УДК: 004.056

МЕТОДИКА ФОРМИРОВАНИЯ БАЗЫ КЛАССИФИКАЦИЙ КОМПЬЮТЕРНЫХ АТАК НА ОСНОВЕ ПРИМЕНЕНИЯ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА СИГНАТУР КОМПЬЮТЕРНЫХ АТАК

METHODOLOGY FOR CREATING A DATABASE OF COMPUTER ATTACK CLASSIFICATIONS BASED ON THE USE OF INTELLIGENT ANALYSIS OF COMPUTER ATTACK SIGNATURES

Канд. воен. наук А.Н. Стадник, Е.В. Алпеев, д-р техн. наук С.В. Скрыль

PhD A.N. Stadnik, E.V. Alpeev, DPhil. S.V. Skryl'

Краснодарское высшее военное училище им. С.М. Штеменко

Представлена структура типовой системы обнаружения компьютерных атак, функционирующая на основе базы сигнатур компьютерных атак. Разработана концептуальная модель системы обнаружения компьютерных атак с применением интеллектуального анализа данных, которая основана на применении сигнатурного, поведенческого и интеллектуального анализа событий информационной безопасности. Предложен подход интеллектуального анализа данных, заключающийся в обучении на основе существующих сигнатур компьютерных атак с целью дальнейшего формирования базы классификаций компьютерных атак. Разработана методика формирования базы классификаций компьютерных атак на основе применения интеллектуального анализа сигнатур компьютерных атак.
Ключевые слова: система обнаружения компьютерных атак, интеллектуальный анализ данных, дерево принятия решений, сигнатуры компьютерных атак, компьютерная атака.

The structure of a typical computer attack detection system based on a database of computer attack signatures is presented. A conceptual model of the system for detecting computer attacks using data mining has been developed, which is based on the use of signature, behavioral and intelligent analysis of information security events. An approach of data mining is proposed, which consists in training on the basis of existing signatures of computer attacks in order to further form a database of classifications of computer attacks. The method of forming a database of computer attacks based on the use of intelligent analysis of computer attack signatures is developed.
Keywords: computer attack detection system, data mining, decision tree, computer attack signatures, computer attack.

Введение

Глобальное единое информационное пространство, на современном этапе развития информационного общества, позволило обеспечить доступ к информационным ресурсам практически из любой точки земного шара и стало неотъемлемой частью жизнедеятельности любой страны, высшего военно-политического руководства, общества и всех сфер жизнедеятельности людей. Невозможно представить современного человека без доступа к сети «Интернет». В настоящее время данная сеть охватывает уже практически людей всех возрастов. В связи с этим глобальное

информационное пространство является и сферой разделения влияния между государствами, где ведется информационное противоборство с целью достижения политических, экономических, психологических, военных и иных целей. Вопросы противостояния, в том числе и доминирования в этой сфере, проработаны на уровне государственных доктрин, концепций и иных основополагающих документов. Киберпреступность, кибертерроризм также активно эксплуатируют глобальную сеть для достижения своих целей.

Система обнаружения вторжения (система обнаружения компьютерных атак) — это программное или программно-техническое средство, реализующее функции автоматизированного обнаружения (блокирования) действий в информационной системе, направленных на преднамеренный доступ к информации, специальные воздействия на информацию (носители информации) в целях ее добывания, уничтожения, искажения, и блокирования доступа к ней [1].

Детектор атак (сенсор) — это программный или программно-аппаратный компонент комплекса средств системы обнаружения вторжений, осуществляющий выявление компьютерных атак на основе анализа сетевого трафика [2]. Анализ данных с целью обнаружения вторжений осуществляется с использованием сигнатурного и эвристического методов. Метод сигнатурного анализа основан на применении набора решающих правил, предварительно загруженных в базу данных.

Обобщенная модель системы обнаружения атак

Как уже было сказано ранее, ключевым элементом системы обнаружения компьютерных атак является детектор атак. С целью построения концептуальной модели, основанной на сочетании существующих подсистем и добавления подсистемы обнаружения компьютерных атак, основанной на применении интеллектуального анализа данных, необходимо представить формальное описание существующей системы обнаружения компьютерных атак. Обобщенная структурно-логическая модель системы обнаружения атак (СОА) представлена на рис. 1.

Основными подсистемами в данной модели являются:

1. Подсистема авторизации;
2. Подсистема управления;
3. Подсистема обеспечения;
4. Подсистема обнаружения компьютерных атак;
5. Подсистема анализа событий информационной безопасности;
6. Подсистема захвата трафика.

Все указанные подсистемы выполняют следующие функции и задачи: функции мониторинга и анализа трафика; выявление признаков сетевых атак и аномалий; формирование сообщений о событиях информационной безопасности; обновление базы решающих правил; обеспечение экспорта данных; формирование реакции на события информационной безопасности [3].

Основой функционирования СОА является база сигнатур компьютерных атак. В настоящее время большинство систем обнаружения работают по принципу обнаружения компьютерных атак на основе сравнения их с сигнатурами компьютерных атак.

Сигнатура — характерные признаки вторжения (атаки), используемые для его (её) обнаружения [1].

С целью обнаружения компьютерных атак, которые отсутствуют в базе сигнатур компьютерных атак, а также которые не выявлены в ходе обнаружения признаков аномального поведения объектов сети, целесообразно дополнить существующую модель подсистемой интеллектуального обнаружения компьютерных атак.

Концептуальная модель системы обнаружения компьютерных атак с применением интеллектуального анализа данных

Для того, чтобы повысить вероятность обнаружения компьютерных атак предлагается разработать подсистему интеллектуального обнаружения компьютерных атак. Структурно, данная подсистема будет состоять из следующих модулей:

– модуль интеллектуального анализа сигнатур компьютерных атак;

– модуль прогнозирования компьютерных атак;

– модуль параллельного анализа данных.

Концептуальная модель системы обнаружения компьютерных атак представлена на рис. 2.

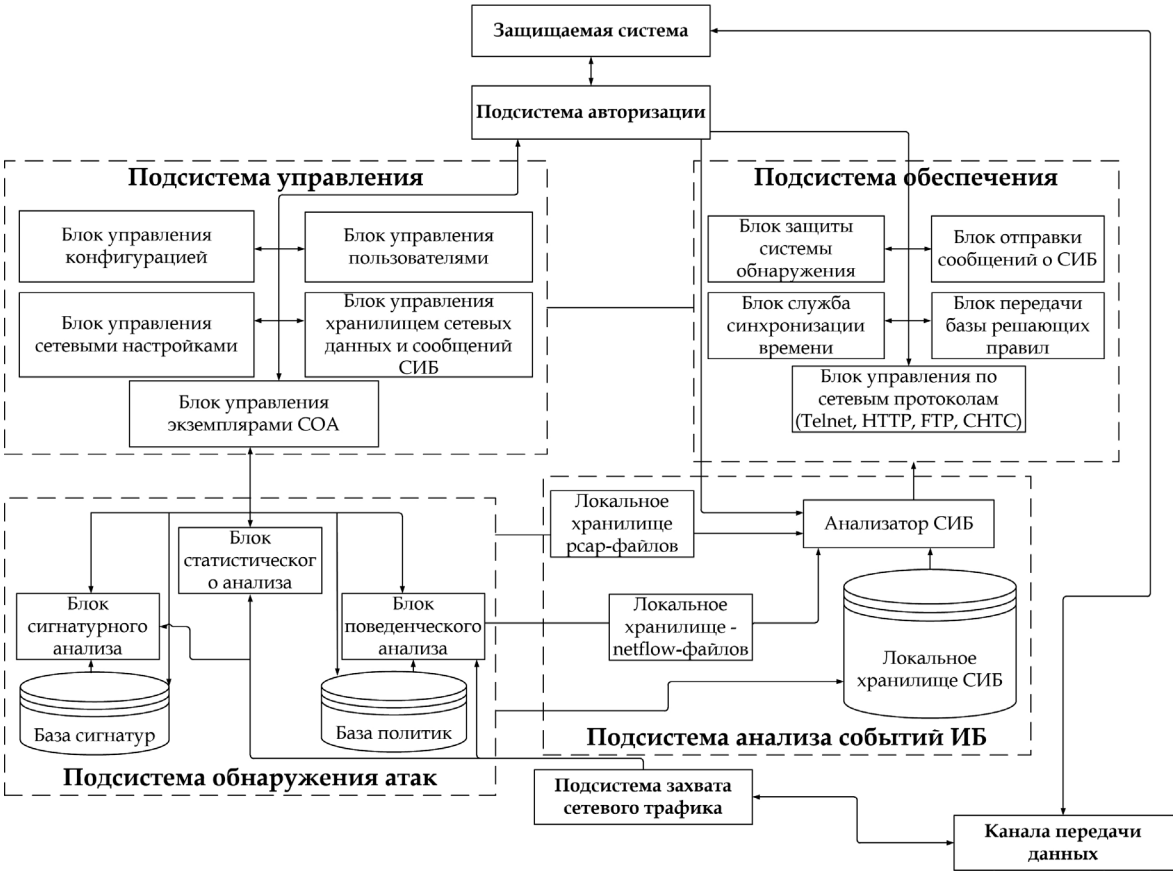

Рис. 1. Структурно-логическая модель системы обнаружения компьютерных атак

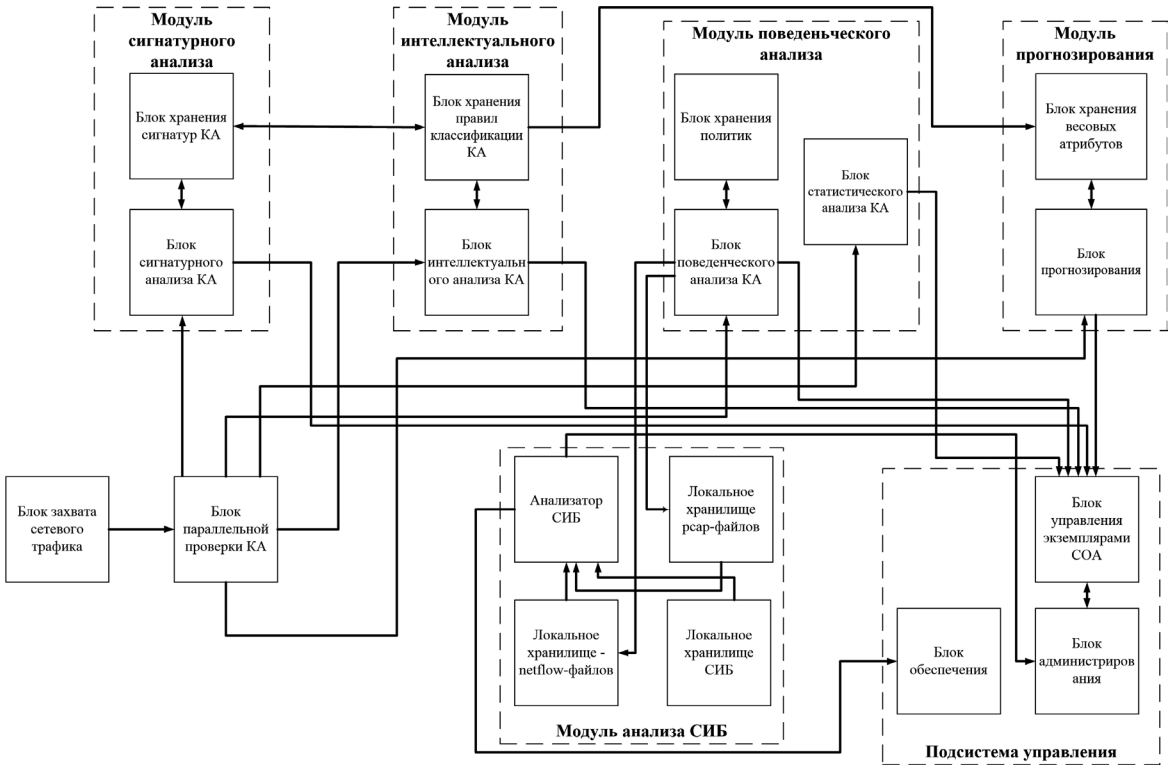

Рис. 2. Структурно логическая модель системы обнаружения компьютерных атак

В данной статье предлагается первая часть решения вопроса построения системы обнаружения компьютерных атак с применением интеллектуального анализа данных, а именно процесс функционирования модуля интеллектуального анализа.

Методика формирования базы классификаций компьютерных атак на основе применения интеллектуального анализа сигнатур компьютерных атак

Дерево принятия решений.

В общем смысле обнаружение компьютерных атак в настоящее время производится преимущественно с применением существующих сигнатур компьютерных атак. Сигнатурный метод ориентирован на уже известные образы компьютерных атак существующих в базах сигнатур. В том случае, когда сценарий компьютерной атаки модифицирован, либо поведение системы не претерпевает существенных изменений в результате воздействия атаки, то пропуск новых (либо модифицированных) компьютерных атак вполне вероятен. В данном случае необходимо разработать механизмы, которые будут способны обнаружить компьютерную атаку в случае, когда атака отсутствует в базе сигнатур и поведение системы находится в допустимых пределах.

В связи с возникающей необходимостью постоянно совершенствовать системы обнаружения, имеет смысл реализации механизмов, основанных на интеллектуальном анализе. Методов интеллектуального анализа данных в настоящее время существует значительное количество, в данной статье более подробно рассмотрим интеллектуальный анализ данных, основанный на дереве принятия решений.

Дерево принятия решений, в первую очередь, предназначено для классификации данных. Дерево принятия решений — это связный ациклический граф, представляющий правила классификации в иерархической последовательной структуре. Дерево состоит из узлов и листов. Исходными данными (база знаний) для дерева принятия решений является база данных, в которой каждому классу сопоставлены определенные свойства (признаки). На основе обучающей выборки и правил построения дерева принятия решений производится построение дерева для каждого класса из обучающей выборки. Построенные деревья, которые являются классификаторами каждого класса, помещаются в базу классификаций. В момент, когда поступает объект с определенными свойствами, его свойства сопоставляются с каждым деревом и если все его свойства совпали, то его соотносят к определенному классу [4–6].

Пример описания реализации компьютерной атаки.

Для того, чтобы описать признаки (свойства) компьютерных атак приведем один из примеров ее реализации для дальнейшего описания порядка ее обнаружения с помощью интеллектуального анализа сигнатур [7].

Рассмотрим достаточно распространённый пример атаки с непредвиденными параметрами сетевых пакетов. Если обнаружен сетевой пакет с установленными битами SYN и ACK (второй этап установления виртуального TCP-соединения), и при этом не было никакого предшествующего пакета с битом SYN (первый этап установления виртуального TCP-соединения), то это может скрывать под собой несанкционированное вторжение. Указанный метод получил применение при так называемом stealth-сканировании. Помимо битов SYN/ACK в заголовке TCP-пакета использовались также биты FIN (FIN-сканирование), RESET (RESET-сканирование). Листинг Stealth-сканирования с помощью SYN/ACK (фрагмент журнала регистрации TCPdump):

06:41:24.067330 stealth.mappem.com.113 > 172.21.32.83.1004- S

4052190291:4052190291(0) **ack** 674711610 win 8192

06:42:08.063341 stealth.mappem.com.113 > 192.168.83.15.2039: S

2335925210:2335925210(0) **ack** 674711610 win 8192

14.582943 stealth.mappem.com.113 > 172.21.64.120.2307: S

2718446928:2718446928(0) **ack** 674711610 win 8192.

Любой пакет, который не соответствует стандартам RFC, может привести к выходу из строя коммуникационного оборудования, которое этот пакет обрабатывает. Причем к такому оборудованию относятся не только маршрутизаторы или коммутаторы, но и межсетевые экраны

и системы обнаружения атак. Многие атаки используют запрещенные комбинации TCP-флагов в сетевых пакетах. Некоторые комбинации приводят к выходу из строя узла, осуществляющего обработку таких пакетов, а благодаря иным комбинациям другие пакеты остаются незамеченными некоторыми системами обнаружения атак или межсетевыми экранами. В RFC 793 описано, как должны реагировать различные системы на нормальные TCP-пакеты. Но в этом (и других) документе не сказано, как система должна реагировать на неправильные TCP-пакеты, в результате этого различные устройства и ОС по-разному реагируют на пакеты с запрещенными комбинациями TCP-флагов. Существуют 6 флагов, которые могут встретиться в TCP-пакете: SYN, ACK, FIN, RST, PSH, URG. Запрещенные комбинации можно обнаружить по хотя бы одному из перечисленных ниже признаков. Рассмотрим вариант сочетания флагов SYN + FIN. Поскольку это два взаимоисключающих флага. Первый устанавливает соединение, а второй завершает его. Такая комбинация очень часто используется различными сканерами, например, N_{map}. Некоторое время назад большое число систем обнаружения атак не могло обнаружить такого рода сканирования. Однако сейчас ситуация изменилась к лучшему, многие системы обнаружения атак отслеживают подобные комбинации флагов. Но добавление еще одного флага к данной комбинации (например, SYN + FIN + PSH, SYN + FIN + + RST, SYN + FIN + RST + PSH) опять приводит к тому, что некоторые системы обнаружения атак не распознают видоизмененное сканирование. Некоторые аналитики называют такого рода комбинации «шаблоном рождественского дерева» («Christmas Tree Pattern»), листинг приведен ниже [7]:

01/23-01:15:22.237103 195.11.212.180:30975 -> 192.0.97.80:49708

TCP TTL:49 TOS:0x0 ID:12207 DF

SERPAU21 Seq: 0x78FFC22C Ack: 0x78FFC22C Win: 0xC22C

TCP Options => Opt 120 (40): C22C 78FF C22C 78FF C22C 78FF

0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 78 FF C2 2C 78 FF x..,x.

01/23-01:15:43.538590 195.11.212.180:30975 -> 192.0.97.80:49708

TCP TTL:49 TOS:0x0 10:13449 DF

SFRPAU21 Seq: 0x78FFC22C Ack: 0x78FFC22C Win: 0xC22C

TCP Options => Opt 120 (40): C22C 78FF C22C 78FF C22C 78FF

0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 78 FF C2 2C 78 FF x..,x.

Данный пример говорит о том, что компьютерные атаки обладают определенными признаками и свойствами, что позволяет обнаруживать их с помощью сигнатурного метода. Однако небольшие изменения или дополнения этих признаков могут повлиять на необнаружение компьютерной атаки в конечном итоге. В связи с этим существует необходимость применения интеллектуальных методов обнаружения компьютерных атак.

Реализация механизма дерева принятия решений для применения в системе обнаружения компьютерных атак

Исходными данными для обучения нашей подсистемы интеллектуального анализа данных, основанного на дереве принятий решений, является база сигнатур компьютерных атак. В данной базе описаны сценарии атак, которые в свою очередь относятся к определенным классам атак, которые обладают определенными свойствами.

Чтобы классифицировать атаку, необходимо спуститься по дереву, описанному для определенного класса атаки, до листа и выдать значение метки класса. Общий принцип построения деревьев решений, заключается в рекурсивном разбиении множества объектов из обучающей базы сигнатур компьютерных атак на подмножества (классы компьютерных атак), содержащие объекты, относящиеся к таковым классам.

Относительно обучающей выборки в виде базы сигнатур компьютерных атак, которую обозначим, как B_{sign} и множества классов компьютерных атак $C_1, C_2, C_3, ..., C_k$ возможны три ситуации:

1. Множество B_{sign} может включать в себя одну или более сигнатур компьютерных атак, относящихся к одному из классов компьютерных атак C_i. Тогда дерево принятия решений для B_{sign} — это лист, определяющий класс компьютерной атаки B_{sign};

2. Множество сигнатур компьютерных атак B_{sign} является пустым множеством, несодержа-

щим в себе ни одной сигнатуры компьютерной атаки, что означает, что множество сигнатур B_{sign} — это лист и, соответственно, класс, состоящий из одного листа, который выбирается из другого множества сигнатур, отличающегося от B_{sign}, например, из множества, ассоциированного с родителем;

3. Множество сигнатур компьютерных B_{sign} включает в себя объекты, относящиеся к разным классам компьютерных атак. В этом случае следует разбить множество B_{sign} на некоторые подклассы. Для этого выбираются атрибуты (свойства) $A_1, A_2, A_3, ..., A_k$ и B_{sign} разбивается на подмножества $B_1, B_2, B_3, ..., B_j$, по выбранному атрибуту и соответствующему ему условию.

Указанные шаги выполняются рекурсивно, в результате чего должны получиться подмножества, которые в конечном итоге являются листами.

Алгоритмы автоматического построения деревьев принятия решений относятся к индуктивным алгоритмам машинного обучения — «обучения с экспертом».

Ключевым аспектом построения деревьев принятия решений для обнаружения неизвестных компьютерных атак является выбор таких условий в узлах, которые обеспечивают наилучшее разбиение базы сигнатур компьютерных атак на классы компьютерных атак. Это необходимо для того, чтобы получившиеся деревья были компактным, с минимальным количеством узлов и обеспечивали максимальную точность при классификации компьютерной атаки.

Математический аппарат построения дерева принятия решения.

Необходимо реализовать алгоритм, который строит дерево принятия решений и подбирает из всех возможных вариантов такую проверку в узле, которая обеспечивает максимальное снижение информационной двоичной энтропии и равномерно распределяет компьютерные атаки по соответствующим данным атакам классам.

Если перемешать объекты из множества B_{sign} и расположить их в ряд, то мы реализуем операцию перестановки, при условии, что сигнатуры компьютерных атак с одинаковыми метками класса идентичны. Тогда количество уникальных перестановок для множества рассчитывается, как:

$$N = \frac{B_{\text{sign}}!}{B_1! \cdot B_2! \ldots} = \frac{B_{\text{sign}}!}{\prod_{i=1}^{k} B_i!},$$

где B_i — количество объектов с меткого класса C_i в множестве B_{sign}, B_{sign} — количество объектов в обучающей выборке.

Если все уникальные перестановки пронумеровать числами от 0 до $W-1$, то количество бит необходимое для кодирования каждого уникального варианта перестановки равно $\log_2(B_{\text{sign}})$. Среднее количество бит на каждую перестановку называется комбинаторной энтропией и рассчитывается как:

$$E = \frac{\log_2(N)}{B_{\text{sign}}} = \frac{1}{B_{\text{sign}}} \cdot \log_2\left(\frac{B_{\text{sign}}!}{\prod_{i=1}^{k} B_i!}\right). \quad (1)$$

Ключевым правилом комбинаторной энтропии при построении дерева принятия решения является — чем меньше энтропия, тем однороднее множество. Самыми однородными множествами в дереве принятия решений являются листы, не содержащие примесей других классов компьютерных атак. Самым неоднородным множеством является база сигнатур компьютерных атак.

Формулу (1) можно преобразовать, применив формулу Стирлинга:

$$\ln B_{\text{sign}}! = B_{\text{sign}} \cdot \ln B_{\text{sign}} - B_{\text{sign}} + O(\ln N) \approx$$
$$\approx B_{\text{sign}} \cdot \ln B_{\text{sign}} - B_{\text{sign}}.$$

Тогда энтропию для множества сигнатур компьютерных атак B_{sign} с учетом упрощений можно рассчитать так:

$$E(B_{\text{sign}}) = -\sum_{i=1}^{k}\left(\frac{B_i}{N} \cdot \log_2 \frac{B_i}{N}\right).$$

Так как значение энтропии *E* находится в зависимости от множества сигнатур компьютерных атак B_{sign}, ее можно выразить через функцию:

$$E(B_{\text{sign}}) = -\sum_{i=1}^{k}\left(\frac{f(B_{\text{sign}}, C_i)}{|B_{\text{sign}}|} \cdot \log_2 \frac{f(B_{\text{sign}}, C_i)}{|B_{\text{sign}}|}\right). \quad (2)$$

где $f(B_{\text{sign}}, C_i)$ — функция, которая определяет количество сигнатур с меткой класса компьютерной атаки C_i в множестве сигнатур компьютерных атак B_{sign}.

При формировании дерева принятия решения в каждом узле размещается условие, которое делит исходное множество B_{sign} на несколько j классов компьютерных атак:

$$B_{\text{sign}} = B_1 \bigcup B_2 \bigcup ... \bigcup B_j.$$

На основании этого для каждого класса компьютерной атаки $B_1, B_2, ..., B_j$ по формуле (2) можно вычислить энтропию. Тогда энтропию множества сигнатур компьютерных атак B_{sign} после разбиения на классы рассчитывается как:

$$E_0(B_{\text{sign}}, B_1, B_2, ..., B_j) = \sum_{i=1}^{j} \left(\frac{|B_i|}{|B_{\text{sign}}|} \cdot E(B_i) \right).$$

На следующем этапе необходимо рассчитать прирост информации, обеспечивающийся данным условием в узле, который рассчитывается как разность энтропии $E(B_{\text{sign}})$ и энтропии разбиения $E_0(B_{\text{sign}}, B_1, B_2, ..., B_j)$:

$$I(B_{\text{sign}}, B_1, B_2, ..., B_j) = \\ = E(B_{\text{sign}}) - \sum_{i=1}^{j} \left(\frac{|B_i|}{|B_{\text{sign}}|} \cdot E(B_i) \right).$$

Однородность класса компьютерной атаки, сигнализирует о том, что большой прирост информации обеспечивает данное разбиение.

Если постоянно проводить проверку условий в узле, которые обеспечивают максимальный прирост информации — это приводит к проблеме их «переобучения». В результате ассиметричных разбиений, листы будут получаться путем отсеивания в них по несколько условий, а это означает, что конечное дерево будет включать в себя большое количество листов, к каждому из которых будет относиться только несколько экземпляров компьютерных атак. Такие деревья будет работать с низкой точностью, на примерах, не содержащихся в обучающейся выборке:

$$I(B_{\text{sign}}, B_1, B_2, ..., B_j) \to M_{\text{porog}},$$

где M_{porog} — максимальный прирост информации.

Чтобы избежать проблемы «переобучения» необходимо ввести параметр нормирования L:

$$L(B_{\text{sign}}, B_1, B_2, ..., B_j) = \\ = -\sum_{i=1}^{j} \left(\left(\frac{|B_i|}{|B_{\text{sign}}|} \right) \cdot \log_2 \left(\frac{|B_i|}{|B_{\text{sign}}|} \right) \right).$$

Чем более равномерные по количеству объектов получаются классы, тем меньше значение L. Нормировка прироста информации осуществляется следующим образом:

$$I_{\text{norma}}(B_{\text{sign}}, B_1, B_2, ..., B_j) = \\ = \frac{I(B_{\text{sign}}, B_1, B_2, ..., B_j)}{L(B_{\text{sign}}, B_1, B_2, ..., B_j)}.$$

Таким образом, для каждого узла дерева принятия решений, помещаемое в него условие должно обеспечивать максимальное значение нормированного прироста информации:

$$I_{\text{norma}}(B_{\text{sign}}, B_1, B_2, ..., B_j) = G_{\text{porog}},$$

где G_{porog} — максимальное значение нормированного прироста информации.

Данный выбор условий для каждого узла обеспечивает максимальное снижение энтропии при сбалансированном (по количеству экземпляров компьютерных атак) разбиении исходного множества сигнатур компьютерных атак на классы компьютерных атак.

Энтропия будет нулевая, если в результате построения дерева принятия решения удалось достигнуть листов без содержания примесей посторонних классов.

Для каждого категориального атрибута существует только один вариант разбиения — когда исходное множество сигнатур компьютерных атак B_{sign} делиться на классы в количестве равном значениям данного атрибута.

Для каждого параметра (свойства, признака) компьютерной атаки существует несколько вариантов разбиений, количество которых равно мощности множества порогов. Множество порогов находится путем записи всех уникальных значений данного параметра (свойства, признака) в обучающей выборке в порядке возрастания без повторов с отбрасыванием наибольшего зна-

Таблица

Результаты обнаружения атак и определения ошибок I и II рода

Класс атаки	Количество атак	Ошибки I рода	Ошибки II рода
Normal	95778	75	15
DoS	397508	9	19
Prob	5538	9	55
R2L	1117	6	18
U2R	59	30	19
Общее количество:	500000	129	126

чения. Затем каждое значение порога H выступает в роли точки деления исходного множества сигнатур компьютерных атак B_{sign}, состоящего из объектов m, на подмножества B_1 и B_2, что указано в следующем выражении:

$$\begin{cases} B_1 = m \in B_{\text{sign}} \mid \alpha(m) \le H; \\ B_2 = m \in B_{\text{sign}} \mid \alpha(m) > H, \end{cases} \qquad (3)$$

где α — значение атрибута компьютерной атаки, по которому производится разбиение. В конечном итоге из всех возможных вариантов разбиения выбирается то, которое соответствует критерию (3). Построение дерева ведется до тех пор, пока не будут найдены все листы.

Реализация предложенных решений.

В ходе выполнения исследований в среде моделирования Matlab была разработана и спроектирована программа для построения деревьев принятия решений в соответствии с математическим описанием указанном ранее.

На основе разработанной программы проведен эксперимент, исходными данными, которого явились данные, полученные из общедоступной базы компьютерных атак KDD-99Cup [8, 9]. Для эксперимента было задействовано 500000 компьютерных атак из тестового набора указанной базы данных. Результаты функционирования построенных на основе обучающей выборки деревьев принятия решений были проверены с помощью генератора атак и определения ошибок I и II рода. Полученные результаты приведены в таблице [10].

Выводы

Полученная методика позволяет на основе существующей базы сигнатур компьютерных атак, которая используется как обучающая выборка для построения деревьев принятия решений, которые являются классификатором компьютерных атак и которые формируются базу классификаций компьютерных атак, разрешить следующие проблемы:

– повысить вероятность обнаружения компьютерных атак, в том случае, когда в базе сигнатур компьютерных атак отсутствует в чистом виде ее сигнатура, но сверив признаки аномалии, появляется возможность ее идентифицировать, опираясь на базу классификаций, полученную с помощью деревьев принятия решений;

– выступить, как резервный источник обнаружения компьютерных атак, в том случае, когда по какой-либо причине база сигнатур компьютерных атак вышла из строя, либо к ней отсутствует доступ.

Таким образом, представленная методика закладывает перспективу для разработки методики прогнозирования компьютерных атак на основе установления весов атрибутам компьютерных атак, для тех случаев, когда удается идентифицировать часть атрибутов компьютерной атаки, а не все полностью, а также перспективу разработки модели параллельной идентификации компьютерной атаки на основе сигнатурного и интеллектуального методов с целью повышения скорости идентификации компьютерной атаки.

Литература

1. Методический документ ФСТЭК России. Профиль защиты системы обнаружения вторжений уровня сети пятого класса защиты ИТ. СОВ. С5. ПЗ. — Москва. 2012. 66 с.

2. Руководство администратора. Система обнаружения вторжений. Аппаратно-программный комплекс шифрования Континент Версия 3.7 (исполнение 2). — Москва. 2017. 9 с.

3. ООО «Центр Специальной Системотехники» / Аппаратно-программный комплекс обнаружения компьютерных атак «Аргус» версии 1.5. URL: http://www.ssec.ru/2014-argus-1.5.htm (дата обращения: 01.12.2020).

4. Loginom — аналитика / Деревья решений: общие принципы. URL: http://www.loginom.ru/decision-tree-p1 (дата обращения: 01.12.2020).

5. Хабр / Открытый курс машинного обучения. Тема 3. Классификация, деревья решений и метод ближайших соседей. URL: https://habr.com/ru/company/ods/blog/322534/ (дата обращения: 01.12.2020).

6. Машинное обучение / Деревья решений: полное введение. URL: https://www.machinelearningmastery.ru/decision-trees-60707-f06e836/ (дата обращения: 01.12.2020).

7. Лукацкий А.В. Обнаружение атак. — Санкт-Петербург. БХВ. 2003. 60 с.

8. KDD Cup 1999 Data / Учебная база данных компьютерных атак. URL: https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (дата обращения: 01.12.2020).

9. SecurityLab.ru — информационный портал, оперативно и ежедневно рассказывающий о событиях в области защиты информации, интернет правах и новых технологиях / Датасеты по ИБ для машинного обучения URL: https://www.securitylab.ru/ blog/personal/Business_without_danger/345789.php (дата обращения: 01.12.2020).

10. Стадник А.Н., Алпеев Е.В., Конышев Е.А. Верификация ложных срабатываний в системе обнаружения компьютерных атак с применением комплексного способа на основе интеллектуального анализа данных и дерева принятия решений // Сборник XXVI Международной научно-технической конференция «Информационные системы и технологии». — Нижний Новгород. 2020. С. 462–467.

References

1. Methodological document of the FSTEC of Russia Protection profile of the intrusion detection system of the network level of the fifth class of IT security. C5. PZ. — Moscow. 2012. 66 p.

2. Administrator's Guide Intrusion Detection System Hardware and Software Encryption complex Continent Version 3.7 (version 2). — Moscow. 2017. 9 p.

3. LLC «Center for Special System Engineering» / Hardware and software complex for detecting computer attacks «Argus» version 1.5. URL: http://www.ssec.ru/2014-argus-1.5.htm (date accessed: 01.12.2020).

4. Loginom – analytics) / Decision trees: general principles. URL: http://www.loginom.ru/decision-tree-p1 (date accessed: 01.12.2020).

5. Habr / Open machine learning course. Topic 3. Classification, decision trees and the nearest neighbor method. URL: https://habr.com /ru/company/ods/blog/322534/ (date accessed: 01.12.2020).

6. Machine Learning / Decision Trees: a complete introduction. URL: https://www.machinelearningmastery.ru/ decision-trees-60707-f06e836/(date accessed:01.12.2020).

7. Lukackiy A.V. Detecting attacks. — St. Petersburg. BHV. 2003. 60 p.

8. KDD Cup 1999 Data / Training database of computer attacks. URL: https://kdd.ics.uci.edu/ databases/kddcup99/kddcup99.html (date accessed: 01.12.2020).

9. SecurityLab.ru — information portal that promptly and weekly tells about events in the field of information protection, Internet law and new technologies / IB datasets for machine learning. URL: https://www.securitylab.ru/blog/personal/Business_without_danger/345789.php (date accessed: 01.12.2020).

10. Stadnik A.N., Alpeev E.V., Konyshev E.A. Verification of false positives in the system of detecting computer attacks using a complex method based on data Mining and decision Tree // Collection of the XXVI International Scientific and Technical Conference «Information Systems and Technologies». — Nizhny Novgorod. 2020. P. 462–467.