УДК 004.056 doi: 10.26102/2310-6018/2019.24.1.011

В.И. Васильев, Р.Р. Шамсутдинов

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА АНАЛИЗА ИНЦИДЕНТОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ (НА ОСНОВЕ МЕТОДОЛОГИИ SIEM-СИСТЕМ С ПРИМЕНЕНИЕМ МЕХАНИЗМОВ ИММУНОКОМПЬЮТИНГА)

ФБГОУ ВО «Уфимский государственный авиационный технический университет», Уфа, Россия

Статья посвящена проблеме интеллектуального анализа инцидентов информационной безопасности с применением методологии, используемой в системах управления информационной безопасностью и событиями безопасности. Проанализирована сущность таких систем, состав основных модулей и порядок их взаимодействия, возможность интеграции с методами искусственного интеллекта. Описана разработанная распределенная система анализа инцидентов информационной безопасности, синтезирующая механизмы искусственной иммунной системы и корреляционного анализа данных для выявления известных и неизвестных аномалий, анализа их критичности и определения приоритетов в реагировании. Представлена схема взаимодействия модулей разработанной системы, математическая составляющая применяемого метода корреляционного анализа данных. Подробно описана серия проведенных вычислительных экспериментов, показавших высокий уровень эффективности системы в обнаружении аномалий и возможности дополнительного обучения друг друга клиентскими модулями, а также успешное выполнение серверной компонентой агрегации и корреляционного анализа данных, поступающих от клиентов, в заданном интервале времени, выделении наиболее существенных инцидентов за последний проанализированный интервал, а также за все время, как в комплексе, так и для каждой группы инцидентов. Графическое отображение сервером статистических данных позволяет наглядно оценить критичность тех или иных инцидентов и определить приоритеты в реагировании на них.

Ключевые слова: SIEM-система, иммунокомпьютинг, корреляционный анализ, информационная безопасность, безопасность сети.

Введение

Роль информационных технологий (ИТ) в настоящее время невозможно переоценить. Широкое применение ИТ в государственных органах, а также в огромном числе организаций требует обеспечения безопасности в информационной сфере с целью защиты процессов, реализуемых ИТ, а также обеспечения максимальной защищенности конфиденциальности, целостности и доступности сведений, независимо от формы их представления от внешних и внутренних угроз [1, 2].

С целью защиты информации (ЗИ) в области ИТ применяются различные программно-аппаратные комплексы и системы, такие как межсетевые экраны; средства обнаружения и предотвращения вторжений;

средства контроля съемных носителей; средства контроля и обеспечения целостности; хостовые средства защиты информации от несанкционированного доступа; средства идентификации, аутентификации, контроля и управления доступом; средства антивирусной защиты; защиты от утечек информации и др.

Однако каждое средство защиты обеспечивает анализ и выявление (блокировку) неправомерных действий только в контекстах своих анализируемых данных [3-5]. SIEM-системы же анализируют данные всех средств защиты в комплексе, сопоставляют их между собой и выявляют угрозы, не обнаруживаемые более узкоспециализированными средствами.

Жесткие правила корреляции SIEM-систем позволяют хорошо обнаруживать известные угрозы, однако неспособны выявлять угрозы, похожие на известные или неизвестные вовсе [6]. Спроектированная система способна обнаруживать и их.

Методология построения интеллектуальной системы

На первый взгляд работа SIEM-систем довольно проста: система собирает информацию, анализирует ее «на лету» (и генерирует предупреждающее сообщение), сохраняет в базы данных, анализирует поведение на основании предыдущих наблюдений. На практике схема реализуется с помощью соответствующих компонентов: агенты (сбор данных из различных источников); серверы-коллекторы (аккумуляция информации, поступившей от агентов); сервер баз данных (хранение информации); сервер корреляции (анализ информации).

Порядок работы SIEM системы:

1) агрегация;
2) нормализация;
3) корреляция;
4) протоколирование;
5) оповещение.

Среди множества существующих методов корреляции можно выделить две большие группы – сигнатурные и бессигнатурные. Сигнатурный анализ осуществляется на основе некоторых определяемых экспертами правил, а бессигнатурные методы действуют согласно обучающему набору данных [7].

Среди основных методов корреляции, используемых в SIEM-системах, можно выделить:

– статистический;
– на основе правил;
– на основе матрицы;
– метод моделирования;

– на основе графа зависимости;
– байесовский метод;
– на основе нейронных сетей [7-9].

Нейросетевой метод основывается на принятии решений обученной искусственной нейронной сетью, имитирующей работу человеческого мозга, способной классифицировать данные, подобные тем, на которых она обучена.

Аналогичным образом возможно применение искусственных иммунных систем в основе корреляционного анализа данных, однако было принято решение о двухэтапной процедуре анализа.

Агенты разрабатываемой системы, каждый из которых представляет собой доработанный хостовой модуль обнаружения атак, основанный на искусственной иммунной системе (ИИС), разработанный нами ранее и представленный в [10, 11], будут не просто пересылать логи на сервер, а осуществлять их первичный анализ на основе иммуннокомпьютинга.

Корреляционный анализ разрабатываемой системы осуществляется по предложенному в [12] методу. Согласно которого анализ выполняется в два этапа. Первый – вычисление частных коэффициентов корреляции между типами событий. Второй – вычисление общих коэффициентов корреляции между типами событий.

Вычисление частных коэффициентов корреляции осуществляется на двух показателях. Первый из которых – временная задержка между двумя событиями в секундах – dT. Второй – относительный вес прямых связей между событиями, определяемый по формуле [12]:

$$w_i = \frac{N_i^{iden}}{N_i^{direct}}, \quad (1)$$

где N_i^{iden} – число идентичных значений между двумя анализируемыми строками, N_i^{direct} – размерность строки.

Частные коэффициенты корреляции между двумя событиями рассчитываются по формуле [12]:

$$r_{priv} = \frac{n\sum_{i=1}^{n-1} w_i dT_i - \sum_{i=1}^{n-1} w_i \times \sum_{i=1}^{n-1} dT_i}{\sqrt{n\sum_{i=1}^{n-1}(w_i)^2 - \left(\sum_{i=1}^{n-1} w_i\right)^2} \times \sqrt{n\sum_{i=1}^{n-1}(dT_i)^2 - \left(\sum_{i=1}^{n-1} dT_i\right)^2}}, \quad (2)$$

где n – число произошедших ранее событий

Общий коэффициент корреляции по типу инцидента рассчитывается как среднее арифметическое частных коэффициентов корреляции. Пороговое значение общего коэффициента корреляции по группе инцидентов для определения серьезности инцидента задается администратором на основе экспертных знаний.

Разрабатываемая интеллектуальная система должна включать все подсистемы классической SIEM. Схема взаимодействия модулей разработанной системы представлена на Рисунке 1. Сервер осуществляет непрерывную агрегацию данных, поступающих от агентов, параллельно выполняя их корреляционный анализ, агенты осуществляют непрерывный перехват данных о сетевом трафике, их нормализацию, заключающуюся в выделении выбранных параметров сетевого трафика и их сведении в строку размером в 13 байт. Затем агенты осуществляют первичный анализ данных с помощью, интегрированной в агенты ИИС.

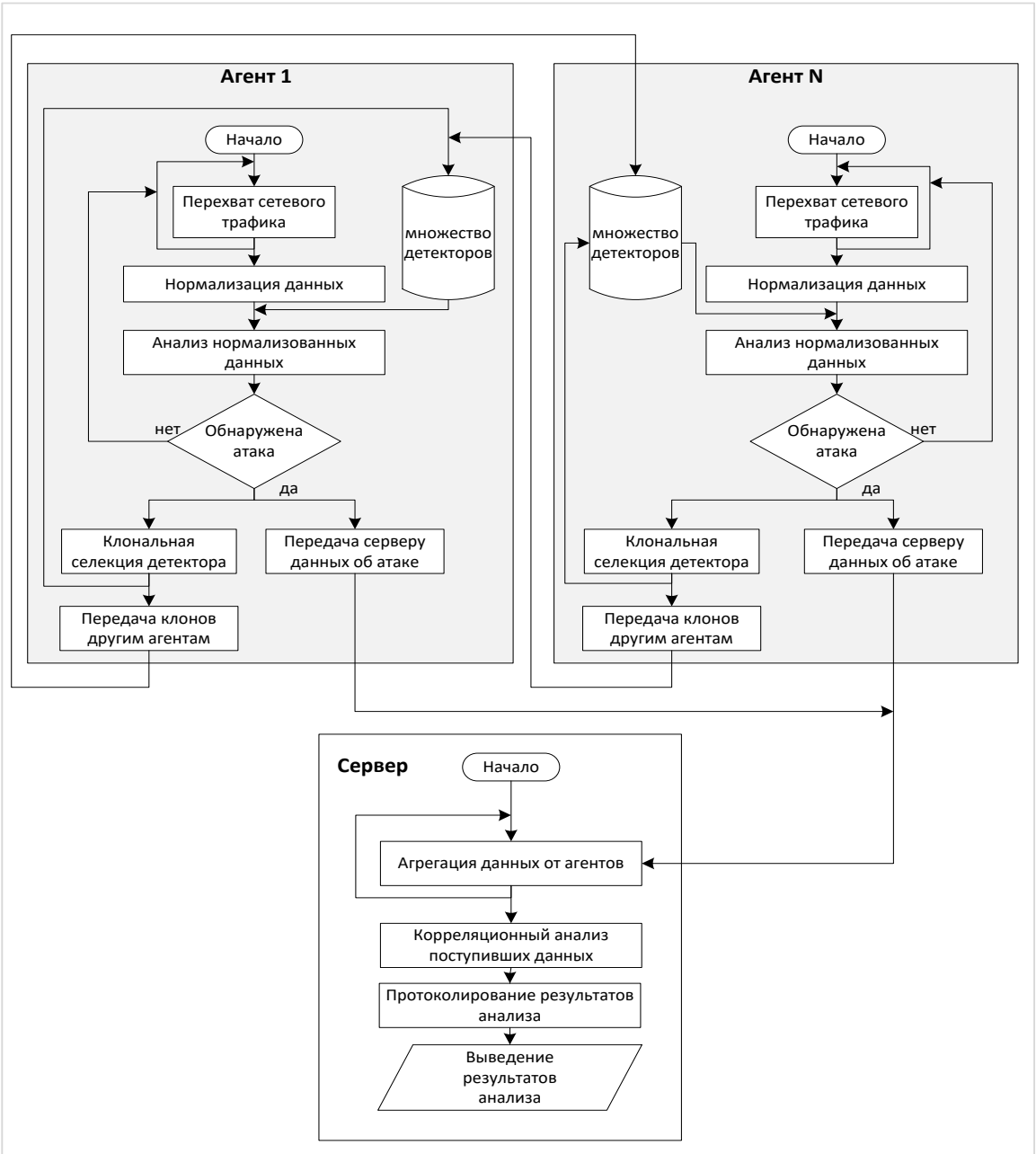

Рисунок 1 – Схема взаимодействия модулей разработанной системы

При выявлении аномалии данные о ней передаются на сервер. Также всем другим агентам передается информация дополнительного обучения выявлению аномалий, подобных обнаруженной, что позволяет распределённому комплексу модулей обнаружения постоянно взаимно обучаться. Агенты во время запуска отправляют 1 байт информации на сервер, для добавления своего адреса в массив адресов агентов, сервер возвращает список адресов других агентов. Также сервер отправляет всем другим агентам информацию о только что зарегистрированном новом. Оценка эффективности системы проводилась на основе десятипроцентного набора данных о сетевых соединениях KDD 99 [13]. Данные были нормализованы и разделены на группы: для обучения, для тестирования. Каждая группа данных была передана на каждый агент.

Результаты тестирования системы

Первая процедура анализа была запущена на хосте 172.17.1.9 (хост *A*). Агент считал обучающие данные нормальной активности, формальные детекторы ИИС, зарегистрировал себя на сервере, получил от него информацию о существовании другого агента 172.17.1.35 (хост *B*). Перед началом процедуры анализа агент считал из ПЗУ анализируемые данные об аномалиях и анализируемые данные нормальной сетевой активности для определения процента ошибок первого рода. Агент, запущенный на хосте *A*, выполнил анализ данных, результат представлен на Рисунке 2.

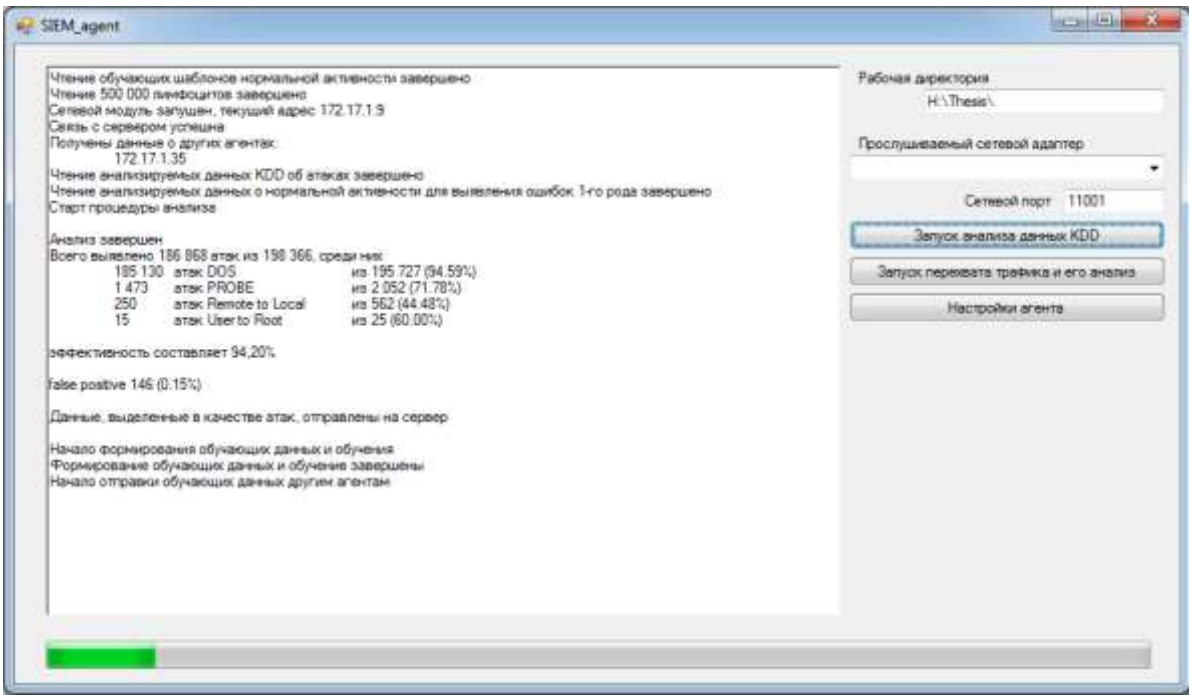

Рисунок 2 – Результат анализа данных хостом *A*

По данным Рисунка 2 видно, что эффективность обнаружения атак, число которых в обучающих данных невелико, значительно хуже, чем, к примеру, атак «DOS», но все же они выявляются. Благодаря особенностям ИИС, процент ошибок первого рода является низким. По завершении анализа агент отправил на сервер данные о выявленных аномалиях, осуществил дополнительное обучение себя выявлению подобных атак и передал обучающие данные агенту на хосте *B*. После чего агент, запущенный на хосте *B*, дополнительно обучившись на данных, полученных от хоста *A*, выполнил процедуру анализа, результат представлен на Рисунке 3.

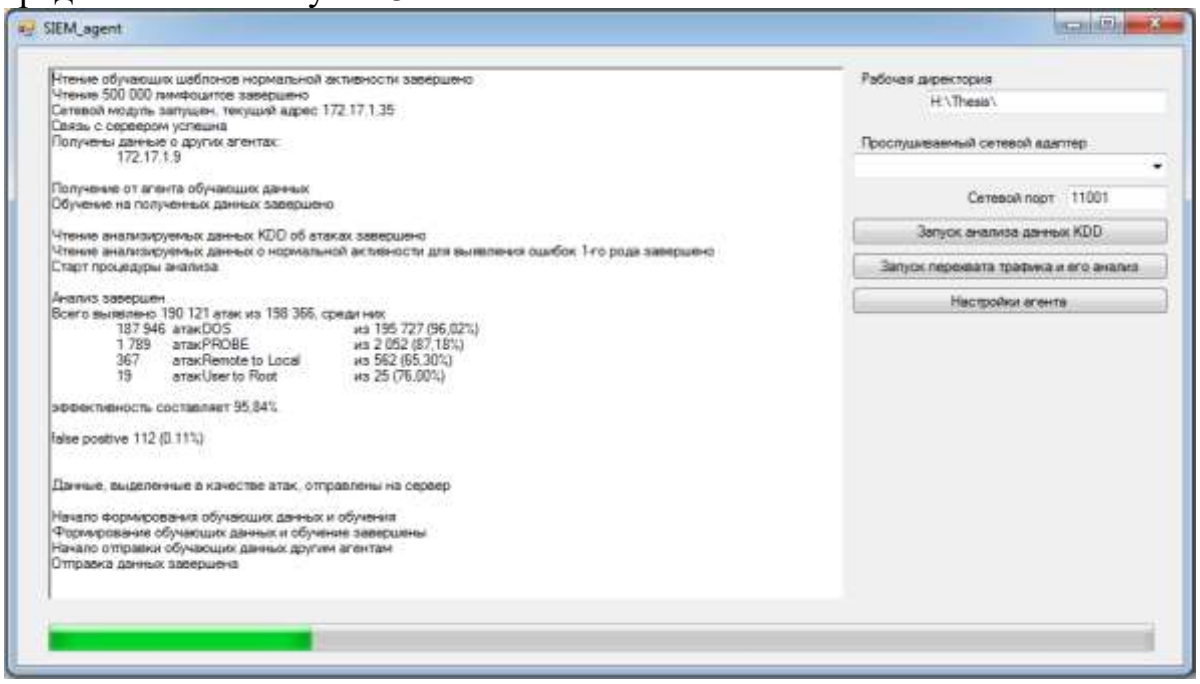

Рисунок 3 – Результат анализа данных хостом *B*

По результатам анализа видно, что, благодаря полученным обучающим данным, эффективность обнаружения хостом *B* выше, чем хостом *A*. Всего было проведено 10 процедур анализа и взаимного обучения, и от итерации к итерации эффективность обнаружения росла, процент ошибок первого и второго рода уменьшался.

Тестирование работы сервера. На хостах *A* и *B* было одновременно запущено выполнение анализа данных, Сервер принимал информацию об аномалиях и осуществлял ее корреляционный анализ совместно с предыдущей информацией, поступившей не более 5 сек. назад. Данный временной интервал регулируется администратором. Пороговое значение коэффициента корреляции, также задается администратором. Сервер отображает статистическую информацию о коррелированных инцидентах информационной безопасности в левой части за выбранный интервал

времени, в правой – за все время. Общая статистика по итогам 13 пятисекундных интервалов представлена на Рисунке 4.

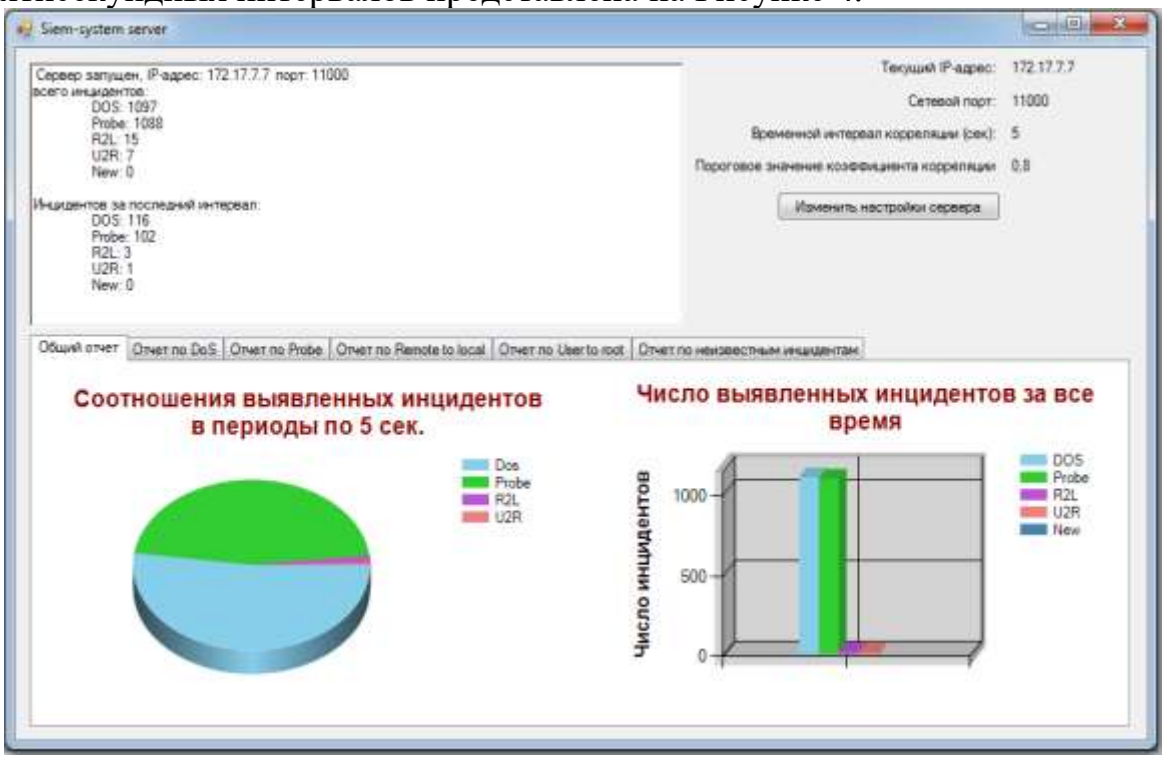

Рисунок 4 – Общая статистика корреляционного анализа

Статистика инцидентов «DOS» представлена на Рисунке 5.

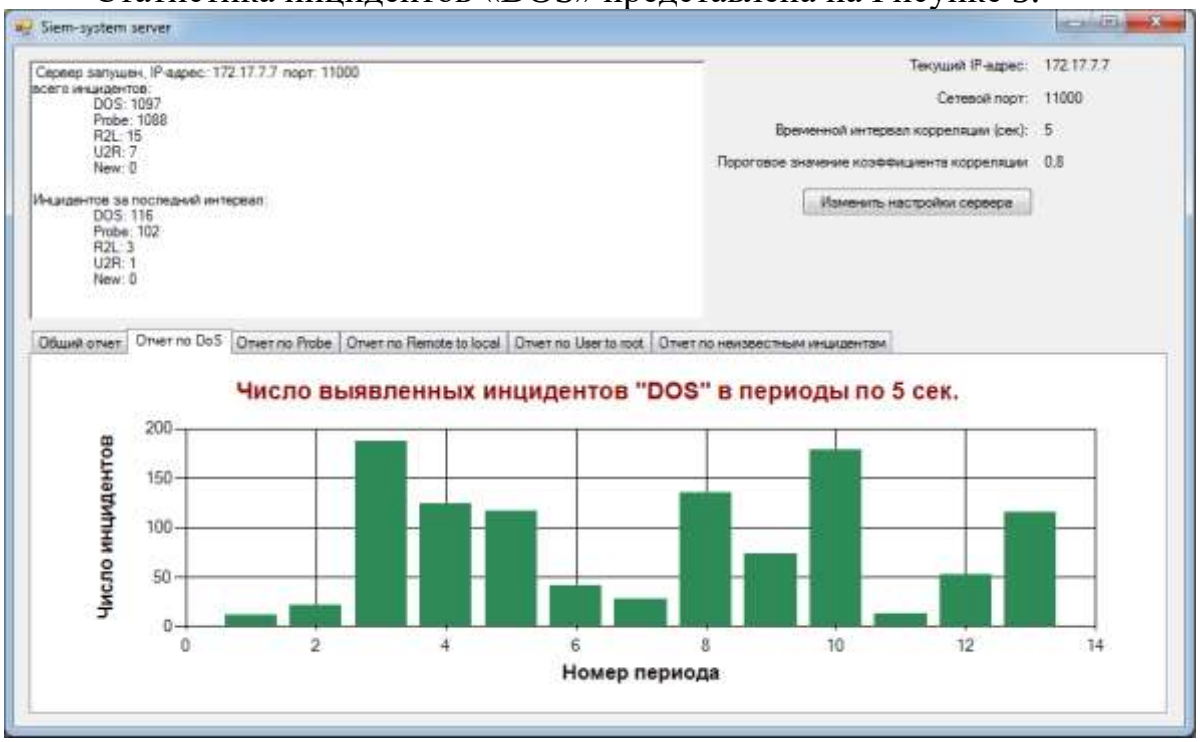

Рисунок 5 – Статистика инцидентов «DOS»

Эффективность обнаружения агентами аномалий, а также процент ошибок первого и второго рода в зависимости от порядкового номера итерации анализа и взаимного обучения представлена в Таблице 1.

Таблица 1 – Результаты обнаружения аномалий агентами

№ итерации анализа и взаимного обучения	Процент ошибок первого рода	Процент ошибок второго рода
1	0,150%	5,796%
2	0,115%	4,156%
3	0,101%	1,747%
4	0,077%	1,459%
5	0,052%	0,734%
6	0,044%	0,684%
7	0,038%	0,556%
8	0,030%	0,498%
9	0,020%	0,363%
10	0,013%	0,276%

Сервер успешно принимал данные о потенциальных инцидентах информационной безопасности от агентов, осуществлял их агрегацию, и корреляционный анализ в заданном интервале времени, а также за все время. Выводил статистические данные анализа за последний проанализированный интервал и за все время, как в комплексе, так и для каждого вида инцидентов. Пороговое значение коэффициента корреляции, достижение которого достаточно для принятия решения о серьезности инцидента и включения его в статистику задается администратором на основе экспертных знаний.

Статистические данные отображаются графически, что позволяет наглядно оценить число тех или иных инцидентов и принимать решение о реагировании. Еще один пример графического представления сервером отображен на Рисунке 6. По данным Рисунка видно значительное число инцидентов DOS, в сравнении с другими, что позволяет определить приоритеты реагирования.

Рисунок 6 – Графическое представление сервером статистических данных за все время

Заключение

Таким образом, была разработана распределенная интеллектуальная система анализа инцидентов информационной безопасности, объединившая в себе преимущества механизмов искусственной иммунной системы в их выявлении, а также методологии систем управления информационной безопасностью и событиями безопасности в их корреляционном анализе, выделении наиболее значимых инцидентов и определении приоритетов в реагировании на них.

Разработанная система при выявлении аномалии одним агентом, обучает лучшему выявлению подобных аномалий все другие агенты. Сервер осуществляет корреляционный анализ выявленных аномалий во временном промежутке, заданном администратором и принимает решение о значимости каждого поступающего события информационной безопасности, формирует статистику по заданным временным интервалам и за все время, как по всем группам инцидентов в целом, так и для каждой группы инцидентов отдельно.

Работа выполнена при поддержке гранта РФФИ №17-48-020095.

ЛИТЕРАТУРА

1. Демидов А. А. Проблемы контроля безопасности информации на объектах телекоммуникационных систем органов государственного управления: учебное пособие. – СПб: Университет ИТМО, 2015. – 70 с.
2. ГОСТ Р 27000-2012 Информационная технология. Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности. Общий обзор и терминология. – Москва: Стандартинформ, 2014. – 16 с.
3. Kostrecova E., Bínova H. Security Information and Event Management // PARIPEX – Indian Journal of Research. – Vol 4. – No. 2. – 2015. – pp. 119-120.
4. Goldstein M., Asanger S., Reif M., Hutchison A. Enhancing Security Event Management Systems with Unsupervised Anomaly Detection // ICPRAM. – No 3. – 2013. – pp. 530-538.
5. Shan Z., Liao B. Design and Implementation of a Network Security Management System // Cornell University Library [Electronic resource]. URL: https://arxiv.org/ftp/arxiv/papers/1609/1609.00099.pdf (accessed 20.11.2017). – p. 1-12

6. Kotenko I., Polubelova O., Chechulin A. Design and Implementation of a Hybrid Ontological-Relational Data Repository for SIEM Systems // Future Internet. – No. 5. – 2013. – pp. 355-375.
7. Шелестова О. Корреляция SIEM – это просто. Сигнатурные методы. // Securitylab [Электронный ресурс] URL: http://www.securitylab.ru/analytics/431459.php (дата обращения: 30.03.2018).
8. Hanemann, A., Marcu, P. Algorithm Design and Application of Service-Oriented Event Correlation // ResearchGate [Electronic resource]. URL: http://www.researchgate.net/publication/221033552_Algorithm_design_and_application_of_service-oriented_event_correlation (accessed: 25.05.2018).
9. Muller, A. Event Correlation Engine // Computer Engineering and Networks Laboratory [Electronic resource]. URL: ftp://ftp.tik.ee.ethz.ch/pub/students/2009-FS/MA-2009-01.pdf (accessed 25.05.2018).
10. Шамсутдинов Р. Р. Разработка подсистемы анализа данных и выявления аномалий на основе концепции искусственной иммунной системы // Материалы VII Всероссийской заочной Интернет-конференции «Проблемы информационной безопасности», Ростов-на-Дону, 20-21 февраля, 2018. С. 239-243
11. Васильев В.И., Шамсутдинов Р.Р. Распределенная система обнаружения атак на основе механизмов иммунной системы // Труды VI Всероссийской научной конференции «Информационные технологии интеллектуальной поддержки принятия решений» (с приглашением зарубежных ученых) Т. 1, Уфа, 28-31 мая, 2018. С. 237-244.
12. Kotenko I., Fedorchenko A., Saenko I., Kushnerevich A. Big Data Technologies for Security Event Correlation Based on Event Type Accounting // Вопросы кибербезопасности. – № 5(24). – 2017. – С. 2-16
13. KDD Cup 1999 Data [Electronic resource]. URL: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (accessed 05.02.2018).

V.V. Vasilyev, R.R. Shamsutdinov

INTELLIGENT SYSTEM OF INFORMATION SECURITY INCIDENT ANALYSIS (BASED ON THE METHODOLOGY OF SIEM-SYSTEMS USING IMMUNOCOMPUTING MECHANISMS)

Ufa State Aviation Technical University, Ufa, Russia

The article is devoted to the problem of information security incidents intelligent analysis using the security information and event management system methodology. The essence of such systems, and its ability to interact with the methods of artificial intelligence

were analyzed. The developed distributed information security incident analysis system was described, which synthesized the mechanisms of the artificial immune system and the correlation analysis of data to identify known and unknown anomalies, analyze their criticality and determine priorities in response. The modules interaction diagram of the developed system and the mathematical component of the applied method for correlation analysis of data were presented. A series of computational experiments was conducted, which showed a high level of system efficiency in detecting anomalies and the possibility of additional training of each other by client modules, as well as the successful implementation of correlation analysis of data from clients in a given time interval, highlighting the most significant incidents for last analyzed interval, as well as for all the time, both in the complex and for each group of incidents. A graphical display of statistical data by the server allows you to visually assess the criticality of certain incidents and to determine priorities in responding to them.

Keywords: SIEM-system, immunocomputing, correlation analysis, information security, network security.

REFERENCES:

1. Demidov A. A. Problemy kontrolya bezopasnosti informacii na ob'ektah telekommunikacionnyh sistem organov gosudarstvennogo upravleniya [Information Security Control Problems at the Objects of Telecommunication Systems of Government Bodies], ITMO University Publ., St. Petersburg, 2015, 70 p. (in Russian).
2. GOST R 27000-2012 Information technology. Security techniques. Information security management systems. Overview and vocabulary, Moscow, Standartinform Publ, 2014, 16 p.
3. Kostrecova E., Bínova H. Security Information and Event Management, PARIPEX – Indian Journal of Research, 2015,vol 4, no. 2, pp. 119-120.
4. Goldstein M., Asanger S., Reif M., Hutchison A. Enhancing Security Event Management Systems with Unsupervised Anomaly Detection, ICPRAM, 2013, no 3, pp. 530-538.
5. Shan Z., Liao B. Design and Implementation of a Network Security Management System, Cornell University Library, available at: https://arxiv.org/ftp/arxiv/papers/1609/1609.00099.pdf (accessed 20.11.2017). – p. 1-12
6. Kotenko I., Polubelova O., Chechulin A. Design and Implementation of a Hybrid Ontological-Relational Data Repository for SIEM Systems, Future Internet, 2013, no. 5, – pp. 355-375.
7. Shelestova O. SIEM Correlation Is Easy. Signature Methods, Securitylab, available at: http://www.securitylab.ru/analytics/431459.php (accessed: 30.03.2018).
8. Hanemann, A., Marcu, P. Algorithm Design and Application of Service-Oriented Event Correlation, ResearchGate, available at:

http://www.researchgate.net/publication/221033552_Algorithm_design_and_application_of_service-oriented_event_correlation (accessed: 25.05.2018).

9. Muller, A. Event Correlation Engine, Computer Engineering and Networks Laboratory, available at: ftp://ftp.tik.ee.ethz.ch/pub/students/2009-FS/MA-2009-01.pdf (accessed 25.05.2018).
10. Shamsutdinov R. R. Development of a Subsystem for Data Analysis and Anomalies Detection Based on the Concept of an Artificial Immune System, Materialy VII Vserossijskoj zaochnoj Internet-konferencii «Problemy informacionnoj bezopasnosti» [Proceedings of the VII All-Russian Correspondence Internet Conference «Problems of Information Security»], Rostov-on-Don, 20-21 February, 2018, pp. 239-243. (in Russian).
11. Vasilyev V. I., Shamsutdinov R. R. Distributed Intrusion Detection System Based on Immune System Mechanisms, Information Technologies for Intelligent Decision-Making Support, vol. 1, Ufa, 28-31 May, 2018, pp. 237-244. (in Russian).
12. Kotenko I., Fedorchenko A., Saenko I., Kushnerevich A. Big Data Technologies for Security Event Correlation Based on Event Type Accounting, Voprosy kiberbezopasnosti [Cyber Security Issues], 2017, no. 5, pp. 2-16
13. KDD Cup 1999 Data, available at: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (accessed 05.02.2018).