STUDIA INFORMATICA 2002
Volume 23 Number 2B (49)

Katarzyna TRYBICKA-FRANCIK
Politechnika Śląska, Instytut Informatyki

WYKORZYSTANIE WIZUALIZACJI ALGORYTMÓW W KRYPTOANALIZIE*

Streszczenie. Przedmiotem artykułu są zagadnienia kryptoanalizy liniowej i różnicowej szyfrów blokowych ze szczególnym uwzględnieniem algorytmu DES. Na przykładzie tego algorytmu przedstawione zostały zasady aproksymacji rund i konstruowania wyrażeń liniowych dla całego algorytmu. Ponadto omówiono możliwość wykorzystania wizualizacji algorytmów do przeprowadzenia obserwacji pozwalających stwierdzić, czy wspomniane metody kryptoanalizy będą skuteczne w przypadku badanych algorytmów kryptograficznych.

APPLICATION OF SOFTWARE VISUALIZATION IN CRYPTANALYSIS

Summary. The paper presents the problems of linear and differential cryptana-lysis of block ciphers. The rules of approximation of the algorithm rounds and formulating the linear representation have been presented on the example of DES algorithm. The paper presents applications of software visualization in estimating potential effectiveness of linear and differential cryptanalysis methods in case of analyzed cryptographic algorithms.

1. Wstęp

Symetryczne szyfry blokowe to jedno z podstawowych narzędzi współczesnej kryptografii. Stosowane są do zapewniania poufności, integralności i uwierzytelniania. Tak szerokie spektrum zastosowań wymaga dużego zaufania do ich bezpieczeństwa. Ponieważ nie są znane szyfry blokowe oferujące bezwarunkowe bezpieczeństwo, ani też nie ma technik, które by w jednoznaczny sposób pozwoliły udowodnić obliczeniowe bezpieczeństwo szyfru, konieczne jest heurystyczne

* Praca finansowana z funduszu Badań Statutowych Instytutu Informatyki w roku 2002.

podejście do oceny takich szyfrów. W związku z tym szyfr uznawany jest za bezpieczny, gdy przestrzeń klucza szyfru jest na tyle duża, by atak metodą brutalną (przeglądania całej przestrzeni klucza) był ze względu na czas nie do zrealizowania oraz nie są znane inne, skuteczne ataki przeciwko niemu. Skuteczność ataku określa się mierząc jego złożoność (czasową i pamięciową) i porównując ze złożonością ataku metodą brutalną. W ocenie bezpieczeństwa uwzględniane są te ataki, które są znane w chwili dokonywania oceny. Jednymi z najważniejszych ataków branych pod uwagę w ocenia bezpieczeństwa szyfrów są kryptoanaliza liniowa i kryptoanaliza różnicowa.

2. Kryptoanaliza liniowa

Kryptoanaliza liniowa [1, 4-6] zaprezentowana została przez Matsui w 1992 roku do analizy szyfru FEAL-8. Metoda ta okazała się skuteczna także wobec algorytmu DES [1-3], zwłaszcza gdy mamy do czynienia z atakiem ze znanym tekstem jawnym (można złamać 16-rundowy DES z 2^{47} znanymi tekstami jawnymi). Kryptoanaliza liniowa jest atakiem wykorzystującym zależności między bitami danych wejściowych rundy, bitami klucza oraz wynikami rundy. Oczywiście, związki te zachodzą tylko statystycznie dla stosunkowo dużej lub stosunkowo małej liczby danych wejściowych. Kluczową rolę odgrywają tu formuły aproksymujące liniowo pojedyncze s-boksy.

W pierwszym kroku należy skonstruować równania opisujące zależności pomiędzy wejściowymi i wyjściowymi bitami każdego z S-boksów. Chociaż S-boksy nie obliczają łatwych do przedstawienia funkcji, to jednak istnieją proste formuły, które w przybliżeniu to czynią. Dla pojedynczego S-boksu S, gdzie S=1,...,8 rozważamy formuły postaci:

$$a_0 s_0 \oplus a_1 s_1 \oplus a_2 s_2 \oplus a_3 s_3 \oplus a_4 s_4 \oplus a_5 s_5 = b_0 s_0^* \oplus b_1 s_1^* \oplus b_2 s_2^* \oplus b_3 s_3^*$$

gdzie s_i oraz s_j^* oznaczają bity danych wejściowych i wyjściowych, a a_i i b_i są współczynnikami przyjmującymi wartości *0* lub *1*. Zasady konstrukcji S-boksów sprawiają, że równość taka nie zachodzi dla wszystkich przypadków danych wejściowych. Do przeprowadzenia skutecznego ataku wystarczy jednak, by równość taka była spełniona stosunkowo często lub stosunkowo rzadko. Jeżeli tak jest, to mówimy, że formuła ta *liniowo aproksymuje* S-boks S.

Na przykład S-boks 5 charakteryzuje formuła:

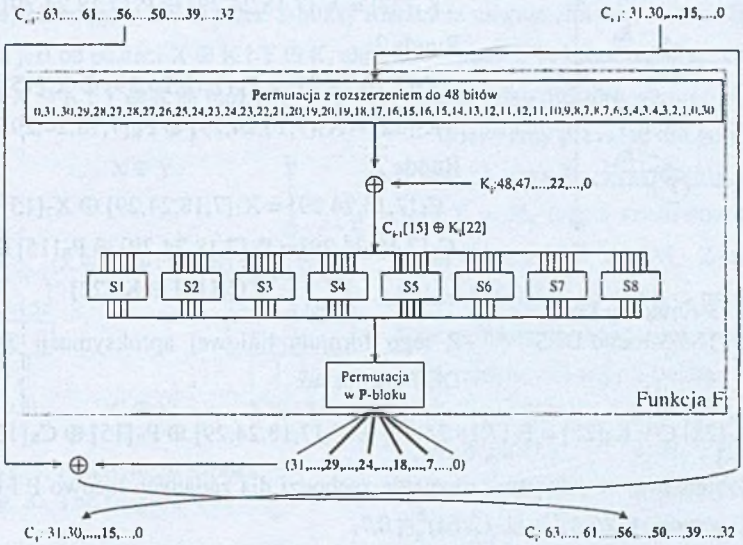
$$s_4 = s_0^* \oplus s_1^* \oplus s_2^* \oplus s_3^*$$

Równość ta zachodzi dla 19% danych wejściowych.

Z formuł aproksymujących S-boksy można zbudować formuły opisujące związki pomiędzy danymi wejściowymi rundy, bitami klucza oraz bitami wyjściowymi rundy. Podobnie jak w przypadku S-boksów, mowa jest o związkach zachodzących statystycznie dla stosunkowo

dużej lub stosunkowo małej liczby danych wejściowych. W ostatnim kroku wiążemy ze sobą formuły dla poszczególnych rund.

Rozważmy przykład dla pojedynczej rundy i S-boksu 5. Rysunek 1 przedstawia aproksymację rundy na podstawie aproksymacji S-boksu 5.



Rys. 1. Pojedyncza runda algorytmu DES
Fig. 1. The single round of the DES algorithm

Z liniowej aproksymacji S-boksu 5 wynika, że wejściowy $C_{i-1}[15] \oplus K_i[22]$ koreluje z wszystkimi bitami wyjściowymi z S-boksu 5.

Biorąc pod uwagę permutację z rozszerzeniem do 48 bitów, a także permutację w P-bloku w funkcji F widzimy, że poniższe równanie zachodzi z prawdopodobieństwem p=0,19 dla ustalonego K (klucza) i losowego X (ciągu wejściowego)

$$X[15] \oplus F(X,K)[7,18,24,29] = K[22],$$

gdzie X[i] to i-ty bit ciągu X, K[j] to j-ty bit ciągu K.

Spróbujmy rozszerzyć liniową aproksymację funkcji F na 3-rundowy DES (rys.2).

Wprowadźmy następującą notację:

P: 64-bitowy tekst jawny, C: 64-bitowy szyfrogram,

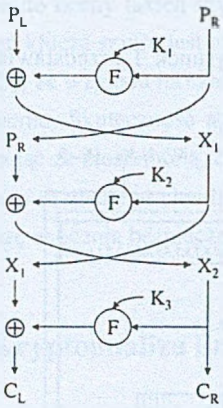
P_L: lewa 32-bitowa część P, C_L: lewa 32-bitowa część C,

P_R: prawa 32-bitowa część P C_R: prawa 32-bitowa część C

X_i: 32-bitowa wartość pośrednia w i-tej rundzie

K_i: 48-bitowa wartość klucza w i-tej rundzie

A[i]: i-ty bit A; A[i, j, ..., k]: A[i] A[j] ... A[k]



Rys. 2. 3-rundowy DES
Fig. 2. The 3-round DES

Wykorzystując równanie aproksymujące pojedynczą rundę zbudujmy zależności dla 3-rudowego DES:

Runda 1:

$$X_1[7,18,24,29] = P_L[7,18,24,29] \oplus P_R[15] \oplus K_1[22]$$

$$K_1[22] = X_1[7,18,24,29] \oplus P_L[7,18,24,29] \oplus P_R[15]$$

Runda 2:

$$X_2[7,18,24,29] = P_R[7,18,24,29] \oplus X_1[15] \oplus K_2[22]$$

$$K_2[22] = X_2[7,18,24,29] \oplus P_R[7,18,24,29] \oplus X_1[15]$$

Runda 3:

$$C_L[7,18,24,29] = X_1[7,18,24,29] \oplus X_2[15] \oplus K_3[22]$$

$$C_L[7,18,24,29] = P_L[7,18,24,29] \oplus P_R[15] \oplus K_1[22] \oplus \oplus C_R[15] \oplus K_3[22]$$

Z tego formuła liniowej aproksymacji 3-rundowego DES ma postać:

$$K_1[22] \oplus K_3[22] = P_L[7,18,24,29] \oplus C_L[7,18,24,29] \oplus P_R[15] \oplus C_R[15]$$

Prawdopodobieństwo, że powyższe równanie zachodzi dla zadanych losowo P i odpowiadających im C wynosi: $(12/64)^2 + (1-12/64)^2 \approx 0{,}7$.

W analogiczny sposób można wyznaczyć liniową aproksymację dla 16-rundowego DES. Najkorzystniejsze prawdopodobieństwo w tym przypadku wyniesie $p = ½ - 1{,}49 * 2^{-24}$.

3. Kryptoanaliza różnicowa

Kryptoanaliza różnicowa [7, 8] jest znana publicznie od 1990 roku za sprawą dwóch kryptografów Bihama i Shamira. W istocie brana była już pod uwagę przy projektowaniu szyfru DES w 1974 roku [1]. Metoda ta polega na szyfrowaniu par tekstów jawnych różniących się w określony sposób (stąd nazwa metody) i analizie uzyskanych szyfrogramów. Podobnie jak w przypadku kryptoanalizy liniowej, jest to atak o charakterze statystycznym.

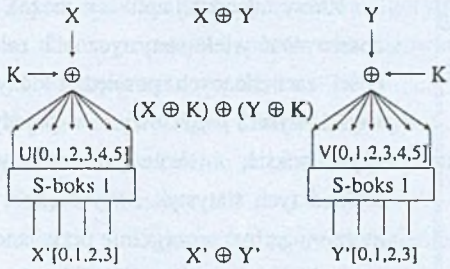
Z kryptograficznego punktu widzenia, tak jak i w poprzednim przypadku, istotne są dwie następujące po sobie fazy:

1) operacja xor z bitami klucza,
2) przekształcenia wykonywane w S-boksach.

Istota kryptoanalizy różnicowej opiera się na następującym spostrzeżeniu. Mimo że nie znamy wartości bitów klucza i nie jesteśmy w stanie określić danych wejściowych do S-boksów, to potrafimy wyznaczyć różnicę ciągów $X \oplus Y$, gdzie X i Y są parą danych wejściowych, a operacja xor wykonywana jest oddzielnie na poszczególnych bitach. Przyjmijmy,

że U = X ⊕ K, V = Y ⊕ K, gdzie U i V to 48-bitowe wartości pośrednie będące danymi wejściowymi do S-boksów, a K to 48-bitowy klucz. Wykonajmy następującą operację U ⊕ V = (X ⊕ K) ⊕ (Y ⊕ K) = X ⊕ Y. Jak widać różnica pomiędzy danymi wejściowymi do S-boksów jest taka sama jak pomiędzy X ⊕ Y.

Po przejściu ciągów U i V przez S-boksy różnica ta ulegnie zmianie. Nowa różnica jak widać zależna jest od postaci X ⊕ K i Y ⊕ K, ale z konstrukcji S-boksów wynika, że tylko niektóre wartości X ⊕ K i Y ⊕ K są możliwe, a co za tym idzie tylko niektóre wartości K.



Rys. 3. Pojedynczy S-boks
Fig. 3. The single S-box

Rozważmy przykład dla pojedynczego S-boksu 1 (rys.3). Przyjmijmy, że różnica X ⊕ Y = 34_x (zapis szesnastkowy), tym samym różnica U ⊕ V = 34_x. Z konstrukcji S-boksu 1 wynika, że w takim przypadku możliwe są następujące różnice pomiędzy danymi wyjściowymi tego S-boksu:

1_x dla 8 par U,V,	2_x dla 16 par U,V,
3_x dla 6 par U,V,	4_x dla 2 par U,V,
7_x dla 12 par U,V,	8_x dla 6 par U,V,
D_x dla 8 par U,V,	F_x dla 6 par U,V.

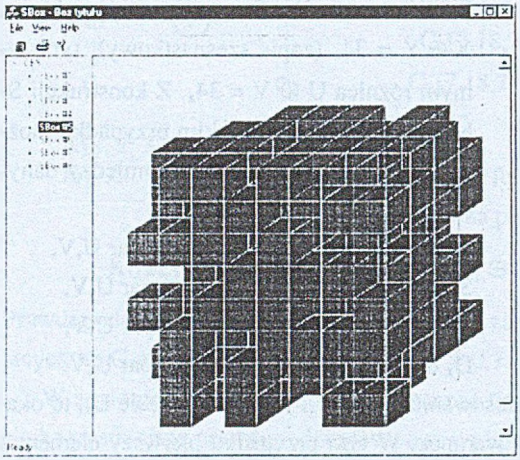
Jeżeli na przykład przyjmiemy, że po zastosowaniu S-boksu 1 różnica wyniesie D_x, to okaże się, że tylko osiem par spełnia powyższy warunek. W tym przypadku pierwszy element z tych par to jeden z ciągów: 07_x, 33_x, 11_x, 25_x, 17_x, 23_x, $1D_x$, 29_x. Drugi element różni się, oczywiście, od pierwszego o 34_x. Mając te informacje możemy zawęzić obszar poszukiwań wyznaczając wartość K ze wzoru K = X ⊕ (X ⊕ K), w naszym przypadku do ośmiu kombinacji. Jeżeli przeprowadzimy podobne postępowanie dla innych wartości X i Y, otrzymamy nowy zbiór możliwych kluczy K. Interesujący nas klucz będzie, oczywiście, należał do obu zbiorów. Procedurę należy powtarzać tak długo, aż zbiór możliwych kluczy ograniczymy do jednego.

Kryptoanaliza różnicowa DES jest uogólnieniem powyżej przedstawionego postępowania dla pojedynczej rundy.

4. Wizualizacja algorytmów - wykorzystanie

Z powyższych rozważań wynika, że krytycznymi fazami pracy algorytmu są dwa typy przekształceń: operacja xor danych wejściowych z bitami klucza oraz przekształcenia realizowane w S-boksach. Pierwsza z tych faz wiąże ze sobą informację znaną (tekst jawny) z informacją poszukiwaną (klucz). Operacje wykonywane w S-boksach zawężają obszar poszukiwań.

Właściwie w przypadku doskonałej większości algorytmów kryptograficznych podstawieniowo-permutacyjnych można wyróżnić podobne dwie fazy, tj. powiązanie klucza z danymi wejściowymi i pozostałe przekształcenia. Nie zawsze udaje się w łatwy sposób zauważyć, które z operacji mogą mieć istotne znaczenie dla przeprowadzenia skutecznego ataku. Techniki analityczne są bardzo często zbyt żmudne i pracochłonne. Narzędziem pomocnym w takich sytuacjach jest wizualizacja. Można to było zaobserwować na rysunku 1, który został wygenerowany z zastosowaniem rozbudowanej aplikacji, której częścią jest również moduł służący do wizualizowania stanów S-boksów algorytmu DES (rys.4).



Rys. 4. Trójwymiarowa wizualizacja piątego s-bloku algorytmu DES
Fig. 4. Three-dimensional visualization of the fifth s-box of DES algorithm

Korzystając z tej aplikacji można zaobserwować wiele statystycznych zależności zachodzących pomiędzy danymi wejściowymi i wyjściowymi w pojedynczym boksie, a także przeanalizować rozkład tych statystyk i wyodrębnić te, które mogą być szczególnie przydatne w procesie kryptoanalizy. Wydaje się, że metodę ataku liniowego da się jeszcze usprawnić, na przykład wprowadzając aproksymację więcej niż jednego S-bloku w jednej rundzie. Obserwując graficzny rozkład statystyk wyjść z S-bloku można zauważyć, że choć formuły aproksymujące skonstruowane z ich wykorzystaniem nie są tak dobre, jak w przypadku S-bloku 5, to jednak wykorzystanie więcej niż jednego równania do wyznaczenia tych samych bitów klucza daje znaczne ograniczenie na ilość wymaganych par tekst jawny/szyfrogram (w przypadku DES liczba wymaganych tekstów jawnych spadnie do 2^{38}). W procesie tworzenia formuł aproksymujących warto też zwrócić uwagę na fakt, że tekst jawny, zapisany w języku naturalnym, składający się ze znaków będących kodami ASCII, ma też swoją specyficzną strukturę. Zwróćmy uwagę na fakt, że co ósmy bit wejściowy w przypadku kodu ASCII jest zerem, a co za tym idzie, wyznaczenie aproksymacji wykorzystującej wyłącznie te bity tekstu jawnego, które są zerami, może dać znacznie lepsze rezultaty, aniżeli ograniczanie się do najlepszej aproksymacji jeśli chodzi o prawdopodobieństwo, co można łatwo zaobserwować korzystając z graficznej reprezentacji rozkładu statystyk wyjść z S-bloków dla języków naturalnych. Wzięcie pod uwagę tej własności w przypadku kryptoanalizy różnicowej ułatwi proces doboru właściwych danych testowych. Celowe też wydaje się wykorzystanie elementów nieliniowych i zespolenie ze sobą obu technik kryptoanali-

tycznych liniowej i różnicowej. Szczegółowe omó-wienie wyników tych prac można znaleźć w [10]. Wyniki badań dla algorytmu DES potwierdziły tezę, że techniki wizualizacyjne są pomocne w procesie selekcjonowania cech algorytmu, które czynią go podatnym na ataki typu kryptoanaliza liniowa i kryptoanaliza różnicowa.

5. Wnioski

Zaproponowana technika badania algorytmów wymaga określenia przez obserwatora tego, co, dlaczego i w jaki sposób wizualizować. Działania te można dość dobrze usystematyzować. Proponowana metodyka:

1. Zapoznanie się z ideą działania algorytmu. Techniki wizualizacyjne są bardzo pomocne w procesie zrozumienia działania algorytmu. Pomagają odpowiedzieć na pytanie *jak*, ale również *dlaczego*.
2. Analiza funkcjonalna z wyodrębnieniem oddzielnych jednostek, ze szczególnym uwzględnieniem dwu faz:
   a) powiązania klucza z danymi wejściowymi,
   b) pozostałe przekształcenia.
3. Analiza statystyk dla danych wejściowych i wyjściowych z uwzględnieniem ich wzajemnych relacji.
4. Obserwacja działania wydzielonych jednostek funkcjonalnych ze szczególnym uwzględnieniem pojedynczego cyklu algorytmu i procedury wyznaczania podkluczy dla cykli. Należy zwrócić szczególną uwagę na przepływ danych pomiędzy poszczególnymi elementami algorytmu.

Rezultatem tych działań powinno być wyodrębnienie zależności pomiędzy danymi wejściowymi (tekst jawny) a danymi wyjściowymi (szyfrogram) oraz wyznaczenie danych testowych, które pozwolą wykazać słabe punkty w konstrukcji algorytmu. Wynikiem jest też ustalenie, czy zaprezentowany algorytm faktycznie w sposób jednoznaczny przekształca tekst jawny w szyfrogram przy stałym kluczu.

LITERATURA

1. Schneier B.: Kryptografia dla praktyków. Protokoły, algorytmy i programy źródłowe w języku C. Wydawnictwa Naukowo-Techniczne, Warszawa 1995.
2. Praca zbiorowa pod redakcją Grzywaka A.: Bezpieczeństwo systemów komputerowych. Wydawnictwo Pracowni Komputerowej Jacka Skalmierskiego, Gliwice 2000.

3. Denning D. E.: Kryptografia i ochrona danych. Wydawnictwa Naukowo-Techniczne, Warszawa 1992.
4. Matsui M.: Linear Cryptanalysis Method for DES Cipher. Advances in Cryptology Eurocrypt'93.
5. Matsui M.: On Correlation Between the Order of S-boxes and the Strength of DES. Advances in Cryptology Eurocrypt'94, Springer Verlag, Berlin 1994.
6. Matsui M.: The First Experimental Cryptanalysis of Data Encryption Standard. Advances in Cryptology Crypto'94, Springer Verlag, Berlin 1994.
7. Biham E., Shamir A.: Differential Cryptanalysis of DES-like Cryptosystems. Advances in Cryptology Crypto'90, Springer Verlag, Berlin 1991.
8. Biham E., Shamir A.: Differential Cryptanalysis of the Data Encryption Standard. Springer Verlag, Berlin 1993.
9. Francik J., Trybicka-Francik K.: Zastosowanie wizualizacji algorytmów w kryptologii. VII Konferencja Sieci Komputerowych, Studia Informatica, seria Informatyka nr 1 (39), v. 21, Zakopane 2000.
10. Trybicka-Francik K.: Rozprawa doktorska – Wybór algorytmów szyfrowania i deszyfro-wania informacji z uwzględnieniem oceny ich jakości.

Recenzent: Dr inż. Andrzej Białas

Wpłynęło do Redakcji 30 kwietnia 2002 r.

Abstract

In the paper problems of linear and differential cryptanalysis are analyzed. A special stress has been given to the DES algorithm. Methods of linear attack to simplified versions of DES – single (fig. 1) as well as triple-round (fig. 2) – have been described in detail. The concepts of differential cryptanalysis have been characterized on the example of a single DES s-box structure (fig. 3).

Further, possibilities of applying software visualization techniques in cryptanalytic research are pointed. Some forms of visual data representation are introduced (fig. 4). Data and algorithm visualization facilitates both introductory and more advanced analyses. It allows drawing conclusions concerning mutual relations between input (plaintext) and output (cipher) bits. The complete review of experiments have been collected and presented in [10].