Обнаружение DoS атак с применением ансамбля классификаторов

Расим Алгулиев[1], Рамиз Алыгулиев[2], Ядигар Имамвердиев[3], Людмила Сухостат[4]

[1,2,3,4]Институт Информационных Технологий НАНА, Баку, Азербайджан

[1]rasim@science.az, [2]r.aliguliyev@gmail.com, [3]yadigar@lan.ab.az, [4]lsuhostat@hotmail.com

***Аннотация*— За последние два десятилетия в мире, ориентированном на «большие» данные, их обработка и аналитика стали важным инструментом обеспечения информационной безопасности. Таким образом, повышение уровня сетевой безопасности является одним из приоритетов исследователей. Чтобы противостоять атакам в сети, были успешно применены ансамбли классификаторов. Хотя существует множество подходов на основе ансамблей классификаторов, остается сложной задачей найти нужную конфигурацию ансамбля для конкретного набора данных. В этой статье предлагается новый метод построения ансамбля классификаторов. Эксперименты проводятся на наборе данных NSL-KDD. Экспериментальные результаты показывают, что предлагаемый подход может генерировать ансамбли классификаторов, превосходящие единичные классификаторы с точки зрения точности.**

Ключевые слова— информционная безопасность, сетевые атаки, DoS, классификация, ансамбль классификаторов, Big data.

I. Введение

Анализ "больших" данных (Big data) при обнаружении вторжений и решении проблем сетевой безопасности привлекает все большее внимание, поскольку он способствует изучению больших объемов сложных и разрозненных данных и обнаруживает сетевые вторжения и способствует борьбе с кибератаками [1].

Сетевые атаки являются одной из причин аномальных явлений наблюдаемых в работе сетевого оборудования, а также при передаче трафика по сетевым каналам. Аномалии сетевого трафика могут стать причиной некорректной работы одного канала или целых сегментов сети, привести к отказу в работе оборудования обслуживающего данную сеть. Сетевые атаки постоянно изменяются, поскольку злоумышленники используют индивидуальные подходы, а также в связи с регулярными изменениями в программном обеспечении и аппаратных средствах компьютерных систем [2].

Решение проблемы обнаружения сетевых атак далеко не тривиально, так как природа самих атак изменчива. В контексте компьютерной сети, обеспечение всестороннего определения аномального или даже нормального поведения достаточно слабое [3, 4]. Другая причина заключается в том, что несколько методов обнаружения атак требуют маркировки шаблонов нормальных и/или аномальных поведений, которые нелегко получить [5, 6]. К тому же, выбрать подходящий инструмент для обнаружения атак не просто. Назначенный инструмент может хорошо подходить только для одного вида атак, но не для всех, что приводит к снижению производительности, что приводит к высоким ложным срабатываниям [7]. Таким образом, когда типы атак не известны априори, что является весьма реалистичным предположением, выбор метода обнаружения атак не прост. Кроме того, масштаб сети является проблемой: при обнаружении атак необходимо учитывать распределение процесса выполнения заданий между несколькими серверами сети с целью увеличения общей производительности и возможность работы системы при отказе отдельных её элементов с учетом роста размеров сетей [8, 9].

Уязвимости в стеке протоколов связи (TCP/IP) приводят к намеренным или непреднамеренным атакам распределенного отказа в обслуживании (Distributed Denial of Service, DDoS). Атаки DDoS могут быть обнаружены с использованием существующих методов машинного обучения.

Последние исследовательские работы были основаны на методах двоичной классификации, которые могут различать два состояния («нормальное» или «аномальное»). В случае конфликта между бинарными классификаторами окончательное решение достигается путем сравнения их точности. Альтернативное решение возможно путем разработки ансамбля классификаторов. Для объединения таких классификаторов в ансамбль требуется новый подход.

В данной статье предлагается новый метод обнаружения DoS атак на основе ансамбля классификаторов.

II. Литературный Обзор

Ряд исследований и обзорных статей посвящен технологиям обнаружения вторжений [10, 11] или интеллектуального анализа данных в конкретных приложениях [12]. Поскольку принципы обнаружения

DOI: 10.25045/NCInfoSec.2017.02

вторжений были впервые введены Деннингом в 1987 году, было разработано большое количество систем реактивной защиты [13-15].

В работе [16] методы обнаружения вторжений категоризируются по 3 категориям, а именно, единичные (single), гибридные и ансамбли. Машины опорных векторов и искусственные нейронные сети являются наиболее популярными подходами среди классификаторов. Несколько классификаторов объединяются в ансамбль классификаторов с целью повышения эффективности классификации [17]. Методы «Большинство голосов», бэггинг и бустинг - некоторые общие стратегии для объединения классификаторов [18]. Хотя известно, что недостатки компонентов классификаторов накапливаются в ансамбле классификаторов, но он работает очень эффективно в той или иной комбинации. Таким образом, исследователи с каждым днем становятся все более заинтересованными в применении ансамбля классификаторов.

В [19] выделяются важные проблемы в области кибербезопасности для математических и статистических решений.

Метод повышения точности обнаружения с помощью ансамбля двухслойных машин опорных векторов (Support Vector Machines, SVMs) на основе ротации леса (rotation forest) был представлен в работе [20]. Эксперименты проводились на наборе данных KDD CUP 1999. Выход ансамбля получен методом большинства голосов (majority voting).

Точность классификации была улучшена путем объединения мнений многих экспертов, используя ансамбль [21]. Метод построения ансамбля использует весовые коэффициенты, полученные методом роя частиц (particle swarm optimization, PSO), для повышения точности обнаружения вторжений.

Статья [22] направлена на определение мультиклассовых моделей SVM для задачи обнаружения вторжений. Новый подход (weighted one-against-rest SVM, WOAR-SVM) основан на наборе оптимальных весовых коэффициентов, которые определяют взаимосвязь между правилами принятия решений для бинарных классификаторов SVM.

В [23] была предложена общая архитектура для автоматического обнаружения DDoS атак с использованием методов машинного обучения. Основной целью данной статьи было минимизировать ошибки классификации при обнаружении вторжений. Предлагаемый алгоритм классификации RBPBoost является ансамблем классификаторов и включает стратегию минимизации затрат Неймана-Пирсона для окончательного классификационного решения.

III. Методы Классификации Данных

Больше информации для обнаружения вторжений может быть получено посредством применения методов классификации данных. Теоретически алгоритмы классификации могут получить высокую производительность, т.е. могут минимизировать уровень ложных тревог и максимизировать точность обнаружения. Одна из наиболее привлекательных особенностей алгоритмов состоит в способности различать нормальное поведение от аномального [9]. В контексте обнаружения вторжений, алгоритмы классификации, как правило, представляют отображение, которое адаптируется к невидимым сетевым аномалиям [24].

Формально, пусть S будет определен таким образом, что $S=\{0,1\}$, и предположим, что экземпляр данных может быть в одном из двух состояний: нормальном (т.е. 0) или аномальном (т.е. 1). Каждый экземпляр данных есть вектор признаков X , измеренный в момент времени t и обозначается как $X(t)$.

Алгоритмы классификации направлены на обучение функции, которая отображает все образцы $X(t)$ к их собственным состояниям из S . Для того, чтобы достичь своей цели, они используют набор экземпляров данных в пределах данной компьютерной сети. Этот набор известен как обучающий набор данных. Некоторые алгоритмы изучают отображающую функцию путем использования меченых обучающих наборов, где каждый образец в обучающем наборе помечается одним из состояний в S . Эти алгоритмы также известны как алгоритмы обучения с учителем. Цель их использования состоит в достижении высокой точности классификации.

Самые популярные методы классификации данных включают в себя SVM, деревья решений, байесовские сети, метод K-ближайших соседей (K-Nearest Neighbors, KNN) и др.

A. Машины опорных векторов

Цель SVM заключается в классификации точек данных X n-мерного пространства с помощью $(n-1)$-мерной гиперплоскости. Любую гиперплоскость можно записать в виде множества точек X , удовлетворяющих $w^T x+b=0$, где вектор w - нормальный вектор, перпендикулярный гиперплоскости и b - смещение гиперплоскости $w^T x+b=0$ от первоначальной точки вдоль направления w.

Расстояние от точки данных до разделяющей гиперплоскости $w^T x+b=0$ может быть вычислено как $r=(w^T x+b)/\|w\|$, и точки данных, наиболее близкие к гиперплоскости называются опорными векторами. Линейный SVM решается путем формулировки задачи квадратичной оптимизации следующим образом:

$$\arg\min_{w,b}\left(\frac{1}{2}\|w\|^2\right), \quad (1)$$
$$y(w^T x+b)\geq 1.$$

SVM может точно найти линейные, нелинейные и сложные границы классификации, даже при небольшом размере обучающей выборки.

SVM широко используется для передачи разнотипных данных путем включения ядерной функции для отображения в пространство данных. В качестве таких функций чаще всего используют линейное ядро, полиномиальное ядро, гауссово ядро с радиальной базовой функцией и сигмоидное ядро.

Однако при выборе ядерной функции и подгонке соответствующих параметров посредством SVM по-прежнему пользуются процедурой проб и ошибок. SVM быстр, но его продолжительность увеличивается в 4 раза, когда размер выборки данных удваивается. К сожалению, корень SVM алгоритмов заключен в двоичной классификации. Для решения задач мультиклассовой классификации, несколько SVM для двоичных классов могут быть объединены путем классификации каждого класса и всех других классов или классификации каждой пары классов.

B. Деревья решений

Дерево решений (Decision tree, DT) является древовидной структурной моделью, которая имеет листья, которые представляют классы или решения, и ветви, представляющие конъюнкции признаков, которые приводят к этим классификациям.

Древовидная классификация входного вектора выполняется путем обхода дерева, начиная с корневого узла, и заканчивая листом. Каждый узел дерева вычисляет неравенство на основе одной из входных переменных. Каждый лист присваивается определенному классу. Каждое неравенство, которое используется, чтобы разделить входное пространство, основано только на одной из входных переменных. Линейные DT подобны бинарным DT, за исключением того, что неравенство, вычисленное в каждом узле, имеет произвольный линейный вид, который может зависеть от нескольких переменных. DT зависит от правила «если-то», но не требует никаких параметров и метрик. Это простая и интерпретируемая структура позволяет деревьям решений рассматривать проблемы атрибутов различных типов. DT также может управлять отсутствующими значениями или зашумленными данными. Тем не менее, они не могут гарантировать оптимальную точность в отличие от других методов машинного обучения. Хотя DT легко узнать и реализовать, они не кажутся популярными методами обнаружения вторжений. Возможной причиной отсутствия популярности является то, что нахождение наименьшего DT является NP-трудной задачей.

C. Байесовские сети

Байесовская сеть основана на правиле Байеса, которое дает гипотезу H классов и данных x, что

$$P(H \mid x) = \frac{P(x \mid H)P(H)}{P(x)}, \qquad (2)$$

где $P(H)$ обозначает априорную вероятность каждого класса без информации о переменной x, $P(H \mid x)$ – апостериорную вероятность переменной x над возможными классами, $P(x \mid H)$ – условную вероятность x на данном правдоподобии H.

Узлы Байесовской сети представлены случайными величинами и дугами. Узел всегда вычисляет апостериорные вероятности, давая доказательства наследования для выбранных узлов.

Наивный Байес (Naïve Bayes, NB) является простой моделью Байесовской сети, которая предполагает, что все переменные являются независимыми. Используя правило Байеса для классификации методом NB необходимо найти гипотезу максимального правдоподобия, которая определяет метку класса для тестируемых данных x. Учитывая наблюдаемые данные x и группу меток класса $C = \{c_j\}$, наивный классификатор Байеса может быть решен путем гипотезы максимальной апостериорной вероятности (maximum a posteriori probability, MAP) следующим образом:

$$\arg\max_{c_j \in C} P(x \mid c_j)P(c_j). \qquad (3)$$

NB является эффективным для задач с логическим выводом и основывается на предположении о независимости переменных.

D. Метод К-ближайших соседей

Число ближайших соседей k и меры расстояния являются ключевыми компонентами для алгоритма KNN. Выбор числа k должен быть основан на проведении кросс-валидации. Как правило, большое число k уменьшает эффект шума в данных при классификации, а это может стереть различия между классами. Преимущество метода проб и ошибок в том, что k должно быть меньше, чем квадратный корень из общего количества обучающих образов.

В случае мультиклассовой классификации метод k-ближайших соседей основан на измерении расстояния от одного образца данных до каждого другого обученного образца [25]. Вычисляются k-наименьших расстояний, и наиболее распространенный класс на основе этих k-ближайших соседей считается меткой выходного класса.

KNN легко реализовать и интерпретировать. Тем не менее, KNN классификация имеет высокую вычислительную сложность.

IV. Предлагаемый Подход

Вводятся следующие обозначения (Таблица I): $x_i \in R^n \ (i = \overline{1,n})$ – точка из набора данных, где n – общее

число точек, $M = \{M_1, M_2, ..., M_m\} \in R^m$ – методы классификации, впоследствии объединяемые в ансамбль, m – число классификаторов, a_{ij} - оценка классификатора для каждой точки набора данных, k – число классов.

ТАБЛИЦА I. ОЦЕНКА КЛАССИФИКАТОРОВ ДЛЯ КАЖДОЙ ТОЧКИ

Data point	Classifiers			Ensemble score
X	M_1	…	M_m	
x_1	a_{11}	…	a_{1m}	$\max_j(a_{1j})$
⋮	⋮	⋱	⋮	⋮
x_n	a_{n1}	…	a_{nm}	$\max_j(a_{nj})$

Предлагаемый алгоритм, указывающий на вероятности принадлежности к определенным классам, возвращает вектор оценок классификаторов для каждой точки.

Особенность предлагаемого подхода состоит в том, что для каждой точки из набора данных предсказанная метка класса соответствует максимальному значению среди всех оценок, полученных методами кластеризации для данной точки.

Алгоритм предлагаемого подхода по обнаружению сетевых атак на основе ансамбля классификаторов представлен ниже:

Вход: $x_i \in R^n$ - набор точек

n – число точек в наборе данных

m – число классификаторов

$M = \{M_1, M_2, ..., M_m\}$ – набор классификаторов

k – число классов.

$A = \{a_{ij}\}_{n \times m}$ – оценки классификаторов для каждой записи набора данных

Выход:

P – вектор оценок ансамбля классификаторов

for i **=1 to** n **do**

for j **=1 to** k **do**

Вычисление значения оценки a_{ij} для M_j

End

$P_i = \max_j a_{ij}$

End

V. ОПИСАНИЕ БАЗЫ ДАННЫХ

Для проведения экспериментов была рассмотрена база данных сигнатур NSL-KDD [26], построенная на основе базы KDD-99 по инициативе американской Ассоциации перспективных оборонных научных исследований DARPA [27]. Для проведения исследований в области обнаружения вторжений был собран набор данных о соединениях, который охватывает широкий спектр различных вторжений, смоделированных в среде, имитирующей сеть Военно-воздушных сил США.

Статистический анализ показал, что существуют важные проблемы в базах данных, которые высоко влияют на производительность систем, а также приводят к очень плохой оценке подходов обнаружения аномалий. Рассмотренная база NSL-KDD имеет следующие преимущества:

1. Нет избыточных записей в обучающем наборе, так что классификатор не покажет какой-либо предвзятый результат.

2. Нет дубликата записей в тестовом наборе. Он содержит некоторые атаки, которые не присутствуют в обучающем наборе.

3. Количество выбранных записей из каждой группы уровней сложности обратно пропорционально доле записей в исходном наборе данных.

Обучающий набор данных состоит из 21 вида различных атак из 37 присутствующих в тестовом наборе данных. Известные виды атак присутствуют в обучающем наборе. Кроме того, количество записей в обучающем (125973 образцов) и тестовом наборах (22544 выборок) NSL-KDD являются приемлемыми. Это преимущество делает его доступным для проведения экспериментов на полных данных без необходимости случайным образом выбирать небольшую часть. Следовательно, результаты оценки различных научно-исследовательских работ будут согласованными и сопоставимыми.

Все атаки в NSL-KDD поделены на четыре группы:

- *DoS (Denial of Service Attack)* включает атаки: “neptun”, “back”, “smurf”, “pod”, “land” и “teardrop”.
- *U2R (Users to Root Attack)* включает атаки: “buffer_overflow”, “loadmodule”, “rootkit” и “perl”.
- *R2L (Remote to Local Attack)* включает атаки: “warezclient”, “multihop”, “ftp_write”, “imap”, “guess_passwd”, “warezmaster”, “spy” и “phf”.
- *Probe (Probing Attack)* содержит следующие атаки: “portsweep”, “satan”, “nmap” и “ipsweep”.

Основные цели, выдвигаемые при обнаружении вторжений сети, включают распознавание редких типов атак, увеличивая точность обнаружения для подозрительной активности, а также повышая эффективности моделей обнаружения вторжений в режиме реального времени. Каждая запись имеет 41 атрибут, описывающий различные признаки.

VI. МЕТРИКИ ОЦЕНКИ МЕТОДОВ КЛАССИФИКАЦИИ

Для оценки производительности классификаторов используются следующие метрики: «аккуратность» классификации (classification accuracy), полнота (recall), точность (precision) и F-мера. Для любого алгоритма классификации возможны четыре классификационных случая, и это помогает понять разницу между рассматриваемыми метриками: истинно-положительные результаты (True Positives, TP), ложно-положительные результаты (False Positives, FP), истинно-отрицательные

результаты и ложно-отрицательные результаты (False Negatives, FN).

«Аккуратность» классификации может быть определена как доля правильных результатов, которая достигается классификатором.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (4)$$

Точность показывает, какая доля объектов, выделенных классификатором как положительные, действительно является положительной.

$$precision = \frac{TP}{TP + FP}. \quad (5)$$

Полнота показывает, какая часть положительных объектов была выделена классификатором

$$recall = \frac{TP}{TP + FN}. \quad (6)$$

F-мера является показателем, который сочетает в себе меры точности и полноты:

$$F - measure = \frac{2 \times recall \times precision}{recall + precision}. \quad (7)$$

VII. Результаты Экспериментов

Эксперименты проводились в ОС Windows® 10–64 с процессором Core i7 (2.5 ГГц), 8.0 Гб RAM. Предлагаемый подход оценивался на языке R 3.4.1 на наборе данных NSL-KDD, который был разбит на 3 класса (DoS, нормальное состояние (Normal) и другие атаки (Other attacks)).

Был проведен сравнительный анализ предложенного подхода с единичными классификаторами. Ансамбль классификаторов состоял из комбинаций алгоритмов DT, SVM с различными ядерными функциями (SVM(Linear), SVM(Polynom) и SVM(RBF)), NB и KNN.

Результаты классификации показаны в таблицах II-V, что дает сравнительный анализ результатов по метрикам: «аккуратность», полнота, точность и F-мера. Наилучшие результаты были отмечены жирным шрифтом.

ТАБЛИЦА II. Сравнение «аккуратности» классификации предложенного алгоритма с другими известными алгоритмами классификации

Класс / Метод	DoS	Normal	Other attacks
DT	86.32%	77.19%	64.25%
KNN	88.21%	79.67%	65.80%
SVM (Linear)	87.21%	77.96%	66.22%
SVM (Polynom)	86.64%	79.50%	68.31%
SVM (RBF)	87.25%	78.84%	65.38%
NB	80.45%	71.25%	66.50%
DT+KNN+SVM(Polynom)	90.74%	84.77%	74.53%
DT+KNN+SVM(Polynom) +NB	92.19%	**88.63%**	82.76%
DT+KNN+SVM(Polynom) +NB+SVM (Linear)	**92.33%**	88.58%	**83.35%**
DT+KNN+SVM(Polynom) +NB+SVM (Linear)+SVM(RBF)	91.89%	88.60%	82.93%

ТАБЛИЦА III. Сравнение точности методов классификации для каждого класса

Класс / Метод	DoS	Normal	Other attacks
DT	95.94%	63.17%	88.15%
KNN	95.88%	65.87%	86.05%
SVM (Linear)	96.65%	64.37%	84.47%
SVM (Polynom)	96.14%	65.70%	88.12%
SVM (RBF)	85.07%	64.92%	96.17%
NB	73.03%	74.78%	41.80%

ТАБЛИЦА III. Сравнение точности методов классификации для каждого класса

Класс / Метод	DoS	Normal	Other attacks
DT+KNN+SVM(Polynom)	96.74%	71.95%	94.41%
DT+KNN+SVM(Polynom) +NB	97.94%	**77.74%**	95.76%
DT+KNN+SVM(Polynom) +NB+SVM (Linear)	**99.98%**	76.81%	**100%**
DT+KNN+SVM(Polynom) +NB+SVM (Linear)+SVM(RBF)	97.91%	77.60%	96.06%

Из таблицы II можно сделать заключение, что наиболее высокая точность обнаружения DoS атак была достигнута для ансамбля из пяти классификаторов (DT+KNN+SVM(Polynom)+NB+SVM (Linear)) – 92.33%, что превысило результат единичного классификатора (KNN) на 4.12%.

Несмотря на то, что NB показывает наименьший результат (80.45%), при добавлении его к ансамблю классификаторов точность предлагаемого подхода возросла и составила для четырех классификаторов (DT+KNN+SVM(Polynom)+NB) 92.19%, а для шести классификаторов (DT+KNN+SVM(Polynom)+NB+SVM (Linear)+SVM(RBF)) – 91.89%.

Сравнение значений точности и полноты при обнаружении DoS атак показано в таблицах III и IV соответственно.

ТАБЛИЦА IV. СРАВНЕНИЕ ПОЛНОТЫ МЕТОДОВ КЛАССИФИКАЦИИ ДЛЯ КАЖДОГО КЛАССА

Класс / Метод	DoS	Normal	Other attacks
DT	74.19%	97.33%	29.75%
KNN	78.08%	97.62%	33.28%
SVM (Linear)	75.72%	96.19%	34.42%
SVM (Polynom)	74.77%	97.50%	38.23%
SVM (RBF)	76.00%	97.57%	32.54%
NB	74.51%	57.07%	58.51%
DT+KNN+SVM(Polynom)	82.86%	98.62%	49.99%
DT+KNN+SVM(Polynom) +NB	**85.28%**	98.63%	66.44%
DT+KNN+SVM(Polynom) +NB+SVM (Linear)	84.67%	**100%**	66.70%
DT+KNN+SVM(Polynom) +NB+SVM (Linear)+SVM(RBF)	84.66%	98.77%	**66.72%**

В таблице V представлены результаты для F-меры, которая оценивает производительность методов классификации, в общем, объединяя значения точности и полноты.

По результатам проведенного анализа можно сделать вывод, что наилучший результат дает ансамбль из пяти классификаторов.

ТАБЛИЦА V. СРАВНЕНИЕ ЗНАЧЕНИЙ F-МЕРЫ МЕТОДОВ КЛАССИФИКАЦИИ ДЛЯ КАЖДОГО КЛАССА

Класс / Метод	DoS	Normal	Other attacks
DT	83.67%	76.62%	44.48%
KNN	86.07%	78.66%	48.00%
SVM (Linear)	84.91%	77.13%	48.91%
SVM (Polynom)	84.12%	78.50%	53.33%
SVM (RBF)	84.90%	77.97%	47.07%
NB	73.76%	64.74%	48.76%
DT+KNN+SVM(Polynom)	89.27%	83.20%	65.37%
DT+KNN+SVM(Polynom) +NB	91.17%	**86.95%**	78.45%
DT+KNN+SVM(Polynom) +NB+SVM (Linear)	**91.69%**	86.88%	**80.02%**
DT+KNN+SVM(Polynom) +NB+SVM (Linear)+SVM(RBF)	90.81%	86.92%	78.74%

VIII. ЗАКЛЮЧЕНИЕ

Обработка и аналитика «больших» данных в настоящее время важна для обеспечения информационной безопасности. Обнаружение вторжений является одной из серьезных проблем в области сетевой безопасности. В этом исследовании чтобы противостоять атакам в сети, были успешно применены ансамбли классификаторов. Ансамбль классификаторов состоял из комбинаций алгоритмов DT, SVM с различными ядерными функциями, NB и KNN.

В целом, рассмотренные методы классификации показали высокую точность обнаружения DoS атак в ходе проведения экспериментов. При этом наиболее точный результат показал ансамбль из пяти классификаторов DT+KNN+SVM(Polynom)+NB+SVM(Linear). Можно сделать вывод о практической значимости предложенного подхода к обнаружению атак в сети.

БЛАГОДАРНОСТИ

Данная работа выполнена при финансовой поддержке Фонда Развития Науки при Президенте Азербайджанской Республики – **Грант № EİF-KETPL-2-2015-1(25)-56/05/1**.

ЛИТЕРАТУРА

[1] R.M. Aliguliyev, Y.N. Imamverdiyev, M.S. Hajirahimova, “Multidisciplinary problems of big data in information security,” Proc. of the II International scientific and practical conference Information Security and Computer Technologies (InfoSec&CompTech), pp. 10-11, 2017.

[2] R.M. Əliquliyev, Y.N. İmamverdiyev, “İnformasiya təhlükəsizliyi insidentləri,” Bakı, İnformasiya Texnologiyaları, 2012, 219 s.

[3] H. Nallaivarothayan, D. Ryan, S. Denman, S. Sridharan, C. Fookes, “An evaluation of different features and learning models for anomalous event detection,”Proc. of DICTA, pp. 1–8, 2013.

[4] M. Xie, S. Han, B. Tian, S. Parvin, “Anomaly detection in wireless sensor networks: a survey,” J. Net. Comp. App., Vol. 34, pp. 1302–1325, 2011.

[5] J.J. Davis, A.J. Clark, “Data preprocessing for anomaly based network intrusion detection: a review,” Comp. & Sec., Vol. 30, pp. 353–375, 2011.

[6] U. Fiorea, F. Palmieri b, A. Castiglionec, A.D. Santis, “Network anomaly detection with the restricted boltzmann machine,” Neurocomputing, Vol. 122, pp. 13–23, 2013.

[7] V. Chandola, A. Banerjee, V. Kumar, “Anomaly detection: a survey,” ACM Comp. Surv., Vol. 41, pp. 1–58, 2009. http://doi.acm.org/10.1145/1541880.1541882.

[8] E. Anceaume, Y. Busnel, E.L. Merrer, R. Ludinard, J. Marchand, B. Sericola, “Anomaly characterization in large scale networks,” Proc. of the 44th annual IEEE/IFIP international conference on dependable systems and networks (DSN), pp. 68–79, 2014.

[9] S. Dua, X. Du, “Data mining and machine learning in cybersecurity,” Auerbach Publications, 2011, 256 p.

[10] C.A. Catania, C.G. Garino, “Automatic network intrusion detection: current techniques and open issues,” Computers and Electrical Engineering, Vol. 38, No. 5, pp. 1062–1072, 2012.

[11] M. Ahmed, A. Mahmood, J. Hu, “A survey of network anomaly detection techniques,” Journal of Network and Computer Applications, Vol. 60, pp. 19–31, 2016.

[12] S. X. Wu, W. Banzhaf, “The use of computational intelligence in intrusion detection systems: a review,” Applied Soft Computing, Vol. 10, No. 1, pp. 1–35, 2010

[13] V. Chandola, E. Eilertson, L. Ertoz, G. Simon, V. Kumar, “Data mining for cyber security,” in: Singhal, A. (Eds.), Data Warehousing and Data Mining Techniques for Computer Security. Springer, New York, pp. 1–20, 2006.

[14] W. Lee, S.J. Stolfo, “A framework for constructing features and models for intrusion detection systems,” ACM Trans. Inf. Sys. Sec., Vol. 4, pp. 227–261, 2000.

[15] M.V. Mahoney, P.K. Chan, “Learning nonstationary models of normal network traffic for detecting novel attacks,” Proc. of the 8th ACM

SIGKDD international conference on knowledge discovery and data mining, pp. 376–386, 2002.

[16] V. Hodge, J. Austin, “A survey of outlier detection methodologies,” Artificial Intelligence Review, Vol. 22, No. 2, pp. 85–126, 2004.

[17] D.M. Farid, M.Z. Rahman, C.M. Rahman, “*Adaptive Intrusion Detection based on Boosting and* Naive Bayesian Classifier,” Int. J. Comp. App., Vol. 24, pp. 12–19, 2011.

[18] C.A. Laurentys, R.M. Palhares, W.M. Caminhas, “A novel artificial immune system for fault behavior detection,” Exp. Sys. App., Vol. 38, pp. 6957–6966, 2011.

[19] J. Meza, S. Campbell, D. Bailey, “Mathematical and statistical opportunities in cybersecurity,” Paper LBNL-1667E, Lawrence Berkeley National Laboratory, Berkeley, CA, pp. 1–11, 2009.

[20] L. Lin, R. Zuo, S. Yang, Z. Zhang, “SVM ensemble for anomaly detection based on rotation forest,” Proc. Of the third international conference on intelligent control and information processing (ICICIP), pp. 150-153, 2012.

[21] A.A. Aboromman, M.B.I. Reaz, “A novel SVM-kNN-PSO ensemble method for intrusion detection system,” Appl. Soft Comput., Vol. 38, pp. 360–372, 2016.

[22] A.A. Aburomman, M.B.I Reaz, “A novel weighted support vector machines multiclass classifier based on differential evolution for intrusion detection systems,” Inf. Sci., Vol. 414, pp. 225-246, 2017.

[23] P.A.R. Kumar, S. Selvakumar, “Distributed denial of service attack detection using an ensemble of neural classifier,” Computer Communications, Vol. 34, pp. 1328–1341, 2011.

[24] N.T. Nguyen, A. Zgrzywa, A. Czyzewski, “Advances in multimedia and network information system technologies, Springer-Verlag, Berlin, 2010, 318 p.

[25] M.-L. Zhang, Z.-H. Zhou, “A k-nearest neighbor based algorithm for multi-label classification,” Proc. of the International Conference on Granular Computing, pp. 718–721, 2005.

[26] P. Aggarwal, S.K. Sharma, “Analysis of KDD dataset attributes-class wise for intrusion detection,” Proc. Comp. Sci., Vol. 57, pp. 842–851, 2015.

[27] J. McHugh, “Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory,” ACM Trans. Inf. Sys. Sec., Vol. 3, pp. 262–294, 2000.

DOS ATTACKS DETECTION USING AN ENSEMBLE OF CLASSIFIERS

Rasim Alguliyev[1], Ramiz Aliguliyev[2],
Yadigar Imamverdiyev[3], Lyudmila Sukhostat[4]
[1,2,3, 4]Institute of Information Technology of ANAS,
Baku, Azerbaijan
[1]*rasim@science.az*, [2]*r.aliguliyev@gmail.com*, [3]*yadigar@lan.ab.az*,
[4]*lsuhostat@hotmail.com*

Abstract – Over the past two decades, processing and analytics of Big data have become an important tool for ensuring information security. Thus, increasing the level of network security is one of the priorities of researchers. The ensembles of classifiers were successfully applied to resist attacks in the network. Although there are many approaches based on ensembles of classifiers, it remains a challenge to find the required ensemble configuration for a specific dataset. This article proposes a new method for constructing an ensemble of classifiers. Experiments are conducted on the NSL-KDD dataset. Experimental results show that the proposed approach can generate ensembles of classifiers that exceed single classifiers in terms of accuracy.

Keywords – network security, network attacks, DoS, classification, ensemble of classifiers, Big data