are presented.

INFORMATION FOR AUTHORS

The journal *STUDIA INFORMATICA* publishes both fundamental and applied Memoirs and Notes in the field of informatics. The Editors' aim is to provide an active forum for disseminating the original results of theoretical research and applications practice of informatics understood as a discipline focused on the investigations of laws that rule processes of coding, storing, processing, and transferring of information or data.

Papers are welcome from fields of informatics inclusive of, but not restricted to *Computer Science*, *Engineering*, and *Life and Physical Sciences*.

All manuscripts submitted for publication will be subject to critical review. Acceptability will be judged according to the paper's contribution to the art and science of informatics.

In the first instance, all text should be submitted as hardcopy, conventionally mailed, and for accepted paper accompanying with the electronically readable manuscript to:

Dr. Marcin SKOWRONEK
Institute of Informatics
Silesian University of Technology
ul. Akademicka 16
44-100 Gliwice, Poland
Tel.: +48 32 237-12-15
Fax: +48 32 237-27-33
e-mail: marcin.skowronek@polsl.pl

MANUSCRIPT REQUIREMENTS

All manuscripts should be written in Polish or in English. Manuscript should be typed on one side paper only, and submitted in duplicate. The name and affiliation of each author should be followed by the title of the paper (as brief as possible). An abstract of not more than 50 words is required. The text should be logically divided under numbered headings and subheadings (up to four levels). Each table must have a title and should be cited in the text. Each figure should have a caption and have to be cited in the text. References should be cited with a number in square brackets that corresponds to a proper number in the reference list. The accuracy of the references is the author's responsibility. Abbreviations should be used sparingly and given in full at first mention (e.g. "Central Processing Unit (CPU)"). In case when the manuscript is provided in Polish (English) language, the summary and additional abstract (up to 300 words with reference to the equations, tables and figures) in English (Polish) should be added.

After the paper has been reviewed and accepted for publication, the author has to submit to the Editor a hardcopy and electronic version of the manuscript.

It is strongly recommended to submit the manuscript in a form downloadable from web site http://zti.polsl.pl/makiety/.

To subscribe: *STUDIA INFORMATICA* (PL ISSN 0208-7286) is published by Silesian University of Technology Press (Wydawnictwo Politechniki Śląskiej) ul. Akademicka 5, 44-100 Gliwice, Poland, Tel./Fax +48 32 237-13-81. 2011 annual subscription rate: US$60. Single number price approx. US$10-20 according to the issue volume.

INSTYTUT INFORMATYKI prowadzi:

- ❑ Studia stacjonarne I stopnia (inżynierskie)
- ❑ Studia stacjonarne II stopnia (magisterskie)
- ❑ Studia niestacjonarne I stopnia (inżynierskie)
- ❑ Studia niestacjonarne II stopnia (magisterskie)
- ❑ Studia podyplomowe:
  - *Sieci i systemy komputerowe, bazy danych*
  - *Systemy informacji geograficznej*
  - *Teleinformatyka w transporcie lotniczym*
  - *Technologie internetowe i technologie mobilne*
  - *Metody eksploracji baz danych przedsiębiorstw*
- ❑ Studia doktoranckie

Informacje:

POLITECHNIKA ŚLĄSKA
Instytut Informatyki

44-100 Gliwice, ul. Akademicka 16
tel. (032) 237 24 05; 237 21 51;
fax (032) 237 27 33 (czynny całą dobę)

e-mail: rau2@polsl.pl

http://www.inf.polsl.pl (dydaktyka)