ISSN 1992-6502 (Print)
2016. Т. 20, № 1 (71). С. 168–174
УДК 004.056

Вестник УГАТУ

ISSN 2225-2789 (Online)
http://journal.ugatu.ac.ru

ГИБРИДНЫЙ АЛГОРИТМ ОБНАРУЖЕНИЯ СЕТЕВЫХ АНОМАЛИЙ НА ОСНОВЕ СИСТЕМЫ ГОЛОСОВАНИЯ

С. Ю. МИКОВА[1], В. С. ОЛАДЬКО[2], А. А. МЕЛКИХ[3]

[1]sofya_mikova@mail.ru, [2]oladko.vs@yandex.ru, [3]melkih.a.a@gmail.com

[1,2]ФГАОУ ВПО «Волгоградский государственный университет» (ВолГУ)
[3]ФГБОУ ВПО «Московский государственный технический университет им. Н. Э. Баумана» (МГТУ им. Н. Э. Баумана)

Поступила в редакцию 12.02.2016

Аннотация. Рассмотрена актуальная проблема обнаружения сетевых аномалий. Проанализированы с помощью оценочных критериев наиболее распространенные алгоритмы обнаружения сетевых аномалий. С целью повышения качества обнаружения аномалий предложено объединить и модифицировать существующие алгоритмы в рамках одной системы. Разработана и описана гибридная процедура обнаружения сетевой аномалии и предложен алгоритм совместного принятия решения об обнаруженной аномалии, в основе которого лежит система голосования.

Ключевые слова: информационная безопасность, сетевая атака, сетевая аномалия, алгоритм обнаружения сетевой аномалии, сетевой трафик.

ВВЕДЕНИЕ

В настоящее время Интернет является местом высокой деловой активности. Значительное число предприятий и организаций по всему миру использует компьютерные системы для управления производственными процессами и персоналом, для распределения ресурсов и подключения удалённых пользователей. Однако рост информационных технологий вызвал стремительный рост компьютерной преступности, в частности возросло число угроз, связанных с несанкционированным доступом (НСД) внешнего злоумышленника из глобальной сети во внутреннюю сеть предприятий. На долю подобных сетевых атак приходится, по данным [1], 44 % нарушений информационной безопасности (ИБ). Поэтому в настоящее время особо актуальны проблемы, касающиеся обнаружения сетевых атак и аномалий сетевого трафика с целью предупреждения дальнейшего вторжения и снижения рисков от подобных нарушений. Анализ [2, 3] показывает, что проблемы с безопасностью могут возникать не только в результате умышленных угроз, атак злоумышленника, но и в результате отказов программно-аппаратного обеспечения, ошибок пользователей, нарушения функционирования обеспечивающих подсистем и обслуживающей инфраструктуры предприятия.

Данная работа посвящена актуальной проблеме разработки алгоритмов обнаружения и выявления сетевых аномалий, так как именно сетевые аномалии являются одним из основных признаков сбоев в работе сети и/или вторжения злоумышленника.

АНАЛИЗ АЛГОРИТМОВ ОБНАРУЖЕНИЯ СЕТЕВЫХ АНОМАЛИЙ

Анализ литературных источников [4–6] показал, что основными алгоритмами обнаружения сетевых аномалий являются:

- алгоритм на основе дискретного вейвлет-преобразования. Применяется метод скользящих окон W1 и W2, которые перемещаются по оси времени, фиксируя значения трафика в режиме реального времени;
- алгоритм Бродского – Дарховского. Рассматривает «выброс» трафика (аномалии) методом разладки. Для работы алгоритм использует метод проверки совпадения или различия среднего значения в двух движущихся окнах;
- алгоритм на основе суммы квадратов вейвлет-коэффициентов. Это алгоритм

выявления аномалий с использованием вейвлета Хаара или Добеши;
- алгоритм на основе максимума квадратов вейвлет-коэффициентов, производный от предыдущего, но менее эффективный;
- алгоритм на основе критерия согласия Колмогорова – Смирнова. Предназначен для проверки гипотез о принадлежности выборки некоторому заданному закону распределения.

Рассмотренные алгоритмы были проанализированы по оценочным критериям [7, 8], приведенным в таблице.

Таблица

Критерии оценки выявления аномалий

Критерий	Метод определения критерия
Ошибки первого рода	Количество ложных тревог
Ошибки второго рода	Количество пропусков событий, уведомляющих о возникновении аномалии
Размер окна	Интервал времени, за который выявляется наибольшее количество аномалий в системе
Сложность алгоритма	Чем большее время и объем памяти требуются для реализации алгоритма, тем больше его сложность
Точность	Отношение правильно обнаруженных алгоритмом аномалий к сумме правильно обнаруженных алгоритмом аномалий и ошибок второго рода
Полнота	Отношение правильно обнаруженных алгоритмом аномалий к сумме правильно обнаруженных аномалий алгоритмом и ошибок первого рода

В результате анализа, подробно описанного авторами в [9, 10], определено, что в каждом из рассмотренных алгоритмов присутствует вероятность возникновения ошибок первого рода, подразумевающих ложные срабатывания, из-за которых не удаётся полностью автоматизировать борьбу со многими видами угроз. Также в результате анализа литературных источников были выделены следующие алгоритмы: алгоритм на основе дискретного вейвлет-преобразования с применением статистических критериев (ДВП), критерия согласия Колмогорова – Смирнова (КС) и алгоритм обнаружения аномалий Бродского – Дарховского (БД). Помимо этого, выбранные алгоритмы просты в программной реализации. Для оценки эффективности выбранных алгоритмов обнаружения сетевых аномалий авторами был создан программный комплекс, реализующий их экспериментальное исследование при различных значениях параметров алгоритмов (размер окна) и сетевого трафика (плотность аномалий). Также программный комплекс позволяет оценить такие показатели работы алгоритмов, как ошибки первого и второго рода, точность и полноту классификации сетевых аномалий каждым алгоритмом, количество правильно найденных аномалий. При оценке алгоритмов моделировались аномалии трафика, заключающиеся в высокой плотности сетевых пакетов между устройствами. Эксперимент проводился при значениях размера окна в интервале [10;100] и плотности аномалий в интервале [0; 0.15].

Ошибки второго рода, подразумевающие пропуск опасных событий, свойственны алгоритму на основе дискретного вейвлет-преобразования и алгоритму Бродского – Дарховского. Но наиболее точным в обнаружении аномалий является алгоритм Бродского – Дарховского. Также при его использовании обнаруживается меньше ошибок 1-го и 2-го рода, чем при использовании алгоритма на основе дискретного вейвлет-преобразования с применением статистических критериев.

Таким образом, одной из основных проблем при реализации и использовании алгоритмов обнаружения сетевых аномалий является наличие ошибок первого и второго рода, которые возникают в процессе классификации характеристик трафика. Данные ошибки оказывают негативное влияние на такие показатели качества, как полнота и точность результата обнаружения, и приводят либо к большому числу ложных срабатываний систем обнаружения, построенных на базе данных алгоритмов, либо к большому числу пропусков. А это, в свою очередь, может привести к своевременно нераспознанной атаке злоумышленника и увеличению рисков ИБ. Поэтому для решения данной проблемы, в рамках проводимого исследования предлагается создать гибридный алгоритм обнаружения сетевых аномалий, в основе которого лежат три наиболее эффективных статистических метода обнаружения аномалий и система совместного принятия решений, построенная на базе голосования. При этом голос каждого метода будет являться функцией от показателя качества классификации данного метода, т.е. чем будет выше показатель качества классификации ме-

тода, тем больший вес будет иметь его голос при совместном принятии решения о наличии или отсутствии аномалии в сетевом трафике.

РАЗРАБОТКА ГИБРИДНОГО АЛГОРИТМА ОБНАРУЖЕНИЯ СЕТЕВЫХ АНОМАЛИЙ

Для реализации гибридного алгоритма обнаружения аномалий сетевого трафика – Anomalies Security (AS) были отобраны три вида статистических метода обнаружения аномалий: метод БД (X), метод ДВП (Y), метод на основе критерия согласия Колмогорова – Смирнова (Z). Данные методы будут параллельно анализировать сетевой трафик на предмет аномалий и совместно, путем взвешенного голосования, принимать решения о наличии или отсутствии аномалии. Целью предложенной модификации является повышение точности и полноты классификации путем минимизации ошибок первого и второго рода. Минимизировать ошибки можно за счет совместного принятия решений о возможной аномалии с учетом знаний о текущем качестве классификации каждого из использованных методов. В качестве показателя качества классификации аномалий в работе будет использована F-мера (формула (1)), которая является функцией от точности (формула (2)) и полноты (формула (3)) классификации.

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

$$Precision = \frac{TP}{TP + FP}, \qquad (2)$$

$$Recall = \frac{TP}{TP + FN}, \qquad (3)$$

где ошибки второго рода – FP, количество правильно идентифицированных аномалий – TP, ошибки первого рода – FN.

Таким образом, алгоритм AS можно описать в виде следующей последовательности шагов, которые представлены в виде блок-схемы на рис. 1.

Блок №1: Начало работы алгоритма.

Блок №2: Происходит прием, обработка трафика со стороны клиента (источника трафика) для последующей классификации.

Блок №3-5: Производится вызов каждого из трех методов (X, Y, Z), которые анализируют трафик на предмет аномалий. Если по результатам классификации, по мнению алгоритма X, аномалия есть, то x=+1. В противном случае x=-1. Действия повторяются для методов Y и Z. В результате каждый метод голосует за наличие/отсутствие аномалии x = ±1, y = ±1, z = ±1.

Блок №6. Происходит поиск ближайшего трафика в файлах тестовой выборки. Так как при различных значениях плотности аномалий и размеров окон статистические методы отличаются эффективностью, то нужно обеспечить различные веса голосов каждого метода при голосовании. Режим формирования тестовой выборки предназначен для того, чтобы выяснить компетентность каждого метода при различных параметрах трафика и назначить соответствующий вес голоса. Вес голоса вычисляется алгоритмом AS исходя из данных файлов тестовой выборки, которые определяются следующим образом: формируются различные параметры плотности аномалий и размеров окон, исходя из которых, создается модель, отправляется на сервер, а затем ответ сервера анализируется клиентом. Результаты анализа, а именно: верное нахождение аномалии, ошибки первого и второго рода, фиксируются в файле для каждого метода. Эти данные используются далее для работы алгоритма голосования на стороне сервера. Таким образом, режим формирования файлов тестовой выборки предназначен для более эффективной работы алгоритма AS.

Блок №7. Для установки веса голоса при принятии общего решения о наличии или отсутствии в проанализированном трафике аномалии для каждого метода X, Y, Z вычисляется показатель качества классификации на основе F-меры. Вычисление F-меры осуществляется по формулам (1)–(3) для каждого метода X, Y, Z.

Блок №8. В том случае, если все методы проголосовали за отсутствие аномалии (x=-1, y=-1, z=-1), то голосование не требуется и принимается первичное совместное решение о результатах классификации – аномалия в сетевом трафике отсутствует (блок №9). Если хотя бы один метод обнаружил аномалию, то требуется голосование при совместном принятии решений и осуществляется переход к блоку 10.

Блок №10. Запуск процедуры голосования S о наличии или отсутствии аномалии. Результат голосования рассчитывается по формуле

$$S = xF(X) + yF(Y) + zF(Z), \qquad (4)$$

где F(X), F(Y), F(Z) – F-мера методов БД, ДВП и КС соответственно.

Блоки № 11–13. Принятие совместного решения на основе результатов проведенного голосования. Если S≥0, то результатом голосования является то, что сетевая аномалия

есть. Если S < 0, то результатом голосования является то, что сетевой аномалии нет, см. формулу

$$AS = \begin{cases} \text{аномалия обнаружена, если } S \geq 0, \\ \text{аномалия не обнаружена, если } S < 0. \end{cases} \quad (5)$$

Блок № 14-16. Если пакет, подлежащий анализу алгоритма, последний, то производится завершение работы алгоритма. В противном случае блоки №8-13 повторяют свои действия.

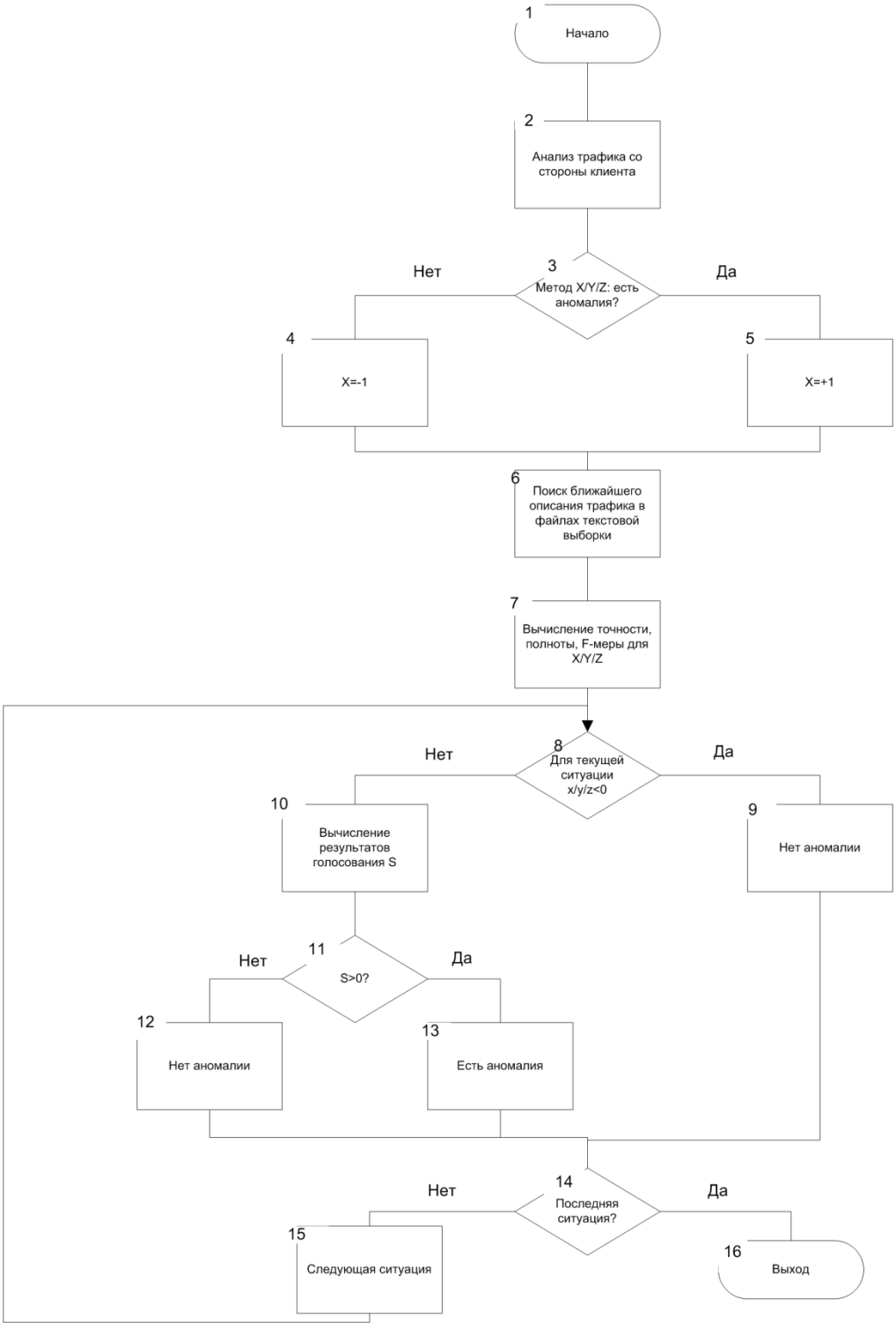

Рис. 1. Блок-схема алгоритма AS

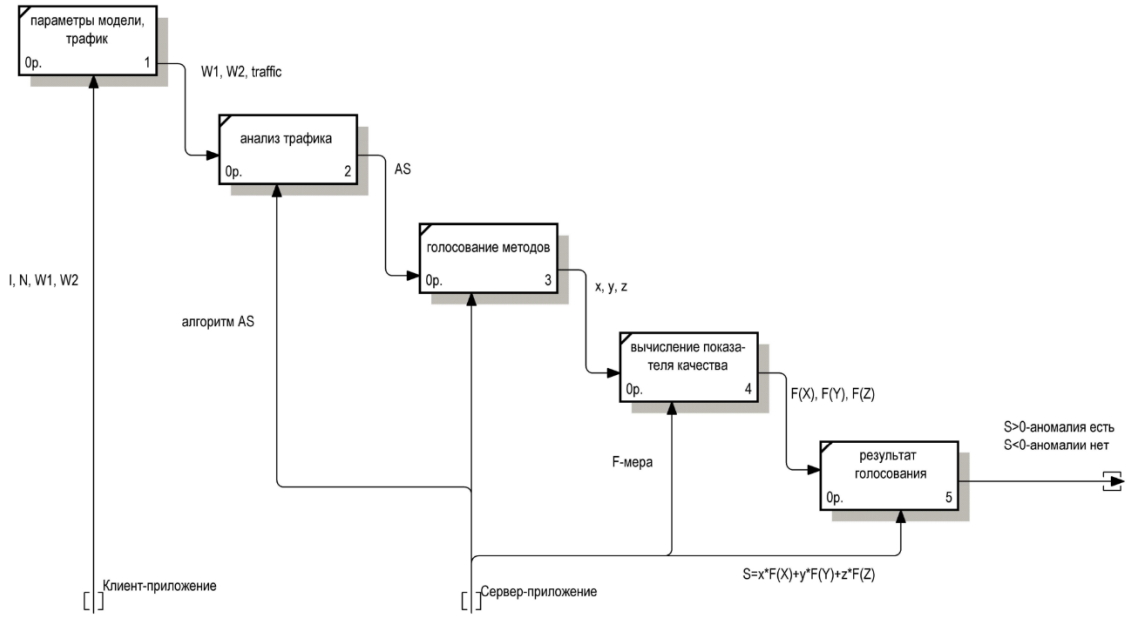

Рис. 2. Процедура обнаружения аномалий с использованием гибридного алгоритма AS

ПРОЦЕДУРА ОБНАРУЖЕНИЯ СЕТЕВЫХ АНОМАЛИЙ С ПОМОЩЬЮ РАЗРАБОТАННОГО АЛГОРИТМА

Процедура мониторинга и обнаружения аномалий в сетевом трафике с использованием разработанного гибридного алгоритма может быть описана посредством 4 основных шагов, представленных в виде функциональной диаграммы в нотации IDEF0 (рис. 2).

Шаг 1. На сервер, предназначенный для обнаружения аномалий от клиента (источника трафика), приходят параметры методов (размер скользящего окна 1 (W1), размер скользящего окна 2 (W2)) и модель исследуемого трафика (traffic).

Шаг 2. Обработка, анализ и классификация полученного сетевого трафика методами БД, ДВП, КС на предмет наличия аномалий.

Шаг 3. Голосование каждого метода о наличии или отсутствии в проанализированном сетевом трафике аномалии.

Шаг 4. Вычисление показателей качества (формулы 1-3) классификации для каждого из трех методов с целью установки значений веса, характеризующего значимость голоса алгоритма при совместном принятии решения об аномалии. Значения показателей находятся с помощью проведения предварительного тестового прогона работы каждого метода голосования (БД, ДВП, КС). Тестовый прогон осуществляется следующим образом:

- Пользователем фиксируется количество интервалов (I), минимальное и максимальное значение окна W2, шаг изменения окна W2, коэффициенты K и B для нахождения значения окна *W1*, максимальная степень (*n*) для подсчета плотности (*P*).
- При фиксированном количестве интервалов изменяется размер окон и количество аномалий (a), где $a=2^n$.
- Вычисляется плотность аномалии $P=a/I$.
- Клиентское приложение производит генерацию модели по параметрам I, a. Сгенерированная модель передается на сервер.
- Серверное приложение обрабатывает принятую им модель выбранным методом обнаружения аномалий, находит значение параметров *TP, FN, FP* и отправляет результаты клиенту.
- Клиент анализирует полученные результаты и записывает их в файл.

Перечисленные действия повторяются для каждого метода (БД, ДВП, КС). После того, как файлы тестового прогона для каждого статистического метода сформированы на стороне клиента, они передаются серверу для дальнейшей работы в качестве параметров. Если сервер обрабатывает пришедший трафик алгоритмом голосования, он обращается к данным файлов, полученным при тестовом прогоне, и извлекает параметры TP, FN, FP для вычисления точности и полноты. Когда значения оценки точности и

полноты известны, для каждого алгоритма рассчитывается значение F-меры.

Шаг 5. Анализ результатов голосования и принятие совместного решения о наличии или отсутствии аномалии.

Таким образом, в результате проведения вышеизложенной процедуры, будет получено количество обнаруженных сетевых аномалий, вычисленное с учетом компетентностей алгоритмов.

ЗАКЛЮЧЕНИЕ

Описанный выше алгоритм обнаружения сетевых аномалий AS был реализован в виде программного комплекса. В настоящее время, в соответствии с разработанной процедурой, проводятся экспериментальные исследования для того, чтобы оценить возможности применения разработанного программного комплекса для обнаружения аномалий сетевого трафика и определить рациональные параметры, при которых разработанный программный комплекс предоставляет лучшие результаты.

Предварительно было проведено исследование по сравнительному анализу алгоритма AS с алгоритмом БД. При значении окна на интервале [25;65] алгоритм AS в среднем показал лучшее значение на 4 %, при значении размера окна 30 – на 17 %. Оценка производилась по значению F-меры алгоритма обнаружения.

Кроме того, планируется провести сравнительный анализ точности полноты и качества существующих алгоритмов обнаружения аномалий и разработанной модификации с целью оценки эффективности предложенного решения.

СПИСОК ЛИТЕРАТУРЫ

1. **Positive Research 2015**. Сборник исследований по практической безопасности // Positive Technologies. – 2015. [Электронный ресурс]. URL: http://security.ru/download/PT_Positive_Research_2015_RU_web.pdf (дата обращения 05.08.2015) [*Positive Research 2015. Collection of Practical Security Research* [Online]. Available: http://security.ru/download/PT_Positive_Research_2015_RU_web.pdf]

2. **Багров Е. В**. Мониторинг и аудит информационной безопасности на предприятии // Вестник Волгоградского государственного университета. Серия 10: Инновационная деятельность. 2011. №5. С. 54–56. [E.V. Bagrov, "*Monitoring and auditing of information security in the enterprise",* (in Russian) in *Vestnik* VolGU. Series 10. Innovation activities, no.5, pp.54-56, 2011.]

3. **Аткина В. С.** Использование программного комплекса для исследования катастрофоустойчивости информационных систем // Вестник Волгоградского государственного университета. Серия 10. Инновационная деятельность. 2011. №5. С. 14–17. [V. S. Atkina, "*Using a software package for the study of information systems disaster recovery",* (in Russian), in *Vestnik* VolGU. Series 10. Innovation activities, no.5, pp. 14–17, 2011.]

4. **Шелухин О. И., Иванов Ю. А., Ригов В. Ю.** Обнаружение DOS и DDOS- атак методом дискретного вейвлет-анализа // T-Comm-Телекоммуникации и Транспорт. 2011. №1. С. 44–46. [O. I. Shelukhin, Y. A. Ivanov, V. Y. Rogov, "*Detection of DOS and DDOS- attacks by discrete wavelet analysis",* (in Russian), in T-Comm-Telekommunikatsii i Transport, no.1, pp.44–46, 2011]

5. **Шелухин О. И., Филинова А. С.** Обнаружение сетевых аномальных выбросов трафика методом разладки Бродского-Дарховского // T-Comm - Телекоммуникации и Транспорт. 2013. №10. Т. 7. С. 116–118. [O. I. Shelukhin, A. S. Filinova, "*Detection of abnormal network traffic emissions by discord Brodsky-Darhovsky",* (in Russian), in T-Comm-Telekommunikatsii i Transport. Vol. 4, no.10, pp. 116–118, 2013]

6. **Вострецов А. Г., Богданович В. А., Гундарева М. В.** Применение критериев согласия для решения задач обнаружения сигналов неизвестной формы // Доклады Томского государственного университета систем управления и радиоэлектроники. 2012. С. 19–23 [A. G. Vostretsov, V. A.Bogdanovich, M. V. Goncharova, "*Application criteria for consent to solve problems of detection signals of unknown shape",* (in Russian), in Doklady Tomskogo gosudarstvennogo universiteta sistem upravleniya i radioelektroniki, pp.19–23, 2012]

7. **Шелухин О. И., Панкрушин А. П.** Оценка достоверности обнаружения аномалий сетевого трафика методами дискретного вейвлет-преобразования // T-Comm-Телекоммуникации и Транспорт. 2013. №10. Т. 7. С. 110–113. [O. I. Shelukhin, A. P. Pankrukhin, "*Evaluation of reliability of anomaly detection network traffic methods discrete wavelet transform",* (in Russian), in T-Comm-Telekommunikatsii i Transport. Vol. 7, no.10, pp. 110-113, 2013]

8. **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. [S. Y. Mikova, V. S. Oladko, M. A. Nesterenko, I. A. Kuznetsov, "*Criteria for assessing the quality of network anomaly detection algorithms",* (in Russian), in Mezhdunarodnyy nauchno-issledovatel'skiy zhurnal, no.4(35), pp.87–88, 2015].

9. **Микова С. Ю., Оладько В. С.** Результат исследования алгоритмов выявления сетевых аномалий // Вопросы кибербезопасности. 2015. №4(12). С.38–41. [S. Y. Mikova, V. S. Oladko, "*The result of the study of algorithms to identify network anomalies",* (in Russian), in Voprosy kiberbezopasnosti, no.4 (12), pp.38–41, 2015]

10. **Микова С. Ю., Оладько В.С.** Оценка качества алгоритма обнаружения сетевых аномалий на основе дискретного вейвлет преобразования с помощью F-меры // Вестник УРФО: Безопасность в информационной сфере.2015. №2(16). С.36–40. [S. Y. Mikova, V. S. Oladko, "*Evaluation of the quality of network anomaly detection algorithm based on discrete wavelet transform using the F-measure",* (in Russian), in Vestnik URFO: Bezopasnost' v informatsionnoy sfere, no.2(16), pp.36–40, 2015.]

ОБ АВТОРАХ

МИКОВА Софья Юрьевна, студ. каф. информационной безопасности. Выполняет НИР (курс. работу) по теме «Разработка алгоритма обнаружения сетевых аномалий».

ОЛАДЬКО Владлена Сергеевна, доц. каф. информационной безопасности**.** Дипл. специалист по защите информации (ВолГУ, 2009). Канд. техн. наук по методам и средствам защиты информации, инф. безопасности (ЮФУ, 2013). Иссл. в обл. анализа катастрофоустойчивости информационных систем.

МЕЛКИХ Александр Алексеевич, студ. каф. компьютерных систем автоматизации производств**.** Выполняет НИР по теме «Исследование систем обеспечения информационной безопасности АСУ ТП».

METADATA

Title: Hybrid algorithm for detecting network anomalies based voting system.

Authors: S. Y. Mikova[1], V. S. Oladko[2], A. A. Melkikh[3]

Affiliation:

[1,2] Volgograd State University (VolGU)

[3] Bauman Moscow State Technical University (BMSTU), Russia.

Email: [1]sofya_mikova@mail.ru, [2]oladko.vs@yandex.ru, [3]melkih.a.a@gmail.com.

Language: Russian.

Source: UGATU (scientific journal of Ufa State Aviation Technical University), vol. 20, no. 1 (71), pp.168-174, 2016. ISSN 2225-2789 (Online), ISSN 1992-6502 (Print).

Abstract: The actual problem of detection of network anomalies. Analyzed by using the evaluation criteria of existing algorithms of detection of network anomalies. Created a hybrid procedure of detection of network anomaly. Developed and described a hybrid algorithm for detecting network anomalies based voting system.

Key words: information security, network attack, network anomaly, algorithm for detecting network anomalies, network traffic.

About authors:

MIKOVA, Sofya Yur'evna, student of the department of information security. Performs NIR (course. work) on the theme "Development of algorithm of detection network anomaly."

OLADKO, Vladlena Sergeevna, associate professor of the department of information security. Dipl. Data Protection Specialist (Volgograd, 2009). Kan. t-tehn. Sciences on ways and means of information protection, inf. security (SFU 2013). Inst. in the region. Analysis disaster recovery of information systems.

MELKIKH, Alexander Alekseyevich, student of the department "Computer manufacturing automation systems". Performs research work on "Research of information security systems ICS".