ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 23 МАТЕМАТИКИ Выпуск 4
2016

Ф. А. Дали, **В. О. Миронкин** (Москва, ТК 26, МИЭМ НИУ ВШЭ).
О вероятностных характеристиках одного класса моделей древовидного хэширования.

Введение

В настоящее время активно развиваются подходы к построению новых, более сложных с аналитической точки зрения и реализации процедур хеширования сообщений большой длины за счет распараллеливания и применения моделей итерационных преобразований (например, как в Skein [1], MD6 [2] и Blake2 [3]). Данная работа посвящена изучению одной из таких моделей [4, 5] и оценки ряда ее вероятностных характеристик.

Авторами получены точные выражения для некоторых характеристик, связанных с коллизиями в различных предположениях о структуре хэшируемого сообщения.

1. Теоретико-вероятностная модель древовидного хэширования

Пусть задано произвольное сообщение $X \in V_M$, $M \in \mathbf{N}$.

$H : V_M \to V_N$-хэш-функция, $N \in \mathbf{N}$, $N < M$.

Для вычисления хэш-образа $H(X)$ исходное сообщение X разбивается последовательно на блоки длины T, при этом формируется первый слой дерева хэширования:

$$X = X_1^{(1)} X_2^{(1)} \dots X_{M/T}^{(1)}.$$

Блоки нумеруются слева направо, при этом нижние индексы обозначают номер блока сообщения, а верхние номер слоя дерева хэширования. Соответствующие блоки будем называть узлами.

О п р е д е л е н и е. *Арностью* узла дерева хэширования назовем число узлов предыдущего слоя дерева, используемых при его формировании.

З а м е ч а н и е. *Арность* узлов определяется для второго и последующих слоев. Для узлов первого слоя арность не определяется.

Пусть задана хэш-функция $h : V_T \to V_N$, $N < T$, на основе которой формируются все слои дерева хэширования.

З а м е ч а н и е. В общем случае при формировании узлов очередного слоя могут использоваться различные хэш-функции, меняя тем самым значение арности. В настоящей работе такие случаи рассматриваться не будут.

Формирование узлов очередного слоя осуществляется за счет последовательной конкатенации η образов узлов предыдущего слоя при отображении h:

$$X_j^{(i)} = h\big(X_{(j-1)\eta+1}^{(i-1)}\big) \| \dots \| h\big(X_{jn}^{(i-1)}\big).$$

где $\eta = \frac{T}{N}$ — арность узлов дерева хэширования.

З а м е ч а н и е. Для корректного формирования очередного слоя дерева хэшировантия необходимо, чтобы число узлов в каждом слое дерева хэширования было кратно значению арности. Однако для сообщения произвольной длины в общем случае это не выполняется.

На рис. 1 изображен переход от $(i-1)$-го слоя дерева к i-му при корректном построении бинарного дерева хэширования. Он соответствует случаю, когда отображение h сжимает преобразуемый вектор в два раза (т. е. $T/N = 2$). На рис. 2 представлен общий случай, когда арность узлов дерева хэширования $\eta = T/N > 2$.

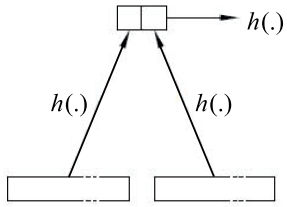

Рис. 1

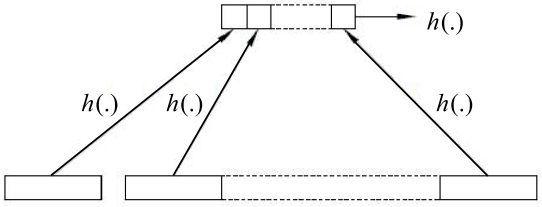

Рис. 2

Результатом корректного последовательного формирования слоев дерева хэширования является слой, состоящий из одного узла. Применение к нему функции h завершает процесс выработки хэш-образа $H(X)$.

С учетом последнего замечания при использовании одной хэш-функции h выделяют ряд методов построения деревьев хэширования по принципу дополнения слоев:

1. *Дополнение исходного сообщения нулевыми битами до степени арности узлов.*

При данном подходе других дополнений при построении дерева хэширования не требуется, что является преимуществом метода. Однако имеет место и существенный недостаток, заключающийся в том, что для ряда сообщений нулевое дополнение может оказаться длиннее самого сообщения, что значительно увеличит трудоемкость выработки хэш-образа $H(X)$. Например, при длине сообщения $\eta^l + 1$ потребуется дополнение длиной $(\eta - 1)\eta^l - 1$, $l \in \mathbf{N}$.

2. *Последовательное дополнение слоев дерева хэширования нулевыми битами до длины кратной арности узлов.*

З а м е ч а н и е. Если при дополнении нулями исходного сообщения его длина становится равной η^l, т о приходим к первому случаю.

Далее будем считать, что $M = 2^l$. Аналитические результаты, представленные в работе, получены в предположении случайного выбора отображения h, заданного на множестве $\Omega_{T,N}$ всех равномерно распределенных отображений $\{V_T \to V_N\}$.

2. Основные понятия и определения

О п р е д е л е н и е. Коллизией в дереве хэширования, построенного на основе отображения $h \in \Omega_{T,N}$, будем называть совпадение значений пары произвольных узлов дерева хэширования.

О п р е д е л е н и е. Пусть Λ — подмножество множества узлов дерева хэширования. Коллизией в подграфе дерева хэширования, построенного на основе отображения $h \in \Omega_{T,N}$, будем называть совпадение значений пары произвольных узлов из множества Λ.

О п р е д е л е н и е. Коллизией в $k \in \overline{2, \log_2 \frac{M}{T}}$ слое дерева хэширования, построенного на основе отображения $h \in \Omega_{T,N}$, будем называть совпадение значений пары произвольных узлов k слоя дерева хэширования.

Через $\xi_{T,N}^{(M)}$ обозначим случайную величину, равную номеру слоя дерева хэширования, в котором впервые возникает коллизия.

Через $\eta_{T,N}^{(M)}$ обозначим случайную величину, равную номеру слоя k дерева хэширования, при формировании которого впервые возникает коллизия в подграфе, состоящего из первых k слоев дерева.

3. Коллизии в заданном слое бинарного дерева хэширования

Будем различать два предположения о структуре хэшируемого сообщения:

1. Блоки сообщения $X \in V_M$ попарно различны.

2. Блоки сообщения $X \in V_M$ с нечетными номерами произвольны и равномерно распределены на множестве блоков длины T, а блоки с четными номерами попарно различны.

Предложение 1. *Пусть $h \in \Omega_{T,N}$. Тогда для случайного сообщения $X \in V_M$ с попарно различными блоками вероятность появления коллизии во втором слое бинарного дерева хэширования равна*

$$\mathbf{P}\{\eta_{T,N}^{(M)} = 2\} = 1 - \sum_{l=1}^{M/2T} \sum_{s=1}^{l} \sum_{\substack{z_1+\cdots+z_l=M/(2T):\\ z_{j_1}=\cdots=z_{j_s}\\ z_i \geqslant 1, j_u \in \overline{1,l}}} \frac{(M/(2T)!}{s! z_1! \dots z_l!} \prod_{j=1}^{l} \Big(1 - \frac{j-1}{2^{T/2}}\Big) \prod_{i=1}^{z_j - 1} \frac{1}{2^{T/2}} \Big(1 - \frac{i}{2^{T/2}}\Big).$$

Предложение 2. *Пусть $h \in \Omega_{T,N}$. Тогда для случайного сообщения $X \in V_M$ с равномерно распределенными нечетными блоками и попарно различными четными блоками вероятность появления коллизии во втором слое бинарного дерева хэширования равна*

$$\mathbf{P}\Big\{\xi_{T,N}^{(M)} = 2\Big\} = 1 - \sum_{l=1}^{M/2T} \sum_{t=1}^{l} \sum_{\substack{z_1+\cdots+z_l=M/(2T):\\ z_{u_1}=\cdots=z_{u_t}\\ z_i \geqslant 1, u_w \in \overline{1,l}}} \frac{(M/(2T)!}{t! z_1! \dots z_l!} \prod_{i=1}^{l} \prod_{j=1}^{z_i - 1} \left(\frac{1}{2^{T/2}} - \frac{j-1}{2^T}\right)$$

$$\times \sum_{r=1}^{l} \sum_{s=1}^{r} \sum_{\substack{y_1+\cdots+y_r=l\\ y_{v_1}=\cdots=y_{v_s}\\ y_i \geqslant 1, v_w \in \overline{1,r}}} \frac{1}{s!} \sum_{\substack{j_1^{(1)} < \cdots < j_{y_1}^{(1)}\\ j_i^{(1)} \in 1,l}} \cdots \sum_{\substack{j_1^{(r)} < \cdots < j_{y_r}^{(r)}\\ j_1^{(r)} \in 1,l \setminus \{j_1^{(u)}, \dots, j_{y_u}^{(u)}\}\\ u \in \{1,\dots,r-1\}}}$$

$$\times \prod_{n=1}^{r} \left(\frac{1}{2^{(y_n-1)\frac{T}{2}}} - \frac{j-1}{2^{y_n \frac{T}{2}}}\right) \prod_{i=1}^{z_{j_1^{(n)}} + \cdots + z_{j_n^{(n)}} - 1} \left(1 - \frac{i-1}{2^{\frac{T}{2}}}\right).$$

СПИСОК ЛИТЕРАТУРЫ

1. *Ferguson N., Bauhaus S. L., Schneier B., Whiting D., Bellare M., Kohno T., Callas J., Walker J.* The skein hash function family (version 1.2), 2009.
2. *Rivest R. L., Agre B., Bailey D. V., Crutch C., eld., Dodis Y., Elliott K., Khan F. A., Krishnamurthy J., Lin Y., Reyzin L., Shen E., Sukha J., Sutherland D., Tromer E., Yin Y. L.* The md6 hash function: A proposal to nist for sha-3, 2008.
3. *Aumasson J.-P., Neves S., Wilcox-O'Hearn Z., Winnerlein C.* Blake2: Simpler, smaller, fast as md5. In Proceedings of the 11th International Conference on Applied Cryptography and Network Security, ACNS'13, p. 119–135, Berlin, Heidelberg, 2013. Springer-Verlag.
4. National Institute of Standards and Technology. FIPS PUB 202: Secure Hash Standard (SHS). Technical report, aug 2015.
5. *Merkle R. C.* Secrecy, Authentication, and Public Key Systems. PhD thesis, Stanford, CA, USA, 1979.