

О Т З Ы В

официального оппонента, кандидата физико-математических наук
ХОЛОДЕНКО Александра Борисовича
на диссертацию Александрова Дмитрия Евгеньевича
**«Сложность распознавания принадлежности слова
регулярному языку в системах обнаружения вторжений»**,
представленную на соискание ученой степени
кандидата физико-математических наук
по специальностям 05.13.17 — Теоретические основы информатики
и 05.13.19 — Методы и системы защиты информации,
информационная безопасность.

Диссертационная работа Д. Е. Александрова посвящена изучению возможности использования аппарата формальных языков в системах обнаружения вторжений.

Актуальность темы обусловлена широким распространением сетевых технологий в различных сферах человеческой деятельности, от развлечений до управления опасными производствами, что ставит задачи обеспечения сетевой безопасности на одно из первых мест. Число атак на сетевые ресурсы постоянно растёт, поэтому средства обнаружения вторжений играют всё большую и большую роль в системах сетевой безопасности. Для обнаружения опасного или вредоносного трафика могут применяться различные подходы, однако нужно понимать, что действующая система безопасности должна обеспечивать очень большую пропускную способность, что либо приводит к существенному усложнению (и, как следствие, удорожанию) системы, либо накладывает существенные ограничения на класс используемых методов и алгоритмов. Одним из методов быстрого распознавания опасного сетевого трафика является использование регулярных языков, который позволяет проводить анализ с очень высокой скоростью, однако при реализации на процессорах общего назначения может требовать очень много оперативной памяти для хранения таблиц переходов.

Оппонируемая диссертация направлена на решение ряда принципиальных задач, связанных с уменьшением сложности конечных автоматов, служащих для обнаружения проблемного трафика. Основной

идеей работы является замена заданного регулярного выражения более простым (в смысле количества состояний реализующего его конечного детерминированного автомата) регулярным выражением, которое задавало бы язык, содержащий исходный, и не слишком сильно от него отличающийся. Решение этой задачи дает новый эффективный инструмент для улучшения действующих систем обнаружения вторжений.

Диссертация состоит из введения, пяти глав, заключения и списка литературы.

В первой главе диссертации описывается содержательная постановка задачи, а также требования к решению, вытекающие из анализа действующих систем обнаружения вторжений. Так, показано, что практически значимыми с точки зрения построения систем обнаружения вторжений являются регулярные языки вида $A^*R_1A^*R_2A^*$.

Во второй главе содержится достаточно полный обзор современного состояния работ по модификациям конечных автоматов. В частности, обсуждаются детерминированные и недетерминированные конечные автоматы и мультиавтоматы, а также автоматы с задержками, двойные автоматы и расширенные автоматы. Все указанные подходы не расширяют класса принимаемых ими языков, однако обладают разными сложностями как с точки зрения их описания, так и с точки зрения вычислительной трудоемкости.

В третьей главе описывается алгоритм объединения двух регулярных выражений специального вида над алфавитом A : выражения $A^*R_1A^*R_2A^*$ и $A^*R_3A^*R_4A^*$ заменяются на выражение $A^*(R_1|R_3)A^*(R_2|R_4)A^*$. Очевидно, что новое выражение задаёт регулярный язык, содержащий объединение двух исходных языков. При этом сложность нового языка (в смысле количества состояний задающего его приведённого автомата) будет не больше сложности автомата, задающего объединение двух исходных языков. Более того, в работе показано, что если алфавит содержит не менее трёх символов, то для любого рационального числа $0 < p/q \leq 1$ можно построить такой набор

регулярных выражений R_i , что отношение сложности нового выражения к сложности объединения двух исходных будет в точности равно p/q . В конце главы приведены результаты применения предложенного подхода к регулярным выражениям из базы сигнатур реальной системы обнаружения вторжений.

В четвертой главе рассматривается вопрос влияния применения операции объединения пары выражений на сложность автомата, задаваемого объединением множества регулярных выражений рассмотренного выше вида. В данной главе приведен ряд оценок для функции Шеннона. Более того, показано, что в случае алфавита, содержащего не менее трёх символов, для любого рационального числа $0 < p/q \leq 1$ можно построить такой набор выражений, что отношение числа состояний модифицированного автомата к числу состояний исходного будет в точности равно p/q .

В пятой главе обсуждаются вопросы роста принимаемого языка при операции объединения выражений. Для сравнения бесконечных языков используется функция роста $G_n(L) = |\{\alpha \in L \mid |\alpha| = n\}|$. В данной главе получена точная формула вычисления функции роста для специального подкласса класса выражений $A^*R_1A^*R_2A^*$, которая применена для оценки регулярных выражений из реальной системы обнаружения вторжений.

Следует заметить, что оппонируемая работа не лишена недостатков.

Во-первых, в Леммах 3 и 4 следовало бы подчеркнуть, что рассматриваемые автоматы – приведенные. Это следует из обозначений, но не отражено в тексте.

Во-вторых, Лемма 9 представляется очевидной, а её присутствие в тексте диссертации в виде отдельного утверждения с доказательством – избыточным.

В-третьих, диссертацию, несомненно, украсили бы оценки для роста языков для других подклассов выражений вида $A^*R_1A^*R_2A^*$. Это можно было бы порекомендовать в качестве направления для дальнейших исследований.

В-четвёртых, в третьей и четвертой главах при оценке уменьшения сложности результирующих автоматов не изучается вопрос о росте языков, получаемых в результате применения операции объединения. Например, тривиальная замена любой пары выражений на язык вида A^* , задаваемый автоматом с единственным состоянием, очевидно, минимизирует количество состояний автомата, но не решает исходную содержательную задачу. Таким образом, в тексте этих глав не хватает оценок качества предложенного решения.

Однако перечисленные недостатки не являются принципиальными и не снижают общую положительную оценку работы. В целом, в диссертационной работе Д. Е. Александрова получены новые интересные результаты, состоящие в разработке новых математических методов исследования регулярных языков специального вида и их применения для улучшения систем обнаружения вторжений.

Достоверность и обоснованность положений диссертации обусловлена строгим математическим доказательством всех утверждений работы.

Диссертация представляет собой завершённую научно-исследовательскую работу, содержащую решение актуальной задачи понижения сложности регулярных языков специального вида. Полученные автором результаты имеют теоретическую и практическую ценность и могут быть использованы при проектировании и реализации систем обнаружения вторжений, вирусной активности, фильтрации трафика и т.д.

Основные результаты диссертации изложены в четырёх публикациях в рецензируемых журналах, входящих в перечень журналов, рекомендованный ВАК, и доложены на ведущих международных и всероссийских семинарах и конференциях по тематике диссертационной работы. Опубликованные работы достаточно полно отражают содержащиеся в диссертации научные результаты.

Автореферат в полной мере отражает содержание диссертации.

Диссертация «Сложность распознавания принадлежности слова регулярному языку в системах обнаружения вторжений» полностью удовлетворяет всем требованиям ВАК РФ, предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук по специальностям 05.13.17 – «Теоретические основы информатики» и 05.13.19 – «Методы и системы защиты информации, информационная безопасность», а ее автор, Александров Дмитрий Евгеньевич, заслуживает присуждения ему ученой степени кандидата физико-математических наук.

Официальный оппонент

кандидат физико-математических наук,
Заместитель генерального директора
ООО «Центр прикладной соционики»,
127055, Москва г, Новослободская ул,
дом 31, стр. 2, тел. +7 (495) 979-78-16.
e-mail: alexander@kholodenko.ru

04.09.2015г.



Холоденко А.Б.