УДК 004.056

С. Ю. Микова, студент, e-mail: sofyamikova@mail.ru,
В. С. Оладько, канд. техн. наук, доц., e-mail: oladko.vs@yandex.ru
Волгоградский государственный университет, г. Волгоград

Сравнение алгоритмов выявления сетевых аномалии с помощью меры Ван Ризбергена

Предложена функциональная модель исследования качества алгоритмов обнаружения сетевых аномалий, использующая в качестве базовой метрики качества меру Ван Ризбергена. Разработан программный комплекс, автоматизирующий предложенную модель. Проведена экспериментальная оценка качества двух алгоритмов — Бродского—Дарховского и алгоритма обнаружения сетевых аномалий на основе дискретного вейвлет-преобразования. Проведенный анализ показал, что лучшим качеством классификации обладает алгоритм Бродского—Дарховского.

Ключевые слова: *точность, полнота, ошибка первого рода, ошибка второго рода, алгоритм Бродского—Дарховского, дискретное вейвлет-преобразование, атака*

Введение

На сегодняшний день важным атрибутом является глобальная информационная интеграция, которая состоит из построения компьютерных сетей и их объединения с помощью глобальной сети Интернет. В соответствии с работами [1, 2] в процессе пользования компьютерными сетями возникает ряд проблем, связанных с эксплуатацией злоумышленниками уязвимостей в приложениях, отказом в обслуживании сетевых ресурсов, утечкой информации, нарушением целостности и доступности данных. Поэтому администраторы должны диагностировать работу сети, подключенных к ней серверов и защищать информационные ресурсы сети от несанкционированной деятельности злоумышленников, воздействий вредоносного программного обеспечения и других видов атак. Одним из подходов к контролю над состоянием сети является регулярный мониторинг аномалий, возникающих в сетевом трафике. Своевременное выявление и подробный анализ сетевой аномалии позволяют администраторам обнаружить атаку злоумышленника на ранней стадии ее проведения. Следовательно, актуальной задачей является поиск наиболее качественного алгоритма обнаружения сетевых аномалий, который с наибольшей точностью и полнотой способен определять их, сводя к минимуму число пропусков и/или ложных срабатываний. В данной работе представлена модель исследования характеристик алгоритмов обнаружения сетевых аномалий, где в качестве основного показателя качества используется мера Ван Ризбергена. На основании разработанной модели проведено сравнение качества двух алгоритмов обнаружения сетевых аномалий — алгоритма Бродского—Дарховского (БД) и алгоритма обнаружения аномалий на основе дискретного вейвлет-преобразования (ДВП).

Показатели качества алгоритма обнаружения сетевых аномалий

Анализ литературных источников [3—7] показывает, что одним из возможных подходов к оценке качества алгоритмов обнаружения сетевых аномалий является вычисление полноты и точности алгоритма, которые напрямую зависят от ошибок первого и второго рода, возникающих в процессе работы алгоритма при решении задач классификации. Мера Ван Ризбергена является функцией от этих показателей и представляет собой гармоническое среднее между точностью и полнотой. Она стремится к нулю, если точность или полнота стремится к нулю, и вычисляется по следующей формуле:

$$F = 2\frac{Precision \cdot Recall}{Precision + Recall}, \qquad (1)$$

где значения полноты *Recall* и точности *Precision* вычисляется по формулам

$$Recall = \frac{TP}{TP + FN}; \qquad (2)$$

$$Precision = \frac{TP}{TP + FP}, \qquad (3)$$

где TP — истинно положительное решение (правильно обнаруженные аномалии); FP — ложно положительное решение (ошибки второго рода); FN — ложно отрицательное решение (ошибки первого рода).

Таким образом, данная мера зависит от числа ошибок первого и второго рода. Чем меньше будет ошибок первого и второго рода при обнаружении аномалий, тем более полным и точным будет алгоритм по их обнаружению.

Модель исследования качества алгоритмов обнаружения сетевых аномалий на основе меры Ван Ризбергена

Задача исследования состоит в том, чтобы, генерируя сетевой трафик с различными характеристиками как условно нормальными, так и аномальными (превышение допустимого объема трафика, частоты передачи пакетов, размеров сетевых пакетов), оценить качество существующих алгоритмов обнаружения сетевых аномалий. Полученные в результате исследования данные не только позволят выбрать наилучший алгоритм, но и могут быть использованы в дальнейшем при разработке новых алгоритмов, учитывающих недостатки и ограничения существующих.

Функциональная модель исследования качества алгоритмов обнаружения сетевых аномалий представлена в виде схемы в нотации IDEF0 на рис. 1. На вход модели подается следующая информация:

- о выбранных для исследования алгоритмах обнаружения сетевых аномалий;
- о параметрах трафика тестовой сети (интенсивность трафика от различных источников (хостов), размер сетевого пакета, число передаваемых за единицу времени сообщений), которые затем будут использованы для формирования "нормального трафика", имитирующего штатный режим работы сети, и трафика, содержащего аномалии, которые в идеальном случае должны быть полностью обнаружены алгоритмом.

На выходе модели должны быть данные о значениях показателей качества исследуемых алгоритмов (точности, полноты и меры Ван Ризбергена), результаты сравнения исследуемых алгоритмов по мере Ван Ризбергена и решение о том, какой алгоритм считается наилучшим по данным показателям.

В обобщенном виде процесс исследования состоит из четырех основных шагов:

1) установка конфигурационных параметров, выбранных для исследования алгоритмов обнаружения сетевых аномалий;

2) генерация нормального и аномального тестового сетевого трафика, который должен быть проанализирован и классифицирован выбранными алгоритмами;

3) классификация алгоритмом поступившего сетевого трафика на нормальный трафик и на трафик, содержащий аномалии;

4) оценка качества алгоритмов на основании данных об обнаруженных алгоритмами аномалиях сетевого трафика и ошибках, совершенных в процессе классификации; сравнение алгоритмов по мере Ван Ризенбергена и принятие решения о лучшем по данному показателю алгоритме.

На рис. 2 представлена декомпозиция процесса оценки качества алгоритмов обнаружения сетевых аномалий, описанного на шаге 4.

Для проведения исследований и оценки качества алгоритмов в соответствии с разработанной моделью был разработан программный комплекс (свидетельство о государственной регистрации программы для ЭВМ № 2015615571 от 19.08.2015), подробно описанный авторами в работе [8]. Программный комплекс имеет клиент-серверную архитектуру,

Рис. 1. Модель исследования качества алгоритмов обнаружения сетевых аномалий

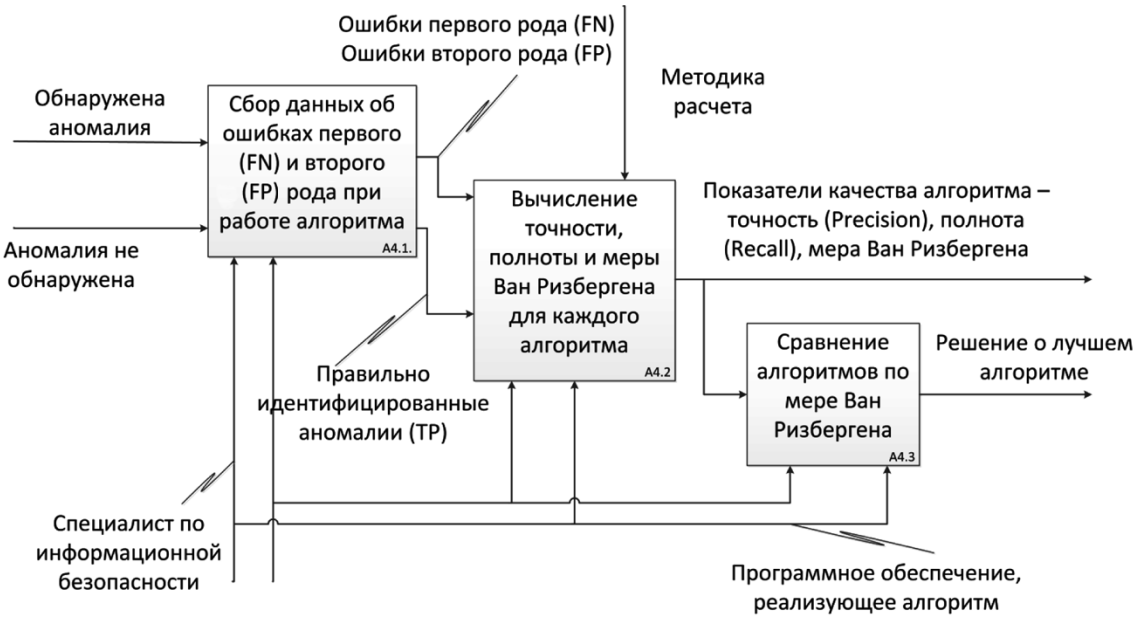

Рис. 2. Декомпозиция процесса оценки качества алгоритмов обнаружения сетевых аномалий

где клиентская часть отвечает за настройку параметров анализа качества алгоритмов и генерацию тестового сетевого трафика с аномальными и нормальными характеристиками. Серверная часть реализует алгоритмы обнаружения сетевых аномалий и собирает информацию о результатах классификации, а также данные, характеризующие алгоритм:

1) ошибки первого рода — *FN*;

2) ошибки второго рода — *FP*;

3) число правильно идентифицированных аномалий — *TP*.

На основании собранных данных программой проводится оценка полноты и точности алгоритма. Затем аналитически вычисляется значение меры Ван Ризбергена для каждого из анализируемых алгоритмов и на основании анализа полученных данных выбирается наилучший.

В качестве объекта исследования авторами были выбраны два наиболее известных и эффективных с вычислительной точки зрения алгоритма обнаружения сетевых аномалий — алгоритм ДВП [7] и алгоритм БД [9].

Обзор алгоритма Бродского—Дарховского и алгоритма на основе дискретного вейвлет-преобразования

Алгоритм обнаружения аномалий БД работает в стандартном режиме анализа сетевого трафика и режиме скользящего окна. При выборе стандартного режима особое влияние оказывают шумы. При выборе алгоритма в режиме скользящего окна совокупное действие помех уменьшается и выбросы, характеризующие начало и конец воздействия, представляются в более явном виде. Для практической реализации лучше использовать алгоритм в режиме скользящего окна. Решение о наличии или отсутствии номинального выброса в трафике принимается в онлайн-режиме, используется скользящее окно $W1$, смещающееся слева направо по мере поступления данных.

Алгоритм ДВП основан на дискретном вейвлет-преобразовании и байесовском анализе. При исследовании сетевого трафика с применением статистических критериев используется техника скользящих окон, двигающихся во времени с определенным шагом, с фиксацией значений трафика в реальном времени, который находится во временных границах каждого окна. Применение данной техники позволяет увеличить надежность обнаружения незначительных аномалий. Конечным критерием обнаружения аномалии является превышение красного порога одним из критериев на этапе реконструкции коэффициентов. Достоинством данного алгоритма является то, что аномалия хорошо обнаруживается на каждом уровне байесовского вейвлет-преобразования декомпозиции. Недостатком данного алгоритма является то, что при начальном уровне разложения он обнаруживает наибольшее число аномалий, но некоторые аномалии могут быть пропущены, если начать разложение с более старших уровней. Более того, на старших уровнях повышается число возникновения ложных тревог, что можно объяснить низкой разрешающей способностью дискретного вейвлет-преобразования во времени. Особенно хорошо этот алгоритм подходит для обнаружения аномалий, связанных с DDos-атаками [7].

Вычисление меры Ван Ризбергена для алгоритма Бродского—Дарховского и алгоритма на основе дискретного вейвлет-преобразования

Входными данными при проведении экспериментов являются размер скользящего окна $W \in [10, 1000]$, число интервалов $I = 2200$ и число поданных аномалий в сгенерированном тестовом сетевом трафике $A \in [1,512]$. При проведении экспериментальных исследований использовали 70 выборок комбинаций входных параметров W и A.

Затем для каждой выборки входных значений N раз был проведен прогон двух анализируемых алгоритмов. В результате была получена информация о среднем числе ошибок первого и второго рода, а также о числе правильно обнаруженных аномалий и рассчитаны по формулам (2) и (3) значения точности и полноты. В табл. 1 представлен пример результирующей оценки частных показателей качества работы алгоритма БД для входного параметра размера окна $W = 10$.

Далее была рассчитана мера Ван Ризбергена [формула (1)] для результирующей оценки двух алгоритмов. Результаты расчета приведены в табл. 2.

Для удобства сравнительного анализа качества работы алгоритмов для каждого размера окна W было рассчитано среднее значение меры Ван Ризбергена (табл. 3).

Проведенные экспериментальные исследования алгоритмов обнаружения сетевых аномалий показали следующее:

- наилучшим качеством классификации по мере Ван Ризбергена на всем выбранном диапазоне значений скользящего окна W обладает алгоритм обнаружения сетевых аномалий БД;
- на краевых значениях диапазона изменения скользящего окна $W \in [10, 1000]$ полученные значения меры Ван Ризбергена алгоритма БД выше значений меры Ван Ризбергена алгоритма ДВП на 30 и 28 % соответственно;
- качество классификации сетевого трафика алгоритмами обнаружения аномалий зависит от размера скользящего окна;
- при увеличении размера скользящего окна W качество классификации обоих алгоритмов снижается, для алгоритма БД значение меры Ван Ризбергена на правой границе диапазона изменения скользящего окна ($W = 1000$) уменьшилось в 5 раз, а для алгоритма ДВП — в 6 раз (рис. 3).

Таким образом, можно выдвинуть гипотезу, что при дальнейших углубленных исследованиях влияния размера скользящего окна на качество классификации будет наблюдаться аналогичная тенденция снижения качества при увеличении размера скользящего окна. Это может быть в первую очередь связано с тем, что аналогично скользящему

Таблица 1

Результирующая оценка частных показателей качества работы алгоритма БД для размера окна $W = 10$ и интервала $I = 2200$

A	1	2	4	8	16	32	64	128	256	512
$TP_{\text{средн}}$	0,4	0,6	2,2	3	6,2	14	28,4	54	100	106
$FN_{\text{средн}}$	3,2	4,4	4,4	7	11,4	18,6	35,4	66	95	112
$FP_{\text{средн}}$	0,6	1,4	1,8	5	9,8	18	35,6	74	156	406
Precision	0,4	0,3	0,55	0,38	0,39	0,44	0,44	0,42	0,39	0,21
Recall	0,11	0,12	0,33	0,3	0,35	0,43	0,45	0,45	0,51	0,49

Таблица 2

Значения меры Ван Ризбергена для алгоритмов ДВП и БД

W	Число аномалий A									
	1	2	4	8	16	32	64	128	256	512
Значение меры Ван Ризбергена алгоритма БД										
10	0,173	0,171	0,415	0,33	0,36	0,43	0,44	0,43	0,44	0,28
15	0,1	0,22	0,32	0,46	0,31	0,38	0,32	0,35	0,32	0,202
20	0,4	0,17	0,1	0,32	0,32	0,31	0,32	0,31	0,24	0,17
25	0,18	0,2	0,48	0,18	0,29	0,32	0,27	0,27	0,21	0,14
30	0,25	0,08	0,139	0,275	0,28	0,22	0,24	0,24	0,21	0,11
500	0,33	0,16	0,09	0,04	0,02	0,012	0,006	0,021	0,004	0,003
1000	0,28	0,16	0,09	0,047	0,02	0,012	0,012	0,018	0,001	0,0007
Значение меры Ван Ризбергена алгоритма ДВП										
10	0,15	0,12	0,15	0,16	0,17	0,16	0,13	0,06	0,01	0,0007
15	0,09	0,02	0,008	0,01	0,002	0,001	0,008	0,005	0,001	0,0007
20	0,73	0,06	0,09	0,08	0,09	0,07	0,05	0,05	0,009	0,0007
25	0,035	0,01	0,006	0,003	0,001	0,001	0,029	0,008	0,001	0,0007
30	0,064	0,06	0,06	0,065	0,05	0,042	0,075	0,035	0,001	0,0007
500	0,003	0,004	0,015	0,02	0,008	0,037	0,019	0,001	0,0015	0,00078
1000	0,004	0,006	0,083	0,041	0,02	0,0123	0,006	0,003	0,0015	0,00078

Таблица 3

Среднее значение меры Ван Ризбергена для исследуемых алгоритмов ДВП и БД

Размер скользящего окна W	Мера Ван Ризбергена	
	Алгоритм ДВП	Алгоритм БД
10	0,114803	0,35098
15	0,016187	0,30158
20	0,061216	0,269707
25	0,010735	0,256468
30	0,046203	0,208267
500	0,011936	0,07118
1000	0,018441	0,066088

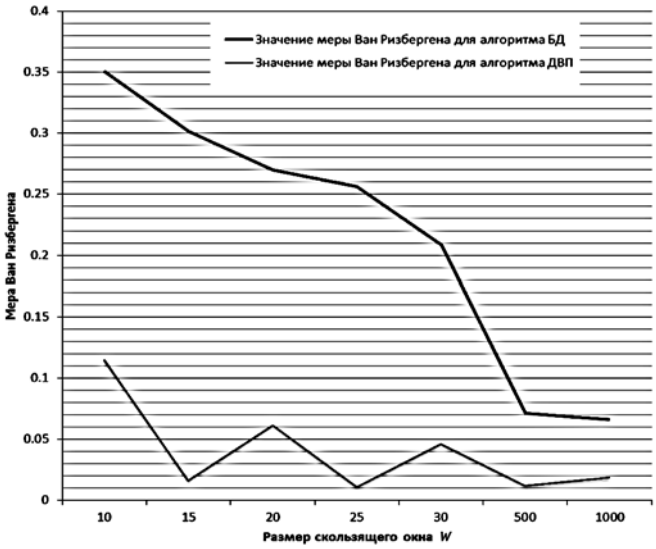

Рис. 3. Зависимость меры Ван Ризбергена от размера окна

окну протокола TCP скользящее окно W алгоритма обнаружения аномалий определяет диапазон подлежащих анализу байтов трафика, которые алгоритм готов принять в настоящее время. Чем больше размер окна, тем больший объем данных должен проанализировать и классифицировать алгоритм. В результате увеличивается неопределенность при классификации, а следовательно, и вероятность ошибок первого и второго рода, что оказывает непосредственное негативное влияние на качество классификации.

Сравнение полученных авторами результатов исследования алгоритмов с результатами, представленными в работах [6—9], показывает следующее.

1. В работе [6] представлен обзор качественных и количественных критериев, которые можно применять при многокритериальном анализе качества алгоритмов обнаружения сетевых аномалий. Однако сама оценка не проводится, поэтому не представляется возможным сравнить результаты данной работы с результатами, представленными в работе [6].

2. В работе [7, с. 114] показано, что при увеличении ширины окна W до 1500 наблюдается увеличение ошибок первого рода, что, в свою очередь, оказывает влияние на такой показатель качества, как полнота классификации, и, опираясь на формулу (2), позволяет сделать вывод о его снижении. А так как полнота участвует в вычислении меры Ван Ризбергена (1), то чем показатель полноты ниже, тем ниже будет и общее качество классификации. Таким образом, можно сделать вывод, что полученные авторами результаты не противоречат результатам исследования, представленным в работе [7].

3. В работе [8] представлено описание архитектуры, алгоритмов функциональных модулей и технических особенностей реализации разработанного авторами программного комплекса, предназначенного для исследования качества алгоритмов обнаружения сетевых аномалий. При этом описание экспериментальных исследований или методики их проведения не приведено, соответственно, по объективным причинам сравнение результатов работ невозможно.

4. В работе [9, с. 116] сделан вывод, что алгоритм разладки Бродского— Дарховского по сравнению с другими алгоритмами анализа сетевого трафика имеет меньшее число ложных тревог (ошибок первого рода) и малое запаздывание в обнаружении изменений сетевого трафика, при этом лучшими показателями обладают реккурентные версии алгоритма. А это соответствует результатам, полученным авторами при сравнении качества классификации двух алгоритмов — БД и ДВП.

Заключение

Была разработана функциональная модель исследования качества алгоритмов обнаружения сетевых аномалий, где базовой метрикой взята гармоническая мера Ван Ризбергена. Данная модель может быть использована для решения задач, связанных с выбором наиболее эффективного алгоритма или средства обнаружения аномалий в сети, что позволило бы с меньшим числом ошибок и ложных срабатываний выявлять признаки возможных атак злоумышленника на сетевые ресурсы организации. Реализован программный комплекс, автоматизирующий предложенную модель, и проведены экспериментальные исследования двух алгоритмов обнаружения сетевых аномалий — алгоритма ДВП и алгоритма БД.

Данные исследования показали, что при сравнении двух алгоритмов по мере Ван Ризбергена лучшим качеством классификации обладает алгоритм обнаружения сетевых аномалий БД. Также анализ результатов экспериментальных исследований позволяет сделать вывод, что для получения лучшего качества классификации необходимо при возможности минимизировать значение такого входного параметра, характерного для обоих алгоритмов, как размер скользящего окна W. А поскольку мера Ван Ризбергена — это метрика, которая гармонически объединяет информацию о точности и полноте анализируемого алгоритма, то при использовании данных алгоритмов для обнаружения

сетевых аномалий необходимо еще в процессе обучения провести калибровку точности и полноты, варьируя значения окна W, так как увеличение данных оценок влияет на повышение качества классификации.

Разработанный программный комплекс может быть применен на практике для оценки качества и выбора наилучшего алгоритма обнаружения аномалий при организации защиты корпоративной сети или в учебном процессе, на лабораторном практикуме, в качестве стенда макета при обучении студентов направления информационная безопасность.

Список литературы

1. **Шелухин О. И., Сакалема Д. Ж., Филинова А. С.** Обнаружение вторжений и компьютерные сети. М.: Горячая линия-Телеком, 2013. 220 с.
2. **Щеглов К. А., Щеглов А. Ю.** Защита от атак на уязвимости приложений // Информационные технологии. 2014. № 9. С. 34—39.
3. **Никишова А. В., Чурилина А. Е.** Обнаружение распределенных атак на информационную систему предприятия // Известия ЮФУ. Технические науки. Тематический выпуск "Информационная безопасность". 2013. № 12 (149). С. 135—143.
4. **Зеленков Ю. Г., Сегалович И. В.** Сравнительный анализ методов обнаружения нечетких дубликатов для Web-документов // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды 9-й Всероссийской научной конференции. Переславль-Залесский, Россия, 2007. С. 84—89.
5. **Никишова А. В., Попов И. С.** Обнаружение сетевых аномалий // Приоритетные технологии: актуальные вопросы теории и практики: Сб. науч. докл. Первого Всерос. конгресса, г. Волгограда, 24—25 апр. 2014 г. — В.: Изд-во ВолГУ, 2014. С. 67—69.
6. **Микова С. Ю., Оладько В. С., Нестеренко М. А., Кузнецов И. А.** Критерии оценки качества алгоритмов обнаружения сетевых аномалий // Международный научно-исследовательский журнал. 2015. № 4 (35). С. 87—88.
7. **Шелухин О. И., Панкрушин А. П.** Оценка достоверности обнаружения аномалий сетевого трафика методами дискретного вейвлет-преобразования // Т-Comm — Телекоммуникации и Транспорт. — 2013. Т. 7, № 10. С. 110—113.
8. **Микова С. Ю., Оладько В. С.** Программный комплекс для исследования качества алгоритмов обнаружения сетевых аномалий // Информационные системы и технологии. 2015. № 5 (91). С. 130—138.
9. **Шелухин О. И., Филинова А. С.** Обнаружение сетевых аномальных выбросов трафика методом разладки Бродского—Дарховского // Т-Comm — Телекоммуникации и Транспорт. 2013. Т. 7, № 10. С. 116—118.

S. Yu. Mikova, student, e-mail: sofya_mikova@mail.ru,
V. S. Oladko, Ph. D., Associate Professor, e-mail: oladko.vs@yandex.ru,
Volgograd State University, Volgograd

Comparison of Algorithms to Identify Network Anomalies Using Measures Van Rizbergena

The problem of selecting algorithms for detecting network anomalies viewed in the article. The aim of this study is to develop an approach to the selection of the best algorithm to detect anomalies in network traffic; this approach will reduce the risk of possible attacks. The authors as a criterion for the selection of the algorithm used the criterion of the quality of the classification of network traffic. Functional model of quality research algorithms to detect network anomalies offered. As a basic measure of quality used Van Rizbergena. Program complex that automates the proposed model developed. The objective of the study is to generate network traffic with different characteristics both normal and abnormal, to assess the quality of existing algorithms for detecting network anomalies. Experimental evaluation of the quality of the two algorithms — Brodsky-Darhovsky algorithm and network anomalies detection algorithm based on discrete wavelet transform performed. These estimate algorithms showed that the best quality classification algorithm has Brodsky — Darhovsky algorithm.

Keywords: *accuracy, completeness, false positive, false negative, Brodsky—Darhovsky algorithm, discrete wavelet transform, attack*

References

1. **Shelukhin O. I., Sakalema D. Zh., Filinova A. S.** *Obnaruzheniye vtorzheniy i komp'yuternyye seti,* Moscow, Goryachaya liniya-Telekom, 2013, 220 p.
2. **Shcheglov K. A., Shcheglov A. Yu.** Zashchita ot atak na uyazvimosti prilozheniy, *Informatsionnyye tekhnologii*, 2014, no. 9, pp. 34—39.
3. **Nikishova A. V., Churilina A. Ye.** Obnaruzheniye raspredelennykh atak na informatsionnuyu sistemu predpriyatiya, *Izvestiya YUFU, Tekhnicheskiye nauki, Tematicheskiy vypusk "Informatsionnaya bezopasnost'"*, 2013, no. 12 (149), pp. 135—143.
4. **Zelenkov Yu. G., Segalovich I. V.** Sravnitel'nyy analiz metodov obnamzheniya nechetkikh dublikatov dlya Web-dokumentov, *Elektronnyye biblioteki: perspektivnyye metody i tekhnologii, elektronnyye kollektsii: trudy 9-oy Vserossiyskoy nauchnoy konferentsii.* — Pereslavl' — Zalesskiy, Rossiya, 2007, pp. 84—89.
5. **Nikishova A. V., Popov I. S.** Obnaruzheniye setevykh anomaliy, *Prioritetnyye tekhnologii: aktual'nyye voprosy teorii i praktiki: sb. nauch. dokl. Pervogo Vseros. kongressa*, g. Volgograda, 24—25 apr. 2014 g. Volgograd, Izd-vo VolGU, 2014, pp. 67—69.
6. **Mikova S. Yu., Olad'ko V. S., Nesterenko M. A., Kuznetsov I. A.** Kriterii otsenki kachestva algoritmov obnaruzheniya setevykh anomaliy, *Mezhdunarodnyy nauchno-issledovatel'skiy zhurnal*, 2015, no. 4 (35), pp. 87—88.
7. **Shelukhin O. I., Pankrushin A. P.** Otsenka dostovernosti obnaruzheniya anomaliy setevogo trafika metodami diskretnogo veyvlet-preobrazovaniya, *T-Comm — Telekommunikatsii i Transport,* 2013, vol. 7, no. 10, pp. 110—113.
8. **Mikova S. Yu., Oladko V. S.** Programmnyy kompleks dlya issledovaniya kachestva algoritmov obnaruzheniya setevykh anomaliy, *Informatsionnyye sistemy i tekhnologii*, 2015, no. 4 (90), sentyabr'—oktyabr' 2015, pp. 130—138.
9. **Shelukhin O. I., Filinova A. S.** Obnaruzheniye setevykh anomal'nykh vybrosov trafika metodom razladki Brodskogo—Darkhovskogo, *T-Comm — Telekommunikatsii i Transport*, 2013, vol. 7, no. 10, pp. 116—118.