Об аппроксимации регулярных языков безопасными языками

А. В. Галатенко

В работе исследуются аппроксимативные свойства безопасных языков, введенных в статье «Автоматные модели защищенных компьютерных систем». Рассматривается приближение произвольных регулярных языков сверху и снизу, а также возможные значения функции роста безопасных языков.

Ключевые слова: безопасные языки, регулярные языки, приближение языков, функция роста.

1. Основные понятия и результаты

Напомним определение безопасного языка, введенного в работе [1]. Под конечным автоматом мы будем понимать четверку $V = (A, Q, \varphi, q_0)$, где A — конечное множество входных символов, Q — конечное множество состояний, $\varphi : A \times Q \to Q$ — функция переходов, $q_0 \in Q$ — начальное состояние. Пусть $Q = S \cup I$, причем $S \cap I = \varnothing$. Состояния из S назовем безопасными, состояния из I — небезопасными. Далее будем предполагать, что начальное состояние является безопасным, все состояния достижимы из начального, а $|Q| > 1$.

Обозначим через A^* множество всех конечных слов в алфавите A. Если не оговорено противное, мы будем рассматривать только слова ненулевой длины. Функция φ может быть продолжена на множество $A^* \times Q$ по мультипликативности.

Подмножество A^* называется языком. Каждому слову $\alpha \in A^*$ соответствует слово $\kappa(\alpha) \in Q^*$, $\kappa(\alpha) = \varphi(\alpha, q_0)$. Назовем слово $\alpha \in A^*$ безопасным, если $\kappa(\alpha) \in S^*$. Назовем язык $\mathcal{A} \subseteq A^*$ безопасным (S-языком), если все слова, составляющие $\mathcal{A}$, безопасны, и не существует безопасных слов, не принадлежащих $\mathcal{A}$.

Пусть L — некоторый регулярный язык в алфавите A. Безопасный язык U называется безопасным приближением снизу для L, если $U \subseteq L$, и не существует безопасного языка U', для которого выполнены включения $U' \subseteq L$ и $U \subset U'$. Содержательно безопасное приближение снизу — это максимальный по включению S-язык, содержащийся в L.

Пусть L — некоторый регулярный язык. По теореме Клини ([2]), существует конечный автомат $V = (A, Q, \varphi, q_0)$, распознающий язык L с помощью подмножества $Q_F \subseteq Q$. Рассмотрим следующее преобразование диаграммы Мура автомата V. Добавим одну дополнительную вершину, соответствующую дополнительному поглощающему состоянию q_i, из которого по любому входному символу автомат переходит в то же состояние. Рассмотрим все переходы в диаграмме Мура автомата V, ведущие в состояние из $Q \setminus Q_F$, и перенаправим соответствующие этим переходам ребра в q_i. Легко увидеть, что получившаяся конструкция будет диаграммой Мура для некоторого автомата V'. В качестве безопасных состояний используем множество $S = \{q_0\} \cup Q_F$. Обозначим соответствующий S безопасный язык через L'.

Теорема 1. *Для любого регулярного языка L существует единственное безопасное приближение снизу U, совпадающее с L'. Для любого безопасного языка SL, отличного от A^*, существует счетное число регулярных языков, для которых SL является приближением снизу. Существует регулярный язык L_0 и его безопасное приближение снизу UL_0, для которых $|L_0 \setminus UL_0| = \infty$.*

Пусть L — некоторый регулярный язык в алфавите A. Безопасный язык W называется безопасным приближением сверху для L, если $L \subseteq W$, и не существует безопасного языка W', для которого выполнены включения $L \subseteq W'$ и $W' \subset W$. Содержательно безопасное приближение сверху — это минимальный по включению S-язык, содержащий L.

Пусть L — некоторый регулярный язык. По теореме Клини ([2]), существует конечный автомат $V = (A, Q, \varphi, q_0)$, распознающий язык L с помощью подмножества $Q_F \subseteq Q$. Обозначим через Q_S подмножество Q, состоящее из всех состояний, достижимых из q_0, из которых достижимо хотя бы одно из состояний из Q_F. Если L непуст, обо-

значим через L'' S-язык, соответствующий автомату V и множеству безопасных состояний Q_S (легко увидеть, что в этом случае $q_0 \in Q_S$). В противном случае обозначим через L'' пустой язык (в [1] доказано, что пустой язык безопасен). Обозначим через PL язык, полученный объединением L и $[L]$, где $[L]$ обозначает множество всех префиксов слов из L ([2]).

Теорема 2. *Для любого регулярного языка L существует единственное безопасное приближение сверху W, совпадающее с L'' и PL. Для любого безопасного языка SL, содержащего бесконечное число слов, существует счетное число регулярных языков, для которых SL является приближением сверху. Существует регулярный язык L_1 и его безопасное приближение сверху WL_1, для которых $|WL_1 \setminus L_1| = \infty$.*

Пусть L — некоторый регулярный язык, $n \in \mathbb{N}$. Обозначим через $L(n)$ слова из L, длина которых равна n. Рассмотрим $G_L : \mathbb{N} \to \mathbb{N} \cup \{0\}$, $G_L(n) = \sum\limits_{i \leqslant n} |L(n)|$. $G_L(n)$ называется функцией роста языка L ([5]). Содержательно она означает число слов в L, длина которых не превосходит n. Функция роста используется для оценки качества приближений языков — в качестве меры близости языков L_1 и L_2 можно рассмотреть, например, отношение $\frac{G_{L_1 \cap L_2}}{G_{L_1 \cup L_2}}$.

Охарактеризуем возможные предельные значения функций роста безопасных языков и точности приближения сверху и снизу.

Теорема 3. *Пусть SL — безопасный язык. Тогда либо существуют $c, C \in \mathbb{R}$, $k \in \mathbb{N} \cup \{0\}$, $\rho \in \mathbb{A}$, $\rho > 1$, такие что $cn^k\rho^n \lesssim GL_{SL}(n) \lesssim Cn^k\rho^n$, $n \to \infty$, либо существуют $N \in \mathbb{N}$ и $c \in \mathbb{N} \cup \{0\}$, такое что $GL_{SL}(n) = c$ для любого $n > N$, либо существуют $p, q \in \mathbb{N}$, $p \geqslant q$, такие что $GL_{SL}(n) = \frac{p}{q}n + o(n)$, либо существуют $p, q \in \mathbb{N}$, $k \in \mathbb{N}$, $k > 1$, такие что $GL_{SL}(n) = \frac{p}{q}n^k + o(n^k)$.*

Будем говорить, что язык имеет полиномиальный рост, если его функция роста асимптотически не превосходит некоторый полином. В противном случае будем говорить, что рост языка экспоненциальный.

Теорема 4. *Пусть L — регулярный язык полиномиального роста, $U(L)$ и $W(L)$ — безопасные приближения L снизу и сверху. Пусть L распознается автоматом с m состояниями. Если L имеет полиномиальный рост, $U(L)$ и $W(L)$ также имеют полиномиальный рост; существуют пределы $c = \lim\limits_{n\to\infty} \frac{GL_{U(L)}(n)}{GL_L(n)}$ и $C = \lim\limits_{n\to\infty} \frac{GL_L(n)}{GL_{W(L)}(n)}$, причем $0 \leqslant c \leqslant 1$, $\frac{1}{m} \leqslant C \leqslant 1$, $c, C \in \mathbb{Q}$; существует регулярный язык LP_0, для которого $c = 0$; существует регулярный язык LP_1, для которого $c = C = 1$; для любых $p, q \in \mathbb{N}$, $p < q$, существует регулярный язык $LP_{\frac{p}{q}}$, для которого $c = \frac{p}{q}$, и регулярный язык $LP'_{\frac{p}{q}}$, для которого $C = \frac{p}{q}$.*

Теорема 5. *Если L имеет экспоненциальный рост, то $U(L)$ может иметь как полиномиальный, так и экспоненциальный рост, $W(L)$ имеет экспоненциальный рост, совпадающий по порядку с ростом L. При этом для любых $p, q \in \mathbb{N}$, $p \leqslant q$, существуют регулярные языки экспоненциального роста L^1, L^2, для которых $\liminf\limits_{n\to\infty} \frac{GL_{L^1}(n)}{GL_{W(L^1)}(n)} = \frac{p}{q}$, $\limsup\limits_{n\to\infty} \frac{GL_{L^2}(n)}{GL_{W(L^2)}(n)} = \frac{p}{q}$.*

Замечание. Предела отношений функций роста в экспоненциальном случае может не существовать. В качестве примера можно рассмотреть язык, состоящий из слов четной длины в алфавите не менее, чем из двух элементов. В этом случае нижний предел равен $\frac{1}{3}$, верхний — $\frac{2}{3}$.

Содержательно теоремы 4 и 5 означают, что безопасное приближение снизу может быть сколь угодно далеким, тогда как безопасное приближение сверху по росту отличается от приближаемого языка в константу раз. При этом константа может быть сколь угодно большой.

Автор выражает глубокую благодарность своему научному руководителю, д.ф.-м.н., проф. В. Б. Кудрявцеву за постановки задач и внимание к работе.

2. Вспомогательные утверждения

Лемма 1 (о полиномиальном росте). *Пусть* $f : \mathbb{N} \to \mathbb{N} \cup \{0\}$, *отлична от тождественного 0 и имеет вид* $\sum_i a_i(n) n^{k_i} \rho_i^n$, *где* $a_i(n)$ — *периодические функции, принимающие алгебраические значения, причем* $a_i(n)$ *не равно тождественному 0,* $k_i \in \mathbb{N} \cup \{0\}$, ρ_i — *неотрицательные алгебраические числа,* $\rho_i \leqslant 1$, *и суммирование ведется по некоторому конечному множеству индексов. Тогда* $\rho_i = 1$, *и все значения* $a_i(n)$ *рациональны.*

Доказательство. Без ограничения общности можно считать, что при различных значениях i либо k_i, либо ρ_i являются различными (в противном случае просто приведем подобные члены — при этом a_i останутся алгебраическими; так как f отлична от тождественного 0, найдется $a_i(n)$, отличная от тождественного 0).

Предположим, что для всех ρ_i выполнено неравенство $\rho_i < 1$. В этом случае все слагаемые есть $o(1)$, $n \to \infty$, так как экспонента растет быстрее любого многочлена. Следовательно, начиная с некоторого $N \in \mathbb{N}$, $f(n)$ станет равна 0. Так как все $a_i(n)$ периодические, можно считать, что имеется общий период π для всех слагаемых. Рассмотрим подмножество длин, отличающихся на величину, кратную π, для которых хотя бы один из коэффициентов a_i не равен 0. Выделим слагаемое, для которого a_i не равно 0, а ρ_i максимально. Если таких слагаемых несколько, выберем из них слагаемое с максимальным значением k_i. Такое слагаемое $S(n)$ выбирается однозначно. Очевидно, что сумма оставшихся слагаемых слагаемых есть $o(S(n))$, $n \to \infty$. Следовательно, начиная с некоторого $N' \in \mathbb{N}$, значение всей суммы для выбранного подмножества слов будет строго положительным или строго отрицательным, в зависимости от знака a_i в $S(n)$. Возникает противоречие с условием, что длин, для которых число слов отлично от 0, конечное число.

Пусть максимальное ρ_i равно 1. Покажем, что в этом случае все ρ_i равны 1. Предположим противное. Рассмотрим все слагаемые, для которых $\rho_i < 1$, и подмножество натуральных чисел, отличающихся на число, кратное π, причем хотя бы одно из выделенных слагаемых на выбранном подмножестве отлично от 0. Если все слагаемые

с $\rho_i = 1$ в выделенном подмножестве равны 0, аналогично предыдущему случаю получаем, что начиная с некоторого N все суммы по модулю будут больше 0, но меньше 1, что невозможно, так как функция может принимать только целые значения.

Пусть хотя бы одно слагаемое с $\rho_i = 1$ отлично от 0. Индукциией по степени многочлена покажем, что в этом случае все $a_i(n)$ для слагаемых с $\rho_i = 1$ в выделенном подмножестве рациональны, и все $a_i(n)$ для слагаемых с $\rho_i < 1$ равны 0. Действительно, для многочленов степени 0 это верно, так как все слагаемые с $\rho_i < 1$ стремятся к 0 при $n \to \infty$. Для многочлена степени 1 это также верно. Действительно, пусть n_0 — период выбранной последовательности, $P(n) = an + b$ — полиномиальная часть f, $E(n)$ — экспоненциальная часть f. Рассмотрим $d(n) = f(n + n_0) - f(n)$ на выбранном подмножестве. Очевидно, что $d(n)$ принимает только целые значения, при этом полиномиальная часть $d(n)$ равна $a \cdot n_0$, а экспоненциальная стремится к 0 при $n \to \infty$. Следовательно, $a \cdot n_0$ целое, и a рационально. Пусть a имеет вид $\frac{p}{q}$, где p целое, а q — натуральное. Домножим $f(n)$ на q. Домноженный многочлен также принимает целые значения, причем линейная часть дает только целые слагаемые. Следовательно, так как экспоненциальная часть стремится к 0, $n \to \infty$, она тождественно равна 0, а $q \cdot b$ целое. Следовательно, b рациональное.

Пусть утверждение доказано для многочленов степени не выше l. Рассмотрим $f(n)$, для которой полиномиальная часть имеет степень $l{+}1$. Рассмотрим $f'(n) = f(n{+}n_0){-}f(n)$. Полиномиальная часть $f'(n)$ будет иметь степень l, а коэффициенты — линейно выражаться через коэффициенты полиномиальной части $f(n)$, причем коэффициенты линейных соотношений будут целыми. По индуктивному предположению, экспоненциальная часть $f'(n)$ равна 0, и все коэффициенты полиномиальной части рациональны. Линейные соотноешния на коэффициенты $f'(n)$ образуют систему линейных уравнений, решив которую, можно найти все коэффициенты полиномиальной части $f(n)$, кроме свободного члена. Несложно увидеть, что матрица системы будет верхнетреугольной, причем диагональные элементы не равны 0. В силу рациональности левых и правых частей, все коэффициенты полиномиальной части $f(n)$ однозначно восстанавливаются и являются рациональными. Равенство нулю экспоненциальной части и ра-

циональность свободного члена доказываются аналогично линейному случаю. Лемма доказана.

Лемма 2 (об оценке роста). *Пусть L — непустой регулярный язык, распознаваемый автоматом с m состояниями, WL — безопасное приближение сверху для L. Тогда $GL_{WL}(n) \lesssim GL_L(n) + GL_L(n+1) + \ldots + GL_L(n+m-1)$, $n \to \infty$. Существует регулярный язык L_1, для которого неравенство превращается в асимптотическое равенство.*

Доказательство. Обозначим через $L(n)$ и $WL(n)$ число слов длины n в соответствующих языках. Поставим в соответствие слову из α WL слово α' из L по следующему правилу. Пусть L распознается автоматом $V = (A, Q, \varphi, q_0, Q_F)$, где Q_F — множество принимающих состояний. Рассмотрим кратчайшее слово β, переводящее V из состояния $\varphi(\alpha, q_0)$ в одно из состояний Q_F. Такое слово существует в силу свойства минимальности безопасного приближения сверху. По принципу ящиков Дирихле $0 \leqslant |\beta| \leqslant m-1$. $\alpha' = \alpha\beta$. Очевидно, что при таком соответствии различным словам длины n из WL соответствуют различные слова из L. Следовательно множество слов длины n в языке WL вложено в объединение множества слов длины от n до $n+m-1$ в языке L, то есть $WL(n) \leqslant L(n) + L(n+1) + \ldots + L(n+m-1)$. Просуммировав неравенства по n и добавив недостающие начальные константы, получаем соотношение $GL_W L(n) \lesssim GL_L(n) + GL_L(n+1) + \ldots + GL_L(n+m-1)$, $n \to \infty$.

Рассмотрим язык $L_1 = ((a)^m)^*$, равный множеству всех слов, состоящих из букв a с длиной, кратной m. Легко увидеть, что этот язык регулярен, распознается автоматом с m состояниями (и не распознается автоматом с меньшим числом состояний), и функция роста равна $[n/m]$. Безопасным приближением сверху по условию префиксности является язык a^* с функцией роста n. Таким образом, асимптотическое неравенство для L_1 превращается в асимптотическое равенство. Лемма доказана.

3. Доказательство теоремы 1

Покажем, что L' является приближением снизу. Сначала докажем, что $L' \subseteq L$. Пусть $\alpha \in L'$. По построению, а также в силу определения безопасного языка, при подаче на вход L' слова α автомат выходит из начального состояния, после чего перемещается по стостояниям из множества $Q_F \cup \{q_0\}$. Если $q_0 \notin Q_F$, то возвращение в q_0 невозможно, так как все переходы в состояния, не принадлежащие Q_F, перенаправлены в q_i. Следовательно, при подаче слова α на вход автомату V автомат перейдет в одно из состояний Q_F, то есть $\alpha \in L$.

Покажем, что L' является максимальным. Предположим противное. Пусть $L' \subset U'$, $U' \subseteq L$, и U' является S-языком. Следовательно, существует слово $\alpha \in U'$, $\alpha \notin L'$. По построению диаграммы Мура для L', у слова α имеется непустой префикс α', при подаче которого на вход автомат, задающий L', переходит в небезопасное состояние, то есть всостояние, не принадлежащее Q_F. Таким образом, $\alpha' \notin L$. Так как $[U'] \subseteq U'$ ([1]), $\alpha' \in U'$, что противоречит определению приближения снизу.

Покажем, что приближение снизу является единственным. Предположим противное. Пусть U_1 и U_2 — два различных приближения снизу для языка L. Рассмотрим язык U $U_1 \cup U_2$. Так как $U_1 \subseteq L$ и $U_2 \subseteq L$, то $U \subseteq L$. Покажем, что язык U является безопасным. Проверим выполнение условий критерия безопасности из работы [1]. Условие непустоты может быть удовлетворено включением пустого слова в U_1 и U_2 (и, следовательно, в U). Условие регулярности следует из регулярности U_1 и U_2 и операции объединения. Включение $[U] \subseteq U$ следует из включений $[U_1] \subseteq U_1$ и $[U_2] \subseteq U_2$. Следовательно, U может быть расширен до приближения снизу U'. Так как U_1 и U_2 различны, и $U_1 \subseteq U'$, $U_2 \subseteq U'$, то либо U_1, либо U_2 не является максимальным безопасным языком, содержащимся в L — противоречие.

Покажем, что для произвольного безопасного языка SL, отличного от A^*, существует счетное число регулярных языков, для которых SL является приближением снизу. Так как $SL \neq A^*$, существует слово $\alpha \in A^*$, $\alpha \notin SL$. Пусть $a \in A$. Рассмотрим последовательность слов $\alpha a, \alpha aa, \ldots, \alpha a^n, \ldots$. Рассмотрим языки $L_{SL}(n)$, $n \in \mathbb{N}$, $L_{SL}(n)$ равен объединению SL и n-того члена построенной последователь-

ности. Очевидно, что $L_{SL}(n)$ регулярны. Приближением снизу для $L_{SL}(n)$ при любом n будет SL, так как подъязыки $L_{SL}(n)$, являющиеся собственными надмножествами SL, не удовлетворяют условию префиксности.

Рассмотрим язык L_0, состоящий из всех слов четной длины. Очевидно, что L_0 регулярен. Безопасное приближение UL_0 состоит только из пустого слова, поэтому $|L_0 \setminus UL_0| = \infty$. Теорема доказана.

4. Доказательство теоремы 2

Покажем, что L'' является приближением сверху. Сначала докажем, что $L \subseteq L''$. Пусть $\alpha \in L$. Следовательно, при подаче α на вход V автомат переходит в состояние из множества Q_F. Следовательно, все состояния, в которые попадет V, окажутся в Q_S, то есть $\alpha \in L''$.

Покажем, что L'' является минимальным. Предположим противное. Пусть $W' \subset L''$, $L \subseteq W'$, и W является S-языком. Следовательно, существует слово $\alpha \in L''$, $\alpha \notin W'$. По построению L'', существует слово β (возможно пустое), такое что при подаче на вход V слова $\alpha\beta$ автомат перейдет в состояние из Q_F. Следовательно, $\alpha\beta$ принадлежит L. Так как $L \subseteq W'$, $\alpha\beta \in W'$. Так как W' является S-языком, для него выполнено свойство префиксности, то есть $\alpha \in W'$ — противоречие.

Покажем, что PL является приближением сверху. По критерию безопасности языка ([1]), PL является безопасным (регулярность следует из теоремы Клини, префиксность — из построения). По построению $L \subseteq PL$. Покажем, что PL является минимальным. Предположим противное. Пусть существует S-язык W'', для которого справедливо $W'' \subset PL$ и $L \subseteq W''$. Следовательно, существует слово $\alpha \in PL$, $\alpha \notin W''$. Так как $L \subseteq W''$, α является префиксом некоторого слова $\beta \in L$. Таким образом, $\beta \in W''$, $\alpha \notin W''$, что противоречит условию префиксности.

Покажем, что приближение сверху является единственным. Предположим противное. Пусть W_1 и W_2 — два различных приближения сверху для языка L. Рассмотрим язык $W' = W_1 \cap W_2$. Легко увидеть, что $L \subseteq W'$. Покажем, что W' является безопасным. Регулярность W' может быть установлена с помощью теоремы Клини и рассмотрения

автомата, являющегося прямым произведением автоматов, распознающих W_1 и W_2. Префиксность следует из префиксности W_1 и W_2. Таким образом, W' является надмножеством некоторого приближения сверху, отличного от W_1 или W_2, что противоречит минимальности W_1 и W_2.

Покажем, что для произвольного безопасного языка SL, состоящего из бесконечного множества слов, существует счетное число регулярных языков, для которых SL является приближением сверху. Так как SL бесконечен, в нем найдется счетное число попарно различных слов, длина которых не меньше 2. Образуем последовательность таких слов $w_1, w_2, \ldots, w_n, \ldots$. Рассмотрим последовательность слов $w'_1, w'_2, \ldots, w'_k, \ldots$, построенную путем отбрасывания последней буквы в словах w_i дополнительной чистки, в результате которой из множества совпадающих слов остается один представитель. Так как $|A| < \infty$, последовательность w'_i состоит из счетного числа членов. Для произвольного $m \in \mathbb{N}$ рассмотрим язык $SL(n) = SL \setminus w'_n$. Все $SL(n)$ регулярны (это несложно показать с помощью теоремы Клини, рассмотрев прямое произведение автоматов, распознающих языки SL и $A^* \setminus w'_n$) и попарно различны (так как все w'_n попарно различны). По построению $[SL(n)] = SL$, так как единственный префикс слова из $SL(n)$, не принаджежащий $SL(n)$, есть w'_n.

В качестве языка L_1 рассмотрим язык, состоящий из всех слов четной длины. Безопасное приближение сверху в этом случае будет состоять из всех конечных слов в алфавите A, поэтому $|WL_1 \setminus L_1| = \infty$. Теорема доказана.

5. Доказательство теоремы 3

В работе [3] показано, что число слов длины n в регулярном языке L может быть вычислено по формуле $\sum\limits_i a_i(n) n^{k_i} \rho_i^n$, где индекс суммирования пробегает некоторое конечное множество, зависящее от L, $a_i(n)$ — периодическая функция, принимающая алгебраические значения, $k_i \in \mathbb{N} \cup \{0\}$, ρ_i — положительное алгебраическое число. Без ограничения общности можно считать, что при различных значениях i либо k_i, либо ρ_i являются различными (в противном случае просто

приведем подобные члены — при этом a_i останутся алгебраическими).

Так как нас интересуют предельные свойства, в случае, когда бесконечное число $a_i(n)$ отлично от 0, можно считать, что $a_i(n)$ строго периодические, так как изменение конечного числа слагаемых в этом случае не повлияет на асимптотику. Действительно, в силу леммы о полиномиальном росте, справедливо одно из следующих условий:

1) периодические части всех $a_i(n)$ равны 0;
2) существует $a_i(n)$ с периодической частью, не равной 0, причем для всех таких $a_i(n)$ $\rho_i = 1$;
3) существует $a_i(n)$ с периодической частью, не равной 0, причем $\rho_i > 1$.

Проанализируем возможные значения функции роста.

Рассмотрим первый случай. Пусть $N \in \mathbb{N}$ — длина максимального предпериода $a_i(n)$. Тогда для любого $n \in mathbbN$, $n > N$, $L(n) = 0$, следовательно, значение $G_L(n)$ становится целой неотрицательной константой. Покажем, что для безопасных языков реализуются все возможные целые неотрицательные константы. Пусть $c \in \mathbb{N} \cup \{0\}$. Рассмотрим автомат с $c+1$ состоянием. Состояние номер 1 является начальным, состояние номер $c+1$ является поглощающим. По первому символу алфавита автомат переходит из состояния номер i в состояние номер $i+1$, $i = 1, 2, \ldots, c$. Остальные переходы ведут в поглощающее состояние. Состояния с номерами от 1 до c являются безопасными. Легко увидеть, что, во-первых, пример обслуживает произвольный входной алфавит, и, во-вторых, безопасных слов длины не более c ровно c, а безопасных слов длины более c нет.

Рассмотрим второй случай. Для каждого i $\sum\limits_{l} = 0^n a_i(l) l^{k_i} \asymp \frac{c}{n+1} n^{k_i+1}$, $n \to \infty$, где c — среднее значение $a_i(l)$ по периоду. Действительно, предпериодическую часть как конечную и не влияющую на асимптотику добавку можно заменить на периодическую. Оценим сумму сверху суммой интегралов $I_t(n) = \int\limits_0^{[(n+1-t)/\pi]} a_i(t)(\pi x + t)^{k_i} dx$, где t пробегает по периоду a_i, обозначенному через π. $I_t(n) \asymp \frac{a_i(t)}{\pi k_i + 1} n^{k_i+1}$, $n \to \infty$. Оценим сумму снизу суммой интегралов $J_t(n) =$

$\int_0^{[(n-t)/\pi]} a_i(t)(\pi x + t)^{k_i} dx$, где t пробегает по периоду a_i. $J_t(n) \asymp \frac{a_i(t)}{\pi k_i+1} n^{k_i+1}$, $n \to \infty$, что доказывает утверждение. Выберем индексы i, для которых степень максимальна. Без ограничения общности можем считать, что такой индекс единственный — в противном случае можно просто привести подобные члены. Очевидно, что асимптотика в этом случае составит $G_L(n) \asymp \frac{p}{q} n^k$, $n \to \infty$, для некоторых $p, q, k \in \mathbb{N}$.

Если $|A| = 1$, то безопасный язык либо попадает в условия первого случая, либо состоит из всех конечных слов, и в этом случае функция роста равна n. В дальнейшем мы будем полагать, что $|A| > 1$.

Пусть $k = 1$. Покажем, что в этом случае для безопасных языков $p \geqslant q$. Действительно, для любого $n \in \mathbb{N}$ из равенства нулю $L(n)$ следует равенство 0 $L(n+1)$, так как в противном случае префикс длины n слова длины $n+1$ принадлежал бы языку. Зафиксируем $p, q \in \mathbb{N}$, $p \geqslant q$. Построим безопасный язык с функцией роста с асимптотикой $\frac{p}{q} n$, $n \to \infty$. В работе [4] показано, что при наличии единственного принимающего состояния q_F элементарный автомат, приведенный на рис. 1, имеет функцию роста, асимптотически равную $\frac{n^k}{l_1 \cdot l_2 \cdot \ldots \cdot l_k \cdot k!}$, $n \to \infty$, где k — число циклов, l_i — длина i-того цикла, все изображенные переходы происходят по одному символу, а все непоказанные переходы ведут в поглощающее состояние. Рассмотрим элементарный автомат с одним циклов длины q. Объявим все состояния, кроме поглощающего, безопасными. Сделаем длину «хвоста» из q_1 в q_F равной $p - q$. Для любого безопасного состояния вне цикла и «хвоста» число слов, заканчивающихся в таком состоянии, конечно и не влияет на асимптотику. Для любого безопасного состояния в цикле или в «хвосте», как показано в [4], число принимаемых этим состоянием слов длины не больше n асмптотически равно $\frac{n}{q}$. Так как принимаемое слово может приводить ровно в одно принимающее состояние, для получения общего числа принимаемых слов нужно провести суммирование по всем принимающим состояниям, дающим вклад в асимптотику. Так как таких состояний p (q в цикле и $p - q$ в «хвосте»), получаем требуемую асимптотику.

Пусть $k > 1$. Зафиксируем $p, q \in \mathbb{N}$ и построим пример безопасного языка с асимптотикой роста $\frac{p}{q} n^k$. Рассмотрим элементарный автомат

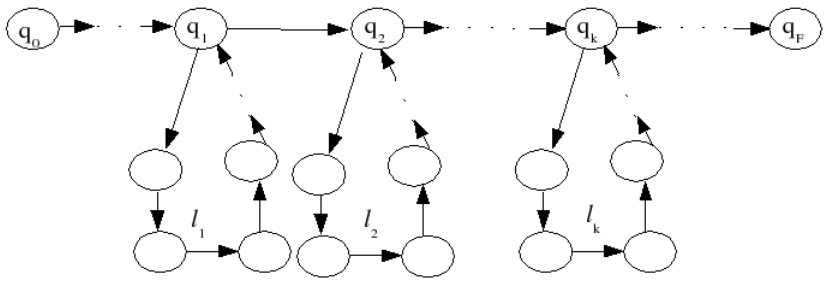

Рис. 1. Диаграмма Мура элементарного автомата из работы [4].

с k циклами, причем длина первого цикла, выходящего из вершины q_1, равна q, длина всех остальных циклов равна 1, длина «хвоста» равна $k! \cdot p - 1$, и все состояния, кроме поглощающего, являются безопасными. По доказанному в [4], вклад в асимптотику функции роста будут давать принимающие состояния от q_k до q_F. Таких состояний $k! \cdot p$, то есть функция роста асимптотически равна $\frac{p}{q} n^k$.

Отметим, что в полиномиальном случае мощность входного алфавита может быть произвольной, начиная с 2.

Рассмотрим третий случай. Выделим члены с максимальным ρ_i. Среди них выделим член с максимальным k_i. Очевидно, только этот член будет давать вклад в асимптотику. Пусть $k_i \neq 0$. Выписывая интегральные оценки аналогично полиномиальному случаю, получаем, что функция роста оценивается снизу функцией $c \cdot n^{k_i+1} \rho_i^n$ для некоторого $c \in \mathbb{R}$, и оценивается сверху функций $C \cdot n^{k_i+1} \rho_i^n$ для некоторого $C \in \mathbb{R}$. В качестве примера безопасного языка с экспоненциальным ростом можно привести язык A^*, функция роста которого равна $|A|^n$. Теорема доказана.

6. Доказательство теоремы 4

Пусть L — регулярный язык полиномиального роста, $U(L)$ — безопасное приближение снизу, $W(L)$ — безопасное приближение сверху. Пусть L распознается автоматом с m состояниями.

Так как $U(L) \subseteq L$, $GL_{U(L)}(n) \leqslant GL_L(n)$, следовательно рост $U(L)$ также полиномиальный. По теореме 3, $GL_{U(L)}(n) \asymp \frac{p_1}{q_1} n^{k_1}$, $GL_L(n) \asymp \frac{p_2}{q_2} n^{k_2}$ для некоторых $p_1, q_1, p_2, q_2, k_1, k_2$. Следовательно, существует рациональный предел $c = \lim\limits_{n\to\infty} \frac{GL_{U(L)}(n)}{GL_L(n)}$. Неравенства на c следуют из того, что $U(L) \subseteq L$, а также из того, что для любого языка его функция роста неотрицательна.

Нижняя оценка в неравенстве достигается, например, на языке, состоящем из одного двухбуквенного слова (в этом случае безопасное приближение снизу пусто, и предел равен 0), верхняя — на произвольном безопасном языке полиномиального роста, очевидным образом совпадающем со своим приближением снизу.

Пусть $p, q \in \mathbb{N}$, $p < q$. Рассмотрим язык, состоящий из объединения множества слов из не более p букв a и множества слов из букв a длины от $p+2$ до $q+1$. Легко увидеть, что безопасное приближение снизу будет содержать только слова первого множества, следовательно отношение функций роста равно $\frac{p}{q}$, начиная с $n = q+1$.

В силу теоремы 3 и леммы об оценке роста, $GL_{W(L)}(n) \lesssim m \cdot GL_L(n)$, следовательно $W(L)$ имеет полиномиальный рост. Применяя теорему 3, получаем, что существует рациональный предел $C = \lim\limits_{n\to\infty} \frac{GL_L(n)}{GL_{W(L)}(n)}$. В силу леммы об оценке роста, $C \geqslant \frac{1}{m}$, причем на языке L_1 из доказательства леммы об оценке роста достигается равенство. Верхнее неравенство вытекает из вложения $L \subseteq W(L)$. Неравенство превращается в равенство на безопасных языках полиномиального роста, совпадающих с со своим приближением сверху.

Пусть $p, q \in \mathbb{N}$, $p < q$. Рассмотрим язык, состоящий из слов из букв a длины от $q-p+1$ до q. Легко увидеть, что безопасное приближение сверху состоит из всех слов из буквы a длины не более q, и отношение функций роста равно $\frac{p}{q}$, начиная с $n = q$. Теорема доказана.

7. Доказательство теоремы 5

В дальнейшем мы будем считать, что входной алфавит состоит не менее, чем из двух символов (в противном случае языки не могут иметь экспоненциальный рост).

Построим примеры языков, доказывающих утверждения, касающиеся приближений снизу. Рассмотрим язык, содержащий все слова длина не менее 2. Легко увидеть, что безопасное приближение снизу пусто, то есть имеет константный рост, тогда как сам язык имеет экспоненциальный рост.

Рассмотрим произвольный безопасный язык полиномиального роста LP. Существование таких языков доказано в теореме 3. Рассмотрим слово $\alpha \notin LP$ и язык $LP' = LP \cup \alpha A^*$. В силу свойства префиксности безопасных языков, безопасное приближение снизу для LP' совпадает с LP. Очевидно, что рост LP' экспоненциальный.

Наконец, безопасным приближением снизу для любого безопасного языка экспоненциального роста LE (например, A^*) служит сам LE.

Перейдем к анализу безопасного приближения сверху. Рассмотрим произвольный регулярный язык экспоненциального роста L, распознаваемый автоматом с m состояниями, и его приближение сверху $W(L)$. По лемме об оценке роста, $GL_L(n) \leqslant GL_{WL}(n) \lesssim GL_L(n) + GL_L(n+1) + \ldots + GL_L(n+m-1)$, $n \to \infty$. В силу экспоненциальности роста без ограничения общности можо считать, что $GL_{WL}(n) > 0$. Поделим неравенство на $GL_{WL}(n)$. Из доказательства теоремы 3 следует, что возможны два случая: либо $GL_L(n)$ имеет порядок $n^k\rho^n$ для некоторых $k \in \mathbb{N}$, $\rho \in \mathbb{A}$, $\rho > 1$, либо $GL_L(n)$ имеет порядок ρ^n для некоторого $\rho \in \mathbb{A}$, $\rho > 1$.

В первом случае имеем:

$$\frac{GL_L(n)}{GL_{WL}(n)} \leqslant 1 \lesssim m\frac{GL_L(n)}{GL_{WL}(n)} + \frac{L(n+1)}{GL_{WL}(n)} + \ldots + \frac{L(n+m-1)}{GL_{WL}(n)}.$$

Так как $L(n)$ имеет порядок $n^{k-1}\rho^n$, при $n \to \infty$ правая часть неравенства асимптотически равна $m\frac{GL_L(n)}{GL_{WL}(n)}$, откуда следует, что $\frac{1}{m} \lesssim \frac{GL_L(n)}{GL_{WL}(n)} \leqslant 1$.

Во втором случае воспользуемся леммой об ограничении роста и оценкой, следующей из порядка $GL_L(n)$. Имеем:

$$GL_{W(L)}(n) \lesssim GL_L(n) + GL_L(n+1) + \ldots + GL_L(n+m-1) \lesssim$$
$$\lesssim C\rho^n(1 + \rho + \ldots + \rho^{m-1}) = C\rho^n\frac{\rho^m - 1}{\rho - 1}.$$

Воспользуемся оценкой, вытекающей из порядка $GL_L(n)$ и получим

$$GL_{W(L)}(n) \lesssim \frac{C}{c}\frac{\rho^m - 1}{\rho - 1} GL_L(n),$$

откуда следует неравенство $\frac{c}{C}\frac{\rho-1}{\rho^m-1} \lesssim \frac{GL_L(n)}{GL_{W(L)}(n)} \leqslant 1$, $n \to \infty$. Утверждение о порядке роста $W(L)$ доказано.

Если в качестве L рассмотреть произвольный безопасный язык экспоненциального роста, $W(L)$ совпадет с L, и верхний и нижний пределы из утверждения теоремы будут равны 1. Пусть $p, q \in \mathbb{N}$, $p < q$. Рассмотрим входной алфавит мощности q, $A = \{a_1, a_2, \ldots, a_q\}$, и язык $L_{\frac{p}{q}} = \bigcup\limits_{i=1}^{p} A^* a_i$. В силу свойства префиксности, безопасное приближение сверху совпадает с A^*. Легко увидеть, что $L_{\frac{p}{q}}(n) = pq^{n-1} = \frac{p}{q} W(L_{\frac{p}{q}})(n)$. Суммируя равенство по n, получаем, что отношение функций роста равно $\frac{p}{q}$. Теорема доказана.

Список литературы

[1] Галатенко А. В. Автоматные модели защищенных компьютерных систем // Интеллектуальные системы. Т. 11, вып. 1–4. М., 2007. С. 403–418.

[2] Кудрявцев В. Б., Алешин С. В., Подколзин А. С. Введение в теорию автоматов — М.: Наука, 1985.

[3] Плесневич Г. С. Оценка среднего времени вычисления на одномерных односторонних итеративных системах // ДАН СССР. 171, № 3. 1966. С. 537–540.

[4] Руденко А. Н., Строгалов А. С. О метрической сложности событий, представимых полиномиальными автоматами // Интеллектуальные системы. Т. 4, вып. 1–2. М., 1999. С. 305–319.

[5] Строгалов А. С. Об ε-моделировании конечных автоматов // Труды Всесоюзного семинара по дискретной математике. М.: изд-во МГУ, 1986.