О криптографии с открытым ключом на основе задачи разложения языков

С.А. Афонин

Современные криптосистемы с открытым ключом, которые широко применяются на практике, основаны на задачах разложения чисел на множители. Данная задача считается вычислительно сложной, однако ее точный класс сложности неизвестен (предполагается, что она относится к классу NP). В целях повышения криптографической стойкости криптосистем в литературе предлагались различные криптосхемы с открытым ключом, основанные на доказано сложных задачах, в комбинаторных NP-полных задачахили алгоритмических задачах алгебры.

Одной из основных проблем при создании криптосистемы с открытым ключом является сложность выбора конкретного экземпляра рассматриваемой вычислительно сложной задачи. Например, если рассматривать NP-полную задачу о рюкзаке, то далеко не каждый выбор весовых коэффициентов порождает действительно вычислительно сложную задачу, хотя универсальный алгоритм, который эффективно решает любую задачу данного типа, неизвестен. В системах с открытым ключом злоумышленник пытается решить один единственный экземпляр задачи, а не любую задачу данного типа.

В данной работе рассматривается возможность построения криптосистемы с открытым ключом на основе задач разложения регулярных языков. Такой выбор связан с высокой алгоритмической сложностью задачи разложения (известные алгоритмы имеют двойную экспоненциальную сложность) и, тем фактом что в настоящее время неизвестно нетривиальных систем языков, допускающих эффективной решение данной задачи. Последнее обстоятельство позволяет надеяться, что при генерации ключей в данном методе можно выбирать произвольные языки с достаточно большим числом состояний соответствующих автоматов.

1 Основные определения

Алфавитом называется конечное непустое множество, элементы которо-

го называются *символами* или *буквами*. Конечная последовательность символов алфавита Σ называется *словом* в этом алфавите. Количество элементов этой последовательности называется *длиной слова*. Слово, не содержащее ни одного символа, называется *пустым* и обозначается ε. Произвольное множество слов в алфавите называется *языком*. *Конкатенацией* языков L_1 и L_2 называется язык $L_1L_2 = \{w_1w_2 \mid w_1 \in L_1\, w_2 \in L_2\}$. Конкатенация языка с самим собой называется степенью: $L^k = LL^{k-1}$. Нулевой степенью языка L по определению является язык содержащий только пустое слово. *Итерацией* языка L называется язык L^*, который является объединением всех его степеней: $L^* = \cup_{k=0}^{\infty} L^k$.

Недетерминированным конечным автоматом (НКА) называется набор $\mathcal{A} = \langle Q, \Sigma, \delta, F, q_\circ\rangle$, где Q — множество состояний, Σ — входной алфавит, $\delta : Q \times \Sigma \to 2^Q$ — функция перехода, $F \subseteq Q$ — множество заключительных состояний, $q_\circ$ — начальное состояние автомата. Последовательность $(q_1, a_1, q_2), \ldots, (q_n, a_n, q_{n+1})$ называется *допустимым путем* в $\mathcal{A}$, если $q_1 = q_\circ$, $q_{i+1} \in \delta(q_i, a_i)$ и $q_{n+1} \in F$. Этому пути соответствует слово $w = a_1 \ldots a_n$. Слово $a_1a_2 \ldots a_n \in \Sigma^*$ *распознается* (допускается) автоматом $\mathcal{A}$, если существует соответствующий этому слову допустимый путь. Каждый автомат *распознает* язык $L(\mathcal{A}) \subseteq \Sigma^*$, который состоит из всех слов, распознаваемых этим автоматом. Язык называется регулярным, если он распознается некоторым автоматом. Множество всех регулярных языков в алфавите Σ обозначается $\mathrm{Reg}(\Sigma)$.

Автоматом с расстояниями над алфавитом Σ называется набор $\mathcal{A} = \langle \Sigma, Q, \rho, q_\circ, F, d\rangle$, где $\langle \Sigma, Q, \rho, q_\circ, F\rangle$ задает недетерминированный автомат, а $d : Q \times \Sigma \times Q \to \{0, 1\}$ — функция расстояния. Расстоянием $d(\pi)$ пути $\pi = (q_1, a_1, q_2), \ldots, (q_n, a_n, q_{n+1})$ является сумма расстояний его элементов. Расстоянием $d(w)$ слова $w \in L(\mathcal{A})$ является минимум расстояний по всем допустимым путям, соответствующим слову w. Автомат с расстояниями $\mathcal{A}$ называется *ограниченным*, если существует такая константа M, что $d(w) < M$ для всех слов $w \in L(\mathcal{A})$.

Теорема 1 ([3, 5]). *Свойство ограниченности автоматов с расстояниями алгоритмически разрешимо. Если автомат ограничен и имеет m состояний, то он ограничен константой $M = 2^{3m^3+m\lg m+m-1}$.*

В заключении данного раздела приведем определение рационального множества регулярных языков. Пусть заданы два непересекающихся алфавита Σ и Δ. Подстановкой регулярных языков назовем отображение $\varphi : \Delta \to \mathrm{Reg}(\Sigma)$, которое сопоставляет каждой букве алфавита Δ некоторый регулярный язык над Σ. Эта подстановка может быть естественным образом расширена до гомоморфизма между свободной полугруппой Δ^+ и конечно порожденной полугруппой регулярных язы-

ков $\varphi : (\Delta^+, \cdot) \to \mathrm{Reg}(\Sigma), \cdot)$. Множество $\mathcal{R}$ регулярных языков над Σ называется *рациональным*, если существует конечный алфавит Δ, регулярный язык $K \subseteq \Delta^+$ и подстановка $\varphi : \Delta^+ \to \mathrm{Reg}(\Sigma)$ для которых $\mathcal{R} = \{\varphi(w) \mid w \in K\}$. Рациональные множества регулярных языков являются рациональными подмножествами конечно порожденных полугрупп регулярных языков относительно конкатенации.

2 Задача разложения регулярных языков

Сформулируем задачу разложения регулярных языков по фиксированному базису. Пусть заданы регулярные языки $E_1, \ldots, E_k$. Существует ли алгоритм, позволяющий проверить, возможно ли представить заданный язык L в виде конкатенации языков $\{E_i\}$. Далее мы будем предполагать, что все языки содержат пустое слово. В противном случае длина искомого разложения ограничена сверху длиной самого короткого слова в языке L. Задача разложения связана с такими задачами, как проверка *свойство конечной степени* регулярного языка, проверка ограниченности автомата с расстояниями, разложение языка в произведение простых и «звездных» языков и теоремой Крона-Роудса. Положительное решение было получено в [4].

Теорема 2 ([4]). *Существует алгоритм, который для любого конечного множества $\mathcal{E}$ регулярных языков над алфавитом Σ, любого набора $T \subseteq \{\cdot, \cup, ^*\}$ языковых операций и регулярного языка L проверяет, возможно ли выразить L через языки $\mathcal{E}$ используя конечное число операций из T.*

Ключевую роль в доказательстве этого утверждения играет разрешимость задачи проверки ограниченности автомата с расстояниями. Алгоритм, который представлен в доказательстве этой теоремы, состоит в построении по языкам L и $\mathcal{E}$ некоторого автомата с расстояниями и проверке всех возможных «коротких» разложений L по базису $\mathcal{E}$, длина которых не превосходит константы, указанной в теореме 1. В [1] алгоритм разложения был расширен для случая рациональных множеств регулярных языков.

Теорема 3 ([1]). *Задача проверки принадлежности регулярного языка L заданному рациональному множеству $\mathcal{R} = \{\varphi(w) \mid w \in K\}$ алгоритмически разрешима.*

Алгоритм проверки принадлежности языка L заданному рациональному множеству, представленный в [1], также основан на проверки свойства ограниченности автомата с расстояниями и имеет двойную экспо-

ненциальную сложность относительно числа состояний недетерминированного автомата, представляющего язык L. Следует отметить, что в работе [6] была доказана экспоненциальная нижняя оценка сложности для рассматриваемой задачи, и неизвестно эффективных алгоритмов решения данной задачи, даже в случае конечных языков.

3 Возможная криптосхема

В предыдущем разделе было показано, что задача проверки принадлежности языка заданному рациональному подмножеству конечно порожденной полугруппы регулярных языков является вычислительно сложной задачей. В данном разделе приводится описание возможной криптосхемы, построенной на основе этой задачи.

В качестве открытого ключа предлагается взять рациональное множество регулярных языков над алфавитом Σ, заданное множеством образующих и регулярным языком в алфавите образующих Δ. Шифротекстом одного бита является автомат, который представляет 1 или 0 в зависимости от принадлежности рациональному множеству.

Закрытый ключ должен позволять эффективно проверить принадлежность, не обращаясь к проверке свойства ограниченности автомата с расстояниями. Возможным выбором закрытого ключа является конечно порожденная подполугруппа F свободной полугруппы Σ^*. Поскольку такие полугруппы обладают достаточно простой алгебраической структурой, то многие алгоритмические задачи имеют в данном случае эффективное решение. В частности, для любой такой полугруппы можно построить регулярный язык нормальных форм ее элементов и семейство конечных автоматов, так называемую *автоматную структуру*, с помощью которых любое слово может быть приведено к нормальной форме [2]. Это означает, что задача равенства слов для таких полугрупп решается за квадратичное, в зависимости от дины слов, время. Таким образом можно использовать следующую схему генерации ключей:

1. выбрать случайным образом конечное множество $F = \{f_1, \ldots, f_n\}$ слов;
2. построить регулярный язык N нормальных форм полугруппы F^+;
3. выбрать регулярный язык $R \subset N$;
4. построит язык $H = Complement(Fact(F^*))$;

5. случайным образом выбрать регулярные языки $H_i \subseteq H$ для $i = 1, \ldots, n$.

Автоматная структура полугруппы F^+ является закрытым ключом, а языки $H_i \cup \{f_i\}$, R образуют открытый ключ. Для шифрования одного бита требуется построить автомат M, соответствующий слову из R (или его дополнения R^c). Дешифровка сводится к нахождению самого длинного слова в $F^+ \cap M$ и проверки его принадлежность рациональному подмножеству полугруппы F^+, которое задано языком R. Эта операция может быть эффективно выполнена с использованием автоматной структуры для F^+. Следует отметить, что поскольку эта проверка выполняется системой конечных автоматов, то она может быть эффективно реализована на аппаратном уровне, например, в реконфигурируемых однородных вычислительных структурах.

4 Заключение

В данной работе описывается возможная криптографическая схема с открытым ключом на основе задачи разложения регулярных языков. В данной схеме декодирование имеет полиномиальную сложность, в то время как злоумышленник должен решать задачу двойной экспоненциальной сложности. В тоже время целый ряд вопросов требует дополнительных исследований. Основным вопросом является оценка «реальной» сложности проверки принадлежности языка заданному рациональному подмножеству полугруппы регулярных языков. Также необходимо исследовать зависимость числа состояний автоматов (элементов полугруппы) от длины соответствующего им слова в языке нормальных форм, поскольку число состояний автомата непосредственно связано с коммуникационной сложностью данной схемы.

Список литературы

[1] *Afonin, S.* Membership and finiteness problems for rational sets of regular languages / S. Afonin, E. Khazova // *International Journal of Foundations of Computer Science.* — 2006. — Vol. 17, no. 3. — Pp. 493–506.

[2] Automatic semigroups / C. M. Campbell, E. F. Robertson, N. Ruškuc, R. M. Thomas // *Theoretical Computer Science.* — 2001. — January. — Vol. 250, no. 1–2. — Pp. 365–391.

[3] *Hashiguchi, K.* Limitedness theorem on finite automata with distance functions / K. Hashiguchi // *Journal of computer and system sciences.* — 1982. — Vol. 24. — Pp. 233–244.

[4] *Hashiguchi, K.* Representation theorems on regular languages / K. Hashiguchi // *Journal of computer and system sciences.* — 1983. — Vol. 27. — Pp. 101–115.

[5] *Leung, H.* The limitedness problem on distance automata: Hashiguchi's method revisited / H. Leung, V. Podolskiy // *Theoretical Computer Science.* — 2004. — January. — Vol. 310, no. 1–3. — Pp. 147–158.

[6] Rewriting of regular expressions and regular path queries / D. Calvanese, G. De Giacomo, M. Lenzerini, M. Vardi // *Journal of Computer and System Sciences.* — 2002. — May. — Vol. 64. — Pp. 443–465.