ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 25 МАТЕМАТИКИ Выпуск 1
2018

В. О. Мironкин, **М. И. Урусов** (Москва, МИЭМ НИУ ВШЭ). **О коллизиях деревьев Меркла.**

В последнее время в научной литературе все чаще появляются публикации связанные с применением древовидных структур в различных криптографических приложениях, например, при построении алгоритмов хэширования [7, 9, 12], проверки цифровой подписи [8, 10] и передачи данных.

Наиболее ярким примером применения подобных моделей является широко используемая технология BlockChain [1, 5, 6] (рис. 1), на основе которой работают платежные системы Bitcoin и Litecoin. К BlockChain весьма активно проявляет интерес и банковский сектор, так, например, в 2016 году Bank of America и Microsoft заявили о начале разработки финансовой BlockChain-платформы. Как следствие становятся все более актуальными исследования, направленные на изучение и описание криптографических свойств древовидных моделей преобразования информации.

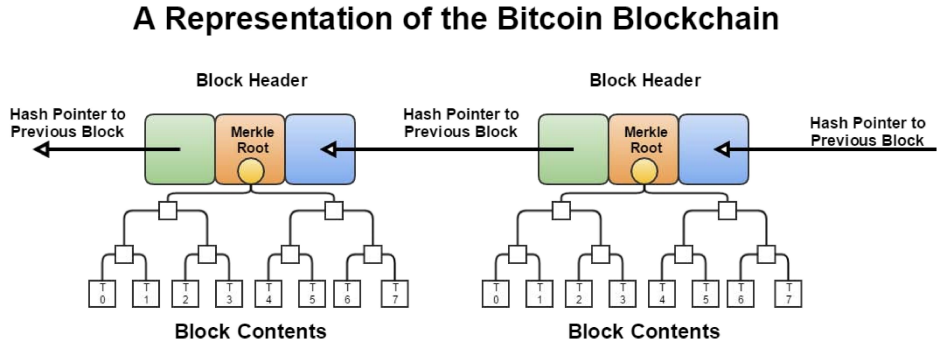

Рис. 1. Схема функционирования BlockChain

В настоящей работе рассматривается простейшая модель полного бинарного дерева, так называемого дерева Меркла [11], и исследуется вопрос, связанный с построением коллизий для соответствующей модели.

Используя терминологию [3, 4], введем ряд обозначений:

- M — преобразуемое сообщение, $M \in V^*$;
- $|M|$ — длина сообщения $M \in V^*$;
- m — размер узлов дерева Меркла в битах;
- $h : V_m \to V_t$ — внутренняя функция, преобразующая произвольный вектор $x \in V_m$ в вектор $h(x) \in V_t$, где $m > t$;
- $H : V^* \to V_t$ — функция сжатия, описывающая процесс преобразования сообщения $M \in V^*$ с использованием дерева Меркла в вектор $H(M) \in V_t$.

Рассмотрим дерево Меркла высоты $k \in \mathbb{N}$, в котором на основе пар узлов каждого слоя с помощью внутренней функции h (функции сжатия или хэш-функции) формируются соответствующие узлы следующего слоя (рис. 2). В этом случае дерево Меркла преобразует векторы из $V_{2^k m}$ в векторы из V_t.

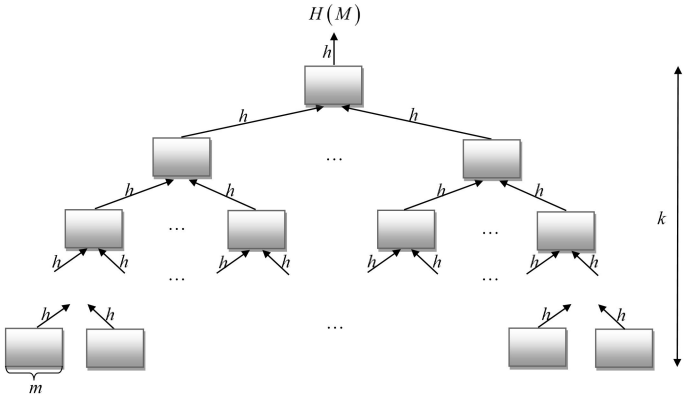

Рис. 2. Дерево Меркла высоты k

О п р е д е л е н и е. Коллизией дерева Меркла, реализующего функцию сжатия $H : V_{2^k m} \to V_t$, называется произвольная пара сообщений $M_1,\ M_2 \in V^* : M_1 \neq M_2$, для которых выполняется равенство $H(M_1) = H(M_2)$.

В рамках рассматриваемой модели построение внутренних коллизий [2] (поиск совпавших значений узлов дерева в процессе его формирования) может повлечь за собой построение коллизий соответствующего дерева в целом. Действительно, при совпадении значений α узлов дерева Меркла (рис. 3) может быть построена коллизия на основе поддеревьев с корнями в соответствующих узлах путем замены листьев меньшего из поддеревьев на соответствующий слой большего поддерева.

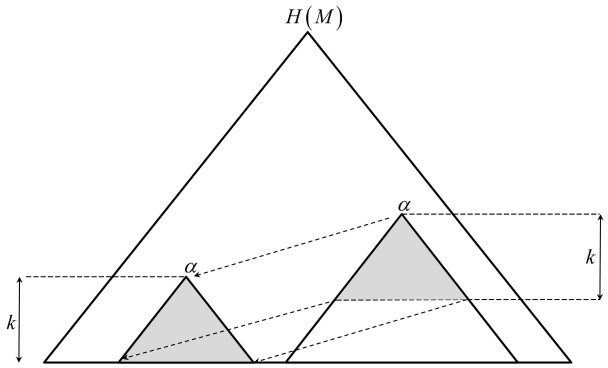

Рис. 3. Построение коллизии дерева Меркла с использованием внутренней коллизии

Следует заметить, что при совпадении значений узлов дерева Меркла коллизия формируется только в том случае, когда соответствующие поддеревья отличаются хотя бы в одном узле (отличие в узлах влечет за собой отличие в листьях, что и требуется для построения коллизии). В противном случае замена поддеревьев оставляет сообщение неизменным.

Через $P_{tree}^{(k,m)}$ обозначим вероятность появления коллизий дерева Меркла высоты k, узлы которого имеют размер m бит.

Предложение 1. *Пусть задано дерево Меркла высоты* k. *Тогда если* $k \geqslant m$, *то* $P_{tree}^{(k,m)} = 1$. *Если же* $k < m$, *то справедлива оценка*

$$P_{tree}^{(k,m)} < 1 - \prod_{i=1}^{k} \prod_{j=0}^{2^{k-i+1}-1} \left(1 - \frac{2^{k+1}\left(1-2^{-i}\right)+j}{2^m}\right).$$

В таблице представлены приближенные значения оценок вероятности $P_{tree}^{(k,m)}$, полученные с использованием предложения 1.

Таблица. Верхние оценки вероятности $P_{tree}^{(k,m)}$

$k\backslash m$	**8**	**10**	**12**	**16**
4	$7.57 \cdot 10^{-1}$	$2.89 \cdot 10^{-1}$	$8.11 \cdot 10^{-2}$	$5.26 \cdot 10^{-3}$
8	1	$1 - 2.41 \cdot 10^{-53}$	$1 - 1.27 \cdot 10^{-11}$	$7.76 \cdot 10^{-1}$
10	1	1	$1 - 2.33 \cdot 10^{-212}$	$1 - 2.91 \cdot 10^{-11}$
12	1	1	1	$1 - 4.05 \cdot 10^{-176}$
16	1	1	1	1

$k\backslash m$	**24**	**32**	**64**
4	$2.06 \cdot 10^{-5}$	$8.04 \cdot 10^{-8}$	$1.88 \cdot 10^{-17}$
8	$5.81 \cdot 10^{-3}$	$2.28 \cdot 10^{-5}$	$5.30 \cdot 10^{-15}$
10	$8.94 \cdot 10^{-2}$	$3.66 \cdot 10^{-4}$	$8.52 \cdot 10^{-14}$
12	$7.77 \cdot 10^{-1}$	$5.84 \cdot 10^{-3}$	$1.37 \cdot 10^{-12}$
16	$1\text{-}5.32 \cdot 10^{-168}$	$7.77 \cdot 10^{-1}$	$3.50 \cdot 10^{-10}$

Так, например, на рисунке 4 представлена зависимость соответствующей оценки вероятности $P_{tree}^{(k,m)}$ от величины m при $k = 8$ и $k = 10$.

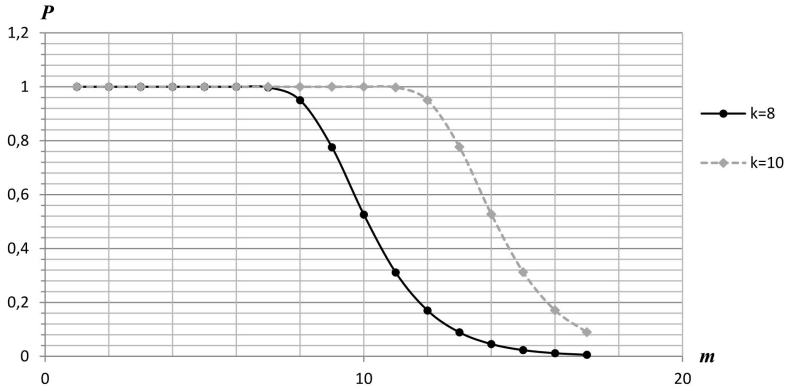

Рис. 4. Дерево Меркла высоты k

С учетом неравенства $1 - x \leqslant e^{-x}$, справедливого при $0 \leqslant x \leqslant 1$, имеет место следующий результат.

Следствие. *В условия предложения* 1 *при* $k < m$ *выполняется неравенство*

$$P_{tree}^{(k,m)} < 1 - \left(1 - \frac{2^{k+1}-1}{2^m}\right)^{2^k k}.$$

Если при этом $k < 2^{m-2k-1}$, *то*

$$P_{tree}^{(k,m)} < \frac{k}{2^{m-2k-1}}.$$

Таким образом, построение коллизий дерева Меркла, основанное на поиске внутренних коллизий, фактически сводится к решению классической задачи парадокса «дней рождений» и дает в среднем корневую оценку трудоемкости — $2^{\frac{m}{2}}$ операций вычисления внутренней функции h, что в свою очередь соответствует деревьям Меркла, построенным на сообщениях со средней длиной $\mathbf{E}|M| \geqslant 2^{\frac{m-2}{2}} m$ бит. В этом случае для используемых на практике значений m порядка 1024 бит $\mathbf{E}|M| \geqslant 2^{521}$ бит, что позволяет сделать вывод о неэффективности соответствующего метода построения коллизий деревьев Меркла.

СПИСОК ЛИТЕРАТУРЫ

1. *Брито Д.*, *Дурадо Э.* Криптовалюты. Mercatus Center, George Mason University. 2014.
2. *Дали Ф. А.*, *Миронкин В. О.* О вероятностных характеристиках одного класса моделей древовидного хэширования. — Обозрение прикл. и промышл. матем., 2016, т. 23, в. 4, с. 345–347.
3. *Дали Ф. А.*, *Миронкин В. О.* О некоторых моделях древовидного хэширования. — Обозрение прикл. и промышл. матем., 2017, т. 24, в. 4, с. 241–244.
4. *Дали Ф. А.*, *Миронкин В. О.* Обзор подходов к построению древовидных режимов работы некоторых хэш-функций. — Проблемы информационной безопасности. Компьютерные системы. СПб.: СПбПУ, 2017, № 2, с. 46–55.
5. *Решевский М.* Золотая лихорадка 21 век. — ComputerBild, 2011, № 17, с. 64–69.
6. *Ammous S.* The Bitcoin Standard: The Decentralized Alternative to Central Banking. New York: John Wiley and Sons Inc, 2018, p. 304.
7. *Bertoni G., Daemen J., Peeters M., Van Assche G.* Sufficient conditions for sound tree hashing modes, Symmetric Cryptography (Dagstuhl, Germany) (Handschuh H., Lucks S., Preneel B., Rogaway P., eds.), Dagstuhl Seminar Proceedings, № 09031, Schloss Dagstuhl Leibniz-Zentrum fuer Informatik, Germany, 2009.
8. *Buchmann J.* CMSS – an improved Merkle signature scheme. — International Conference on Cryptology in India. Springer Berlin Heidelberg, 2006, p. 349–363.
9. *Damgard I.* A design principle for hash function. — Advances in Cryptology – Crypto '89 (G. Brassard, ed.), LNCS, № 435, Springer-Verlag, 1989, p. 416–427.
10. *Garcia LC.-C.* On the security and the efficiency of Merkle signature scheme, Technical Report 2005/192. — Cryptology ePrint Archive, 2005.
11. *Merkle R. C.* Secrecy, authentication, and public key systems, PhD thesis, UMI Research Press, 1982.
12. *Sarkar P., Schellenberg P. J.* A parallelizable design principle for cryptographic hash functions, Cryptology ePrint Archive, Report 2002/031, 2002, http://eprint.iacr.org.