Ataki na kryptograficzne funkcje skrótu

Krystian Matusiewicz

Danmarks Tekniske Universitet

Enigma 2008, 28 maja 2008

Motywacja

W wielu zastosowaniach przydatna jest skrótowa reprezentacja przetwarzanych danych, rodzaj "wizytówki".

- Jeśli dwie "wizytówki" są różne, dane są też *na pewno* różne
- Jeśli są identyczne, z dużym prawdopodobieństwem odnoszą się do tych samych danych

Przykłady:

- Tablice haszowe: w oparciu o wartość "wizytówki" decydujemy o pozycji danych w tablicy
- Wyszukiwanie wzorca metodą Rabina-Karpa: jeśli bieżący podciąg ma tą samą "wizytówkę" co poszukiwany, sprawdzamy go dokładnie
- Sumy kontrolne: jeśli obliczona ponownie "wizytówka" danych się różni, dane zostały zmienione

Funkcje skrótu

Definicja

Funkcja skrótu – funkcja, która mapuje ciągi bitowe dowolnej długości na ciągi o określonej, stałej długości.

$$h : \{0,1\}^* \rightarrow \{0,1\}^n$$

Przykłady:

- operator reszty z dzielenia przez 2^n: ciąg danych przedstawiamy jako dużą liczbę i dzielimy ją przez 2^n – reszta jest skrótem danych
- CRC: ciąg bitów danych przedstawiamy jako wielomian nad $\mathbb{F}_2$ i dzielimy go przez ustalony wielomian stopnia n. Otrzymana reszta jest sumą kontrolną danych.

Znajdowanie przeciwobrazów może być łatwe

Dla skrótu o długości n bitów prawdopodobieństwo, że losowo wybrane dane dają zadany z góry skrót to ok. 2^{-n}.

Nie znaczy to, że działając umyślnie nie można ich łatwo znajdować.

Przykład: Niech s będzie wartością otrzymaną po zastosowaniu funkcji skrótu "reszta modulo 2^n". Wszystkie dane wejściowe dające skrót s mają postać: s, $s + 2^n$, $s + 2 \cdot 2^n$, ..., $s + k \cdot 2^n$, ...

Problem: możliwy atak na system

- Atak na tablicę haszową: jeśli wiemy, jak funkcja oblicza skrót używany do indeksowania w tablicy haszowej i mamy kontrolę nad danymi wejściowymi, możemy "zdegenerować" tablicę – w rezultacie atak DoS
- Atak na sieć p2p: jeśli potrafimy znajdować inne dane dające taki sam skrót możemy podmieniać fragmenty danych i "zatruć" sieć p2p.

Rozwiązanie: kryptograficzne funkcje skrótu

Nieformalnie, kryptograficzna funkcja skrótu to funkcja h posiadająca następujące właściwości:

Odporność na znajdowanie przeciwobrazów

obliczeniowo niewykonalne jest znajdowanie danych x znając tylko $h(x)$.

Odporność na znajdowanie drugich przeciwobrazów

Obliczeniowo niewykonalne jest znajdowanie innych danych wejściowych $x' \neq x$ znając tylko $h(x)$ i x.

Odporność na znajdowanie kolizji

Obliczeniowo niewykonalne jest znajdowanie par różnych wiadomości x, x' takich, że $h(x) = h(x')$.

Ataki ogólne

Pytanie: jaka jest górna granica złożoności ataku znajdującego przeciwobrazy, drugie przeciwobrazy czy kolizje dla funkcji dającej skrót długości n bitów?

Idealnie losową funkcję skrótu można modelować za pomocą wyroczni losowej: Na każde nowe zapytanie x wyrocznia $\mathcal{O}$ odpowiada wylosowanym z rozkładem jednostajnym skrótem $\mathcal{O}(x) \in \{0,1\}^n$. Dla takich samych danych wejściowych wyrocznia zwraca ten sam wynik (zapamiętuje to, co już powiedziała).

Złożoność ataków ogólnych: przeciwobrazy

- Znajdowanie przeciwobrazów wymaga 2^n zapytań (czyli obliczeń funkcji skrótu)
- Znajdowanie drugich przeciwobrazów wymaga 2^n zapytań

Wynika to z tego, że odpowiedzi na zapytania są niezależne od siebie i generowane z rozkładem jednostajnym.

Złożoność ataków ogólnych: kolizje

- Znajdowanie kolizji wymaga $2^{n/2}$ zapytań

Mniejsza złożoność wynika z faktu, że potrzebujemy dowolnej pary wiadomości. Możemy więc obliczać kolejne skróty i zapamiętywać dotychczas obliczone. Dla każdego nowego skrótu porównujemy go ze wszystkimi poprzednimi.

Taka metoda wymaga wykorzystania dużej pamięci rzędu $2^{n/2}$. Są jednak lepsze metody, bazujące na znajdowaniu cykli w grafie funkcyjnym, które wymagają znacznie mniej pamięci.

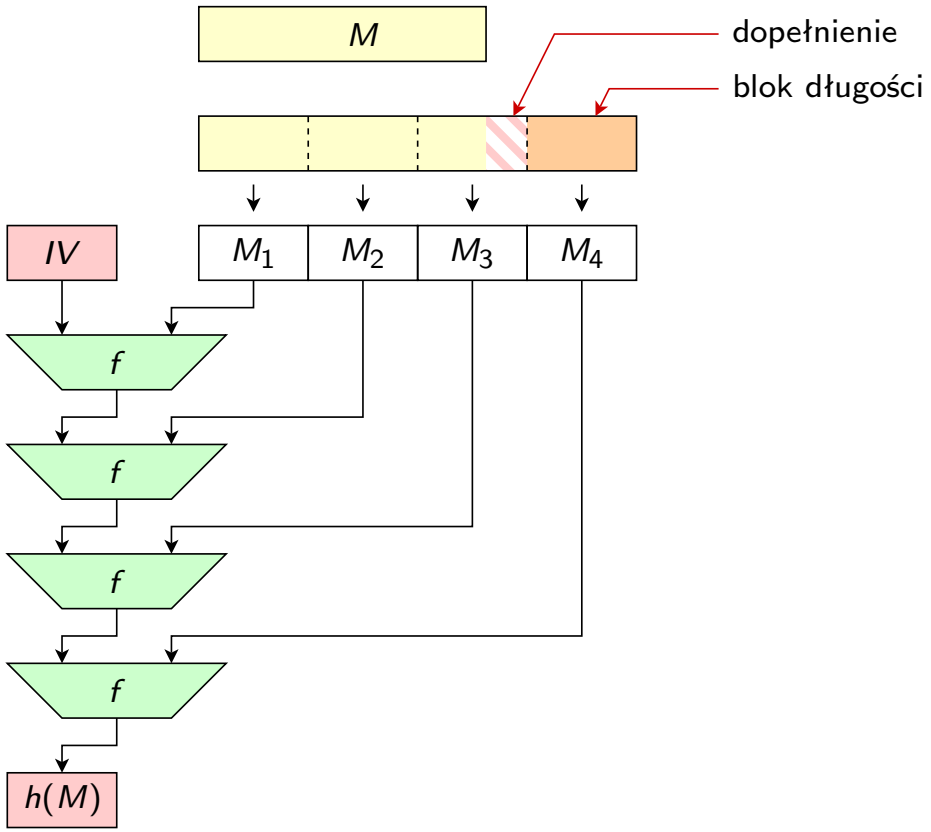
Konstrukcja Merkle-Damgård

Jak zaprojektować funkcję, która może przyjmować dane dowolnej długości?

Użyć wielokrotnie funkcji kompresji:

$$f : \{0,1\}^{n+k} \longrightarrow \{0,1\}^n$$

- Dopełnić wiadomość i dodać blok długości
- Podzielić na bloki rozmiaru k
- $h_{i+1} := f(h_i, M_i)$



Właściwości konstrukcji Merkle-Damgård

Zalety:

- redukcja bezpieczeństwa: funkcja kompresji jest odporna na kolizje $\Longrightarrow$ funkcja skrótu odporna na kolizje
- "strumieniowalność": każdy bit wiadomości jest używany tylko raz

Problemy:

- ataki wydłużenia wiadomości
- multikolizje
- znajdowanie przeciwobrazów dla długich wiadomości
- słabości właściwości losowych

Ataki wydłużania wiadomości

Po znalezieniu dwóch kolidujących wiadomości M, M' o tej samej długości można znajdować wiele kolizji poprzez dołączanie dodatkowych bloków.

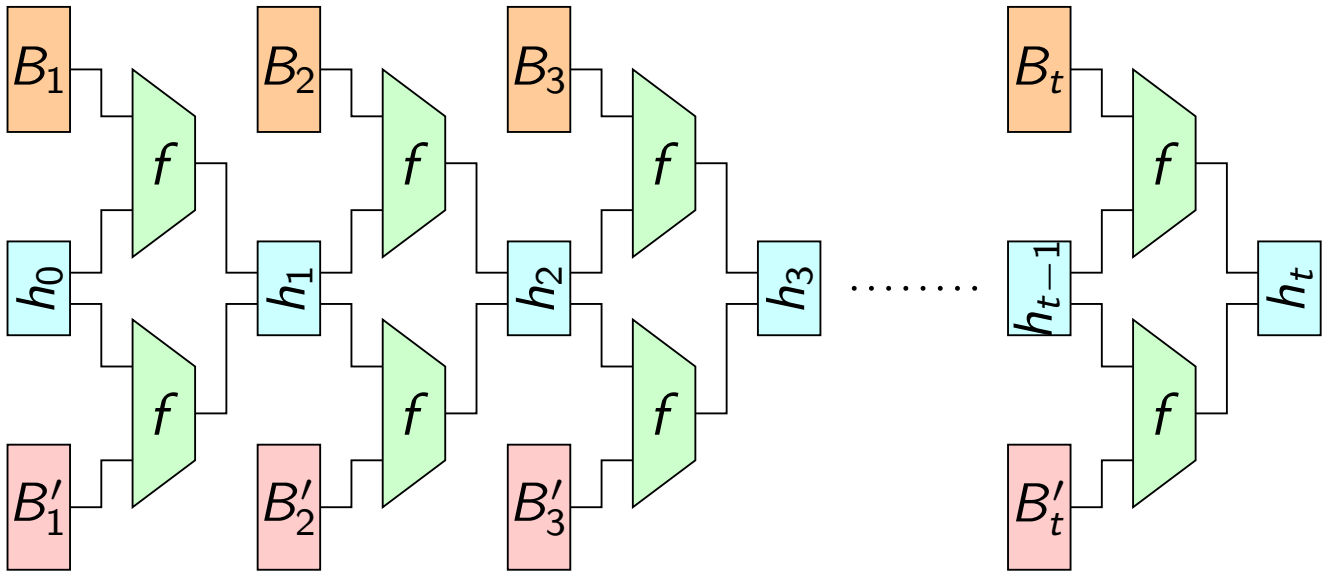
$$h(M) = h(M) \implies h(M||Q) = h(M'||Q)$$

Multikolizje dla iterowanych funkcji skrótu

Multikolizja

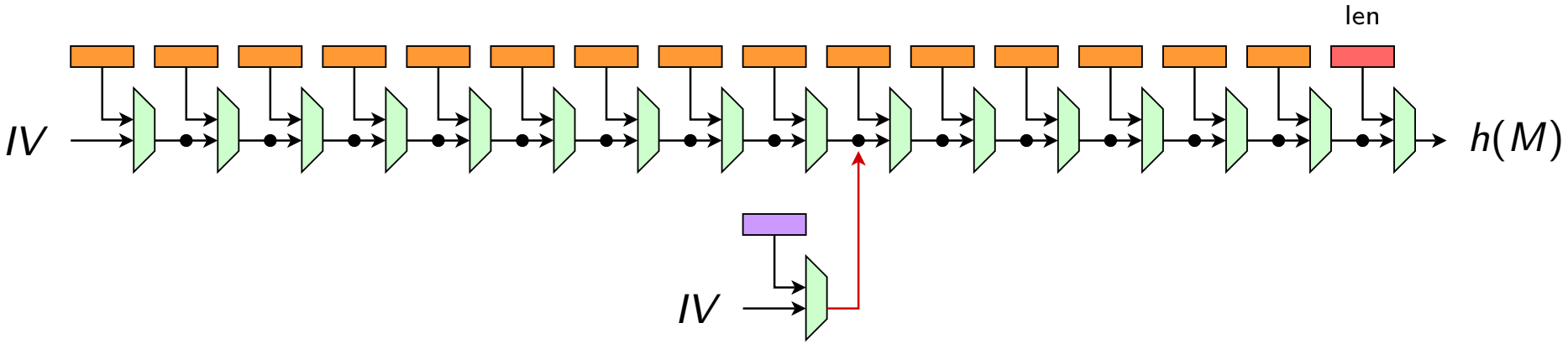
Zbiór $M_1, M_2, \ldots, M_t$ taki, że $h(M_1) = h(M_2) = \cdots = h(M_t)$

Złożoność idealna: $2^{n(t-1)/t}$, dla funkcji iteracyjnych: $\lceil \lg t \rceil \cdot 2^{n/2}$.



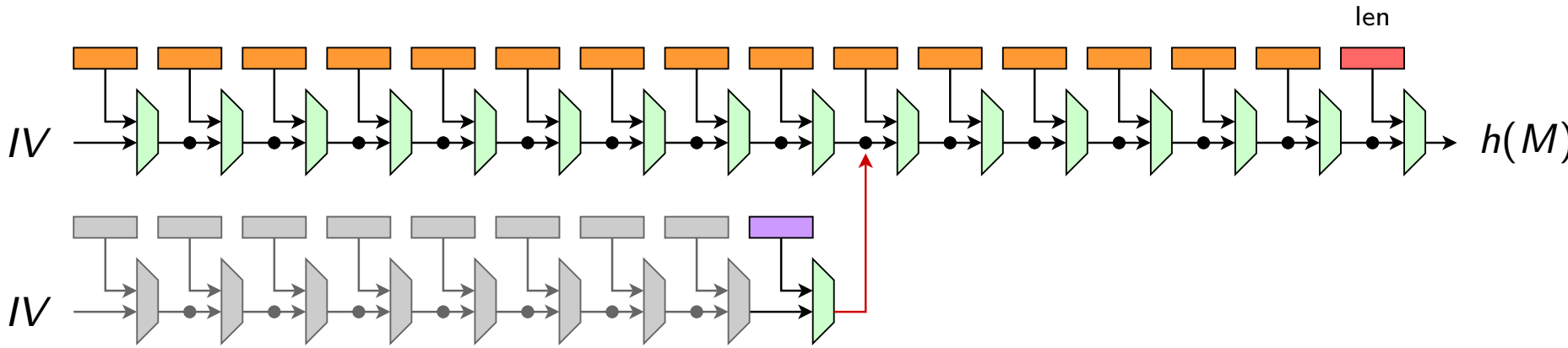
Drugie przeciwobrazy dla długich wiadomości

- Dla wiadomości o długości 2^t bloków, mamy $2^t - 1$ stanów pośrednich $h_1, h_2, \ldots, h_{2^t-1}$
- Dowolna wartość będzie równa któremuś stanowi pośredniemu z prawd. $\approx 2^t/2^n$, złożoność: 2^{n-t}
- "Prawie atak" – bo nie zgadza się długość wiadomości w ostatnim bloku



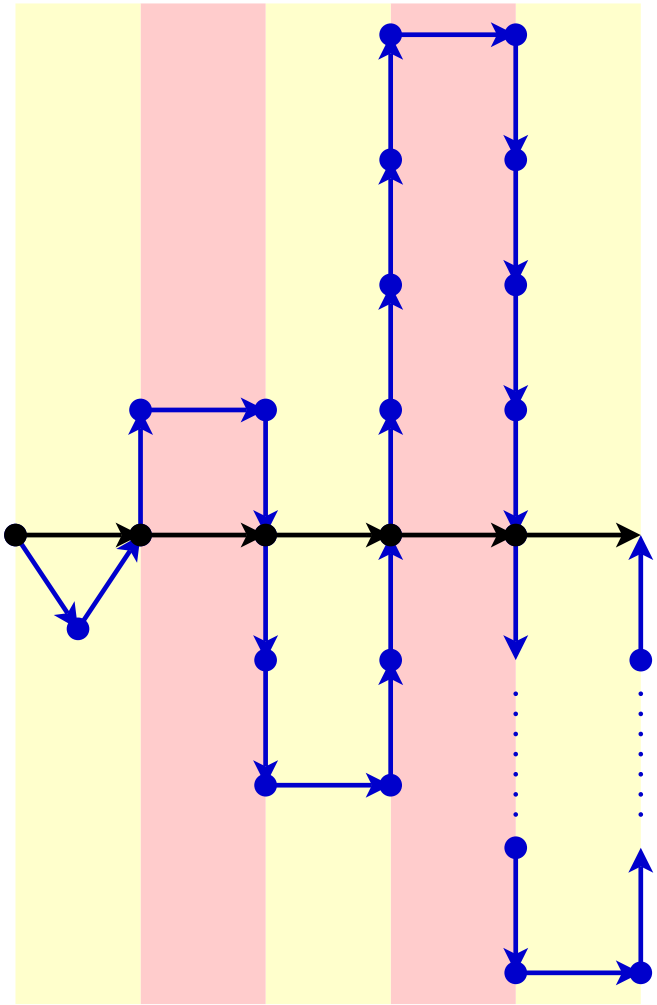
Drugie przeciwobrazy c.d.: wiadomości rozszerzalne

- Aby rozwiązać ten problem, trzeba znaleźć metodę konstruowania wiadomości o zmiennej długości kończącej się zadanym z góry stanem pośrednim

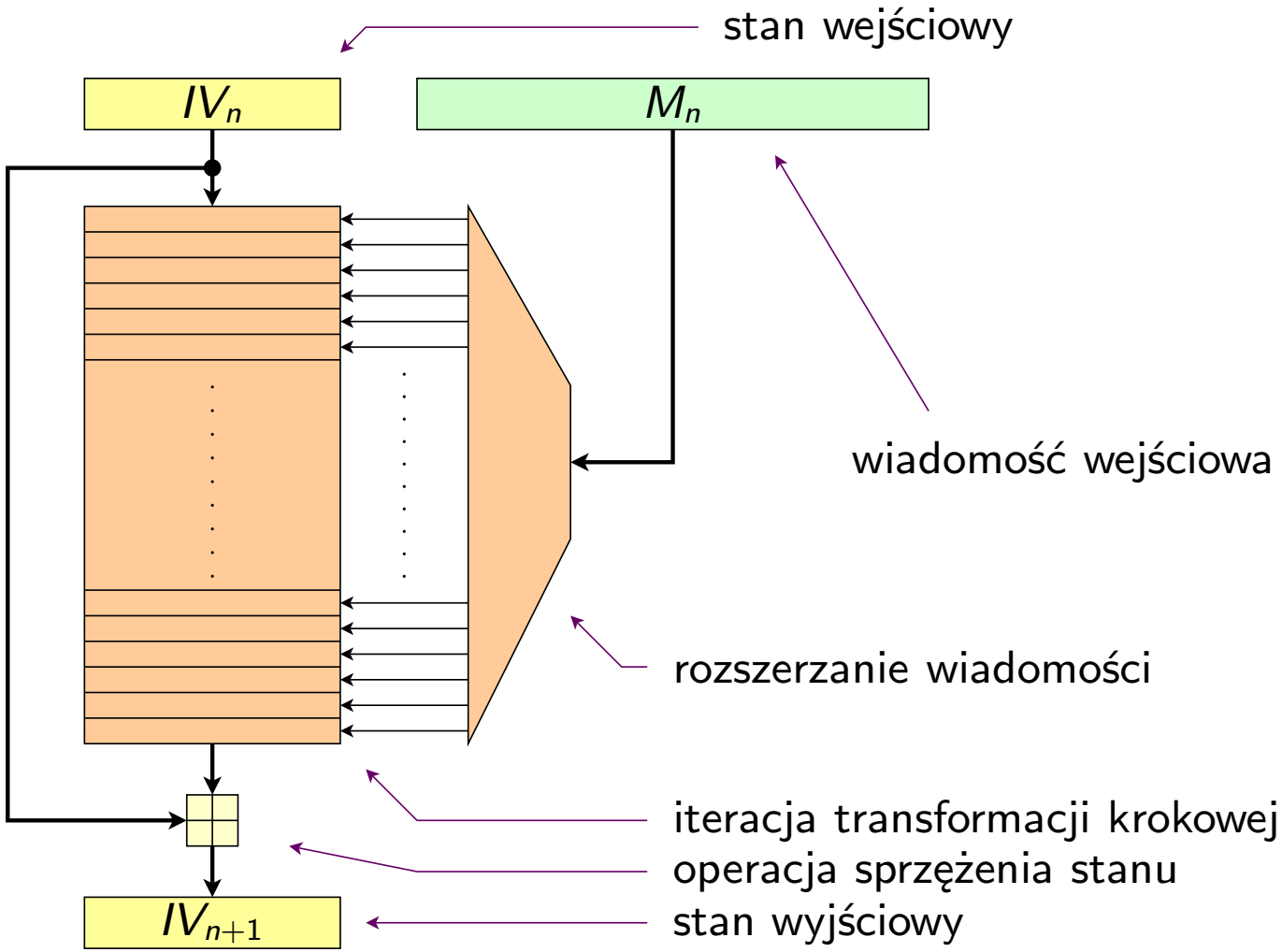


Wiadomości rozszerzalne

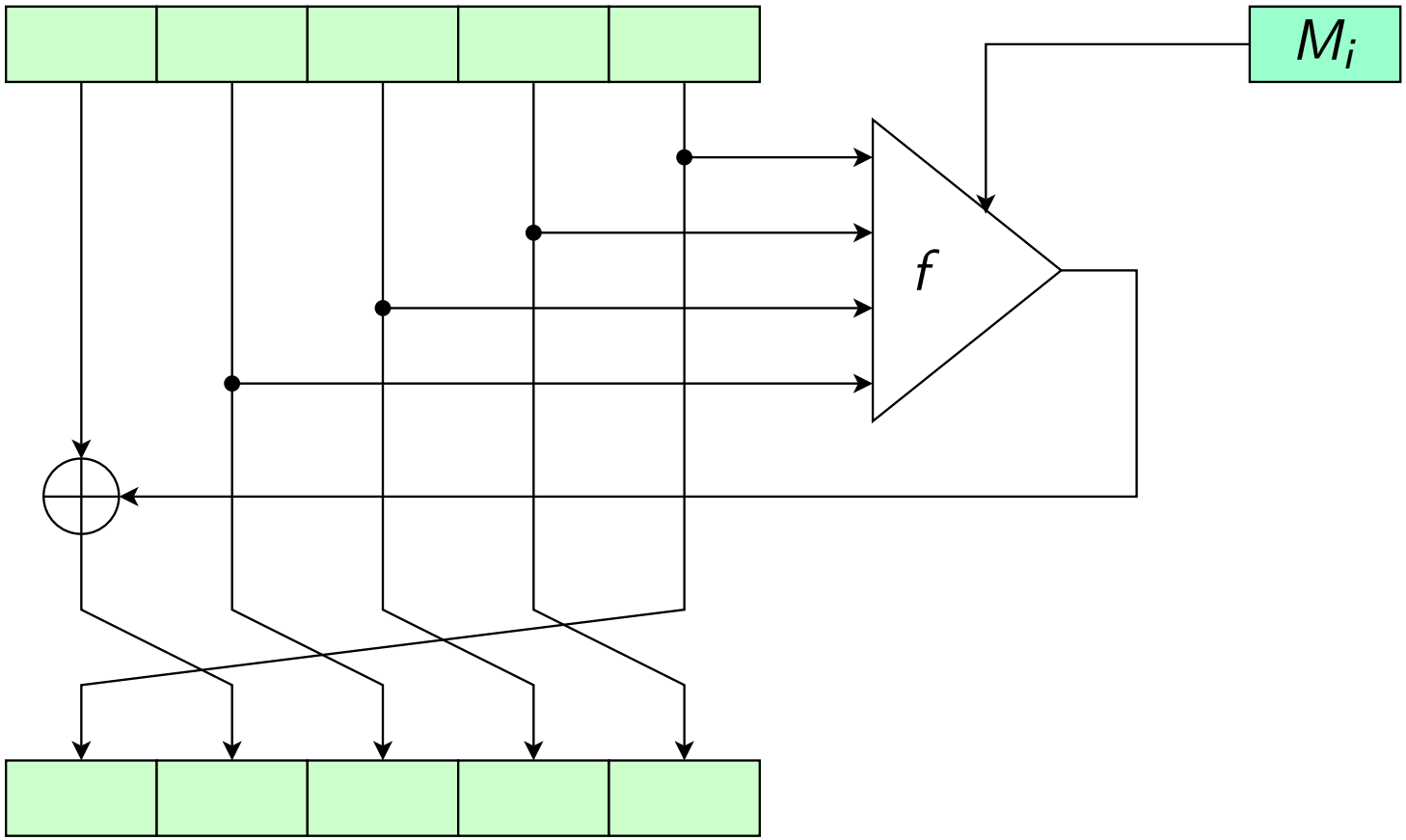
- Idea oparta na multikolizjach, ale w parach znajdowane są wiadomości o różnej ilości bloków.
- w kroku k znajdowana jest kolizja pomiedzy wiadomością o 1 bloku i $2^k + 1$ blokach
- Po znalezieniu t multikolizji (koszt $t \cdot 2^{n/2}$) możemy skonstruować wiadomość o długości od t do $2^t + t - 1$ bloków



Funkcje skrótu z rodziny MD



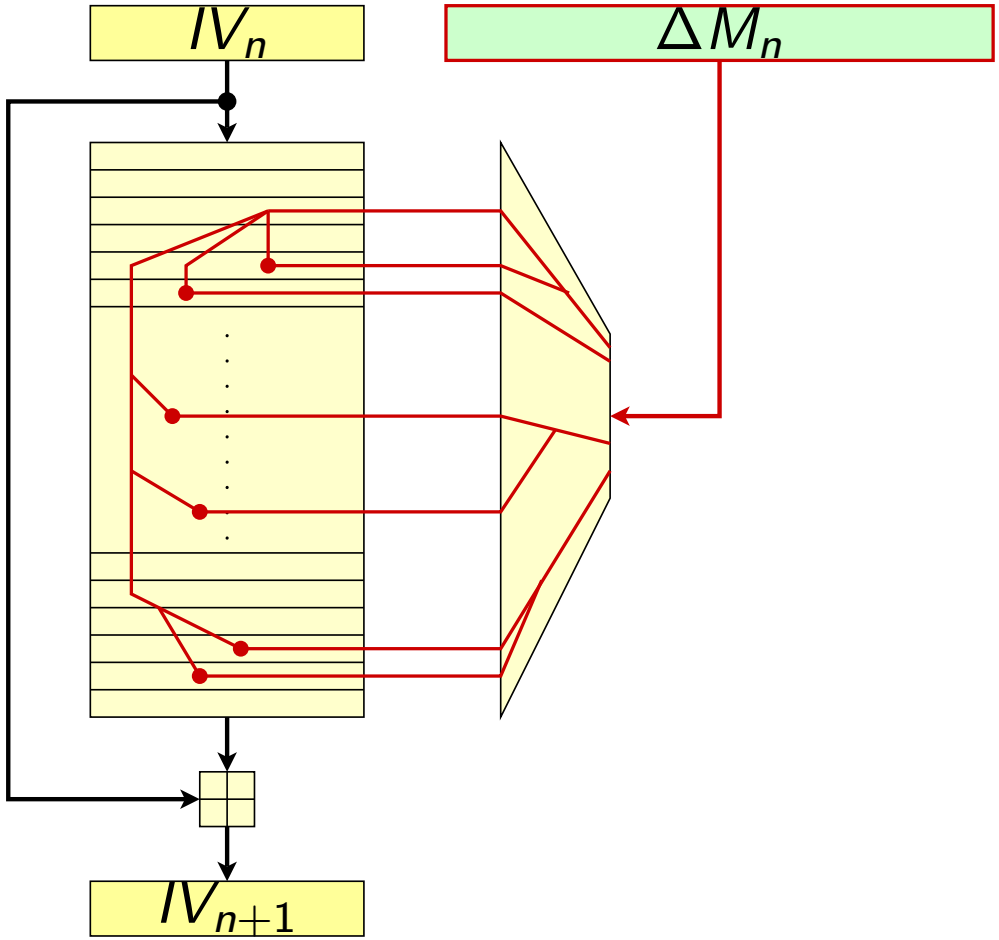
Funkcje rodziny MD: transformacja krokowa



Podstawowa technika: kryptoanaliza różnicowa

- Śledzimy różnice pomiedzy dwoma egzemplarzami funkcji którym podano dwie wiadomości różniące się w określony sposób
- Atak różnicowy podzielony na dwie fazy: znajdowanie ścieżki i znajdowanie warunków do jej zaistnienia
- W pierwszym etapie analizuje się uproszczone modele funkcji i szuka ścieżki różnicowej o dużym prawdopodobieństwie
- W drugim etapie określa się zbiór warunków, które zapewniają propagację różnicy według ścieżki i szuka rozwiązania

Ilustracja kryptoanalizy różnicowej funkcji MD



Nowatorskie podejście Wang et al

- Jednoczesne śledzenie różnic XOR i modularnych (śledzenie różnic binarnych ze znakiem)
- Metoda "modyfikowania wiadomości" która usprawnia znajdowanie wiadomości spełniających warunki na ścieżkę różnicową

Status funkcji z rodziny MD

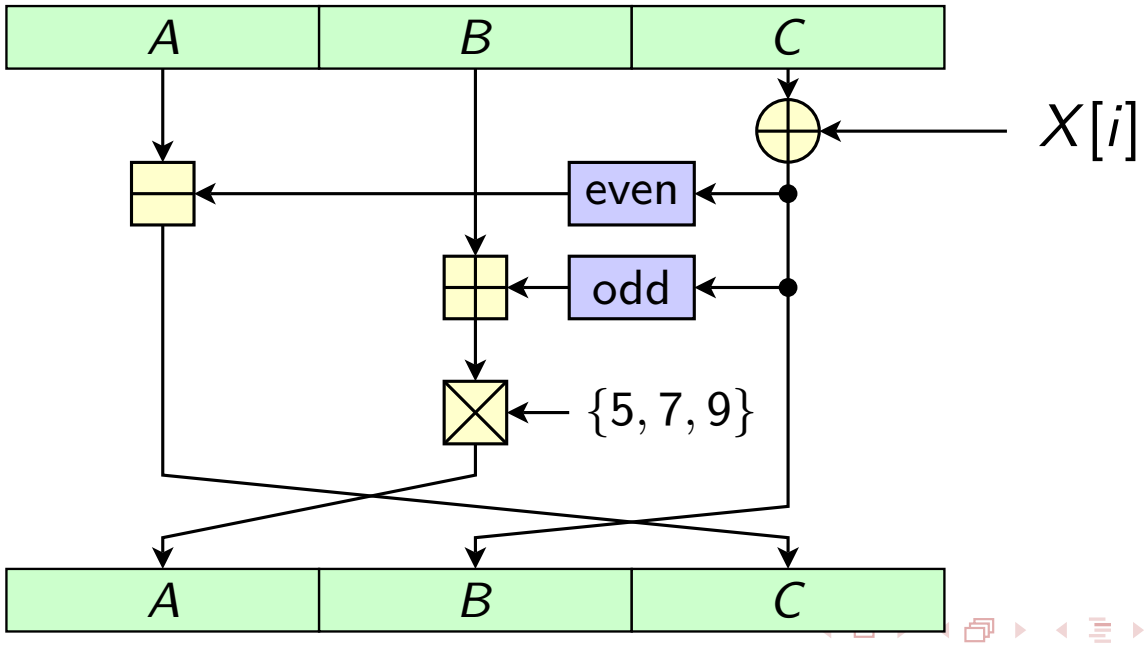
- Ataki na MD4 tak efektywne że znalezienie kolizji niewiele trudniejsze od obliczenia funkcji
- MD4 nie jest jednokierunkowa
- Kolizje MD5 możliwe do znalezienia w ciągu 1 minuty na PC
- Kolizje MD5 o praktycznym znaczeniu: certyfikaty X.509, protokół APOP, etc.
- Kolizje SHA-0 możliwe w czasie ok. 1 godz. na PC
- Atak na 70 (z 80) kroków SHA-1 o złożoności 2^{43}
- Najlepszy atak na SHA-256 to atak na 24 kroki z 64.

Ataki na inne konstrukcje

- Ciekawsze funkcje skrótu znacząco różne od funkcji rodziny MD i ataki na nie.
- Subiektywnie "ciekawe"

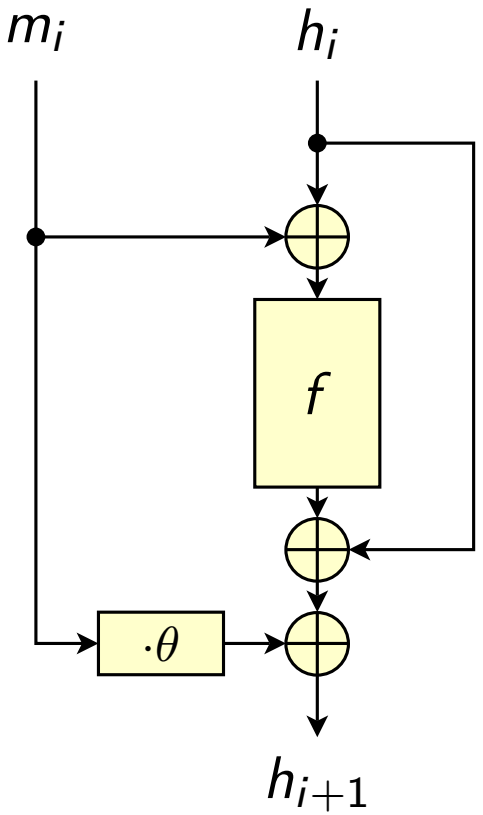
Tiger

- Zaprojektowana przez Rossa Andersona i Eli Bihama w 1995
- 64-bitowe rejestry, “target-heavy” UFN
- Używa S-Boxów (even, odd)
- pseudo-blisko kolizja dla całej funkcji, pseudo-kolizje dla 23 rund z 24



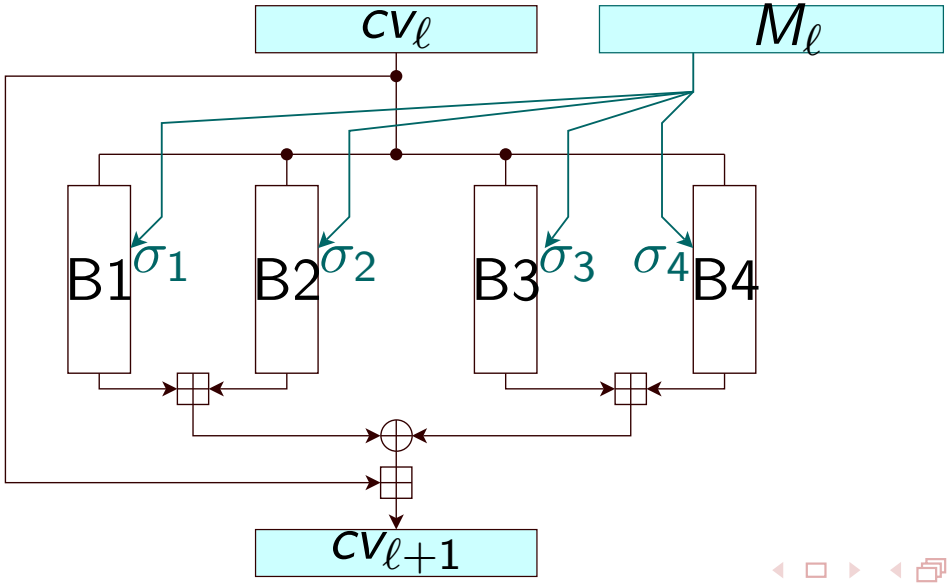
SMASH

- Zaproponowana przez Knudsena [FSE 2005]
- Nowa strategia budowania funkcji kompresji, oparta na jednej, ustalonej permutacji f [np. szyfr blokowy ze stałym kluczem]
- Niestandardowy tryb iteracyjny
- Złamana: znajdowanie kolizji, przeciwobrazów dla wszystkich instancji



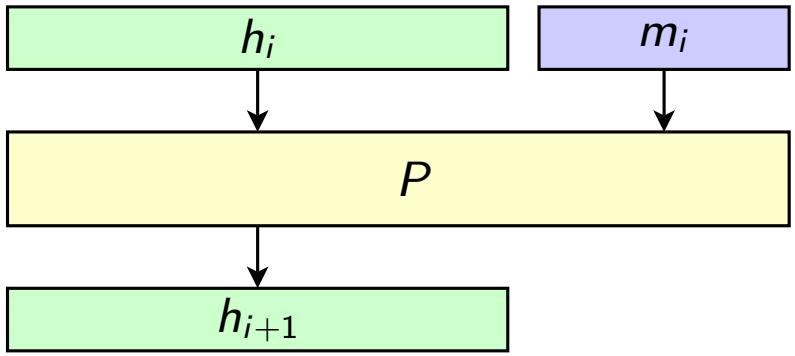
FORK-256

- Zaprojektowana przez Hong et al [FSE 2006]
- Cztery krótkie (8 kroków) równoległe gałęzie
- Tranformacja krokowa oparta o "target-heavy" UFN, tylko XOR, ADD i rotacje
- Złamana: kolizje w czasie 2^{126}, praktyczne blisko-kolizje



Grindahl

- Zaproponowana przez Knudsena, Rechbergera, Thomsena [FSE 2007]
- Filozofia "Concatenate-Permute-Truncate"
- Jako permutacje używa konstrukcji bazującej na komponentach AES
- Duży stan wewnętrzny
- Grindahl-256 złamany, kolizje w czasie 2^{117}



LASH

- Funkcja oparta na ideach algebry liniowej i teorii krat

$$f(r,s) = (r \oplus s) + H \cdot [Rep(r) \oplus Rep(s)]^T$$

- Inspirowana funkcjami z redukcja bezpieczeństwa do problemu SVP
- Jednak heurystyczna
- Złamana: przeciwobrazy ze złożonością $2^{4/7n}$, dla dowolnego IV $2^{7/8n}$

GOST

- Rosyjski standard federalny funkcji skrótu
-
- Kolizje ze złożonością 2^{105}: Kontak, Mendel, Rechberger, Szmidt, CRYPTO'08

Podsumowanie

- Podstawowe właściwości funkcji skrótu
- Ataki na strukturę iteracyjną Merkle-Damgård
- Ataki na funkcje z rodziny MD
- Inne konkstrukcje i ataki

Perspektywy:

- Konkurs Advanced Hash Standard
- Nowe tryby pracy (Haifa, sponge)
- Nowe funkcje kompresji