ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 1999
Seria: INFORMATYKA z. 36 Nr kol. 1414

Adam KAPRALSKI
Politechnika Śląska, Instytut Informatyki

MASZYNA DO GŁĘBOKIEGO WYSZUKIWANIA JAKO MASZYNA KODUJĄCA i DEKODUJĄCA

Streszczenie. W artykule przedstawiono podstawowe koncepcje wykorzystania maszyn do głębokiego wyszukiwania jako maszyn kodujących i dekodujących. Wskazano na właściwości maszyn do głębokiego wyszukiwania czyniące je szczególnie przydatnymi do budowy systemu przesyłania wiadomości szyfrowanych odpornych na złamanie. Przedstawiona tutaj ogólna koncepcja szyfrowania i odszyfrowywania może być również zaimplementowana w architekturach tradycyjnych. Jednakże zastąpienie identyfikacji wykonywanej w maszynach do głębokiego wyszukiwania przeszukiwaniem zupełnym w architekturach tradycyjnych znacznie ogranicza rozmiary możliwych do zastosowania książek szyfrów, z uwagi na wzrastający czas szyfrowania/odszyfrowywania.

DEPTH SEARCH MACHINES AS CODING AND DECODING UNITS

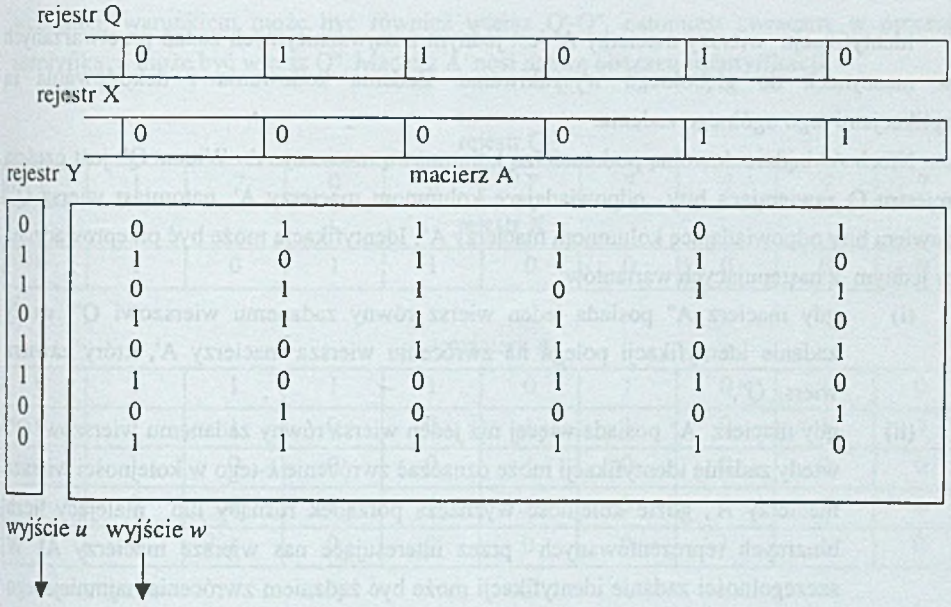
Summary. It is shown that DSMs (Depth Search Machines) can be successfully applied for building communication system using coded mails. Each node of the system contains two DSMs. The main process applied for coding and decoding is identification performed in DSMs. Usage of DSMs gives greater resistance of the system against possible actions of hackers in comparison with traditional architectures applied. This greater resistance is affected by possibility of applying greater books of codes and each book can be used in more ways than it could be done when traditional architectures were used. Usage as big books for traditional architectures as could be used for DSMs would effect much the time of processing

1. Architektura i podstawowe przetwarzanie w maszynach do głębokiego wyszukiwania

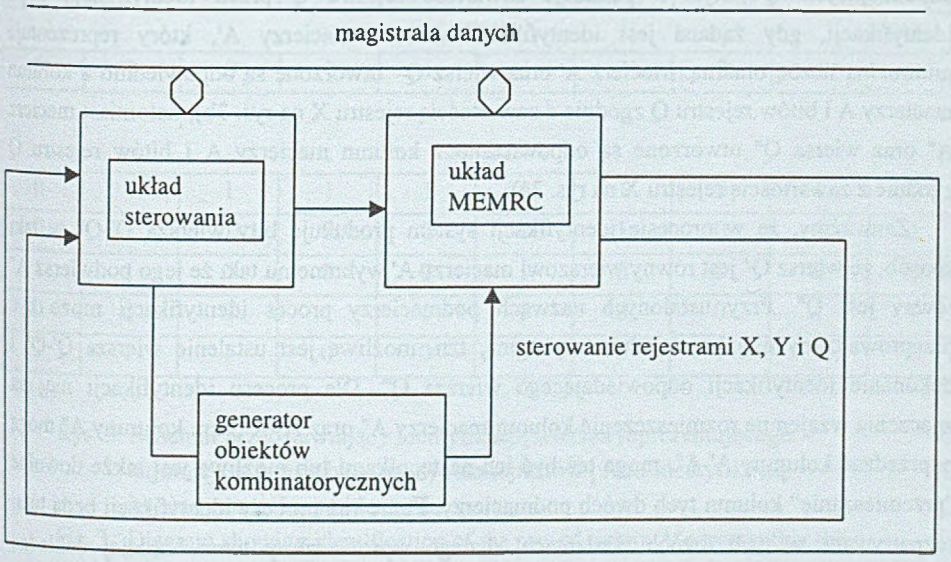
Maszyny do głębokiego wyszukiwania są nowym uniwersalnym modelem maszyny cyfrowej. Podstawowym komponentem tych maszyn jest układ MEMRC, którego schemat blokowy przedstawia rys. 1. Macierz A może być uważana za procesor asocjacyjny o strukturze komórkowej. Komórkę tworzy jeden bit pamięci oraz kilka bramek. Wiersze macierzy A są rejestrami zawierającymi *m* elementarnych komórek. Każda komórka może zawierać bit pamięci dowolnego typu: może to być komórka pamięci z wpisem/odczytem w szczególności RAM, może to być pamięć wyłącznie do odczytu ROM, jak również może to być pamięć programowana PROM. W zależności od stosowanej pamięci układ MEMRC przyjmuje odpowiednio nazwę RAMRC lub ROMRC lub PROMRC.

Rejestry X oraz Y służą do maskowania kolumn i wierszy macierzy A pozwalając wybrać z macierzy A jej podmacierz A'. Ponadto rejestr X służy do maskowania bitów rejestru Q, tak że bity części tego rejestru oznaczone przez Q' odpowiadają kolumnom macierzy A tworzącym macierz A'. Rejestr Q posiada *m* bitów odpowiadających kolumnom macierzy A. Podstawowe funkcje układu MEMRC realizowane są poprzez ustalanie zawartości rejestru Q oraz rejestrów maskujących X, Y, a także poprzez obserwację jednobitowych wyjść *w* i *u*. Jeżeli w macierzy A' istnieje wiersz równy Q', wtedy wyjście *w* przyjmuje wartość 1, w przeciwnym przypadku wyjście *w* przyjmuje wartość 0. Jeżeli wszystkie wiersze podmacierzy A' równe są Q', wtedy także wyjście *u* przyjmuje wartość 1. Komponentem maszyny do głębokiego wyszukiwania może być uproszczony układ MEMRC, różniący się od wersji podstawowej brakiem rejestru maskującego Y oraz wyjścia *u*. W uproszczonym układzie MEMRC wybrana podmacierz A' jest podmacierzą kolumnową macierzy A. Schemat blokowy maszyny do głębokiego wyszukiwania przedstawiono na rys. 2.

W maszynach do głębokiego wyszukiwania można przetwarzać rekordy binarne reprezentowane poprzez wiersze macierzy A', CF bazy danych [1] oraz całe podmacierze binarne A'. Te ostatnie struktury mogą być przetwarzane tylko wówczas, gdy komponentem maszyny do głębokiego wyszukiwania jest oryginalny układ MEMRC, pozostałe dwie struktury danych mogą być przetwarzane również wtedy, gdy podstawowym komponentem jest uproszczony układ MEMRC.



Rys. 1. Schemat blokowy układu MEMRC
Fig. 1. The block diagram of MEMRC



Rys. 2. Schemat blokowy maszyny do głębokiego wyszukiwania
Fig. 2. The block diagram of depth search machine

1.1. Identyfikacja wierszy macierzy A' w maszynach do głębokiego wyszukiwania

Identyfikacja wierszy macierzy A' jest jednym z najważniejszych zadań przetwarzanych w maszynach do głębokiego wyszukiwania. Zadania kodowania i dekodowania są aplikacjami tego ogólnego zadania.

Niech A" będzie dowolną podmacierzą kolumnową macierzy A'. Wiersz Q' jest częścią rejestru Q zawierającą bity odpowiadające kolumnom macierzy A', natomiast wiersz Q" zawiera bity odpowiadające kolumnom macierzy A". Identyfikacja może być przeprowadzona w jednym z następujących wariantów:

(i) gdy macierz A" posiada jeden wiersz równy zadanemu wierszowi Q", wtedy zadanie identyfikacji polega na zwróceniu wiersza macierzy A', który zawiera wiersz Q",

(ii) gdy macierz A" posiada więcej niż jeden wiersz równy zadanemu wierszowi Q", wtedy zadanie identyfikacji może oznaczać zwrócenie *k*-tego w kolejności wiersza macierzy A', gdzie kolejność wyznacza porządek rosnący lub malejący liczb binarnych reprezentowanych przez interesujące nas wiersze macierzy A'. W szczególności zadanie identyfikacji może być żądaniem zwrócenia najmniejszego lub największego wiersza.

Zwracanie wspomnianego wiersza A' odbywa się poprzez wypełnienie bitów Q' dopełniających Q". Rys. 3 pokazuje zawartość rejestru Q przed identyfikacją i po identyfikacji, gdy żądana jest identyfikacja wiersza macierzy A', który reprezentuje minimalna liczbę binarną. Macierz A' oraz wiersz Q' utworzone są odpowiednio z kolumn macierzy A i bitów rejestru Q zgodnie z zawartością rejestru X na rys. 3b), natomiast macierz A" oraz wiersz Q" utworzone są odpowiednio z kolumn macierzy A i bitów rejestru Q zgodnie z zawartością rejestru X na rys. 3a).

Zauważmy, że w procesie identyfikacji system produkuje bity wiersza Q'-Q" w taki sposób, że wiersz Q' jest równy wierszowi macierzy A' wybranemu tak, że jego podwiersz A" równy jest Q". Przy ustalonych nazwach podmacierzy proces identyfikacji może być przeprowadzony niejako w odwrotną stronę, tzn. możliwe jest ustalenie wiersza Q'-Q" i dokonanie identyfikacji odpowiadającego wiersza Q". Dla procesu identyfikacji nie ma znaczenia wzajemne rozmieszczenie kolumn macierzy A" oraz A'-A", tzn. kolumny A" mogą poprzedzać kolumny A'-A", mogą też być ich następnikami lub możliwe jest także dowolne "przemieszanie" kolumn tych dwóch podmacierzy. Ponieważ procesy identyfikacji będą tutaj rozpatrywane w tych dwóch wariantach, dlatego wprowadzimy teraz bardziej dokładną nomenklaturę wierszy Q', Q" i Q'-Q" z uwagi na prowadzony proces identyfikacji. Jeżeli proces identyfikacji jest prowadzony dla zadanego wiersza macierzy Q", wtedy Q" nosi nazwę

warunku, natomiast Q'-Q" **jest zwracany** w procesie identyfikacji. Jak to podkreśliliśmy wcześniej, warunkiem może być również wiersz Q'-Q", natomiast zwracany w procesie identyfikacji może być wiersz Q". Macierz A' nosi nazwę **obszaru identyfikacji.**

rejestr Q

?	1	?	0	0	?	?	?	?	?

rejestr X

0	1	0	1	1	0	0	0	0	0

macierz A

0	1	1	1	1	0	1	0	1	0
0	1	1	0	0	1	1	0	1	1
1	1	0	0	0	1	0	0	0	0
0	0	0	1	1	1	0	0	1	0
1	1	1	0	0	0	0	1	1	0

a)

rejestr Q

?	1	?	0	0	0	0	?	1	0

rejestr X

0	1	0	1	1	1	1	0	1	1

macierz A

0	1	1	1	1	0	1	0	1	0
0	1	1	0	0	1	1	0	1	1
1	1	0	0	0	1	0	0	0	0
0	0	0	1	1	1	0	0	1	0
1	1	1	0	0	0	0	1	1	0

b)

Rys. 3. Schemat przedstawiający identyfikację wiersza reprezentującego najmniejszą liczbę binarną: a) stan rejestrów przed identyfikacją, b) stan rejestrów po identyfikacji

Fig. 3. Diagram showing identification of the row of matrix A' representing the minimal binary number: a) contents of the registers before identification b) contents of the registers after identification

Proces identyfikacji w wariancie (i) jest niezależny od liczby wierszy w macierzy A ani nie jest zależny od ilości powtórzeń warunku Q" w tablicy Q". Złożoność asymptotyczna algorytmu identyfikacji w tym wariancie jest wyłącznie zależna od liczby kolumn macierzy A'-A''. Złożoność czasowa dla tego procesu jest mała w porównaniu ze złożonością czasową jakiegokolwiek innego znanego procesu wyszukiwania. Pozwala to na praktyczne operowanie macierzami A o dużych rozmiarach. Ma to istotne znaczenie dla zwiększenia odporności na złamanie omawianego tutaj systemu komunikacji za pomocą wiadomości szyfrowanych.

2. Sformułowanie zadań kodowania i dekodowania

W sieci istnieje z węzłów mogących komunikować się ze sobą przy pomocy wiadomości szyfrowanych. Każdy z węzłów może być nadawcą lub odbiorcą wiadomości szyfrowanej. Nadawana wiadomość jest odbierana i deszyfrowana przez wszystkie węzły systemu, lecz może być odszyfrowana poprawnie jedynie przez węzły uprawnione przez nadawcę. Nadawca jest niejawny dla wszystkich odbiorców i może być poznany jedynie na mocy poprawnego odszyfrowania przesyłanej wiadomości, jeżeli taka jest intencja nadawcy. Przesłanie zawiera *m*.-bitowy nagłówek oraz zaszyfrowaną wiadomość właściwą. Nadawca wybiera grupę(y) węzłów uprawnionych do rozkodowania wiadomości i formułuje nagłówek, w którym zakodowana jest informacja o uprawnionych do rozkodowania grupach węzłów. Książka szyfrów jest wspólna dla całego systemu i jest macierzą binarną podzieloną na strony. Sposób wypełnienia bitami książki szyfrów nie ma istotnego znaczenia dla możliwości funkcjonowania systemu, można ją wypełnić losowo. Istotną rzeczą jest to, aby wiersze i kolumny każdej strony książki kodowej możliwie najbardziej różniły się pomiędzy sobą. Jednakże odporność systemu na złamanie jest zależna od spełnienia dodatkowych warunków sprowadzających się do możliwie największej różnorodności rozmieszczenia zer i jedynek. Pełna znajomość książki szyfrów jest konieczna, lecz nie jest wystarczająca do rozkodowania przesyłanych wiadomości. Nadawca informuje w sposób zakodowany odbiorców, jaka część przesłanego kodu zawiera informację przeznaczoną dla każdego z odbiorców. Kod pozwalający odczytać wskazane przez nadawcę części książki szyfrów, którymi ma się posługiwać poszczególny węzeł dla odszyfrowania otrzymanej wiadomości, jest indywidualny dla każdego węzła i jest zapisany w jego wnętrzu. Złamanie tego zaszyfrowanego kodu dla jakiegoś węzła nie może powodować złamania całego systemu, a jedynie pozwala na rozkodowanie wiadomości przeznaczonych dla tego węzła. Wymiana książki szyfrów powinna być czynnością prostą. Złamanie kodu przesłanej wiadomości do

danego węzła nie może być wystarczające do rozkodowania innych wiadomości przesyłanych do tego węzła.

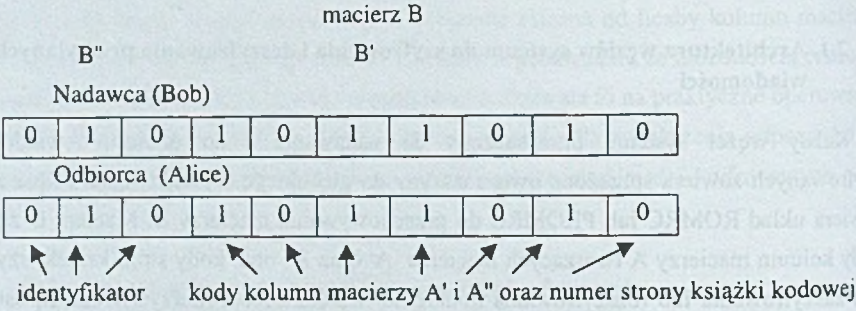
2.1. Architektura węzłów systemu do szyfrowania i deszyfrowania przesyłanych wiadomości

Każdy węzeł systemu przeznaczony do nadawania i do odbioru wiadomości szyfrowanych zawiera sprzężone dwie maszyny do głębokiego wyszukiwania. Jedna z nich zawiera układ ROMRC lub PROMRC do przechowywania macierzy B. Macierz B zawiera kody kolumn macierzy A tworzących macierze A' oraz A" oraz kody stron książki szyfrów. Dla zaszyfrowania lub rozszyfrowania konkretnej wiadomości wykorzystywane są ustalone macierze A' oraz A" oraz ustalona jest strona książki szyfrów. Maszyna do głębokiego wyszukiwania zawierająca macierz B przetwarza wyłącznie nagłówek otrzymywanej lub nadawanej wiadomości. W przypadku gdy węzeł pełni funkcję odbiornika otrzymywanej wiadomości, zawsze mamy do czynienia z pojedynczymi podmacierzami A' i A", w przypadku gdy nadajnik ma wyprodukować nagłówek, efektem może być kilka różnych macierzy A' i odpowiadających im podmacierzy A". Druga sprzężona maszyna do głębokiego wyszukiwania zawiera jako podstawowy komponent układ RAMRC. Jest ona przeznaczona do zapamiętania książki szyfrów oraz do szyfrowania i deszyfrowania otrzymywanych lub nadawanych wiadomości. Bierze ona jako daną wejściową numery kolumn tworzących macierze A' i A" oraz numer strony kodowej. Dane te nie powinny być wyprowadzane na zewnątrz ani nie powinny być możliwe do uzyskania przez osoby obsługujące lub konserwujące dany węzeł systemu, a jedynie powinny być przesłane z maszyny przetwarzającej nagłówek do maszyny dokonującej procesu szyfrowania lub deszyfrowania.

3. Struktura danych

Nadawana wiadomość podzielona jest na m-bitowe ciągi włączając w to nagłówek, zatem każdy ciąg posiada długość rejestrów X i Q oraz wierszy macierzy A lub B. Nagłówek jest pierwszym pojedynczym m-bitowym ciągiem. Jest on podzielony na z-1 pól tworzących macierze B dedykowane dla każdego węzła lub dla stałej grupy węzłów. Zawiera on informacje, która strona książki kodowej oraz które kolumny macierzy A tworzą dla danego węzła obszar identyfikacji. Każda para (nadawca, odbiorca) ma identyczny wiersz w ramach macierzy B. Wiersz ten nie może się pojawić w tablicy B żadnego innego węzła systemu. Zamiast jednego wiersza dla pary (nadawca, odbiorca) może być kilka wspólnych wierszy o

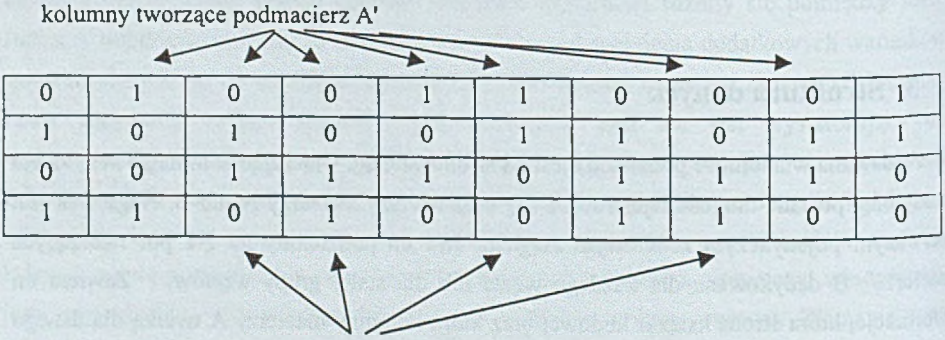
różnych identyfikatorach i kodujących różne kolumny i różne strony książki szyfrów. Na rys. 4 przedstawiono strukturę macierzy B dla danego węzła.



Rys. 4. Struktura wspólnych wierszy dla pary (nadawca, odbiorca) zapisanych w odpowiednich macierzach B

Fig. 4. The structure of the common rows in the corresponding tables B for a pair (Bob, Alice)

Dedykowane kolumny dla danego węzła podzielone są na dwie części tworzące dwie kolumnowe podmacierze A' oraz A" macierzy A. Wiersze macierzy A'-A" odpowiadają kodom ASCII przesyłanych znaków, natomiast odpowiadające wiersze macierzy A" reprezentują szyfry tych kodów. Celem zwiększenia odporności systemu na złamanie ciąg kolejnych numerów kolumn macierzy A nie powinien tworzyć podmacierzy A' i A", lecz poszczególne kolumny tworzące te podmacierze powinny być 'rozstrzelone' i przemieszane. Na rys. 4 pokazano sposób zdefiniowania podmacierzy A' oraz A" dla danego węzła. Kolejność kolumn w podmacierzach A' i A" wyznaczają rosnące ich indeksy z macierzy A.



Rys.5. Sposób zdefiniowania podmacierzy A' oraz A" dla danego węzła

Fig. 5. Defining submatrices A' and A" for a given node

Wybór kolumn tworzących podmacierz A' musi zapewniać na każdej stronie stosowanej książki kodowej spełnienie następujących warunków:

(i) liczba kolumn podmacierzy A'-A" musi być równa liczbom bitów stosowanego kodu ASCII.

(ii) macierz A'-A" musi zawierać co najmniej jeden wiersz odpowiadający każdemu stosowanemu znakowi ASCII ,

(iii) dla zwiększenia bezpieczeństwa systemu liczba wierszy w macierzy A'-A" odpowiadających każdemu stosowanemu znakowi z kodu ASCII powinna być większa od jeden, zapewnia to możliwość takiego szyfrowania przesyłanych wiadomości, że temu samemu znakowi będą mogły odpowiadać w ramach tej samej wiadomości różne szyfry.

4. Kodowanie i dekodowanie jako procesy identyfikacji

W procesie kodowania nadawca zadaje wiersz Q'-Q", a następnie identyfikuje wiersz Q". Zatem warunkiem jest wiersz Q'- Q", a zwracany jest wiersz Q". Zwracany wiersz Q" jest jedynym elementem wspomnianego *m*-bitowego ciągu zapewniającym poprawne rozszyfrowanie przez odbiorcę zaszyfrowanego znaku. Dlatego uzupełnienie wiersza Q" do pełnego wiersza Q nie ma żadnego znaczenia dla poprawności rozszyfrowania reprezentowanego znaku. Jednakże celem większego utajnienia zwracany wiersz Q" powinien być uzupełniony do jakiegokolwiek innego wiersza Q obecnego w książce szyfrów i nie zawierającego warunku Q'-Q", lecz oczywiście zawierającego Q". Wiersz Q jest przesyłany jako komponent zaszyfrowanej wiadomości. W procesie deszyfrowania odbiorca przyjmuje jako warunek wiersz Q" i dokonuje procesu identyfikacji celem zwrócenia wiersza Q'-Q". Proces ten nie wymaga żadnych dodatkowych wyjaśnień, jeżeli macierz A w obrębie wybranej strony książki szyfrów posiada tylko jeden wiersz zawierający Q". W przypadku gdy w macierzy A" dla danej strony kodowej mamy więcej niż jeden wiersz Q", proces identyfikacji musi być dokładniej sprecyzowany, patrz sekcja 1. W podsumowaniu związków pomiędzy identyfikacją a szyfrowaniem i deszyfrowaniem możemy powiedzieć, że zarówno proces szyfrowania, jak i proces deszyfrowania sprowadzają się do procesu identyfikacji przeprowadzanego w maszynach do głębokiego wyszukiwania. W procesie szyfrowania warunkiem jest wiersz Q'-Q", a zwracany jest wiersz Q". Natomiast w procesie deszyfrowania warunkiem jest wiersz Q", a zwracany w czasie identyfikacji jest wiersz Q'-Q".

Zasadniczo wybór kolumn macierzy A" powinien zapewniać różność wierszy w ramach każdej strony stosowanej książki kodowej. W przypadku gdy wiersze macierzy A"

powtarzałyby się, wtedy proces identyfikacji musi być dokładniej sprecyzowany i musi być wspólny dla nadawcy i uprawnionego do poprawnego rozszyfrowania odbiorcy, patrz sekcja 1.

5. Podsumowanie

Podstawową własnością maszyn do głębokiego wyszukiwania jest szybkość identyfikacji niezależna od jakiegokolwiek uporządkowania zarówno wierszy, jak i kolumn tworzących warunek oraz pole identyfikacji. Czas identyfikacji zasadniczo nie zależy od liczby wierszy w tablicy A. Pozwala to z jednej strony stosować duże książki szyfrów, z drugiej strony stosowana książka szyfrów może być wielostronnie wykorzystana do szyfrowania i deszyfrowania, ponieważ każda kolumna i każdy wiersz książki szyfrów może być składnikiem wielu różnych kodów. Implementując przedstawiony system komunikacji za pomocą wiadomości szyfrowanych w architekturach tradycyjnych, niemożliwe jest stosowanie z jednej strony takich dużych książek szyfrów, z drugiej strony niemożliwe jest tak wielostronne ich wykorzystanie, ponieważ ceną, jaką należałoby zapłacić za taką implementację, byłby bardzo długi czas potrzebny na szyfrowanie i deszyfrowanie wiadomości kodowanych. W architekturach tradycyjnych implementacja przedstawionego tutaj procesu identyfikacji sprowadzać się musi do przeszukiwania zupełnego, którego złożoność czasowa zależy liniowo od liczby wierszy stosowanej strony książki szyfrów lub zależy liniowo od liczby wierszy całej książki szyfrów. Ten ostatni przypadek musi mieć miejsce wtedy, gdy stosowana strona książki kodowej powinna zostać możliwie maksymalnie utajniona dla osób posiadających dostęp do wybranego lub do wybranych węzłów systemu.

Przestawiony system komunikacji za pomocą wiadomości szyfrowanych w istotny sposób bazuje na przewadze maszyn do głębokiego wyszukiwania w porównaniu z maszynami tradycyjnymi w obszarze możliwości realizacji procesu wyszukiwania. Z jednej strony pozwala to na bardziej zróżnicowane i bogatsze wykorzystanie istniejącej książki szyfrów, z drugiej strony pozwala na stosowanie znacznie większych książek szyfrów przy krótszym czasie szyfrowania i deszyfrowania.

LITERATURA

1. Jeong-Hyun Park.: Key distribution for secure VSAT satellite communication. IEEE Transactions on Broadcasting, vol.44, no.3, Sept. 1998, pp.274-7.

2. Kapralski A.: Sequential and parallel processing in depth search machines. World Scientific, Singapur, 1994.
3. Kapralski A.: Macierze binarne, ich zastosowania i przetwarzanie w dedykowanych procesorach szeregowych i równoległych. Politechnika Krakowska, Monografia 95, 1989.
4. Palmer M.J., McGlaughlin D.C., Robinson M.J.: Security in data networks. BT Technology Journal, vol.16, no.1, Jan. 1998, pp.52-75.
5. Pohlmann N.: Information security in data networks. Ingenieur der Kommunikationstechnik, vol.48, no.2, March-April 1998, pp.34, 36-8.
6. Pounder C.: First steps towards a European Union policy on the securing of electronic communication. Computers & Security, vol.16, no.7, 1997, pp.590-4.

Recenzent: Dr inż. Andrzej Białas

Wpłynęło do Redakcji 22 marca 1999 r.

Abstract

DSMs (depth search machines) are new models of computations widely applicable in combinatorial computations. The data processed are stored in two-dimensional binary tables. The organization of the table A and of the basic registers of DSMs are given in Fig.1 and in Fig.2. The identification process concerns two arbitrary columnar submatrices A' and A'' of the matrix A and the corresponding subrows Q' and Q'' of the register Q. We fix the subrow Q' and there is retrieved subrow Q" that corresponds to a row of matrix A containing the subrow Q' in the range of A'. In the case where more than one row matches the requirement we can require additionally retrieval of the row that corresponds to the maximal number or to the minimal number and so on. The identification process is applied for coding and decoding binary strings representing characters of the cod ASCII. We develop basic organization of the multi-users system that each node of the network contains two DSMs. The first machine contains DSM possessing ROM or PROM memory components that are used for coding information concerning columns of the matrix A that are to be used for defining the submatrices A' and A''. The second DSMs contains RAM and the table A represents the book of codes partitioned into a number of pages. The submatrices A' and A'' defined by using the first DSM are specified for coding or for decoding the mail massage using the

second DSMs that contains RAM components. The system proposed possess high level of safety by practical possibility of applying long coding strings since the time of coding and decoding is practically not affected by the length of code, moreover the time of processing is very short.