Алексей В. Плугатарев[1], Анатолий Л. Марухленко[2], Михаил А. Бугорский[3],
Андрей С. Булгаков[4], Михаил А. Марченко[5]
[1,2]Юго-Западный государственный университет,
ул. 50 лет октября, 94, Курск, 305040, Россия
[3]в/ч 33310, п. Шайковка, Кировский р-н, Калужская обл, 249455, Россия
[4,5]Национальный исследовательский университет «Московский институт электронной техники»,
площадь Шокина, 1, Зеленоград, Москва, 124498, Россия
[1]e-mail: aplugatarev@bk.ru, https://orcid.org/0000-0002-8549-4382
[2]e-mail: proxy33@mail.ru, https://orcid.org/0000-0002-3575-924X
[3]e-mail: bygorbtc2013@gmail.com, https://orcid.org/0000-0003-1788-0211
[4]e-mail: bulgakov1703@yandex.ru, https://orcid.org/0000-0003-4374-8409
[5]e-mail: marchenko14120@gmail.com, https://orcid.org/0000-0001-6885-8415

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ В СИСТЕМАХ ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

DOI: http://dx.doi.org/10.26583/bit.2021.3.06

Аннотация. Для минимизации ущерба от кибератак, организации используют системы мониторинга своей сетевой инфраструктуры. Такие инфраструктуры собирают огромное количество информации из журналов систем безопасности и систем управления событиями. Структурируя эти данные и анализируя при помощи нейронных сетей, данное исследование ставит цель – разработка алгоритма автономного обнаружения вредоносных хостов в сети. Автоматизация этого процесса, позволит более точно, и оперативно, чем это делают эксперты вычислять вредоносные компьютеры-хосты, что увеличит отказоустойчивость информационной системы предприятия. Для моделирования системы принятия решений используется ограниченная машина Больцмана, в данном исследовании – модель стохастической нейронной сети, определяющей распределение вероятности на входных образцах данных. Эффективность предложенной модели исследуется при помощи показателя AUC – численной характеристики кривой ошибок, которая является площадью, ограниченной ROC-кривой и осью. Даная ось отделяет долю ложных положительных результатов – вероятностей, которые классификатор верно отнёс к безопасным хостам. Представленная модель с применением нейронных сетей, обеспечивает мониторинг и обнаружение вредоносных компьютеров-хостов в информационных системах в реальном времени, позволяет принимать решение о событиях информационной безопасности без участия экспертов, на основе анализа данных различных внутренних систем и служб.

Ключевые слова: система машинного обучения; обнаружение вредоносного хоста; системы информационной безопасности.

Для цитирования: ПЛУГАТАРЕВ, Алексей В. и др. ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ В СИСТЕМАХ ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ. Безопасность информационных технологий, [S.l.], т. 28, № 3, с. 73–80, 2021. ISSN 2074-7136. URL: https://bit.mephi.ru/index.php/bit/article/view/1364. DOI: http://dx.doi.org/10.26583/bit.2021.3.06

Alexey V. Plugatarev[1], Anatoly L. Maruhlenko[2], Mikhail A. Bugorskij[3],
Andrey S. Bulgakov[4], Mikhail A. Marchenko[5]
[1,2]South-West State University,
50 Let Oktyabrya str., 94, Kursk, 305040, Russia
[3]Military unit No. 33310, item Shaikovka, Kirovsky district, Kaluga region, 249455, Russia
[4,5]National Research University of Electronic Technology,
Shokin Square, 1, Zelenograd, Moscow, 124498, Russia
[1]e-mail: aplugatarev@bk.ru, https://orcid.org/0000-0002-8549-4382
[2]e-mail: proxy33@mail.ru, https://orcid.org/0000-0002-3575-924X
[3]e-mail: bygorbtc2013@gmail.com, https://orcid.org/0000-0003-1788-0211
[4]e-mail: bulgakov1703@yandex.ru, https://orcid.org/0000-0003-4374-8409
[5]e-mail: marchenko14120@gmail.com, https://orcid.org/0000-0001-6885-8415

The neural networks application for information security systems

DOI: http://dx.doi.org/10.26583/bit.2021.3.06

Abstract. In order to minimize the damage caused by cyberattacks, most organizations are using the network monitoring systems for its infrastructures. Such infrastructures collect a huge amount of information from security system logs event management systems. By structuring this data and analyzing it with the help of neural networks, the current study sets autonomous detection of malware hosts in a network as a goal. Automatization of this process will allow to detect malware computer hosts more precisely and promptly than experts do. It increases fault tolerance of the information system of a department. In order to design a decision-making system a restricted Boltzmann machine is used, while in the present study a stochastic neural network model determining probability distribution at the input data samples was used. Efficiency of the proposed model is investigated with the help of AUC indicator. This is a numeric characteristic of the error curve, which is an area, limited by a ROC-curve and the axis. The axis defines a share of false positive results which are the probabilities correctly defined by the classifier as safe hosts.The proposed model with the usage of neural networks provides monitoring and detection of malware computer hosts in information systems in real-time. This architecture significantly increases analysis efficiency and improves risk detection in an information environment with the help of data analysis or various internal systems and services. Development of this concept allows for involving a bigger amount of data from corporate networks and adding extra functions for improvement of vulnerabilities detection accuracy and reduction of the number of false alarms.

Keywords: machine learning system; detection of a malicious host; information security systems
For citation: PLUGATAREV, Alexey V. et al. The neural networks application for information security systems. IT Security (Russia), [S.l.], v. 28, n. 3, p. 73–80, 2021. ISSN 2074-7136. URL: https://bit.mephi.ru/index.php/bit/article/view/1364. DOI: http://dx.doi.org/10.26583/bit.2021.3.06.

Введение

В эпоху развитых сетевых технологий для обеспечения информационной защиты предприятиям, как правило, приходится выстраивать систему информационной безопасности, которая занимается оценкой рисков и анализов данных всей сетевой инфраструктуры [1]. Данное исследование посвящено обнаружению опасных компьютеров-хостов внутри сетевой инфраструктуры, на основе анализа данных с помощью нейронных сетей. Аналитики в центре безопасности исследуют предупреждения, чтобы решить, являются ли связанные хосты злонамеренными или нет. Однако, количество предупреждений и событий, как правило, огромно, и превышает возможности аналитиков по обработке информации, а ложное срабатывание может повлечь за собой потраченные впустую ресурсы. Следовательно, существует потребность в максимальном уменьшении количества ложных срабатываний. Данное исследование сосредоточено на обнаружении взломанных хостов в информационных системах интеллектуальной системой с использованием глубокого машинного обучения [2].

Основополагающий принцип классических систем управления событиями безопасности состоит в том, что данные о безопасности информационной системы собираются из разных источников, и результат их обработки предоставляется в едином интерфейсе, доступном для аналитиков безопасности, что помогает изучать характерные особенности, которые соответствуют инцидентам информационной безопасности [3]. Построение таких систем опирается на анализ существующих данных, стараясь улучшить долгосрочную эффективность системы и оптимизировать хранение информационных данных, а также делается акцент на выгрузке из имеющихся данных определенного объема информации, с помощью которого могут быть немедленно выявлены уязвимости [4].

Система, спроектированная по предлагаемому алгоритму, способна самостоятельно анализировать информационные оповещения, журналы безопасности и информацию

аналитиков; готова к интеграции с реальной корпоративной средой для определения хостов с высокой вероятностью компрометации в реальном времени.

1. Материалы и методы

Основное преимущество глубокого машинного обучения перед другими моделями машинного обучения состоит в том, что для анализа используются, как правило, очень сложные линии поведения или особенности компьютера-хоста [5]. Например, хост может посещать разные вредоносные веб-сайты, скачивать/закачивать файлы или иметь различные поражения вредоносным программным обеспечением одновременно. Фактически, создаются сотни функций в день для описания состояния безопасности хоста [6, 7]. Таким образом, главное преимущество глубокого машинного обучения – это способность его послойной стратегии обучения, в которой функции более высокого уровня извлечены из предыдущих, то есть высокоуровневые функции лучше извлекают информацию из входных данных [8, 9].

Ограниченная машина Больцмана – двухслойная стохастическая модель, которая включает скрытый слой и видимый слой. Видимый слой состоит из видимых состояний $V = (v_1, \ldots, v_m)$, а скрытый слой имеет состояния $H = (h_1, \ldots, h_n)$, которые нельзя измерить напрямую. Между двумя уровнями состояния полностью связные. Однако нет никакой связи между состояниями в одном слое, что означает, что состояния в одном слое взаимно независимы.

В рамках данной структуры рассмотрим обучающий набор двоичных векторов, которые будем считать двоичными изображениями для аутентификации хостов. Совместная конфигурация в данной модели состоит из видимых и скрытых единиц (v, h), которые обладают энергией, определяемой:

$$E(v,h) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_i \sum_j v_i w_{i,j} h_j, \quad (1)$$

где v_i, h_j – двоичные состояния видимой единицы i и скрытой единицы j; a_i, b_j – их смещения, а $w_{i,j}$ – вес между ними. Сеть присваивает вероятность каждой возможной паре видимого и скрытого векторов через энергетическую функцию:

$$P(v,h) = \frac{1}{Z} e^{-E(v,h)}, \quad (2)$$

где «статистическая сумма» Z определяется суммированием всех возможных пар видимых и скрытых векторов:

$$Z = \sum_{v,h} e^{-E(v,h)}. \quad (3)$$

Вероятность того, что сеть присваивает доверительное значение видимому вектору v, определяется суммированием всех возможных скрытых векторов:

$$p(v) = \frac{1}{Z} \sum_{v,h} e^{-E(v,h)}, \quad (4)$$

где Z – константа нормализации, которая гарантирует, что сумма всех вероятностей равна единице.

Предложенная сеть глубокого обучения – это вероятностная генеративная модель. Как показано на рис. 1, сеть глубокого обучения в первую очередь построена путем интегрирования ограниченных машин Больцмана и классификатора (верхний слой).

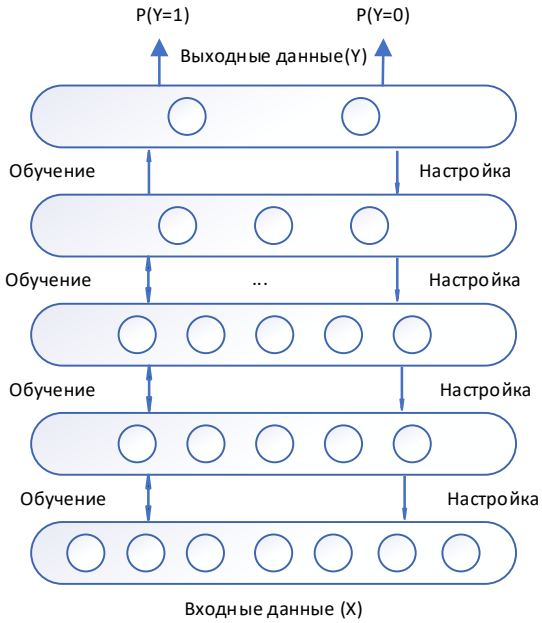

Рис. 1. Структура глубокого машинного обучения
Fig. 1. Structure of Deep Belief Network

Структура обучения имеет две части: объединённые ограниченные машины Больцмана и классификатор (верхний слой). Тренировочный процесс для данной сети включает предварительное обучение и настройку соответственно.

Необработанные данные собираются из журналов служб внутренней безопасности. Данные состоят из предупреждений от систем интеллектуальной защиты, заметок аналитики и журналов из разных источников, включая межсетевой экран, систему обнаружения/предотвращения вторжений, HTTP/FTP/DNS–трафики, DHCP, сканирование уязвимостей, событий безопасности Windows, VPN и т.д. В журналах появляются терабайты данных каждый день. В табл. 1 перечислены данные оповещений ключевых элементов системы интеллектуальной защиты:

Таблица 1. Пример входных данных

ID хоста	Результаты 1 категории	Индикатор 2 категории	Показатель 3 категории	Показатель 4 категории	Метка
Хост1	13	1	0,65	5,17	1 (рискованный)
Хост2	25	0	2,74	9,34	1 (рискованный)
Хост3	4	0	1,33	3,52	0 (нормальный)

Аналитические функции создаются на уровне отдельного хоста, так как основная цель – спрогнозировать риск хоста [10]. Характеристики, которые в данном исследовании выделены для анализа, можно классифицировать в следующие четыре категории:

1) Сводные характеристики: эти функции могут быть сгенерированы из статистических сводок. Например, число событий «Вредоносное ПО: не исправлено» за последние 24 часа, или количество событий высокой важности (с определённой степенью серьёзности) за последние 7 дней.

2) Характеристики индикатора: эти функции представлены в двоичном формате (0 – событие ложно, 1 – событие верно), например, «Вредоносное ПО: не исправлено – ложно»; «событие происходит в выходные дни – верно».

3) Временные особенности: эти события включают временную информацию, например, частота появления оповещения системы безопасности, которая учитывает временной интервал между двумя последовательными событиями.

4) Реляционные функции: эти функции получены из анализа событий, рассчитанные по графу хост-событие, где узлы – это хосты или события. Отношение между хостом и его событиями представлен ребром графа, а вес ребра – это количество определенных событий на хосте. На рис. 2. приведен пример подграфа на один конкретный хост с его событиями безопасности и некоторые взвешенные значения, полученные из более крупных графов с большим количеством хостов и событий. Более высокий показатель подразумевает более подозрительное поведение компьютера-хоста.

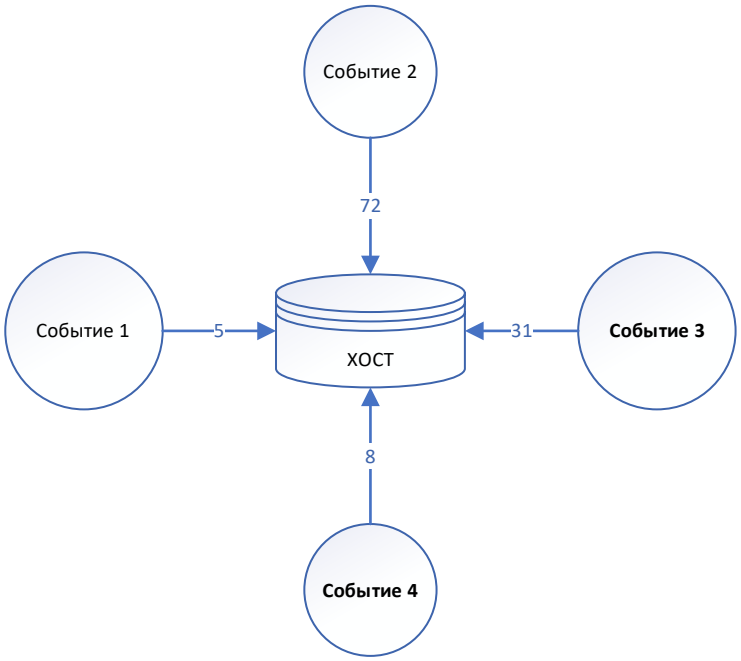

Рис. 2. Пример функции взвешенного рейтинга событий хоста
Fig. 2. Example of a weighted host event rating function

2. Результаты и их обсуждение

В рамках исследований было проведено математическое моделирование для поиска вредоносных хостов [11, 12]. В данных тестированиях были смоделированы 5100 хостов с данными, из которых 60 будем считать потенциально опасными. В данном исследовании будем тестировать разные модели машинного обучения, чтобы определить, могут ли они работать лучше, чем действующая система, основанная на правилах. Как правило, необходимо случайным образом разделить на данные с хостов (75% образцов для мониторинга) и тестовые (оставшиеся 25%). Были опробованы разные соотношения между целевыми выборками и тестами – 50% / 50%, 60% / 40% и 75% / 25%. При моделировании было получено, что коэффициенты разделения мало повлияли на классификацию полученных результатов [13].

Меры оценки определены в уравнениях (5) – (7).

Модель AUC – численная характеристика кривой ошибок – площадь, ограниченная ROC-кривой и осью, которая отделяет долю ложных положительных результатов – вероятностей, которые классификатор верно отнёс к безопасным хостам:

$$\approx \sum_i \frac{y_{i+1}+y_i}{2} \times (x_{i+1} - x_i), \quad (5)$$

где x – вероятность i-го успешного определения хоста к вредоносным или безопасным, а y – классификатор хостов, принимающий значения 0 – если хост был определён к вредоносным и 1 если к безопасным.

Модель оценки эффективности обнаружения вредоносных хостов, определяется как отношение прогнозируемых опасных хостов, к общему количеству опасных хостов в сети:

$$\frac{\text{Прогнозируемое количество опасных хостов}}{\text{Общее количество опасных хостов}} \times 100\%. \quad (6)$$

Мера производительности алгоритма рассчитывается как отношение доли прогнозируемых опасных хостов в сети к общей доле опасных хостов в сети:

$$\frac{\text{Доля прогнозируемых опаных хостов}}{\text{Общая доля опасных хостов}} \times 100\%. \quad (7)$$

На рис. 3 приведены результаты моделирования от количества скрытых нейронов N, при следующих параметрах: количество эпох для точной настройки – 100; четыре слоя с одним входным слоем, двумя скрытыми слоями и один выходной слой; в каждом слое 100, 20, 10, 2 нейрона соответственно.

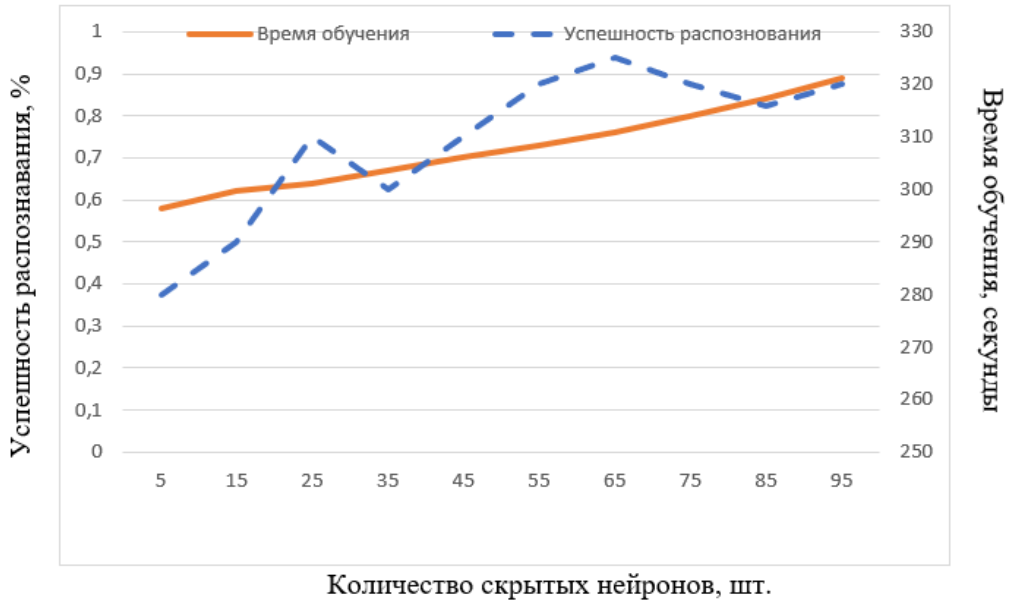

Рис. 3. Время обучения относительно количества скрытых нейронов
Fig. 3. Training time as function of the number of hidden neurons

Из рис. 3 видно, что время обучения модели увеличивается почти линейно с N, что означает, что модель имеет хорошую масштабируемость с большим количеством скрытых нейронов. С другой стороны, показатель AUC на тестовых данных увеличивается, когда $N<25$, и начинается уменьшаться после этого, затем наблюдаются тенденции к росту, однако тенденции нестабильны, что может означать, что обобщаемость нейронной сети снижается из-за слишком большого количества скрытых нейронов. В рамках исследования были проведены исследования с более сложной структурой и с большим количеством нейронов и слоев, но нет очевидного повышения производительности – время обучения продолжает расти линейно, а успешность распознавания колеблется 0,85 – 0,95 (тенденции к этому видны в диапазоне $N>85$). Поэтому, на графике представлен диапазон до 100 скрытых нейронов, который и представляет наибольший интерес. Оптимальным можно

назвать диапазон $60<N<70$, который даёт лучшее соотношение времени обучения и успешности распознавания.

Заключение

Описанный в статье подход, основан на нейросети с глубоким машинным обучением, которая использует данные журналов безопасности различных устройств в информационной системе, предупреждающую информацию и аналитические сведения для идентификации рискованных компьютеров-хостов. Разработанный алгоритм предназначен для работы с информацией из необработанных и неструктурированных массивов данных, чем и отличается от традиционных моделей систем управления событиями безопасности. Алгоритм реализован, таким образом, что готов для программной или аппаратной реализации и последующей интеграции в информационную инфраструктуру, также имеет возможность выявлять опасные хосты в режиме реального времени, обновляя данные, получая их из журналов систем безопасности. Программная реализация алгоритма позволит полностью автоматизировать процесс: от сбора данных, до обновления оценки в реальном времени, что значительно улучшает эффективность аналитики и повышает эффективность обнаружения рисков в информационной среде.

СПИСОК ЛИТЕРАТУРЫ:

1. Marukhlenko A.L., Plugatarev A.V., Bobyntsev D.O. Complex evaluation of information security of an object with the application of a mathematical model for calculation of risk indicators. Lecture Notes in Electrical Engineering. 2020. Vol. 641 LNEE. P. 771–778. DOI: https://doi.org/10.1007/978-3-030-39225-3_84.
2. A.L. Buczak and E. Guven. A survey of data mining and machine learning methods for cyber security intrusion detection, IEEE Communications urveys & Tutorials. Vol. 18, no. 2. P. 1153–1176. DOI: https://doi.org/10.1109/COMST.2015.2494502.
3. M. Bhuyan, D. Bhattacharyya and J. Kalita. Network anomaly detection: Methods systems and tools. IEEE Commun. Surv. Tuts., vol. 16, no. 1. P. 303–336, 2014. DOI: http://dx.doi.org/10.1109/SURV.2013.052213.00046.
4. T.T.T. Nguyen and G. Armitage. A survey of techniques for internet traffic classification using machine learning. IEEE Commun. Surv. Tuts., vol. 10, no. 4. P. 56–76, 2008 DOI: https://doi.org/10.1109/SURV.2008.080406.
5. D.Y. Yeung and Y. Ding. Host-based intrusion detection using dynamic and static behavioral models, Pattern Recognition. Vol. 36, Issua 1, 2003. P. 229–243. DOI: https://doi.org/10.1016/S0031-3203(02)00026-2.
6. B. Li, M.H. Gunes, G. Bebis and J. Springer. A supervised machine learning approach to classify host roles on line using sflow, Proceedings of the first edition workshop on High performance and programmable networking. Association for Computing Machinery, New York, NY, USA, 2013. P. 53–60. DOI: https://doi.org/10.1145/2465839.2465847.
7. Kuleshova E., Marukhlenko A., Dobritsa V., Tanygin M. Formation of Unique Characteristics of Hiding and Encoding of Data Blocks Based on the Fragmented Identifier of Information Processed by Cellular Automata. Computers 9(2), 51 (2020). DOI: https://doi.org/10.3390/computers9020051.
8. Марухленко А.Л., Плугатарев А.В., Таныгин М.О. Вариант разграничения доступа к информационным ресурсам на основе неявной аутентификации и др. Известия Юго-Западного государственного университета. 2020. Т. 24. № 2. С. 108–121. DOI: https://doi.org/10.21869/2223-1560-2020-24-2-108-121.
9. A. Fischer and C. Igel. Training restricted Boltzmann machines: An introduction, Pattern Recognition. 2014. Vol. 47, Issue 1. P. 25–39. DOI: https://doi.org/10.1016/j.patcog.2013.05.025.
10. M. Kang and J. Kang. A novel intrusion detection method using deep neural network for in-vehicle network security, 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring). P. 1–5. DOI: https://doi.org/10.1109/VTCSpring.2016.7504089.
11. Center for Strategic and International Studies (CSIS) and McAfee. Economic Impact of Cybercrime Slowing Down Report. URL: https://www.csis.org/analysis/economic-impact-cybercrime (дата обращения: 02.04.2021).
12. Deep Learning 0.1 documentation: Deep Belief Networks. URL: https://deep-learning-tensorflow.readthedocs.io/en/latest/ (дата обращения: 03.04.2021).

13. A. Fischer and C. Igel. Training restricted Boltzmann machines: An introduction, Pattern Recognition. 2014. Vol. 47, Issue 1. P. 25–39. DOI: https://doi.org/10.1016/j.patcog.2013.05.025.

REFERENCES:

[1] Marukhlenko A.L., Plugatarev A.V., Bobyntsev D.O. Complex evaluation of information security of an object with the application of a mathematical model for calculation of risk indicators. Lecture Notes in Electrical Engineering. 2020. Vol. 641 LNEE. P. 771–778. DOI: https://doi.org/10.1007/978-3-030-39225-3_84.
[2] A.L. Buczak and E. Guven. A survey of data mining and machine learning methods for cyber security intrusion detection, IEEE Communications urveys & Tutorials. Vol. 18, no. 2. P. 1153–1176. DOI: https://doi.org/10.1109/COMST.2015.2494502.
[3] M. Bhuyan, D. Bhattacharyya and J. Kalita. Network anomaly detection: Methods systems and tools. IEEE Commun. Surv. Tuts., vol. 16, no. 1. P. 303–336, 2014. DOI: http://dx.doi.org/10.1109/SURV.2013.052213.00046.
[4] T.T.T. Nguyen and G. Armitage. A survey of techniques for internet traffic classification using machine learning. IEEE Commun. Surv. Tuts., vol. 10, no. 4. P. 56–76, 2008. DOI: https://doi.org/10.1109/SURV.2008.080406.
[5] D.Y. Yeung and Y. Ding. Host-based intrusion detection using dynamic and static behavioral models, Pattern Recognition. Vol. 36, Issua 1, 2003. P. 229–243. DOI: https://doi.org/10.1016/S0031-3203(02)00026-2.
[6] B. Li, M.H. Gunes, G. Bebis and J. Springer. A supervised machine learning approach to classify host roles on line using sflow, Proceedings of the first edition workshop on High performance and programmable networking. Association for Computing Machinery, New York, NY, USA, 2013. P. 53–60. DOI: https://doi.org/10.1145/2465839.2465847.
[7] Kuleshova E., Marukhlenko A., Dobritsa V., Tanygin M. Formation of Unique Characteristics of Hiding and Encoding of Data Blocks Based on the Fragmented Identifier of Information Processed by Cellular Automata. Computers 9(2), 51 (2020). DOI: https://doi.org/10.3390/computers9020051.
[8] Marukhlenko A.L., Plugatarev A.V., Tanygin M.O., Marukhlenko L.O., Shashkov M.Yu. Option of Control of Access to Information Resources Based on Implicit Authentication. Proceedings of the Southwest State University. 2020.Vol. 24. No. 2. P. 108–121. DOI: https://doi.org/10.21869/2223-1560-2020-24-2-108-121 (in Russian).
[9] A. Fischer and C. Igel. Training restricted Boltzmann machines: An introduction, Pattern Recognition. 2014. Vol. 47, Issue 1. P. 25–39. DOI: https://doi.org/10.1016/j.patcog.2013.05.025.
[10] M. Kang and J. Kang. A novel intrusion detection method using deep neural network for in-vehicle network security, 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring). P. 1–5. DOI: https://doi.org/10.1109/VTCSpring.2016.7504089.
[11] Center for Strategic and International Studies (CSIS) and McAfee. Economic Impact of Cybercrime – No Slowing Down Report. URL: https://www.csis.org/analysis/economic-impact-cybercrime (accessed: 02.04.2021).
[12] Deep Learning 0.1 documentation: Deep Belief Networks. URL: https://deep-learning-tensorflow.readthedocs.io/en/latest/ (accessed: 03.04.2021).
[13] A. Fischer and C. Igel. Training restricted Boltzmann machines: An introduction, Pattern Recognition. 2014. Vol. 47, Issue 1. P. 25–39. DOI: https://doi.org/10.1016/j.patcog.2013.05.025.

Поступила в редакцию – 28 апреля 2021 г. Окончательный вариант – 20 августа 2021 г.
Received – April 28, 2021. The final version –August 20, 2021.