Bezpieczeństwo systemów komputerowych

Wprowadzenie do kryptologii

Marcin Peczarski

Instytut Informatyki Uniwersytetu Warszawskiego

23 października 2011

Nazwa dziedziny

- Kryptografia (ang. *cryptography*) – dziedzina zajmująca się utajnianiem danych. Zajmują się nią kryptografowie (ang. *cryptographer*).
- Kryptoanaliza (ang. *cryptoanalysis*) – dziedzina zajmująca się łamaniem zabezpieczeń chroniących utajnione dane. Zajmują się nią kryptoanalitycy (ang. *cryptoanalyst*).
- Kryptologia (ang. *cryptology*) – dziedzina wiedzy i nauki obejmująca kryptografię i kryptoanalizę. Zajmują się nią kryptolodzy (ang. *cryptologist*).

Podstawowe pojęcia

- Tekst jawny (ang. *plaintext*), tekst otwarty (ang. *cleartext*), wiadomość (ang. *message*) – dane poddawane operacjom ochrony kryptograficznej
- Szyfrowanie (kryptaż) (ang. *encryption*, *encipher*) – proces ukrywania (utajniania) tekstu jawnego
- Szyfrogram (kryptogram) (ang. *ciphertext*) – zaszyfrowana (utajniona) postać tekstu jawnego
- Deszyfrowanie (dekryptaż) (ang. *decryption*, *decipher*) – proces odtwarzania tekstu jawnego na podstawie szyfrogramu

Podstawowe oznaczenia

- M – tekst jawny
- C – szyfrogram
- E – funkcja szyfrująca

$$E(M) = C$$

- D – funkcja deszyfrująca

$$D(C) = M$$

- Podstawowa zależność

$$D(E(M)) = M$$

Algorytmy i klucze

- Algorytm kryptograficzny, szyfr – para funkcji: funkcja szyfrująca i funkcja deszyfrująca
- Algorytm ograniczony (ang. *restricted algorithm*) – bezpieczeństwo opiera się na tajności algorytmu.
- Nowoczesna kryptografia zakłada jawność algorytmu kryptograficznego.
- Bezpieczeństwo zapewnia się za pomocą klucza (ang. *key*).
- Obie funkcje szyfrująca i deszyfrująca zależą od tajnego klucza K:

$$E_K(M) = C,$$

$$D_K(C) = M,$$

$$D_K(E_K(M)) = M.$$

Algorytmy i klucze, cd.

- Niektóre algorytmy stosują różne klucze dla operacji szyfrowania i deszyfrowania K_1, K_2:

$$E_{K_1}(M) = C,$$

$$D_{K_2}(C) = M,$$

$$D_{K_2}(E_{K_1}(M)) = M.$$

- Przemienność kluczy – przestawienie roli kluczy z pary:

$$E_{K_2}(M) = C,$$

$$D_{K_1}(C) = M,$$

$$D_{K_1}(E_{K_2}(M)) = M.$$

- Przestrzeń kluczy (ang. *keyspace*) – zbiór wszystkich możliwych kluczy
- Kryptosystem (ang. *cryptosystem*) – szyfr wraz ze zbiorami wszystkich możliwych tekstów jawnych, szyfrogramów i kluczy

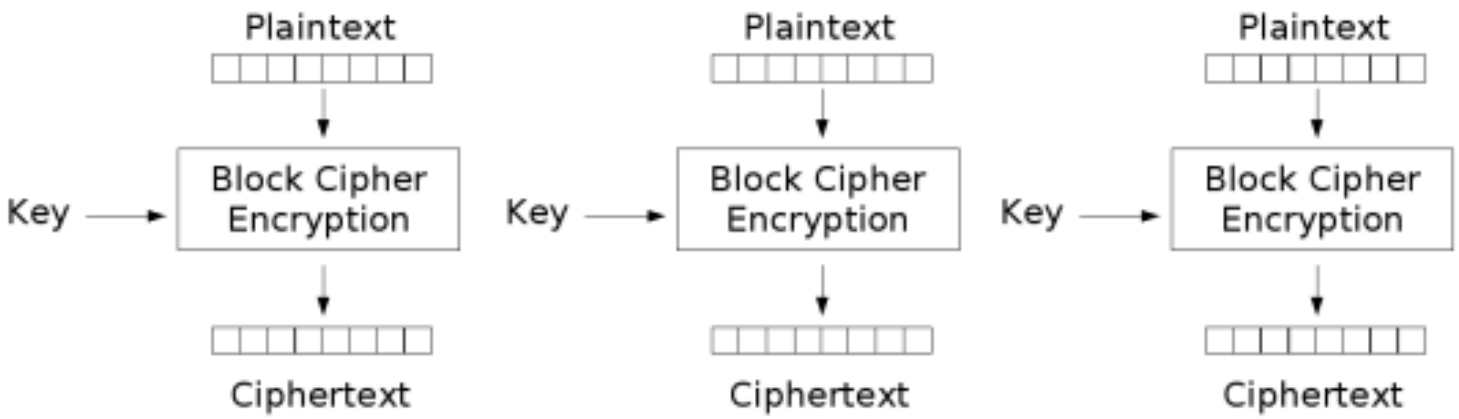
Szyfry symetryczne

- Klucz deszyfrujący może być łatwo wyznaczony z klucza szyfrującego lub oba klucze są identyczne.
- Nazywane są też:
  - szyframi konwencjonalnymi,
  - szyframi z kluczem tajnym (ang. *secret-key cipher*),
  - szyframi z jednym kluczem (ang. *one-key cipher*).
- Wymagane jest uzgodnienie klucza między nadawcą a odbiorcą.
- Klucz musi być utrzymywany w tajemnicy.
- Ten sam klucz może obowiązywać w obu kierunkach komunikacji lub są oddzielne klucze dla każdego kierunku.
- Każda para komunikujących się (a właściwie każda sesja komunikacji) musi używać innego klucza.
- Dla n użytkowników potrzeba $n(n-1)/2$ kluczy.
- Trzeba rozwiązać problem uzgadniania wspólnego klucza „na odległość".

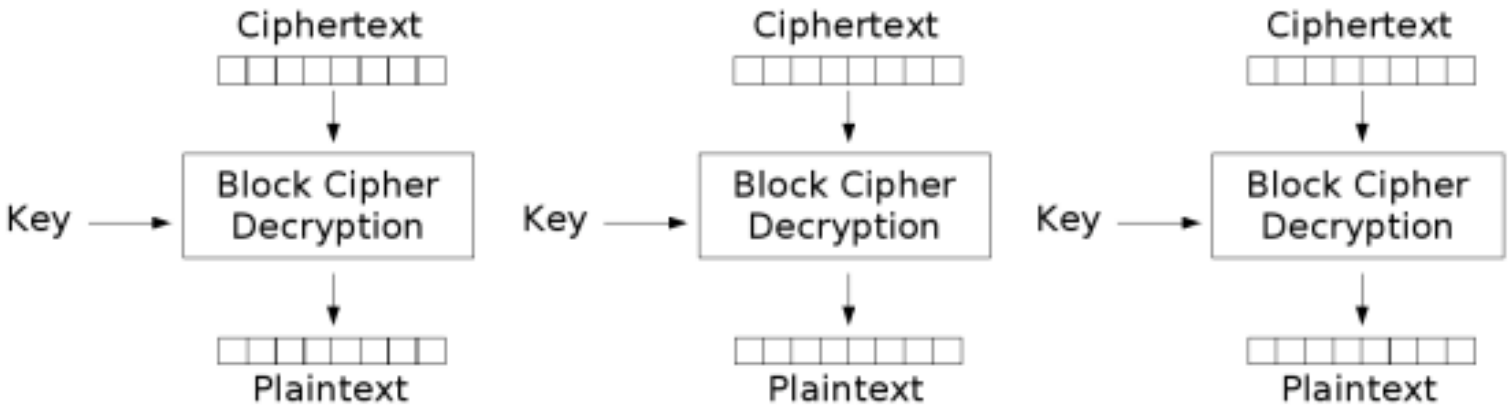
Podział szyfrów symetrycznych

- Szyfr strumieniowy (ang. *stream cipher*) – jednostką przetwarzania informacji jest 1 bit lub 1 oktet (8 bitów, bajt) lub 1 znak tekstu jawnego.
- Szyfr blokowy (ang. *block cipher*) – jednostką przetwarzania informacji jest blok bitów, obecnie najczęściej 64 lub 128 bitów.

ECB – tryb elektronicznej książki kodowej



Electronic Codebook (ECB) mode encryption

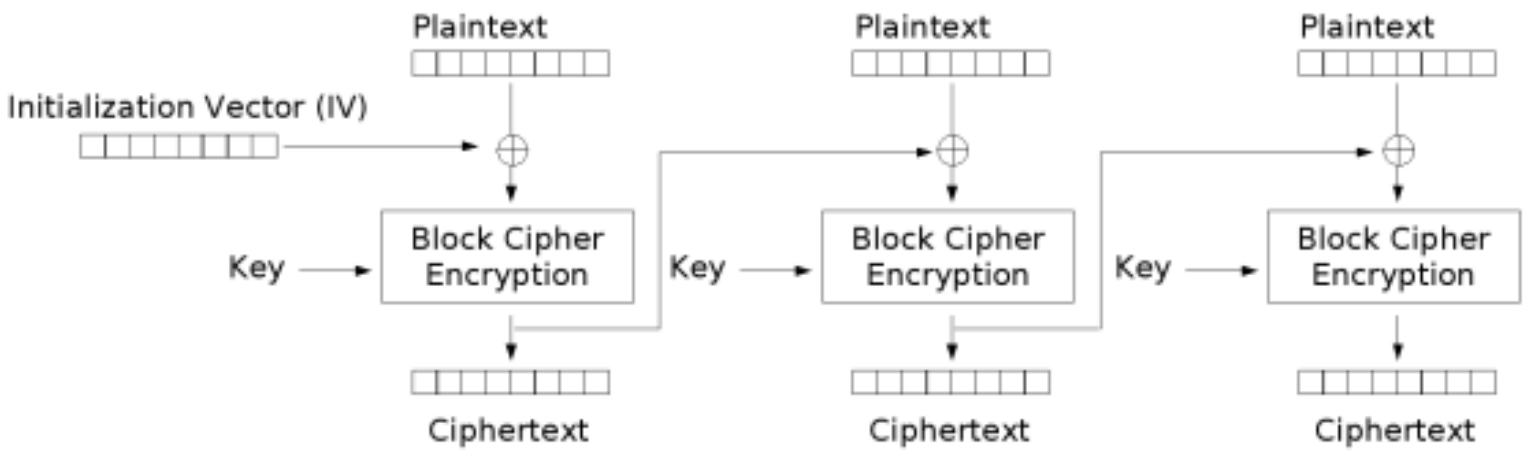


Electronic Codebook (ECB) mode decryption

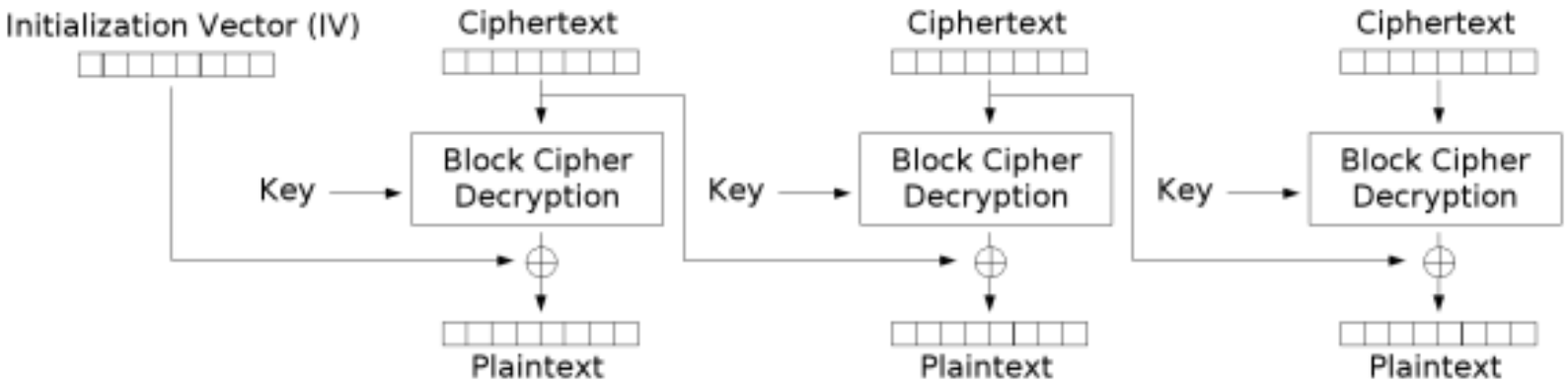
ECB – tryb elektronicznej książki kodowej, cd.

- Tekst jawny jest dzielony na bloki.
- Ostatni fragment jest uzupełniany do pełnego bloku.
- Każdy blok jest szyfrowany niezależnie, z użyciem tego samego klucza.
- Powtarzające się bloki tekstu jawnego dają powtarzające się bloki szyfrogramu, co ułatwia kryptoanalizę.
- Struktura danych (np. sekwencja rekordów bazy danych) nie jest dostatecznie dobrze ukrywana.

CBC – tryb wiązania bloków zaszyfrowanych



Cipher Block Chaining (CBC) mode encryption

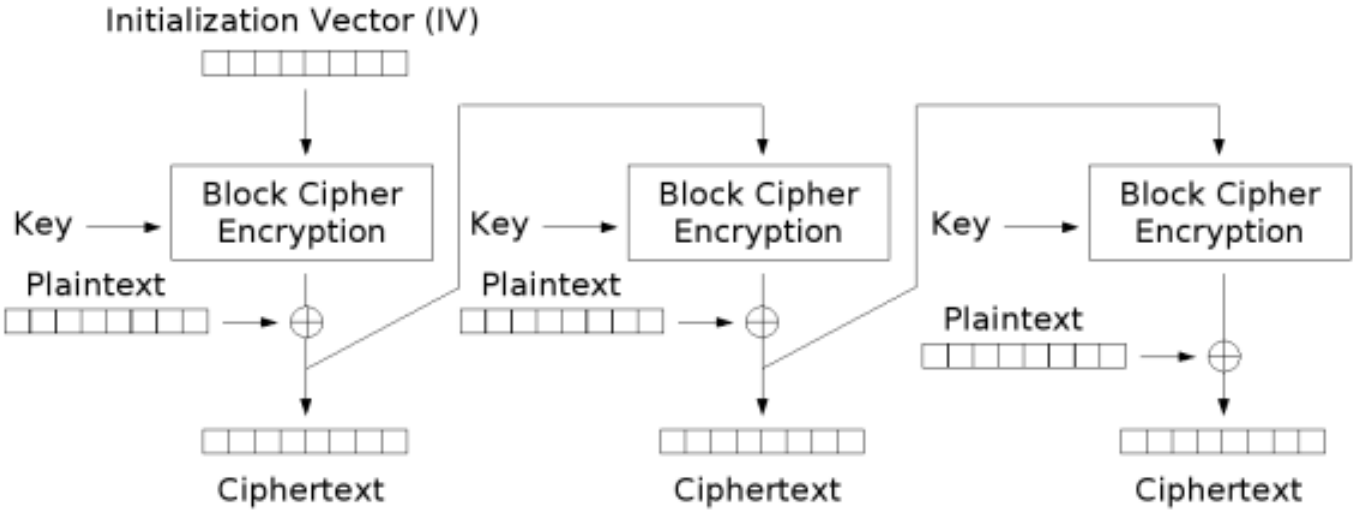


Cipher Block Chaining (CBC) mode decryption

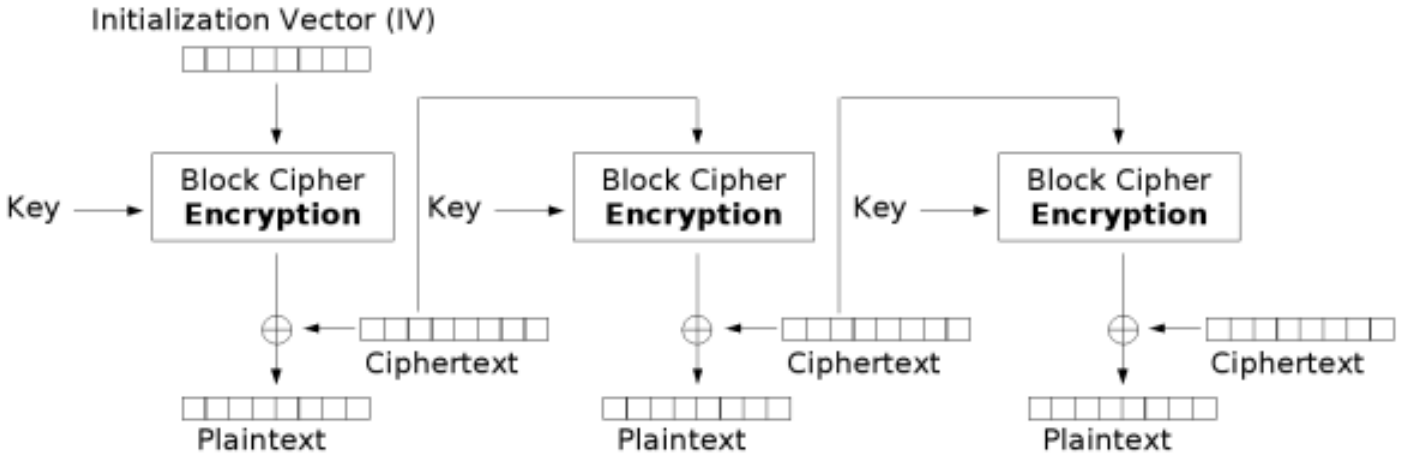
CBC – tryb wiązania bloków zaszyfrowanych, cd.

- Blok szyfrogramu zależy od wszystkich dotychczasowych bloków tekstu jawnego i wektora początkowego IV (ang. *inicializaction vector*).
- Jest to tryb samoodtwarzający (ang. *self-recovering*), błędy w szyfrogramie nie propagują się.
- Szyfrowanie jest trudno zrównoleglić.
- Deszyfrowanie może być zrównoleglone.

CFB – tryb sprzężenia zwrotnego szyfrogramu



Cipher Feedback (CFB) mode encryption



Cipher Feedback (CFB) mode decryption

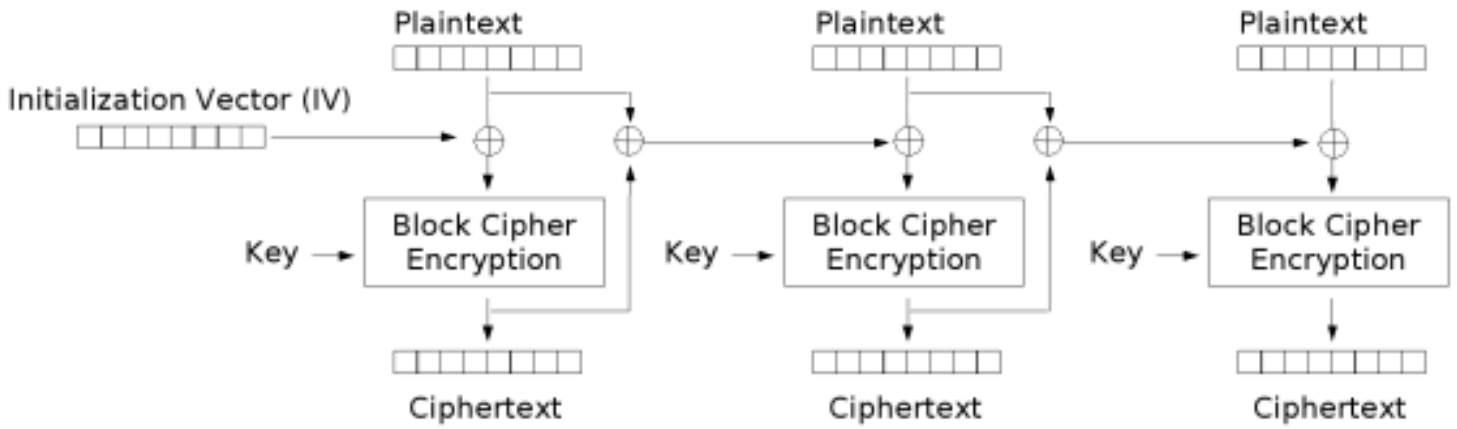
CFB – tryb sprzężenia zwrotnego szyfrogramu, cd.

- Możemy rozpocząć szyfrowanie przed otrzymaniem całości bloku.
- Dane mogą być szyfrowane w mniejszych jednostkach niż cały blok, np. w porcjach po x bitów:

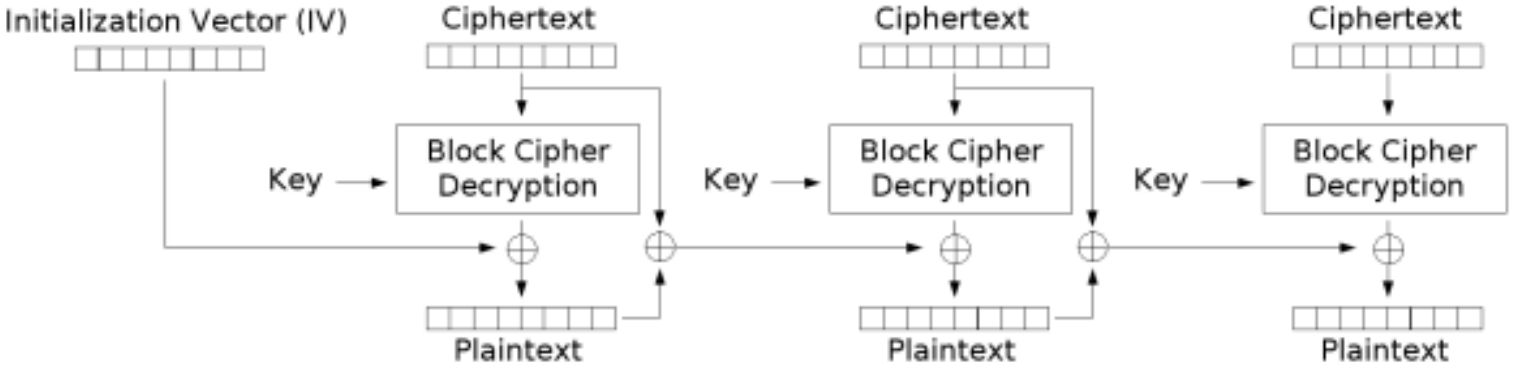
$$\begin{aligned} C_i &:= \mathrm{head}(E_K(S_{i-1}), x) \oplus M_i, \\ M_i &:= \mathrm{head}(E_K(S_{i-1}), x) \oplus C_i, \\ S_i &:= (S_{i-1} \ll x) \parallel C_i, \\ S_0 &:= \mathrm{IV}. \end{aligned}$$

- Używana jest tylko procedura szyfrowania.
- Nie ma konieczności uzupełniania tekstu jawnego do wielokrotności rozmiaru bloku.
- Szyfrogram zależy od całego poprzedzającego tekstu jawnego.
- CFB samosynchronizuje się na poziomie bloków.

PCBC – tryb propagującego wiązania bloków zaszyfrowanych



Propagating Cipher Block Chaining (PCBC) mode encryption

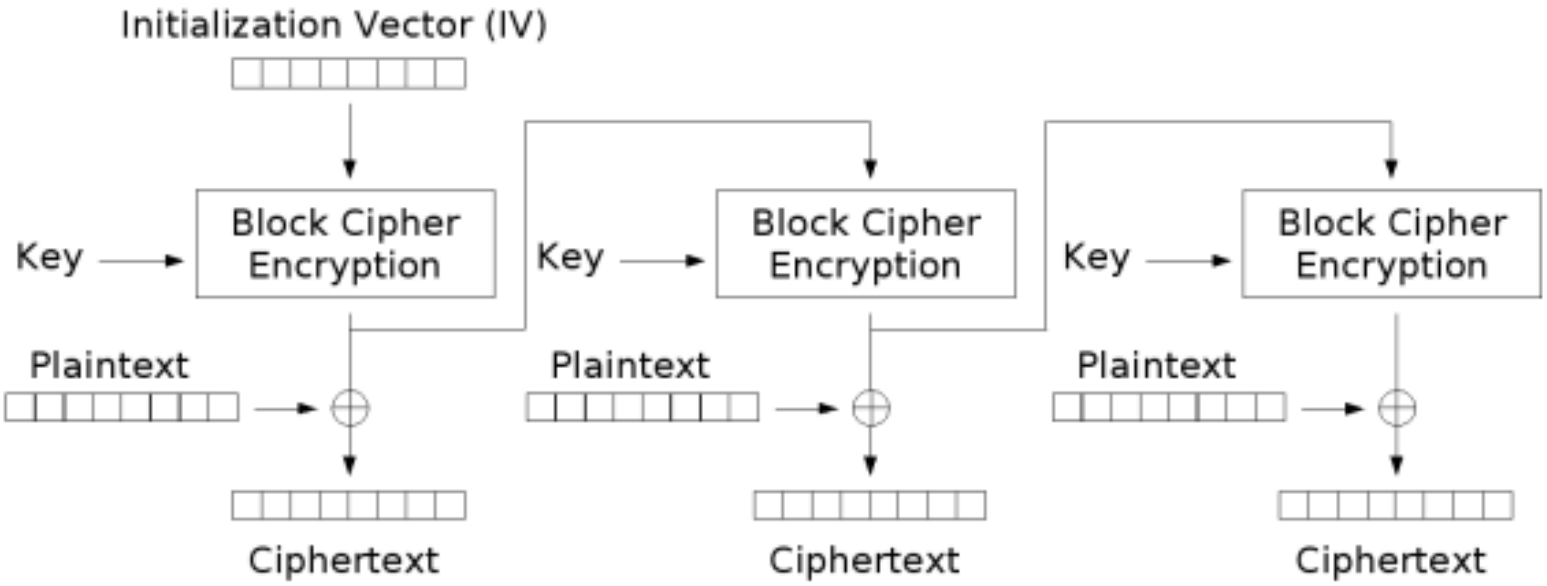


Propagating Cipher Block Chaining (PCBC) mode decryption

PCBC – tryb propagującego wiązania bloków zaszyfrowanych, cd.

- Propaguje w nieskończoność małe zmiany w tekście zaszyfrowanym.

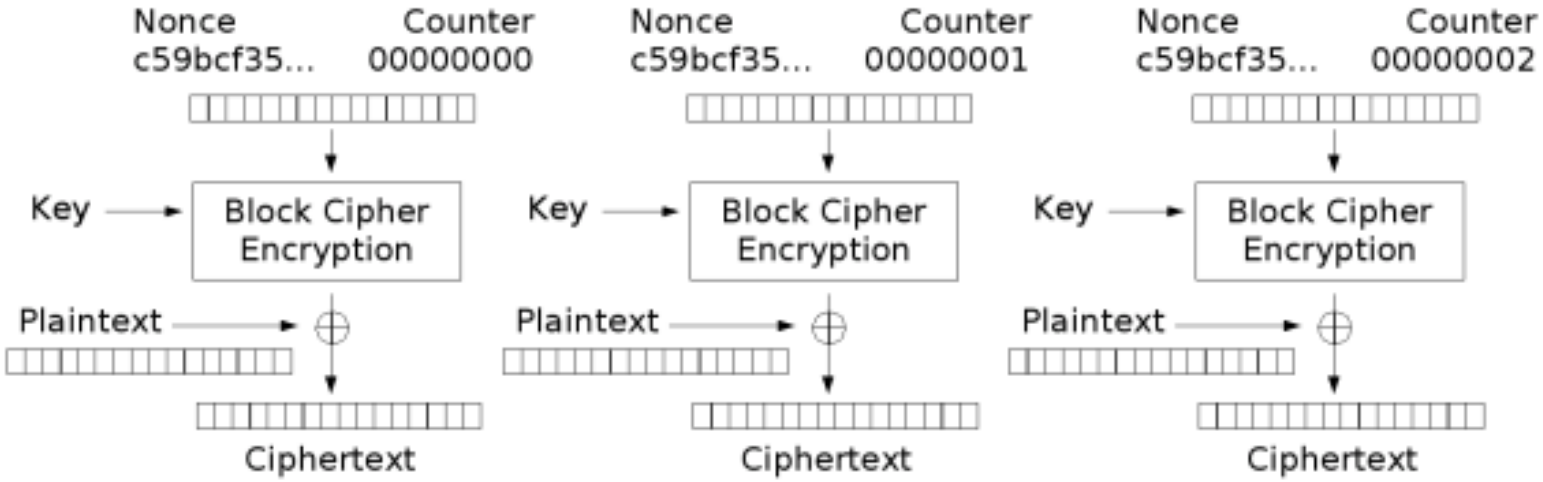
OFB – tryb z wyjściowym sprzężeniem zwrotnym



Output Feedback (OFB) mode encryption

- Generuje strumień klucza, xorowany następnie z tekstem jawnym.
- Deszyfrowanie jest identyczne jak szyfrowanie.
- Umożliwia stosowanie kodów korekcyjnych przed szyfrowaniem.

Tryb licznikowy



Counter (CTR) mode encryption

- Użyteczny do szyfrowania danych o dostępie swobodnym, np. plików.

Szyfry asymetryczne

- Nazywane są też szyframi z kluczem publicznym (ang. *public key cipher*).
- Klucze szyfrowania i deszyfrowania są różne.
- Klucz deszyfrowania nie może być łatwo (w rozsądnym czasie) wyznaczony na podstawie klucza szyfrowania.
- Klucz szyfrowania nazywany jest kluczem publicznym (ang. *public key*).
- Klucz deszyfrowania nazywany jest kluczem prywatnym (ang. *private key*) lub kluczem tajnym (ang. *secret key*).
- Klucz publiczny można ujawnić – każdy może zaszyfrować wiadomość.
- Klucz prywatny musi być utajniony – tylko jego właściciel może odszyfrować wiadomość.
- Wiadomość może być też szyfrowana kluczem prywatnym i deszyfrowana kluczem publicznym, np. w celu zapewnienia autentyczności (podpis cyfrowy).

Podstawy kryptoanalizy

- Kryptoanaliza zajmuje się odtwarzaniem tekstu jawnego bez znajomości klucza lub odtwarzaniem klucza.
- Stosowanie kryptoanalizy nazywa się łamaniem szyfru.
- Bezpieczeństwo systemu kryptograficznego jest oparte wyłącznie na kluczu.
- Kryptoanalityk zna wszystkie szczegóły algorytmu kryptograficznego i jego implementacji.
- Kryptoanalityk zna format tekstu jawnego, np. język, w którym został napisany.

Metody łamania szyfrów

- Tylko tekst zaszyfrowny (ang. *ciphertext-only*) – kryptoanalityk zna tylko pewną liczbę szyfrogramów.
- Znany tekst jawny (ang. *known-plaintext*) – kryptoanalityk oprócz znajomości pewnej liczby szyfrogramów zna odpowiadające im teksty jawne.
- Wybrany tekst jawny (ang. *chosen-plaintext*) – kryptoanalityk może wybrać pewną liczbę tekstów jawnych i otrzymać ich szyfrogramy.
- Adaptacyjnie wybrany tekst jawny (ang. *adaptive-chosen-plaintext*) – kryptoanalityk może wykonywać kolejne próby ataku z wybranym tekstem jawnym.

Metody łamania szyfrów, cd.

- Wybrany szyfrogram (ang. *chosen-ciphertext*) – kryptoanalityk może wybrać pewną liczbę szyfrogramów i otrzymać ich postać jawną.
- Wybrany tekst – kryptoanalityk stosuje jednocześnie łamanie z wybranym tekstem jawnym i wybranym szyfrogramem.
- Wybrany klucz (ang. *chosen-key*) – kryptoanalityk posiada pewną wiedzę o powiązaniach między różnymi kluczami.
- Gumowa pałka (ang. *rubber-hose*) – kryptoanalityk stosuje groźby, szantaż lub tortury.
- Przekupstwo (ang. *purchase-key*) – kryptoanalityk kupuje potrzebne informacje.

Kategorie łamania szyfrów

- Znana jest metoda odtwarzania klucza na podstawie szyfrogramu.
- Znany jest algorytm odtwarzania tekstu jawnego bez znajomości klucza.
- Poznano tekst jawny przechwyconego szyfrogramu.
- Poznano częściową informację o kluczu i tekście jawnym, np. pewne bity lub fragmenty.

Obliczeniowe bezpieczeństwo systemu kryptograficznego

- Informatyka interesują tylko obliczeniowe metody łamania szyfrów. :-)
- Szyfr jest bezwarunkowo bezpieczny (ang. *unconditionaly secure*), jeśli nie jest możliwe odtworzenie tekstu jawnego, nawet przy nieograniczonych zasobach obliczeniowych.
- Tylko szyfr z kluczem jednorazowym jest bezwarunkowo bezpieczny.
- Szyfr jest obliczeniowo bezpieczny lub silny, jeśli nie może być złamany za pomocą dostępnych obecnie i w przyszłości zasobów obliczeniowych.
- Złożoności metody łamania:
  - danych – ilość danych wejściowych niezbędnych do złamania,
  - obliczeniowa – czas niezbędny do złamania,
  - pamięciowa – wielkość pamięci niezbędnej do złamania.

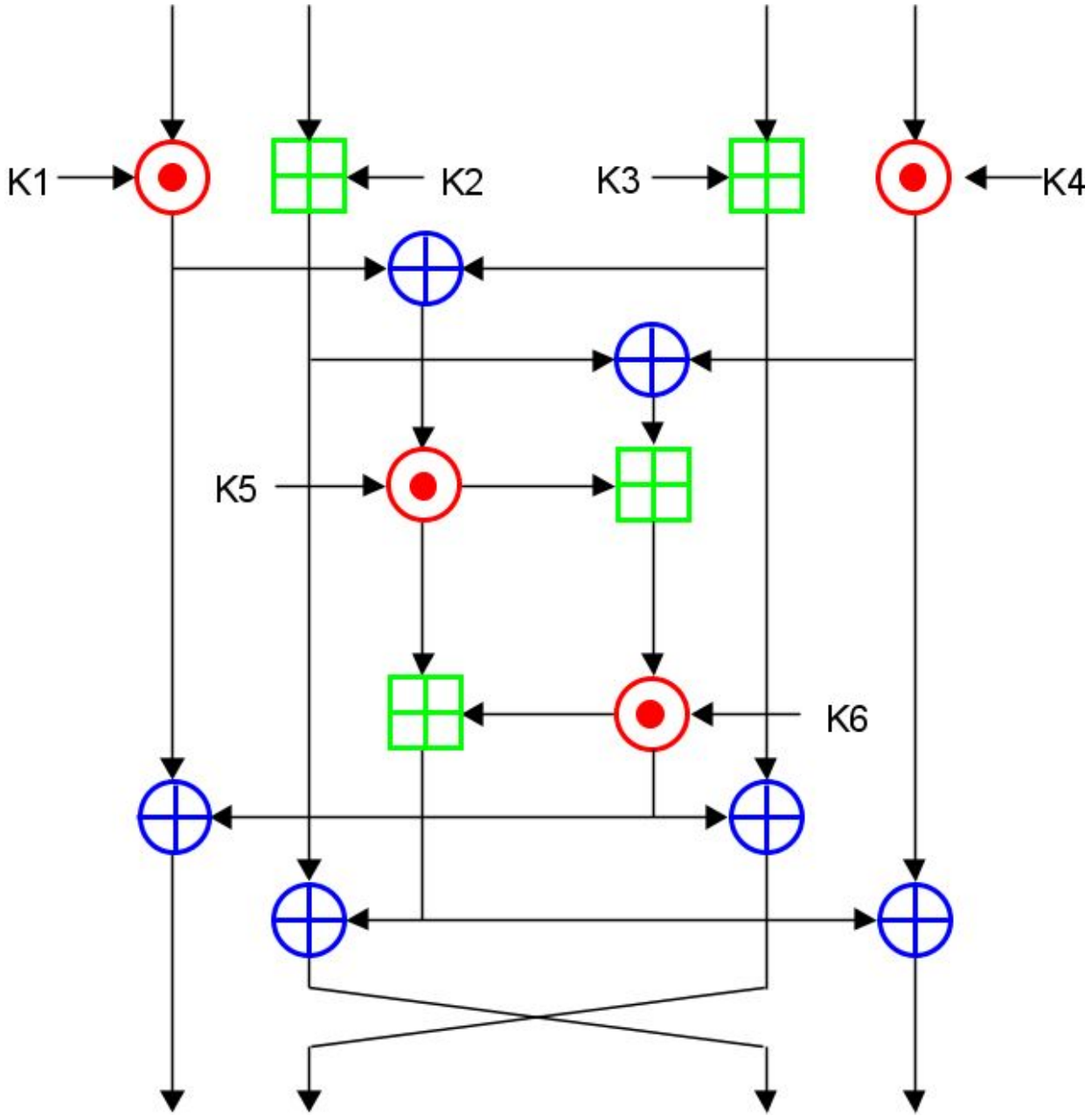
Szyfr z kluczem jednorazowym

- Rozważamy alfabet A z operacjami dodawania i odejmowania modulo rozmiar tego alfabetu.
- Zwykle $A = \{0, 1\}$, a dodawanie i odejmowanie to operacja XOR.
- Klucz to ciąg losowych znaków z alfabetu A o długości niemniejszej niż długość tekstu jawnego.
- Szyfrowanie polega na dodaniu do każdego znaku tekstu jawnego odpowiedniego znaku klucza.
- Deszyfrowanie polega na odjęciu od każdego znaku szyfrogramu odpowiedniego znaku klucza.
- Nadawca wyciąga klucz z sejfu, szyfruje tekst jawny i niszczy klucz.
- Odbiorca wyciąga klucz z sejfu, deszyfruje szyfrogram i niszczy klucz.
- Trzeba rozwiązać problem dystrybucji i przechowywania kluczy.

Przykład szyfru blokowego – IDEA

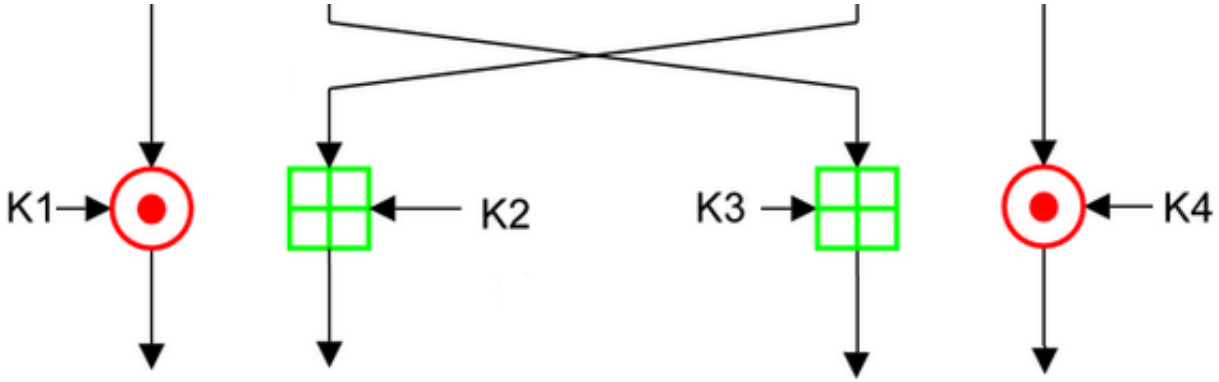
- Opracowany w latach 1990–1992 przez Xuejia Lai i Jamesa Masseya.
- Patenty wygasną do 2012 r.
- 64-bitowy blok danych jest dzielony na cztery 16-bitowe podbloki.
- Używa 128-bitowego klucza, z którego generuje się 52 16-bitowe podklucze.
- Łatwo można go zaimplementować sprzętowo i programowo na procesorach 16-bitowych.
- Używa trzech podstawowych operacji:
  - $\oplus$ – bitowa suma modulo 2 (XOR),
  - $\boxplus$ – dodawanie modulo 2^{16},
  - $\odot$ – mnożenie modulo $2^{16} + 1$.
- Żadna para operacji nie spełnia prawa rozdzielności ani łączności.

IDEA, pojedyncza iteracja



IDEA, cd.

- Iteracja powtarzana jest 8-krotnie.
- Na zakończenie stosowany jest dodatkowy etap przekształcania wyniku.



IDEA, mnożenie

- Opearcję mnożenia modulo $2^{16}+1$ można zaimplementować następująco

$$x = (ab \bmod 2^{16}) - (ab \operatorname{div} 2^{16}),$$

$$ab \bmod (2^{16}+1) = \begin{cases} x, & \text{gdy } x \geqslant 0, \\ x + 2^{16} + 1, & \text{w p.p.} \end{cases}$$

DES

- Data Encryption Standard:
  - opracowany w USA na podstawie projektu IBM przy współudziale NSA;
  - ogłoszony jako standard w 1976;
  - 64-bitowy blok danych;
  - 56-bitowy klucz;
  - obecnie uznawany za słaby.
- 3DES:
  - trzykrotny DES;
  - 112-bitowy klucz;
  - próba uratowania algorytmu DES (istnieją implementacje sprzętowe).
- DES z niezależnymi podkluczami:
  - 768-bitowy klucz.
- DESX:
  - opracowany przez RSA Data Security;
  - dodatkowy 64-bitowy klucz;
  - „wybielanie” uodparniające na atak brutalny.

Następca DES

- CRYPT(3):
  - wariant DES spotykany w systemach uniksowych;
  - wykorzystywany jako funkcja jednokierunkowa dla haseł.
- s^nDES – zmienione S-bloki.
- DES z S-blokami zależnymi od klucza.
- AES – Advanced Encryption Standard
  - powstał w wyniku konkursu ogłoszonego przez NIST w 1997;
  - zaaprobowany jako standard i następca DES w 2002;
  - wykorzystuje algorytm Rijndeal opracowany w 1999 przez Joana Daemena i Vincenta Rijmena;
  - 128-bitowy blok danych;
  - klucz długości 128, 192 lub 256 bitów.

Rivest Cipher, Ron's Code

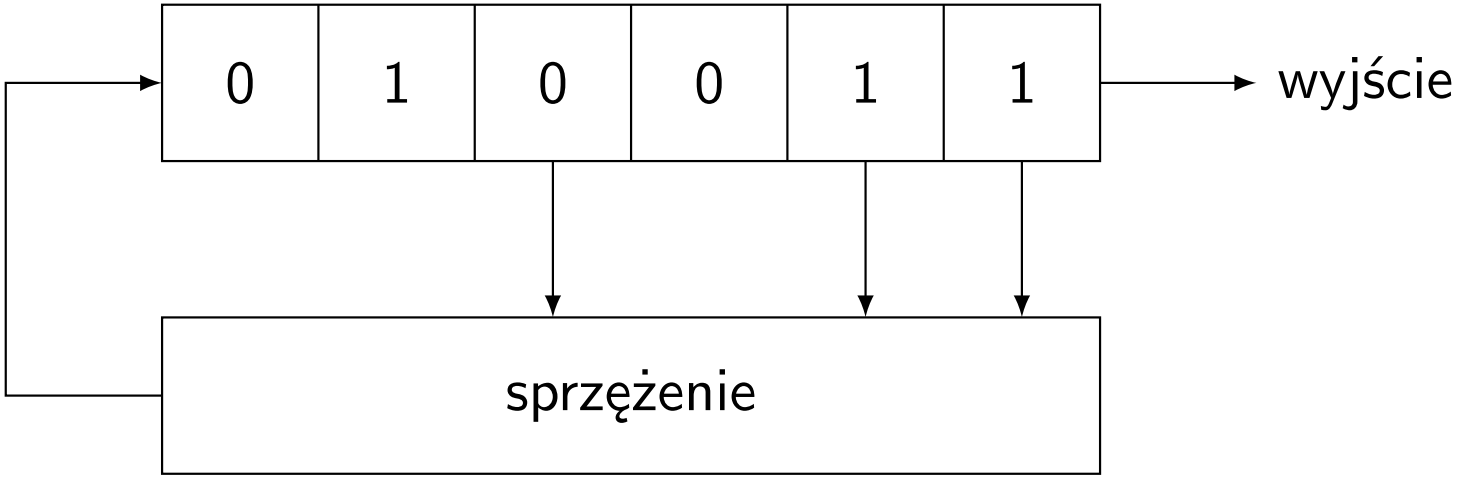
- Opatentowane algorytmy opracowane przez Ronalda Rivesta, pracownika MIT i założyciela RSA Data Security.
- Bardzo wydajne, dużo szybsze od DES.
- RC2:
  - blokowy, długość bloku 64 bity;
  - klucz długości od 8 do 128 bitów;
  - używany m.in. przez firmę Lotus.
- RC4:
  - strumieniowy;
  - klucz długości od 40 do 256 bitów;
  - używany m.in. w SSL, WEP, WPA.
- RC5:
  - blokowy, długość bloku 32, 64 lub 128 bitów;
  - klucz długości do 2040 bitów.
- RC6:
  - opracowany na bazie RC5 na potrzeby konkursu AES;
  - blokowy, długość bloku 128 bitów;
  - klucz długości 128, 192 lub 256 bitów.

Inne szyfry symetryczne

- CAST:
  - rodzina szyfrów zbliżonych do DES o zmiennej długości kluczy i bloków;
  - CAST-128 opisany w RFC 2144;
  - CAST-256 opracowany na potrzeby konkursu AES.
- SAFER:
  - algorytm blokowy opracowany przez Jamesa L. Masseya;
  - wersja z kluczem 64 bitowym (SAFER-K64) obejmująca 6 rund;
  - wersja z kluczem 128 bitowym (SAFER-K128) – do 12 rund (rekomendowane 10).
- BLOWFISH:
  - opracowany przez Bruce'a Schneiera;
  - blok danych ma 64 bity;
  - podstawowy klucz ma długość do 448 bitów;
  - w algorytmie występuje 16 iteracji wykorzystujących 18 kluczy pomocniczych (wyznaczanych każdorazowo przed szyfrowaniem i deszyfrowaniem) i 4 S-bloki 256-elementowe o wartościach zależnych od klucza podstawowego, danych oraz liczby π.

Szyfry strumieniowe, generatory ciągów pseudolosowych

- Wiele szyfrów strumieniowych działa w oparciu o generator ciągu pseudolosowego.
- Szyfrowanie i deszyfrowanie polega na xorowaniu tego ciągu odpowiednio z tekstem jawnym i szyfrogramem.
- Prosta realizacja sprzętową w oparciu o rejestry przesuwające z liniowym sprzężeniem zwrotnym LFSR (ang. *linear feedback shift register*).



- W najprostszym przypadku sprzężenie zwrotne to suma modulo 2.

Przykłady szyfrów strumieniowych

- A5:
  - szyfry stosowane w GSM;
  - 3 rejestry o długościach 19, 22 i 23 bity;
  - w każdej rundzie są zwykle taktowane 2 rejestry;
  - wyjście to suma wyjść rejestrów modulo 2.

RSA

- Opublikowany w 1978 roku przez Ronalda Rivesta, Adi Shamira i Leonarda Adlemana.
- Niedawno wygasła jego ochrona patentowa.
- Dobór kluczy:
  - p, q – losowo wybrane duże liczby pierwsze;
  - $n = pq$ – moduł;
  - e – liczba względnie pierwsza z $(p-1)(q-1)$;
  - d – liczba wyznaczona tak, że zachodzi $ed \bmod (p-1)(q-1) = 1$;
  - (n, d) – klucz prywatny;
  - (n, e) – klucz publiczny;
  - po wygenerowniu kluczy liczby p, q powinny być wymazane i nigdy nie ujawnione.
- Szyfrowanie
$$C = M^e \bmod n.$$
- Deszyfrowanie
$$M = C^d \bmod n.$$

Szyfr ElGamala

- Opublikowany w 1985 roku.
- Niechroniony patentem, brak ograniczeń eksportowych USA – wykorzystuje koncepcję (i patent) Diffiego-Hellmana, lecz ów patent wygasł w 1997 r.
- Szyfrowanie wymaga losowo wybranej wartość, dlatego ten sam tekst jawny każdorazowo daje inny szyfrogram.
- Szyfrogram jest dwukrotnie dłuższy od tekstu jawnego.
- Generowanie kluczy:
  - wybieramy losowo liczbę pierwszą p;
  - wykorzystujemy grupę multiplikatywną $\mathbb{Z}_p^*$;
  - wybieramy element pierwotny (generator) g grupy $\mathbb{Z}_p^*$;
  - wybieramy losowo liczbę $x \in \{0, 1, \ldots, p-1\}$;
  - obliczamy $y = g^x \bmod p$;
  - kluczem publicznym jest trójka (g, p, y);
  - g i p mogą być wspólnie wykorzystywane przez grupę użytkowników;
  - kluczem prywatnym jest x.

Szyfr ElGamala, cd.

- Szyfrowanie:
  - wybieramy losowo liczbę k względnie pierwszą z $p-1$;
  - obliczamy $a = g^k \bmod p$;
  - obliczamy $b = y^k \cdot M \bmod p$;
  - szyfrogram to para (a, b).
- Deszyfrowanie:
  - $M = b/a^x \bmod p$.
- Uzasadnienie:
  - $b/a^x \equiv y^k \cdot M/a^x \pmod p$;
  - $a^x \equiv g^{kx} \pmod p$;
  - $y \equiv g^x \pmod p$;
  - $b/a^x \equiv g^{xk} \cdot M/g^{kx} \pmod p$;
  - $b/a^x \equiv M \pmod p$.

Inne szyfry asymetryczne

- Zastosowanie znajdują:
  - problem plecakowy,
  - liniowe kody korekcyjne,
  - uogólnienia RSA,
  - faktoryzacja złożenia dwóch automatów skończonych.
- Najbardziej obiecujące są krzywe eliptyczne.