Методи та програмні засоби гешування із зав'язуванням даних

Розробив: Комаров А. О.

Науковий керівник: к. т. н., доцент каф. ЗІ
Баришев Ю. В.

Мета та задачі

- Об’єкт дослідження – процес криптографічного гешування даних.
- Предмет дослідження – розпаралелене гешування даних.
- Мета дослідження – підвищення стійкості геш-функцій, які передбачають розпаралелення обчислень, до мультиколізій.
- Задачі:
  - проаналізувати відомі геш-функції та атаки на них;
  - розробити геш-функції для розпаралеленого гешування, які стійкі до мультиколізій;
  - розробити алгоритми гешування;
  - розробити засоби, що реалізують ці алгоритми.

Методи досліджень

- теоретико множинний підхід для розробки конструкції гешування даних;
- методи математичної статистики для оцінки рівномірності розподілу вхідних даних;
- методи дедукції для оцінки рівня покращення стійкості до загальних атак;
- методи економічного аналізу для визначення економічної доцільності дослідження

Терміни, що використовуються

- Геш-функція – необоротна функція $h(\cdot)$, що перетворює вхідні дані M довільної довжини в вихідні дані фіксованої довжини.

$$f: \{0, 1\}^{\omega} \rightarrow \{0, 1\}^{n},\ \omega=\text{var},\ n=\text{const}$$

- Геш-значення – вихідні дані геш-функції
- Колізія – два повідомлення M і M', таких, що $h(M')=h(M)$
- Знаходження першого прообразу – по заданому значенню H знайти таке M, що $h(M)=H$.
- Знаходження другого прообразу – по заданому M знайти таке M', для якого $h(M')=h(M)$.

Конструкції криптографічного гешування

- Конструкція Меркля-Дамгарда:

$$h_i = f(h_{i-1}, m_i),$$
$$f\colon \{0,1\}^{t} \times \{0,1\}^{b} \rightarrow \{0,1\}^{t}$$

де h_i – проміжне геш-значення, отримане після обробки i-го блоку даних; m_i – i-й блок даних.

- Конструкція Девіса-Меєра для певного блокового шифру на основі ключа k $E_k(\cdot)$:

$$f(h_{i-1}, m_i) = E_{m_i}(h_{i-1}) + h_{i-1}$$

Конструкції криптографічного гешування

- Конструкція Люкса для подвійного каналу:

$$f': \{0,1\}^{\omega} \times \{0,1\}^{m} \to \{0,1\}^{\omega}$$

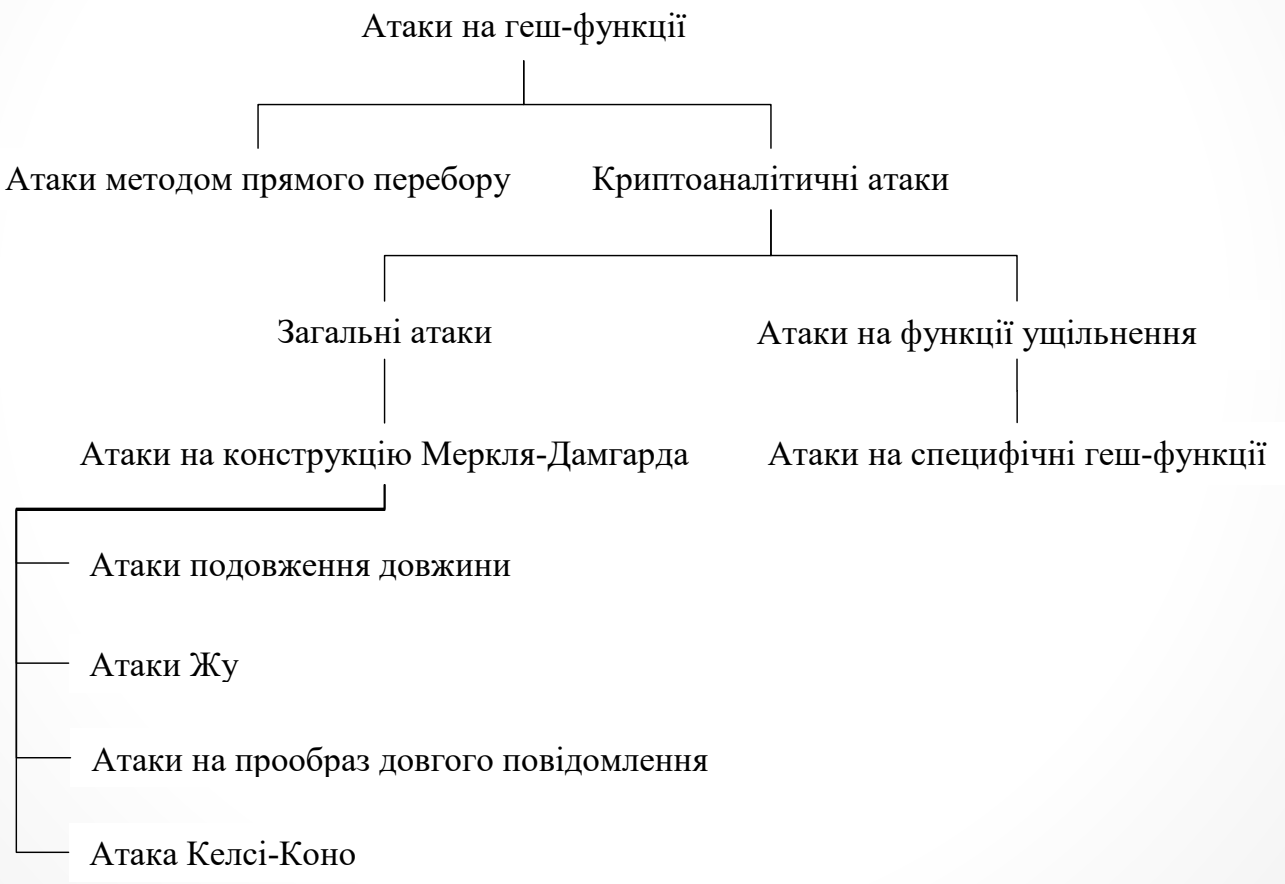
$$f'': \{0,1\}^{\omega} \to \{0,1\}^{n}$$

де вихідні значення $f'(\cdot)$ мають вдвічі більшу довжину, ніж $f''(\cdot)$.

- Конструкція Преніла

$$\begin{cases} h_i^{(1)} = f^{(1)}\left(h_{i-1}^{(1)}, m_i\right); \\ h_i^{(2)} = f^{(2)}\left(h_{i-1}^{(2)}, m_i\right); \\ \dots \\ h_i^{(q)} = f^{(q)}\left(h_{i-1}^{(q)}, m_i\right); \end{cases}$$

Класифікація атак на геш-функції (Гаураварам)

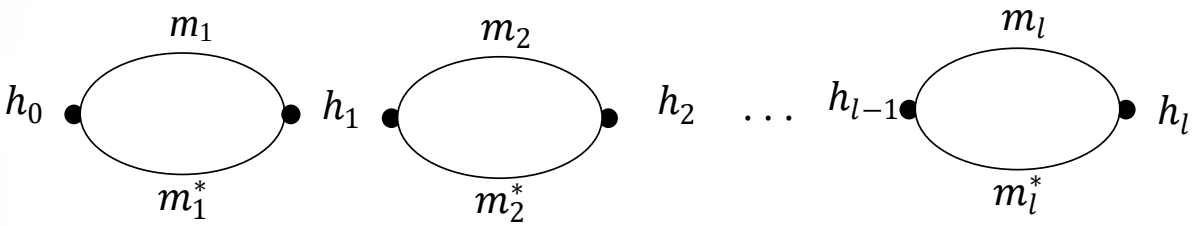


Атаки, що використовують мультиколізії

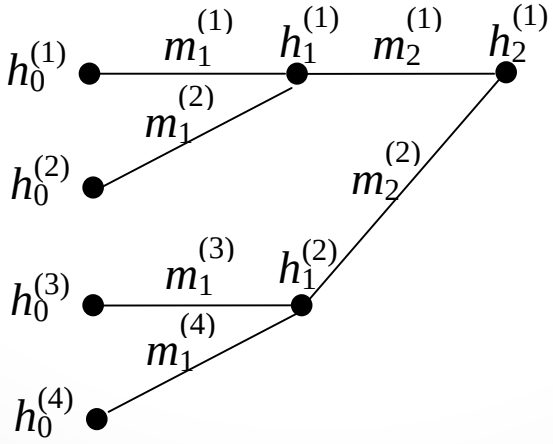
- Атака Жу

2^l *різних повідомлень за*

$l \cdot 2^{n/2}$ *викликів функції ущільнення*



- Атака Келсі-Коно



Конструкції гешування стійкі до мультиколізій

$$\begin{cases} h_i^{(1)} = f^{(1)}\left(h_{i-1}^{(1)}, m_i, m_{i-r_i}\right); \\ h_i^{(2)} = f^{(2)}\left(h_{i-1}^{(2)}, m_i, m_{i-r_i}\right); \\ \dots \\ h_i^{(q)} = f^{(q)}\left(h_{i-1}^{q)}, m_i, m_{i-r_i}\right); \\ r_i = rand(m_i). \end{cases} \quad \begin{cases} h_i^{(1)} = f^{(1)}\left(h_{i-1}^{(1)}, m_i, m_{i-r_i}\right); \\ h_i^{(2)} = f^{(2)}\left(h_{i-1}^{(2)}, m_i, m_{i-r_i}\right); \\ \dots \\ h_i^{(q)} = f^{(q)}\left(h_{i-1}^{q)}, m_i, m_{i-r_i}\right); \\ r_i = rand(m_i, m_{i+1}). \end{cases}$$

Перестановка

- для *j*-го каналу вхідними даними буде послідовність блоків даних:

$$\left\{m_1^{(j)}, m_2^{(j)}, ..., m_l^{(j)}\right\} = p^{(j)}\left(\left\{m_1, m_2, ..., m_l\right\}\right)$$

де $p^{(j)}(\cdot)$ – перестановка.

Метод розпаралеленого гешування

- Визначити конфігураційні значення:
  - s – розмір проміжного стану та вихідного геш-значення;
  - q – кількість каналів паралельного обчислення, причому s кратне q;
  - $base$ – початковий стан генератора;
  - h_0 – початкове значення проміжного стану;
- Додати до повідомлення M розмір самого повідомлення;
- Доповнити повідомлення до розміру кратного s;
- Розбити повідомлення на блоки розміром s;
- Обчислити l значень r_i:

$$r_i = (i - \text{rand}(m_i, m_{i+1})) \bmod l;$$

- Для кожного з 1, 2, …, q паралельних потоків t для кожного із 1, 2, …, l блоків i виконати такі обчислення:

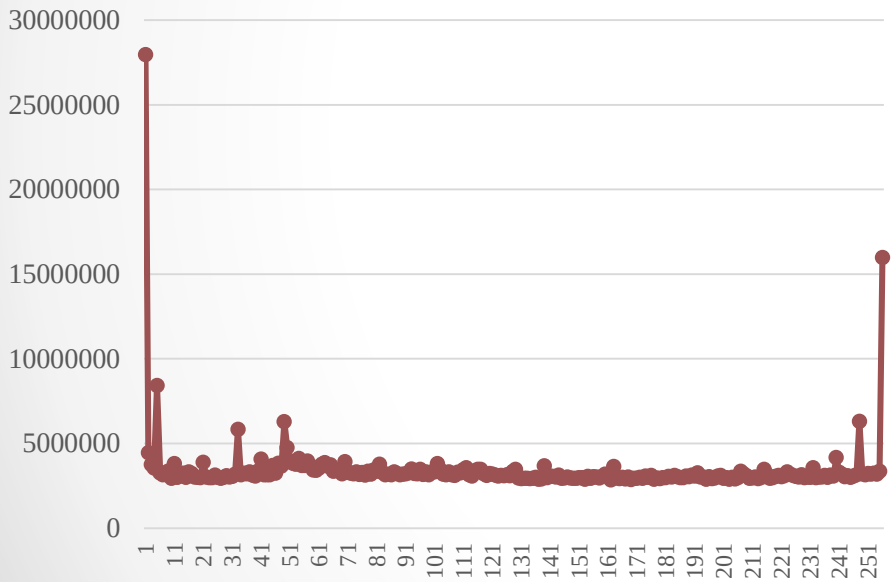
$$h_i^t = f(h_{i-1}^t, m_i^t, m_{r_i}^t)$$

- $H = h^1 | h^2 | \ldots | h^t$, де ‘|’ – операція конкатенації значень;
- Результуючим геш значенням є значення H.

Забілювання

- Розподіл різних байтів вхідних файлів

До забілювання



Після забілювання

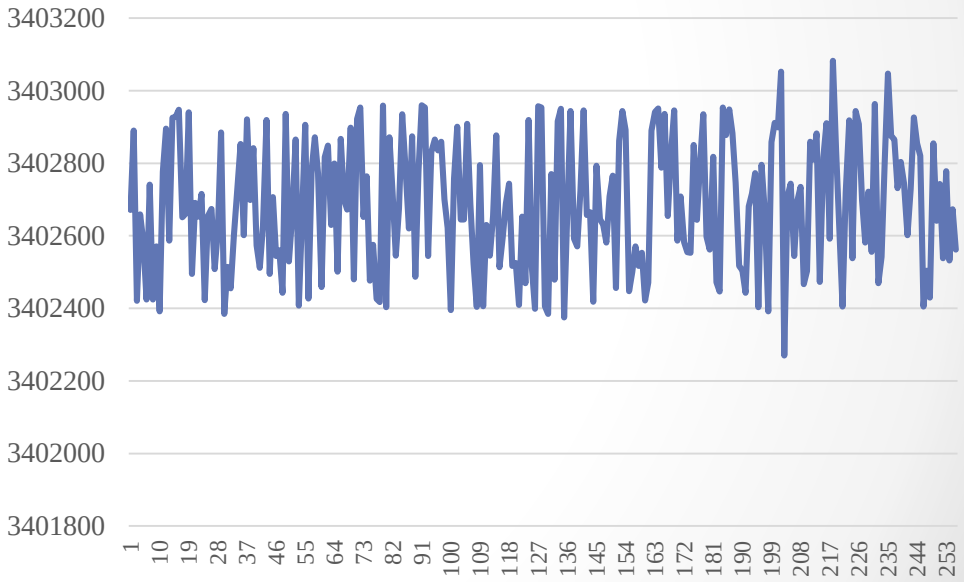
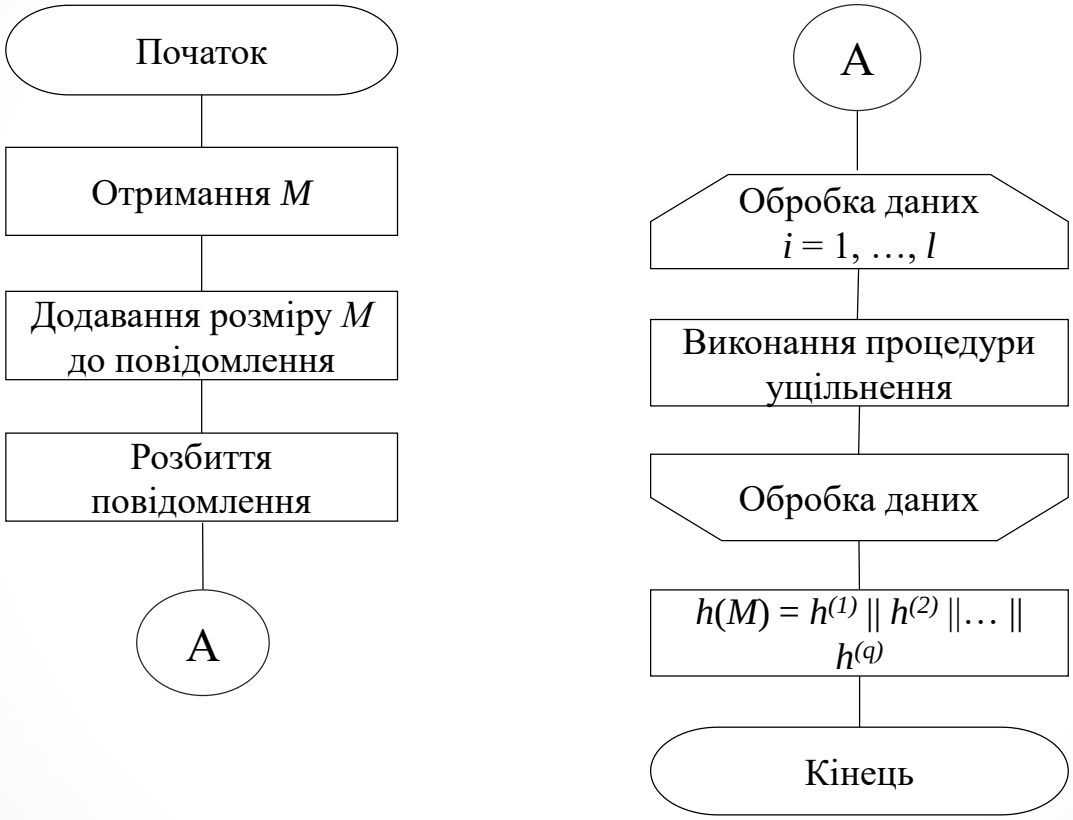


Схема алгоритму гешування



Функції ущільнення

- на основі піднесення до степеня за модулем простого числа (дискретне логарифмування):

$$f(a,b,c) = g^{a+b+c} mod\ p.$$

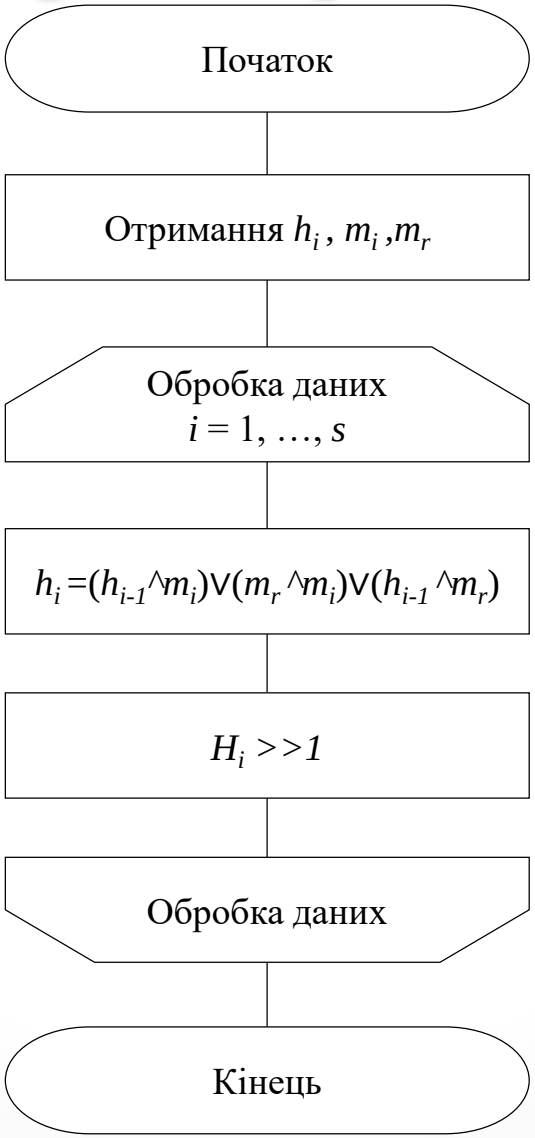
- на основі нелінійних логічних функцій (SHA2):

$$f(a,b,c) = a \wedge b \oplus a \wedge c \oplus b \wedge c;$$

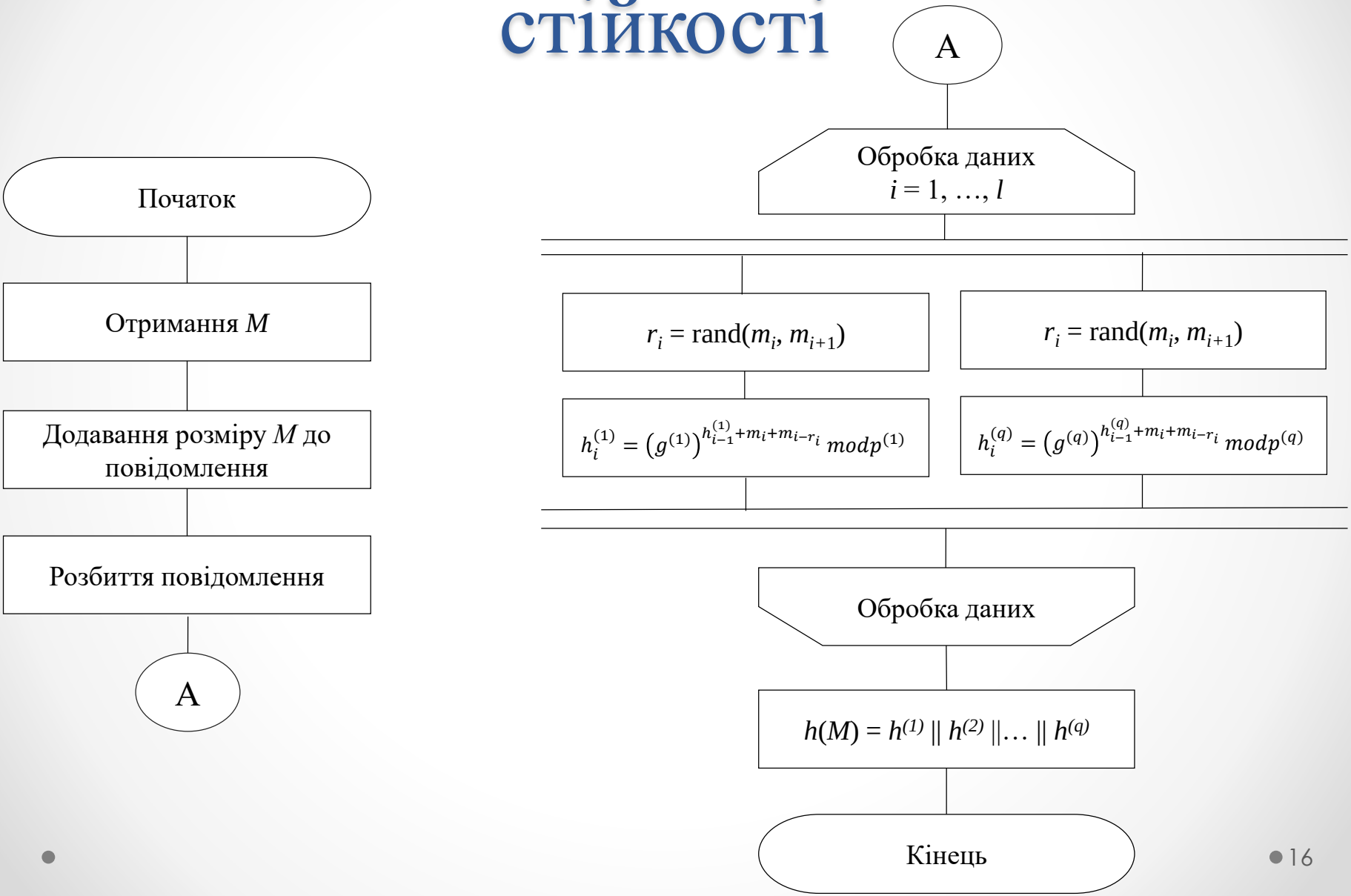
$$f(a,b,c) = a \wedge b \oplus \bar{a} \wedge c.$$

- на основі еліптичних кривих

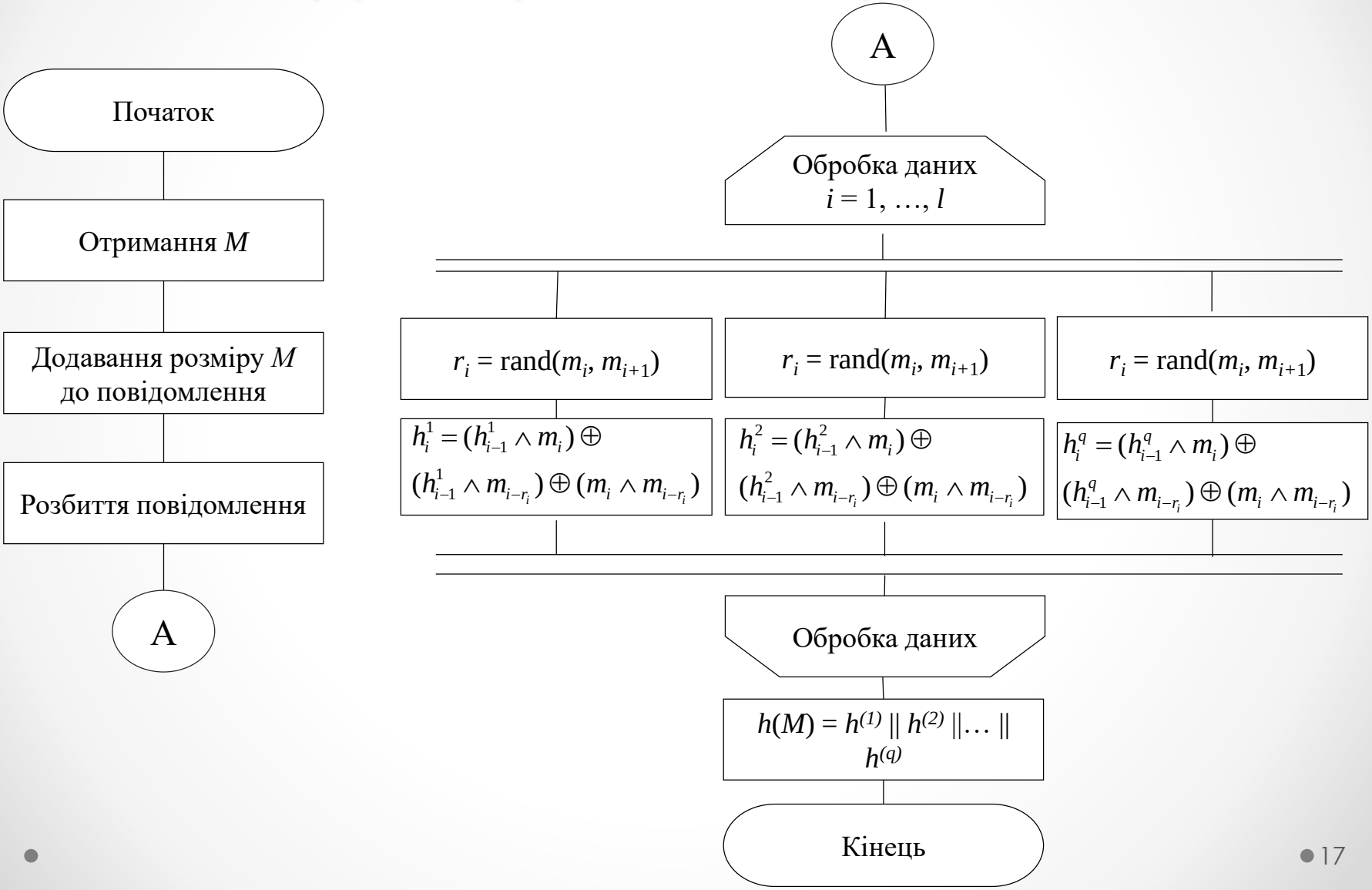
Алгоритм раундового перетворення



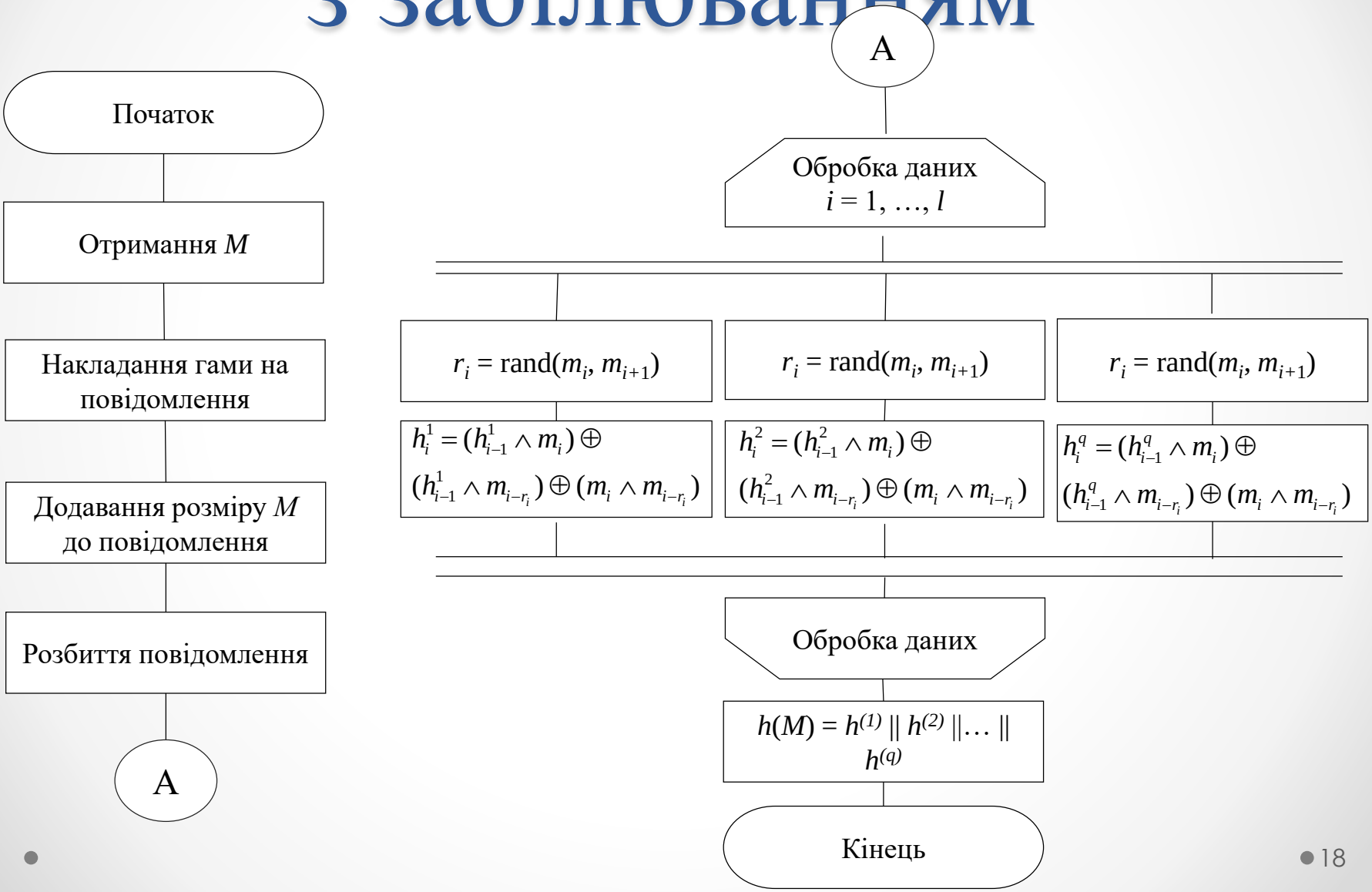
Алгоритм гешування підвищеної стійкості



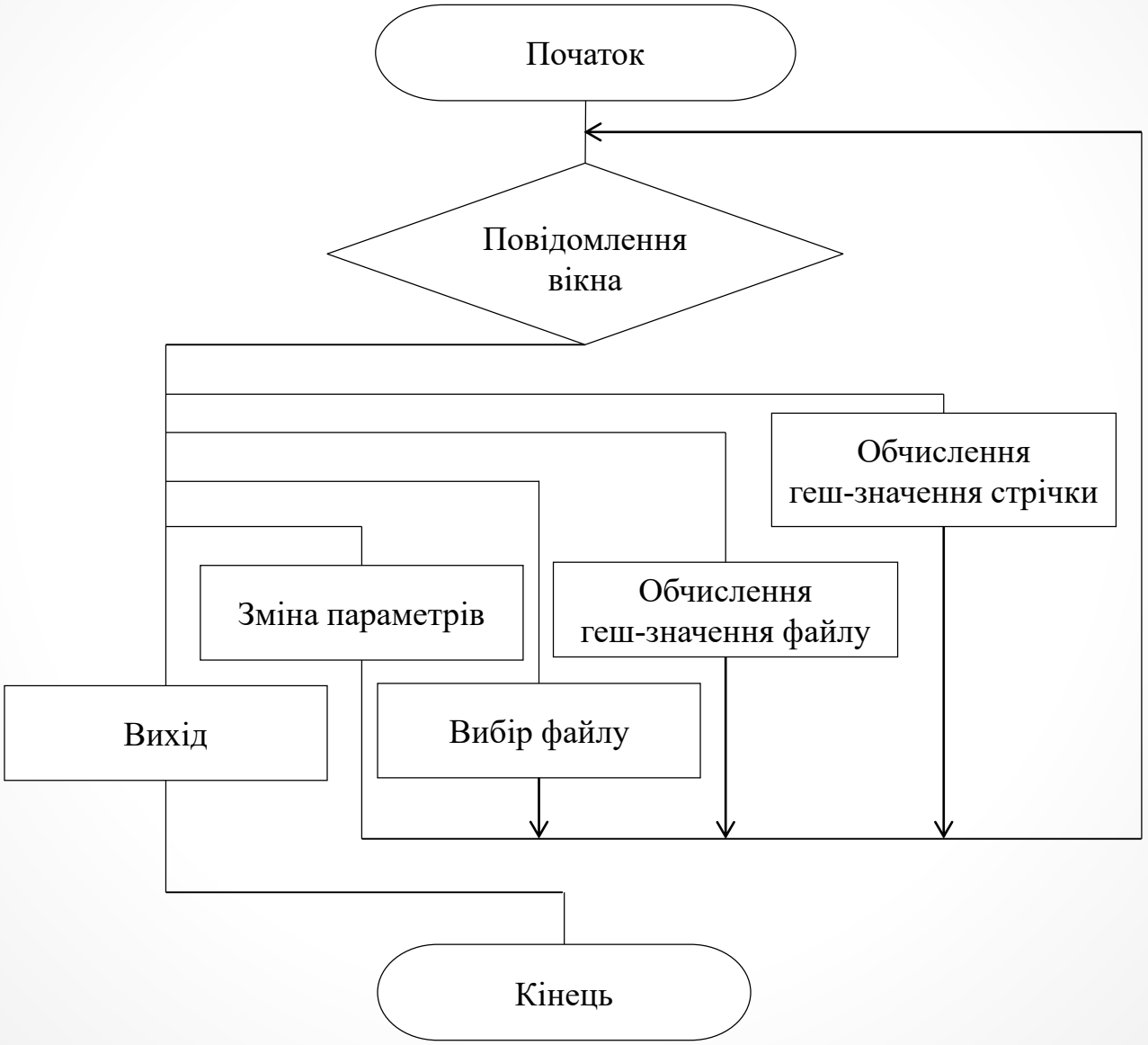
Алгоритм гешування підвищеної швидкості



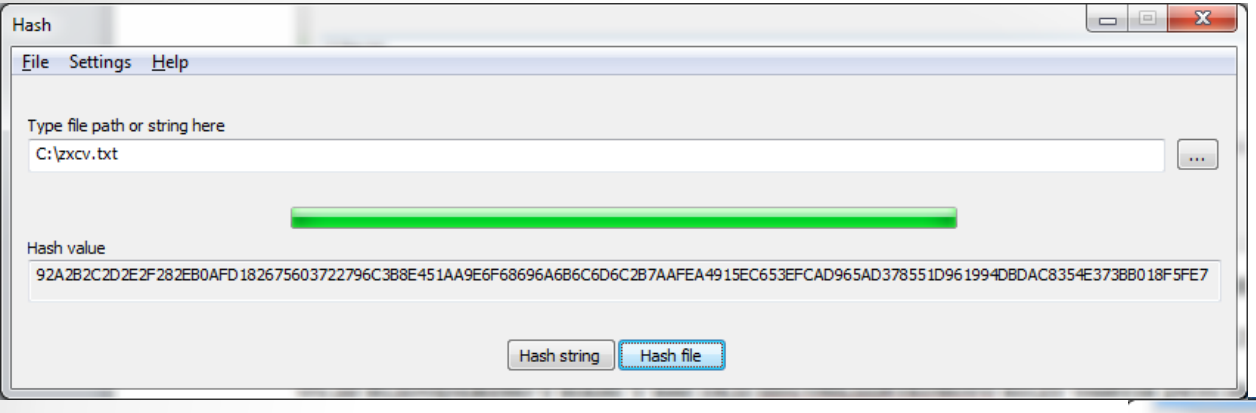
Алгоритм гешування з забілюванням



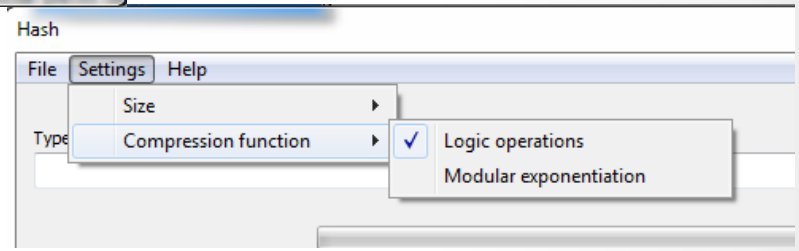
Алгоритм роботи програми



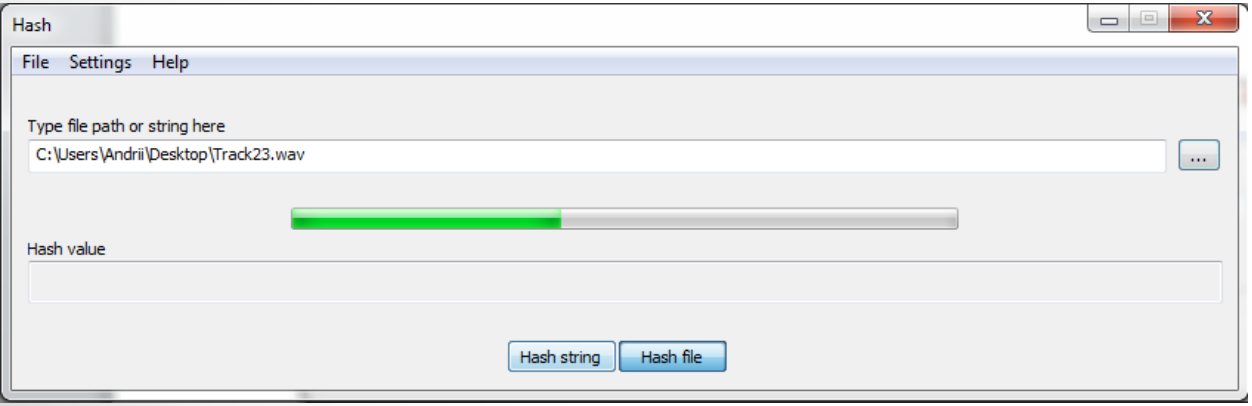
Інтерфейс програмного засобу



Вигляд вікна після обчислення геш-значення файлу



Вигляд вікна з відкритим меню



Вигляд вікна при зміні стану зчитування файлу

Огляд результатів

- ″B″ (01000010):
  7D33E9810179952F4D433131631F457D275B09097B775D553F336161534F352D170
  B79792B270D056F635151033F651D477B29291B177D7119D32723D38D374D

- ″C″ (01000011):
  4939E5EF513BC1495579254B335D012F0F311D036B1579674769557B234D315F3F
  210D331B0569177719456B537D214F2F513D230B35190321895339E38F337F

- ″″ (00000000)
  8736E6A0D634F626D0C91A9D8992D7D0AB49706C8A9A21AA83DBE20CB34C0F93
  21912225C944111879059676E9DCFF3109619C463B166FDA7653F4E48B487BB3

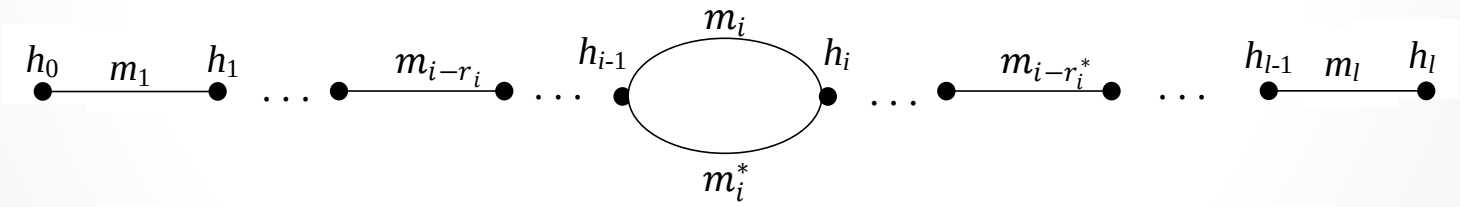
Приклад атаки

Зловмисник відповідно парадоксу «дня народження» за $2^{n/2}$ обчислень знаходить колізію для блоку даних m_i:

$$\begin{cases} h_i = f\left(h_i, m_i, m_{i-r_i}\right) = f\left(h_{i-1}, m_i^*, m_{i-r_i^*}\right); \\ r_i = rand(m_i); \\ r_i^* = rand(m_i^*), \end{cases}$$

де $r_i \in [1; l-1]$, $r_i \in N$.

Тоді граф процесу гешування набуває вигляду:



Така атака здійсненна лише коли $\forall j \in N$, $j \in [1; l]$, $j - r_i \neq i$, що малоймовірно у випадку, коли вихідні значення функції $rand(\cdot)$ підкорюються рівномірному закону розподілу і $l > 2$

Результати експериментальних досліджень

$r_i = \text{rand}(m_i)$

	Експеримент №1	Експеримент №2	Експеримент №3	Експеримент №4	Експеримент №5
Блоки не були зав'язані	361/1000	385/1000	367/1000	367/1000	364/1000
Блоки були зав'язані на 1 ітерації	381/1000	354/1000	478/1000	363/1000	381/1000
Блоки були зав'язані на 2 ітераціях	175/1000	172/1000	171/1000	197/1000	173/1000
Блоки були зав'язані на 3 і більше ітераціях	83/1000	89/1000	84/1000	73/1000	82/1000

Результати експериментальних досліджень

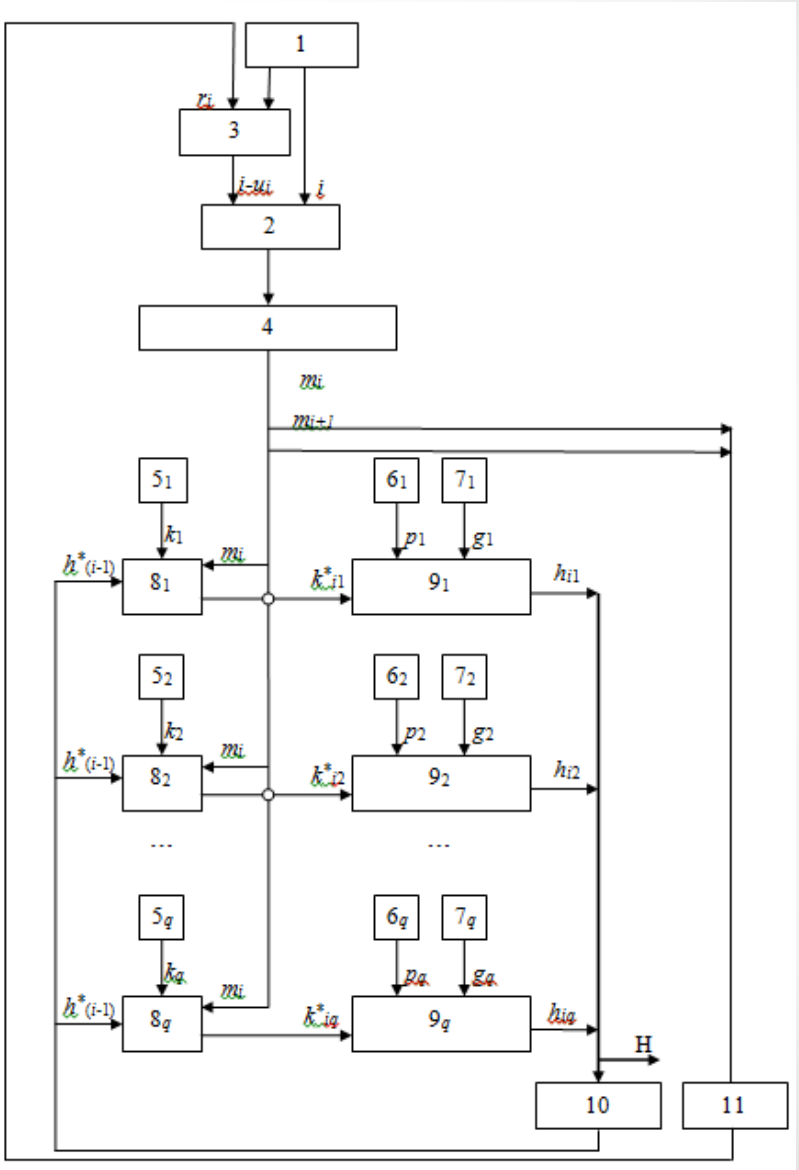
$r_i = \text{rand}(m_i, m_{i+1})$

	Експеримент №1	Експеримент №2	Експеримент №3	Експеримент №4	Експеримент №5
Блоки не були зав'язані	-	-	-	-	-
Блоки були зав'язані на 1 ітерації	1000/1000	1000/1000	1000/1000	1000/1000	1000/1000
Блоки були зав'язані на 2 ітераціях	375/1000	368/1000	371/1000	389/1000	380/1000
Блоки були зав'язані на 3 і більше ітераціях	258/1000	264/1000	255/1000	262/1000	261/1000

Схема пристрою для реалізації способу

Патент України на корисну модель №102987
МПК G09C 1/00

На рисунку показано:
1 – лічильник;
2 – комутатор;
3 – блок додавання;
4 – запам’ятовуючий пристрій, з якого надсилається і-й елемент послідовності.
5j – блок зберігання j-ї частини ключа.
6j – блок зберігання j-го значення модуля.
7j – блок зберігання j-го примітивного елемента.
8j – j-й пристрій додавання.
9j – j-й пристрій піднесення до степеня за модулем.
10 – пристрій додавання.
11 – блок генерації псевдовипадкових чисел.



Економічне обґрунтування витрат

- Загальні витрати на нову розробку 31 304,88 (грн.)
- Розрахунковий чистий прибуток за рік 34 429,16 (грн.)
- Експлуатаційні витрати 276,45 (грн./рік)
- Економічний ефект для споживача 200 (грн./рік.)
- Термін окупності витрат на розробку 0,91 (року)

Наукова новизна одержаних результатів

- Удосконалено методи гешування із зав’язуванням даних, які відрізняються від існуючих тим, що до зволяють підвищити стійкість гешування до загальних атак за рахунок використання попередньої перестановки блоків даних, що дозволяє ускладнити реалізацію атак Жу та Нострадамуса.
- Удосконалено метод гешування даних з використанням декількох блоків даних на кожній ітерації, що, на відміну від існуючих, забезпечує покращення зав’язування блоків даних на кожній ітерації від двох разів.
- Отримали подальший розвиток методи гешування із використанням забілювання даних, які на відміну від відомих передбачають використання забілювання перед зав’язуванням блоків даних, що дозволяє ускладнити попередню підготовку до реалізації загальних атак з використанням мультиколізій.

Практична значимість

- проведені дослідження та отримані наукові результати є основою для підвищення стійкості методів та засобів гешування до загальних атак. На базі цих методів розроблено: програмний засіб для обчислення геш значень тексту чи файлів. Отримано статистичні оцінки ступеня зав’язування блоків даних залежно від кількості блоків даних, які дозволяють виконати вибір необхідного способу зав’язування до конкретної задачі відповідно до критерію швидкість/стійкість

Дякую за увагу