УДК 004.056.2

О. Я. Матов[1], В. С. Василенко[2]
[1]Інститут проблем реєстрації інформації НАН України
вул. М. Шпака, 2, 03113 Київ, Україна
[2]Національний авіаційний університет
вул. Космонавта Комарова, 1, 03058 Київ, Україна

Хеш-функції та цілісність інформаційних об'єктів

Розглянуто можливості застосування відомого механізму хешування для контролю цілісності інформаційних об'єктів.

Ключові слова: *інформаційні об'єкти, хешування, цілісність.*

Вступ

Високі надійність, ефективність і технологічність телекомунікаційних мереж (ТКМ) є можливими тільки за умови надійного та ефективного захисту конфіденційності, цілісності та доступності інформаційних об'єктів телекомунікаційних мереж. Залежно від умов застосування, складності та класу ТКМ, а також характеристик можливих загроз, вага цих функціональних властивостей може змінюватись, але проблеми забезпечення цілісності інформації є одними з основних при розробці та впровадженні будь-яких захищених ТКМ. З цією метою або виключається доступ до первинної, відкритої інформації (задачі управління доступом) на усіх етапах її циркуляції у ТКМ, або використовуються обчислені та включені до складу інформаційних об'єктів хеш-функції, так звані ознаки цілісності.

Контроль, а можливо і поновлення цілісності інформації в умовах впливу як природних, так і штучних (найчастіше цілеспрямованих) спотворень забезпечується, найчастіше, шляхом порівняння початкових ознак цілісності (хеш-функцій), що відповідають цим об'єктам, і ознак цілісності, сформованих для інформаційних об'єктів, цілісність яких контролюється наразі (вторинних ознак цілісності). При цьому вторинні ознаки цілісності інформаційних об'єктів, цілісність яких контролюється, обчислюються за тими ж механізмами, що і початкові. У разі їхнього збігу робиться висновок про відсутність порушень цілісності [1−4].

Постановка задачі

Нехай передаванню чи збереженню підлягає деякий інформаційний об'єкт A, інформаційна довжина якого не перевищує m розрядів. Необхідно забезпечити його цілісність чи, хоча б контроль цієї цілісності. Для забезпечення контролю та поновлення цілісності інформаційних об'єктів в умовах тих чи інших руйнуючих

впливів до складу інформації, яка захищається, включають надмірну інформацію — ознаку цілісності або контрольну ознаку (залежно від прийнятої в задачах контролю цілісності або завадостійкого кодування термінології) — своєрідний образ, відображення цієї інформації, процедура формування якого відома, і який із дуже високою вірогідністю відповідає інформації, що захищається [1]. Обчислення такої надмірної інформації — ознаки цілісності (хеш-функції) — $h(A)$ цього об'єкта здійснюється із використанням механізмів хешування. Інформаційний об'єкт та його ознака цілісності зберігаються чи передаються єдиним блоком у вигляді конкатенації $A \| h(A)$ (у цьому виразі позначка $\|$ означає конкатенацію A та $h_i(A)$).

Потенційний порушник, як правило, приховує порушення цілісності (модифікації) таких інформаційних об'єктів. Для цього він має або формує нову ознаку цілісності (хеш-функцію), яка відповідає модифікованій інформації, або здійснює таку модифікацію, при якій ознака цілісності збігається з початковою (використати колізії хеш-функцій).

Для запобігання підробки порушниками вторинних ознак цілісності необхідно або виключити доступ неавторизованих користувачів до інформації (задачі управління доступом, розгляд яких виходить за межі роботи), або авторизованим користувачам сформувати із використанням секретних ключів хеш-функції, що мають мінімальну кількість (а краще — зовсім не мають) колізій. Для порушення цілісності порушник може використати колізії відповідних хеш-функцій чи визначити ключі перетворення. Розгляд способів протистояння використанню колізій і є задачею цієї роботи.

Вирішення задачі

Нагадаємо [1−5], що для розрахунку хеш-функцій застосовується операція визначення лишку від ділення вихідного інформаційного об'єкта A (що розглядається як деяке число) на ключ p — застосування операцій визначення лишку числа A за модулем (ключем) p:

$$h(A) = A \bmod p = A - \lfloor A / p \rfloor \cdot p = A - lp \,, \qquad (1)$$

де позначка $l = \lfloor A / p \rfloor$ (див. рис. 1,*а* та 1,*б*) означає обчислення цілої частини від розподілу A на p.

Колізією хеш-функцій $h(x)$ та $h(y)$ називається ситуація, коли два різних вхідних блоки даних x та y мають однакові хеш-функції, тобто $h(x) = h(y)$:

$$h(x) = x \bmod p = x - \lfloor x / p \rfloor \cdot p \,;\; h(y) = y \bmod p = y - \lfloor y / p \rfloor \cdot p \,.$$

Наприклад, при $x = p + h(A)$, $y = 2p + h(A)$, … (рис. 1,*в*, 1,*г*). Неважко зрозуміти, що колізії хеш-функцій вигляду (1) існують, коли вхідні блоки x та y пов'язані співвідношенням

$$y = x \pm k \cdot p \,, \qquad (2)$$

де $k = 0, 1, 2, \ldots$, оскільки у цьому випадку:

$$h(y) = y \bmod p = (x \pm k \cdot p) \bmod p = x \bmod p \pm k \cdot p \bmod p = x \bmod p \pm 0 = x \bmod p = h(x).$$

Тоді, виходячи із викладеного, слід зробити висновок, що в разі, коли будь-який вхідний блок даних A характеризується хеш-функцією $h(x)$, то числове значення цього блоку може бути визначеним як:

$$A = x \bmod p + \left[A / p \right] \cdot p = h(x) + l \cdot p .$$

У цьому виразі l — кількість колізій хеш-функцій блоку даних A, при обраному значенні модуля — p. Якщо ж такий блок даних представляється в деякому діапазоні $0 \le A \le R = 2^m$, а колізії хеш-функцій для такого блоку пов'язані співвідношенням (2), то максимальна кількість колізій цього блоку даних (рис. 1,*д*) дорівнює $n_k = \left[R / p \right]$.

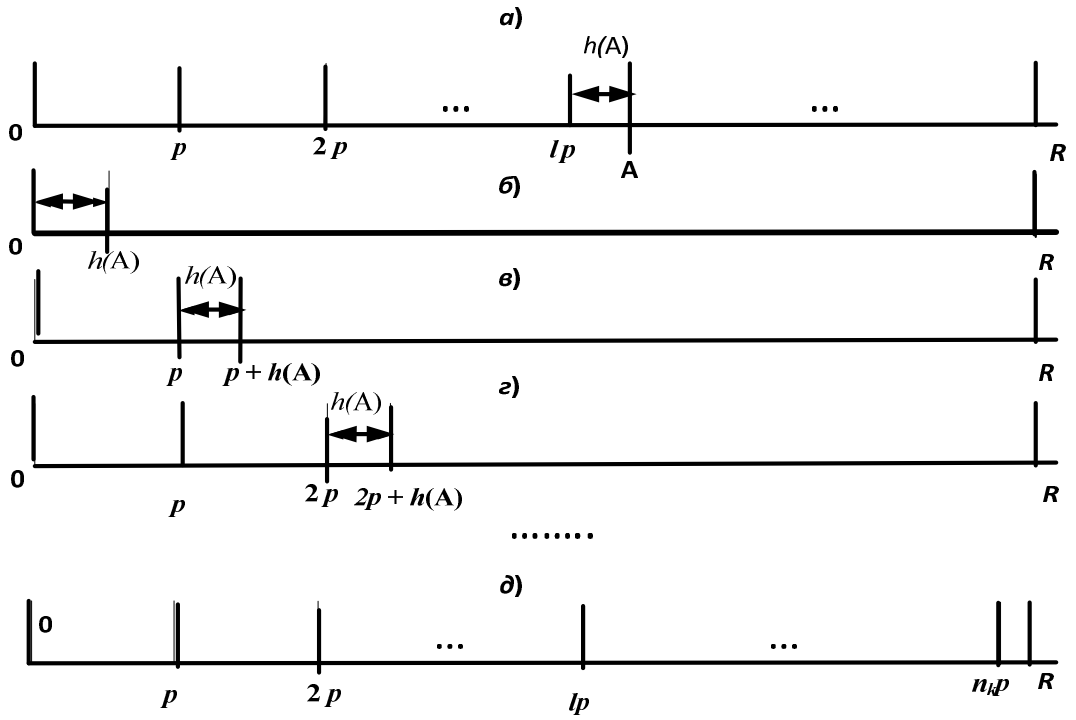

Рис. 1. Хеш-функції та їхні колізії

Із викладеного неважко зробити висновок, що досвідчений порушник у разі необхідності порушити цілісність інформаційного об'єкта x може здійснити таку його модифікацію, яка залишиться непоміченою. З цією метою йому слід скористатися перетворенням $x \to y$ з використанням виразу (2). Кількість таких модифікацій уже визначена та дорівнює кількості можливих колізій n_k. Отже для протистояння можливим порушенням цілісності інформаційних об'єктів кількість колізій слід зменшувати.

Зрозуміло, що для цього слід зменшувати величину $n_k = [R/p]$, мінімальне значення якої $n_k = 0$ досягається при $p > R$. Дійсно, у цьому випадку співвідношення $[R/p] < 1$, а отже, $n_k = 0$. У цьому випадку, тобто при $p > R$ хеш-функція вигляду (1) набуває значення

$$h(A) = A \bmod p = A - [A/p] \cdot p = A - 0 \cdot P = A;$$

результуючий об'єкт для зберігання чи передавання набуває вигляду:

$$A_{рез} = A \| A,$$

тобто здійснюється дублювання первинного інформаційного блоку.

Таке дублювання іноді може мати сенс, але вкрай обмежено, по-перше, через велику надлишковість, що часто є неприпустимим. По-друге, такий спосіб при такому значення лише одного ключа перетворення ніякої конфіденційності навіть самої хеш-функції не забезпечує. Це свідчить про те, що за цих умов обчислення хеш-функції, як в задачах контролю цілісності, так і, особливо, в задачах криптографічних перетворень, значною мірою втрачає сенс.

Таким чином, задача має протиріччя, яке полягає в тому, що для зменшення кількості колізій величину ключа (модуля, за яким обчислюється хеш-функція) слід обирати якомога більшою (аж до $p > R$), а отже, ключ слід мати один, а для забезпечення властивостей із криптографічної стійкості результатів хешування слід мати множину ключів.

Розв'язання цього протиріччя забезпечується шляхом утворення єдиного ключа перетворення у вигляді добутку *n* його складових [5]:

$$P = \prod_{i=1}^{n} p_i > R.$$

Такий підхід є еквівалентним застосуванню до інформаційного об'єкта *A* декількох (*n*) перетворень (1). Але, на відміну від (1), відповідна сукупність ключів розглядається як один складний модуль (основа), який дорівнює добутку із набору *n* часткових основ. Унаслідок цього одержується кількість лишків $h_i(A)$, яка є відповідною кількості основ p_i, та зменшеною, відповідно, кількістю колізій: при $A < R$ та $P > R$ кількість колізій дорівнює нулю.

Тоді результуючий інформаційний блок може буди записаним у вигляді:

$$A_{рез} = A \| h_1 \| h_2 ... \| h_n. \qquad (3)$$

При цьому згаданого дублювання первинного інформаційного блоку не відбувається, а, оскільки в цьому випадку $P > R$, то кількість колізій дорівнює нулю: $n_k = 0$. Окрім того, слід звернути увагу на те, що контроль цілісності здійснюєть-

ся за модулем (основою) значної величини. Як відомо, ймовірність пропуску спотворення у цьому випадку $p_{пр} \approx 1/P$ і є вкрай незначною. Отже, *такий підхід забезпечує надійний контроль цілісності інформаційних об'єктів.*

Звернемо увагу, що сукупність основ p_i та відповідних хеш-функцій (лишків) h_i (i = 1, 2, …, n) утворює певну систему числення із діапазоном представлення $P = \prod_{i=1}^{n} p_i > R$. Ця система числення має назву системи лишкових класів, а сукупність лишків $h_1 \| h_2 ... \| h_n$ є представленням сукупного лишку $h(A)$ у діапазоні P цієї системи числення. Тобто вираз (3) може бути наданим у вигляді

$$A_{рез} = A \| A_{СЛК} ,$$

де $A_{СЛК} = h_1 \| h_2 ... \| h_n$ є представленням вихідного числа A в системі лишкових класів.

Тоді, в разі, коли здійснене порушення зосереджене лише в межах первинного інформаційного блоку A, його цілісність може бути поновленою шляхом відомого алгоритму переведення надлишкової інформації $A_{СЛК}$ в початкову систему числення

$$A = \sum_{i=1}^{n} h_i B_i \,(\mathrm{mod}\, P) ,$$

де B_i — так звані ортогональні базиси введеної системи лишкових класів.

Звернемо, однак, увагу на досить обмежену можливість такого поновлення цілісності, оскільки ймовірність того, що досвідчений порушник обмежиться лише модифікацією первинного інформаційного блоку A чи не модифікує й надлишкової інформації, є незначною. Отже можливість поновлення цілісності, як і в будь-якому варіанті дублювання (використання однієї резервної копії) інформаційних об'єктів є примарною. Розгляд механізмів надійного поновлення цілісності виходить за межі цієї роботи.

Висновок

Запропонованій підхід за рахунок відсутності колізій і контрольного модуля значної величини забезпечує виконання вимог щодо надійного контролю цілісності інформаційних об'єктів, але не забезпечує можливість її поновлення.

1. *Матов О.Я.* Целостность и доступность информационных объектов (монографія) / В.С. Василенко, А.Я. Матов. — Saar-Brucken, Deutschland: LAMBERT Academic Publishing, 2013. — 95 с. — ISBN 978-3-659-47333-3.

2. *Василенко В.С.* Код условных вычетов (монографія) / В.С. Василенко. — Saarbrucken, Deutschland: LAMBERT Academic Publishing, 2013. — 129 с. — ISBN 978-3-659-48203-8.

3. *Корченко О.Г.* Сучасні квантові технології захисту інформації / О.Г. Корченко, Є.В. Васіліу, С.О. Гнатюк // Захист інформації. — 2010. — № 1. — С. 77–89.

4. *Українець І.В.* Аналіз і дослідження криптографічних засобів захисту інформації на базі «Укргазбанк» / І.В. Українець. — Рівне: МЕГУ, 2011. — 150 с.

5. *Василенко В.С.* Хеш-функції та криптографічні перетворення / В.С. Василенко // Матеріали X Міжнар. наук.-практ. конф. «Праці академічних наук − 2014» 30 серпня − 7 вересня 2014 р.». — Йоркшир, Англія: Science and Education LTD, 2014. — С. 26–29. — ISBN 978-966-8736-05-6.

Надійшла до редакції 19.11.2014