МЕТОДИКА ИДЕНТИФИКАЦИИ АТАК НА БЕСПРОВОДНЫЕ СЕНСОРНЫЕ СЕТИ НА ОСНОВЕ АНАЛИЗА ПОВЕДЕНИЯ СЕТИ

AN ATTACK IDENTIFICATION TECHNIQUE BASED ON BEHAVIOUR ANALISYS FOR WIRELESS SENSOR NETWORK

V. Korzhuk

Summary. This article presents a technique for identifying attacks on wireless sensor networks. The developed technique allows using a certain set of features to identify 14 types of network-level attacks. The article describes and justifies the use of computer programs developed during the study, which are necessary to increase the efficiency of identification of attacks on wireless sensor networks. The random forest algorithm and the probabilistic classifier are used together for identification. Experiments on mesh and cluster topology networks were carried out. Recommendations on the application of the technique are offered. The results can be used as part of intrusion detection systems to improve the availability and integrity of information.

Keywords: cyber-physical systems, cyber security, information security, wireless sensor networks, attack identification, identification methodic.

Коржук Виктория Михайловна
Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики
vika@cit.ifmo.ru

Аннотация. В данной статье представлена методика идентификации атак на беспроводные сенсорные сети. Разработанная методика позволяет использовать определенный набор признаков для идентификации 14 типов сетевых атак. Описывается и обосновывается применение разработанных в рамках исследования компьютерных программ, необходимых для повышения эффективности идентификации атак на беспроводные сенсорные сети. Для идентификации совместно применяются алгоритм «случайный лес» и вероятностный классификатор. Проведены эксперименты на сети с ячеистой и кластерной топологией. Предложены рекомендации по применению методики. Полученные результаты могут применяться в составе систем обнаружения вторжений для повышения уровня доступности и целостности информации.

Ключевые слова: на русском языке: киберфизические системы, кибер-безопасность, информационная безопасность, беспроводные сенсорные сети, идентификация атак, методика идентификации.

Введение

Современный этап развития информационных технологий характеризуется повсеместным внедрением и использованием различных киберфизических систем (КФС). В качестве основы для таких систем нередко используются беспроводные сенсорные сети (БСС), и защита информации в этих сетях является новой и актуальной задачей. Важность решения задачи идентификации атак на БСС обусловлена спецификой БСС в КФС и непрерывным ростом количества разнообразных сетевых угроз, реализация которых может привести к финансовым, репутационным и даже человеческим потерям.

БСС состоят из маломощных беспроводных устройств, чаще всего не подключенных к постоянному источнику питания. В соответствии с этими особенностями, формируется ряд ограничений для систем аудита и мониторинга состояния таких сетей [7]. По данным, полученным в результате анализа текущего состояния области, представленных в предыдущих работах автора, было выявлено, что наиболее опасными с точки зрения простоты реализации и возможного ущерба являются сетевые атаки. В данной статье приводится возможное решение таких задач, как моделирование поведения сети, выбор признаков идентификации и использование машинного обучения для повышения уровня доступности и целостности информации, циркулирующей в БСС на основе анализа поведения сети.

Составляющие элементы

В качестве составляющих элементов методики используются такие разработки автора, как модель профиля поведения БСС и метод идентификации атак. Особенность модели профиля поведения заключается в использовании новой комбинации признаков идентификации атак, составленной на основе стандарта 802.15.4 и спецификации ZigBee [10]; особенность метода идентификации состоит в совместном использовании алгоритма «случайный лес» [2], вероятностного классификатора [1] и введения параметра степени уверенности. Также, в процессе проведения исследо-

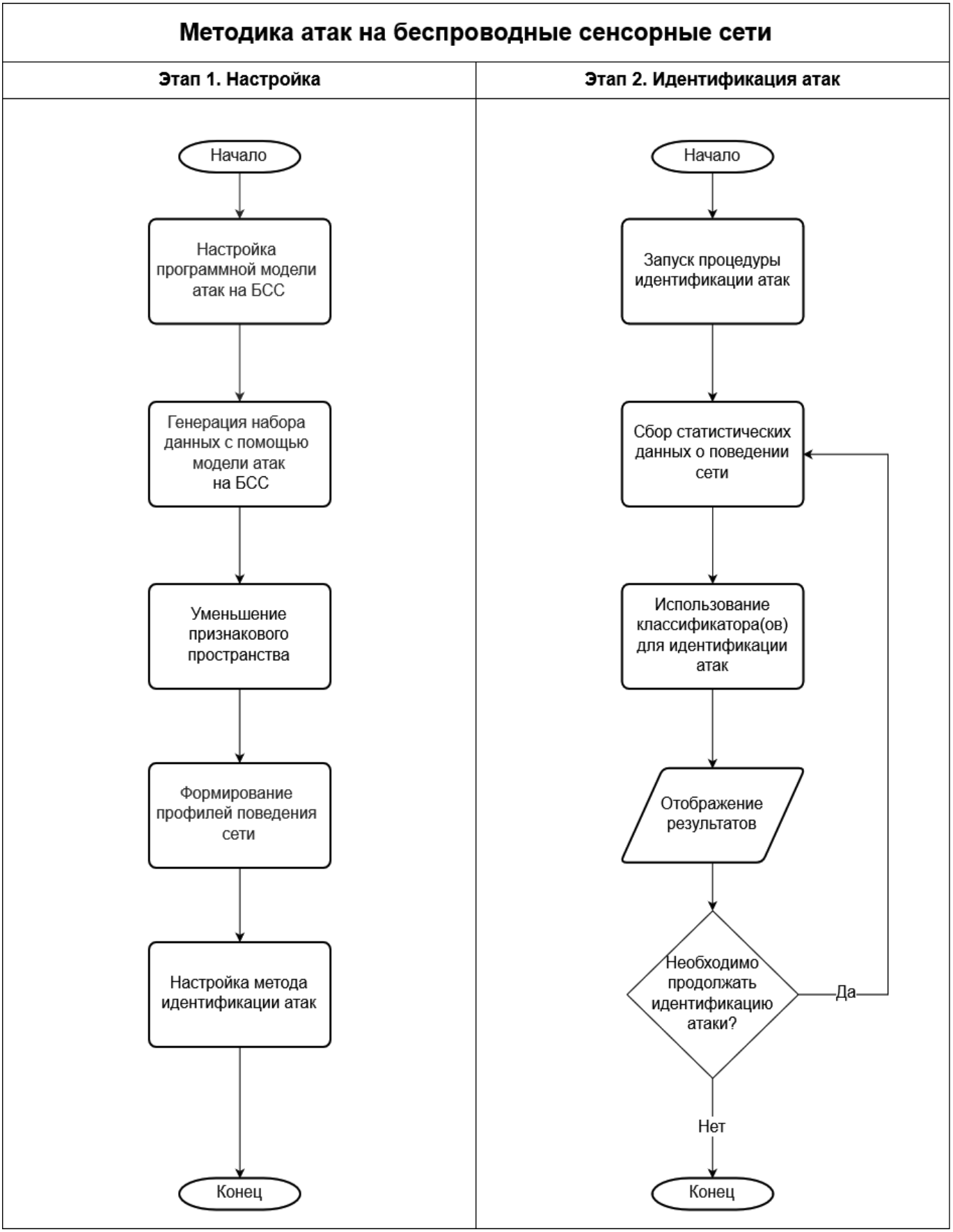

Рис. 1. Общая схема методики идентификации атак на БСС

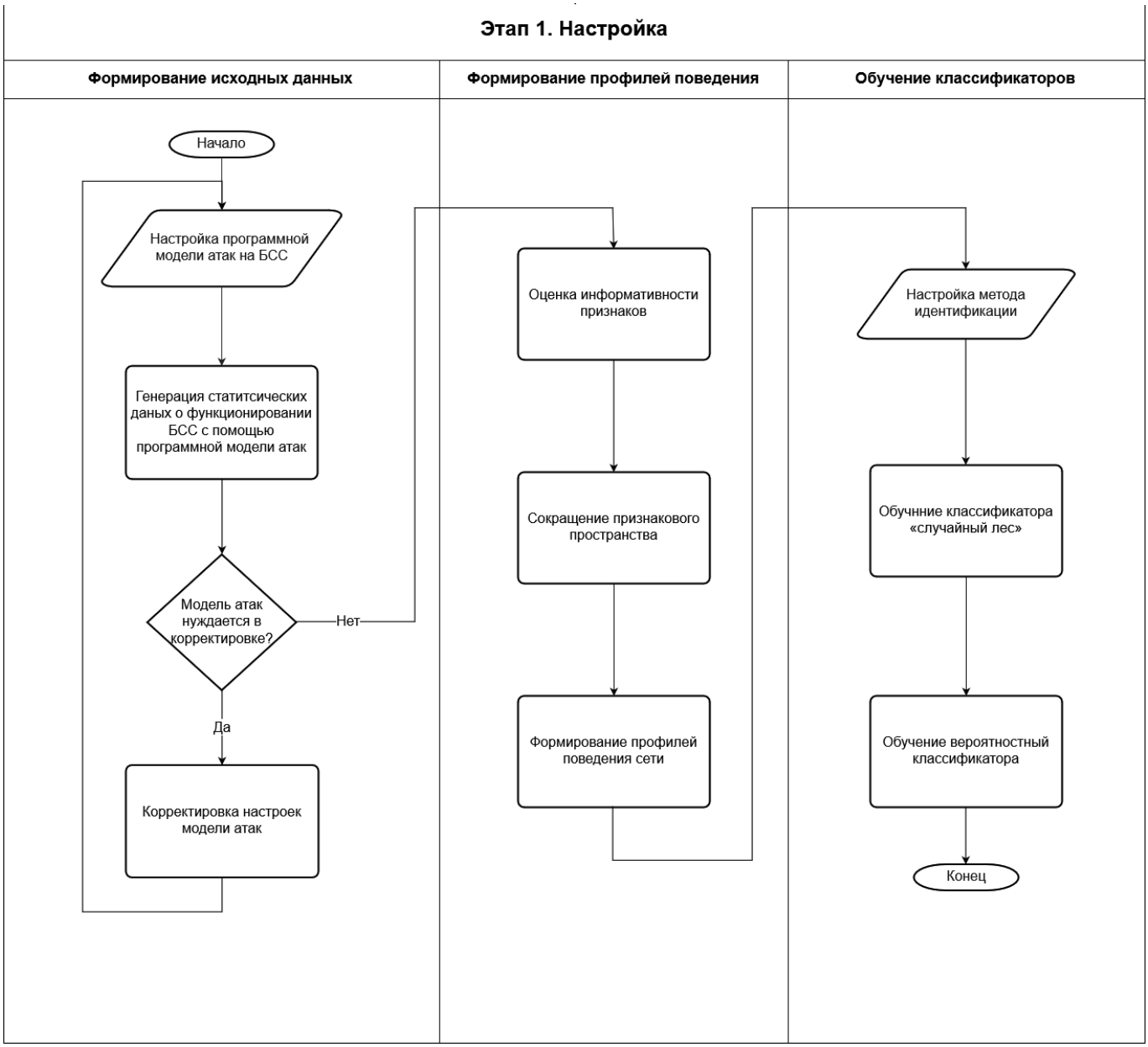

Рис. 2. Этап настройки методики идентификации

вания были разработаны: программная модель проведения атак на БСС (№ 2018617190 от 20.06.2018) [5], позволяющая получить набор данных для анализа; программа подсчета информативности признаков статистической выборки (№ 2018618975 от 24.07.2018), необходимая для автоматизированного сокращения признакового пространства идентификации [8]; и программа вычисления апостериорных распределений дискретного параметра распределения многомерной случайной величины по статистической выборки (№ 2018619014 от 25.07.2018), используемая на этапе обучения вероятностного классификатора и формирования рекомендаций для его применения. Более подробные сведения возможно найти в предыдущих работах автора.

Методика идентификации атак

Разработанная методика описывает процесс идентификации атак на БСС на основе анализа поведения сети [6].

Общая схема методики представлена на рисунке 1.

Подготовительный этап методики начинается с настройки программной модели атак на беспроводные

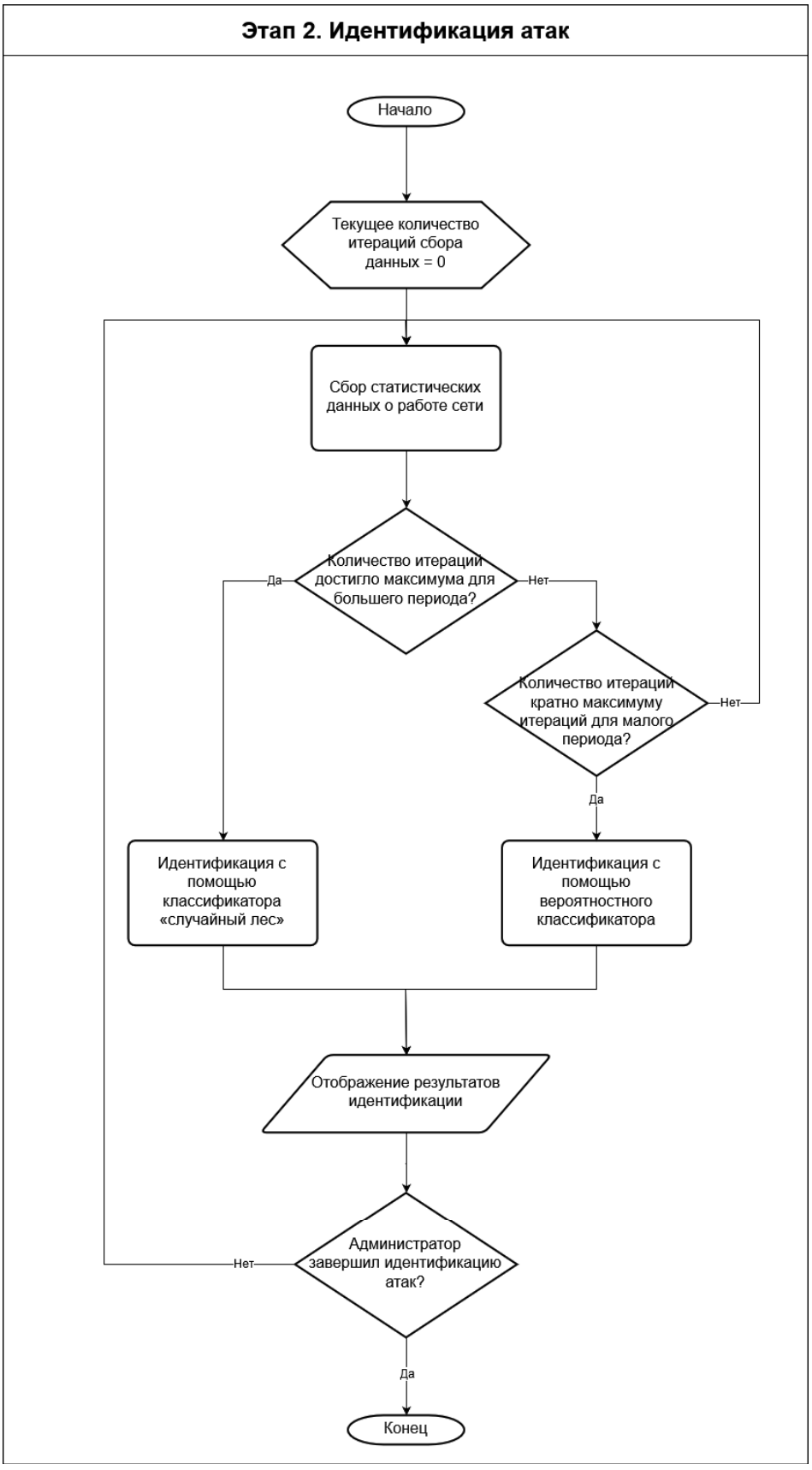

Рис. 3. Этап идентификации атак

сенсорные сети в соответствии с характеристиками защищаемой сети (рис. 2). При этом, возможно корректировка и изменение таких параметров, как: топология сети (ячеистая или кластерная); количество узлов в сети; профили поведения сети (нормального поведения и атак, которые необходимо идентифицировать); и набор признаков, характеризующих поведение сети.

После настройки программной модели осуществляется генерация статистических данных о нормальном поведении сети и поведении сети под атаками. Возможна корректировка модели поведения сети в соответствии с реальными данными в случае необходимости. На основе полученных данных далее формируется модель профиля поведения сети. Данный подпункт включает в себя и снижение размерности набора признаков, необходимых для идентификации. Это реализуется с помощью вышеуказанной программы, которая подсчитывает информативность признаков по формулам Шеннона, Кульбака и методу накопленных частот. После оценки корреляции между признаками, применяется алгоритм машинного обучения из библиотеки SciKitLearn для финального, минимального достаточного для идентификации набора признаков.

После формирования модели профиля поведения и оценки информативности, происходит формирование определенных профилей поведения сети и, при необходимости, занесение сформированных профилей в базу данных.

Следующий шаг — настройка метода идентификации атак на БСС. На данном шаге указываются следующие:

- параметра степени уверенности;
- период сбора данных с узлов сети;
- больший и меньший временные периоды для сбора данных о состоянии сети в целом.

Следует отметить, что период характеризует количество итераций сбора данных о работе сети. Рекомендуется для большого временного периода выбирать $T\_b = 360T$, а для малого $T\_s = 60T$. T задается администратором сети (предыдущие работы автора предлагают $T=10$ c). Два периода стоит выбирать только в том случае, если выбран подход с последовательным применением вероятностного классификатора и «случайного леса». В другом случае можно использовать один временной период. Имеет смысл выбирать величину малого периода такой, чтобы она была кратна большему периоду.

Выбор априорной вероятности нормального состояния. В результате проведения рядя экспериментов (для обеспечения высокой точности классификации нормального поведения сети и 14 атак) установлено, что необходимо, чтобы значение параметра степени уверенности было выше значения апостериорной вероятности нормального состояния на ~30% и нет необходимости в использовании апостериорной вероятности нормального состояния, превышающей 20%.

После этого происходит обучение алгоритма «случайный лес» и вероятностного классификатора.

Непосредственно этап идентификации атак представлен на рисунке 3.

Предполагается, что для идентификации можно использовать любой из упомянутых классификаторов или же использовать их совместно. Вероятностный классификатор стоит использовать в тех случаях, когда важную роль играет энергоэффективность: он позволяет идентифицировать атаки на неполном наборе признаков с заданной точностью (экспериментально было выявлено, что возможно достижение точности 95% при использовании только одного признака). Этот вариант применим, когда, например, узлы сети не обладают постоянным источником питанием (в частности, узлом, осуществляющим мониторинг поведения, может быть шлюз, собирающий данные о поведении сети, также работающий от собственного источника питания). Классификатор «случайный лес» менее энергоэффективен, но в то же время показывает более высокую точность работы: он демонстрирует хорошие результаты на наборе статистических данных, собранных за более длительный промежуток времени (экспериментально было выявлено, что длительность периода сбора статистики, при которой достигается максимальная точность, равняется $360T$ (одному часу при соответствующих настройках)). В связи с этим, предполагается целесообразным совместное использование двух классификаторов: на длительном промежутке «случайный лес», а на более коротких — вероятностный — для примерной оценки поведения сети. Следующую конфигурацию методов идентификации атак следует рассматривать в качестве рекомендаций.

Анализ неполного набора признаков из профиля поведения сети вероятностным классификатором производится каждые T_s. В этом случае вероятностный классификатор будет работать с растущим объемом набора данных: T_s, $2*T\_s$, $3*T\_s$. nT_s, где $nT\_s < T\_b$.

Результаты анализа предоставляются администратору сети. При соответствии показателя уверенности и результатов сравнения признака(ов) выводится результат идентификации (поведение под атакой или наиболее вероятные варианты атак, близкие к идентифицируемой по исследуемому признаку). В случае, если результатов достаточно и идентификация является положительной

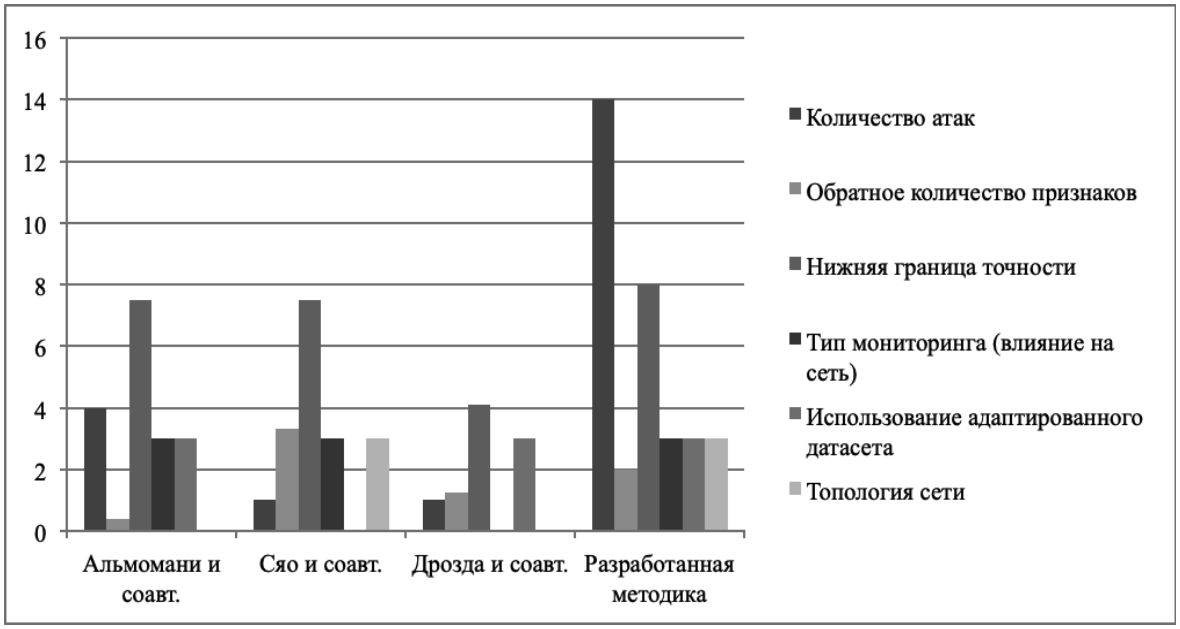

Рис. 4. Сравнение эффективности разработанной методики с существующими исследованиями.

(вероятность конкретного поведения больше заданного параметра степени уверенности), администратор может завершить этап идентификации атак и, в случае необходимости, перейти к мерам по противодействию атаке. Если полученные результаты меньше или равны показателя степени уверенности, то запрашивается дополнительный признак из признакового пространства и происходит следующая попытка идентификации. При несоответствии показателя уверенности и использовании всех признаков состояния сети происходит завершение анализа поведения вероятностным классификатором. Возможно также завершение процесса идентификации, если администратор сети предполагает, что полученных результаты являются удовлетворительными.

Необходимо отметить, что классификация изначально осуществляется на небольших по сравнению с T_b периодах сбора данных. Поэтому на начальном этапе работы стоит выбирать не столь большие величины параметра степени уверенности: при большом значении параметра степени уверенности, например, 85–99%, классификатор будет использовать в среднем больше признаков для классификации, что увеличит нагрузку на устройства. Результаты на небольших наборах данных, как показывает практика, менее информативны, поэтому расходование энергии устройств на этом этапе нецелесообразно. Тем не менее, на каждой последующей итерации работы вероятностного классификатора набор данных и, соответственно, информативность признаков будут расти. В связи с этим предлагается использование прогрессирующего параметра степени уверенности от 60% до 90%.

После этого производится анализ полного набора признаков профиля поведения, собранного за T_b и выполняется классификация с помощью классификатора «случайный лес». При положительной идентификации выводится результат (тип атаки), при невозможности идентификации выводится сообщение о неизвестном аномальном поведении.

Оценка качества разработанной методики производилась в сравнении с существующими исследованиями [3, 4, 9]. Диаграмма оценки представлена на рисунке 4. Шкала слева отражает наиболее важный параметр — количество идентифицируемых атак. Также числовыми являются обратное количество признаков и нижняя граница точности, умноженные на 10 для облегчения визуального восприятия. Тип мониторинга (пассивный = 1, активный = 0), топология сети и использование адаптированного набора данных являются бинарными признаками (увеличены в 3 раза для облегчения восприятия).

В соответствии с диаграммой, можно судить о повышении эффективности идентификации атак сетевого уровня на беспроводные сенсорные сети.

Выводы

В статье сформулирована проблема обеспечения доступности и целостности беспроводной сенсорной сети, связанная с анализом поведения сети.

Кратко представлены составные элементы методики, более подробное описание которых приводится в предыдущих работах автора.

Предложенная методика позволяет применять предложенную модель профиля поведения сети, разработанное программное обеспечение и метод идентификации атак сетевого уровня на основе поведенческого анализа для повышения уровня целостности и доступности КФС, в основе которых лежат БСС.

ЛИТЕРАТУРА

1. Грозных А. В. Разработка алгоритма обнаружения вторжений в беспроводных сенсорных сетях на основе вероятностного классификатора // Выпускная квалификационная работа бакалавра. — 2018. — Университет ИТМО.
2. Шилов И. М. Оценка аномального поведения узлов беспроводной сенсорной сети на основе статистических методов // Выпускная квалификационная работа бакалавра. — 2017. — Университет ИТМО
3. Almsman I., Al-Kasasbeh B., AL-Akhras M. WSN-DS: A Dataset for Intrusion Detection Systems in Wireless Sensor Networks // Journal of Sensors. Volume 2016, Article ID4731953, 16 pages
4. Drozda M., Bate I., Timmis J. Bio-inspired error detection for complex systems, in: 17th Pacific Rim International Symposium on Dependable Computing, Pasadena, CA, USA, 2011, pp. 154–163.
5. Korzhuk V., Krivtsova I., Shilov I. The Model of the Attack Implementation on Wireless Sensor Networks // Proceedings of the 20th Conference of Open Innovations Association FRUCT — 2017, pp. 187–194
6. Korzhuk V., Groznykh A., Menshikov A., Strecker M. Identification of Attacks against Wireless Sensor Networks Based on Behaviour Analysis // Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications — 2019, Vol. 10, No. 2, pp. 1–21
7. Shams B., Alrajeh N. A., Khan S. Intrusion detection systems in wireless sensor networks: A review. Inter- national Journal of Distributed Sensor Networks, 9(5), 2013.
8. Shilov I., Korzhuk V., Torshenko J. Reduction of the Feature Space for the Detection of Attacks of Wireles Sensor Networks // Proceedings of the 20th Conference of Open Innovations Association FRUCT — 2017, pp. 195–201
9. Xiao Z., Liu C., Chen C. An anomaly detection scheme based on machine learning for wsn, in: 1st International Conference on Information Science and Engineering, Nanjing, China, 2009, pp. 3959–3962
10. Zikratov I.A., Korzhuk V., Shilov I., Gvozdev A. Formalization of the Feature Space for Detection of Attacks on Wireless Sensor Networks // Proceedings of the 20th Conference of Open Innovations Association FRUCT — 2017, pp. 526–533

Журнал «Современная наука: актуальные проблемы теории и практики»

Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики