УДК 004.056.5

ИССЛЕДОВАНИЕ ОБНАРУЖЕНИЯ DDoS-АТАК С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

Джабиа Нзи Жан Максим, О. А. Сафарьян
Донской государственный технический университет (г. Ростов-на-Дону, Российская Федерация)

В статье представлено сравнение алгоритмов обнаружения сетевых атак типа DDoS-атак (отказ в обслуживании) для различных сервисов хранения, обработки и передачи данных через Интернет. Особое внимание уделяется применению алгоритмов машинного обучения, таких как гауссовская смешанная модель для максимизации ожиданий (GMM-EM), линейная регрессия (LR), SVM (машина опорных векторов) (с линейным, RBF (радиальная базисная функция) или полиномиальные ядра), алгоритмы дерева решений (Decision Tree), наивный Байеса (Naive Bayes) [4] и рандом Forest (Random Forest) для обнаружения такого типа атак. В конце статьи оцениваются перечисленные выше алгоритмы машинного обучения и тщательно сравнивается их производительность. Все экспериментальные результаты показывают, что более 99,7 % двух видов DOS-атак успешно обнаруживаются. Этот подход не снижает производительность и может быть легко распространен на более широкие DOS-атаки.

***Ключевые слова*:** DDoS-атака, машинное обучение, облачная платформа, DNS, ICMP, алгоритм.

INVESTIGATION OF DDoS ATTACKS DETECTION USING MACHINE LEARNING

Djabia Nzi Jean Maxim, Olga A .Safaryan
Don State Technical University (Rostov-on-Don, Russian Federation)

The article presents a comparison of algorithms for detecting network attacks such as DDoS attacks (denial of service) for various data storage, processing and transmission services over the Internet. Particular attention is paid to the application of machine learning algorithms such as the Gaussian mixed model for expectations maximization (GMM-EM), linear regression (LR), SVM (support vector machine) (with linear, RBF (radial basis function) or polynomial kernels), Decision Tree algorithms, Naive Bayes [4] and Random Forest to detect this type of attacks. At the end of the article, the machine learning algorithms listed above are evaluated and their performance is carefully compared. All experimental results show that more than 99.7% of the two types of DOS attacks are successfully detected. This approach does not reduce performance and can be easily extended to broader DOS attacks.

***Keywords*:** DDoS attacks, machine learning, cloud platform, DNS, ICMP, algorithm.

Введение. Сегодня существует множество атак на сетевые инфраструктуры. К ним относятся атаки на доступность сети, а также на конфиденциальность и целостность сетевых пакетов и их источники назначения.

Атаки типа «отказ в обслуживании» (DOS) представляют собой серьезную угрозу сетевой безопасности. Эти атаки часто осуществляются с виртуальных машин в облачной платформе [1], а не с собственной машины злоумышленника, чтобы обеспечить анонимность и более высокую пропускную способность сети. Большинство исследований сосредоточено на анализе трафика на стороне назначения (жертвы) с предопределенными пороговыми значениями. Эти подходы имеют существенные недостатки [2]. Они являются лишь пассивной защитой после атаки и не могут

использовать внешние статистические характеристики атак. С этими подходами трудно отследить злоумышленника.

В данной статье предлагается реализация системы обнаружения DOS-атак на стороне источника в облачной платформе, основанной на методах машинного обучения.

Целью данной статьи является обнаружение DDoS DNS-атак и ICMP-флуд с использованием машинного обучения.

Основная часть

DNS-атака (Domain Name Server) является неотъемлемой частью сетевой инфраструктуры и отвечает за преобразование доменных имен в IP-адреса. DNS-атаки DDoS, основанные на отражении, представляют большую угрозу для службы доменных имен и безопасности сети [3].

Мониторинг атаки с отражением DNS представлен на рис. 1. Во время атаки с отражением DNS злоумышленник отправляет большое количество DNS-запросов на DNS-серверы с исходным и поддельным IP-адресом жертвы.

Эти DNS-запросы требуют ответов с высокой нагрузкой, таких как «Дайте мне все ваши записи DNS» или «Дайте мне IP-адреса нескольких доменных имен». Эти ответы подавляют жертву и истощают ее ресурсы.

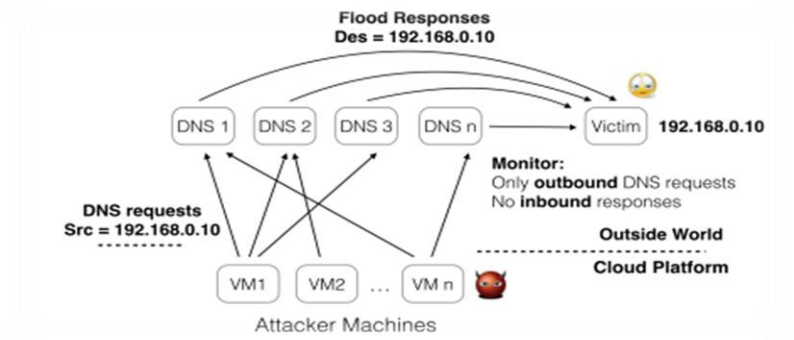

Рис. 1. Мониторинг соотношения входящих и исходящих пакетов DNS к обнаружению атаки отражения DNS

Для атак с отражением DNS идея заключается в мониторинге входящего и исходящего трафика. Для обычных DNS-запросов входящий трафик приблизительно пропорционален исходящему трафику.

Маловероятно, что будет отправлено огромное количество запросов, а ответов не будет. Однако, во время атаки с отражением DNS взломщик подделывает исходный IP-адрес и перенаправляет ответ жертве, что приводит к бóльшему количеству запросов, чем пакетов ответов. Поэтому необходимо использовать соотношение входящих и исходящих пакетов DNS для обнаружения этой атаки. На рис. 2 можно наблюдать две атаки. Обе атаки имеют низкое соотношение входящего/исходящего трафика, близкое к нулю, то есть только запросы, но нет ответов. В периоды (в минутах) без приступов этот коэффициент очень близок к 1, то есть каждый запрос имеет ответ.

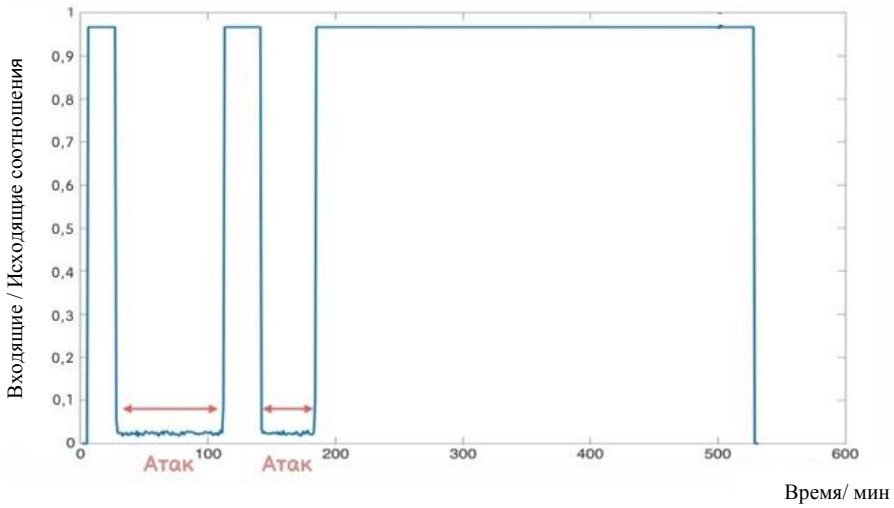

Рис. 2. Соотношение входящих/исходящих DNS-пакетов с атаками отражения DNS или без атак

ICMP-флуд. Интернет-протокол управляющих сообщений (ICMP) предназначен для передачи управляющей информации, такой как индикаторы ошибок. Это один из основных протоколов в наборе интернет-протоколов (IP). ICMP отличается от протоколов UDP и TCP тем, что ICMP не используется для передачи данных. ICMP-флуд — это атака, при которой злоумышленник отправляет огромное количество ICMP-пакетов и перегружает жертву.

В отличие от SSH или DNS, в обычных ситуациях существует не так много пакетов ICMP. Поэтому напрямую используется скорость пакетов ICMP в качестве индикатора ICMP-флуда.

Обычно ICMP-пакетов очень мало, поэтому большое количество ICMP-пакетов за короткий промежуток времени очень подозрительно. Помимо количества ICMP-пакетов с одной виртуальной машины, получаются более точные результаты путем совместного анализа состояния ICMP нескольких виртуальных машин.

Алгоритм обнаружения DDoS DNS-атак. Для данной системы обнаружения DDoS был предложен алгоритм генерации признаков и классификаций (рис. 3), где n — количество статистических признаков, k — количество серверов, li — количество виртуальных машин (VMs) на сервере i. После объединения всех векторов F_{ij} на всех интересующих серверах был получен и отправлен длинный вектор признаков F в механизм машинного обучения. Этот двигатель использует предварительно обученный модуль машинного обучения M_0 для обнаружения DDoS-атак.

Система также может иметь механизм онлайн-обучения. Например, другие модули машинного обучения, M_1, M_2...M_r в фоновом режиме также могут классифицировать вектор признаков F. Если предопределенное число k этих модулей классифицирует этот вектор признаков F как безопасный или злонамеренный вектор, F (с его выходной меткой отсутствия атаки или атака соответственно) используется для обновления основного модуля тестирования M_0.

Select monitored features $f^1, f^2 \ldots f^n$
for *Server* S_i, $i = 1, 2 \ldots k$ **do**
  for VM_{ij} *on server* S_i, $j = 1, 2 \ldots l_i$ **do**
    VMM_i reports monitored statistical features F_{ij} of VM_{ij}
    where $F_{ij} = \{f_{ij}^1, f_{ij}^2 \ldots f_{ij}^n\}$
  end
end

Рис. 3. Алгоритм генерации признаков и классификаций

Облачная платформа. Был реализован прототип системы обнаружения с реальными облачными настройками, в облаке находится четыре сервера (S0...S3), где на каждом сервере работает несколько виртуальных машин. Все облачные серверы и запущенные на них виртуальные машины работают под управлением Ubuntu 20.04.

Было проведено два эксперимента для проверки защитных механизмов системы. В первом эксперименте были запущены DDoS-атак (отражение DNS и ICMP-флуд) на виртуальные машины сервера S0. Жертва — виртуальная машина на сервере S1, на котором работает веб-сервис. Разработанная система защиты на сервере запускает виртуальные машины, выполняющие атаки. Виртуальные машины на других серверах (кроме S0 и S1) запрашивают веб-сервис, имитируя легитимных пользователей.

Во втором эксперименте атаки были запушены из трех виртуальных машин S0, S2 и S3 для имитации распределенных DoS-атак. Жертва —виртуальная машина на S1. Системы защиты развернуты на S0, S2 и S3. Все эксперименты безопасны, поскольку они выполняются за VPN-маршрутизатором, поэтому пакеты атаки никогда не уходят за пределы Интернета.

Сбор данных и алгоритмы машинного обучения. В данных экспериментах были собраны сетевые пакеты, входящие и исходящие из виртуальных машин злоумышленника в течение 3 часов.

Основная цель — обнаружение атак, независимо от того, к какой категории относится атака. Оцениваются алгоритмы обучения как с учителем, так и без учителя. Для контролируемой классификации оцениваются линейная регрессия (LR), SVM (машина опорных векторов), RBF (радиальная базисная функция) или полиномиальные ядра, алгоритмы дерева решений (Decision Tree), наивного Байеса (Naive Bayes) [4] и рандом Forest (Random Forest). Также тестируются неконтролируемые алгоритмы обучения, k-средних (K-means) и гауссовская смешанная модель для максимизации ожиданий (GMM-EM).

Интервал времени для сбора статистических признаков составляет 60 секунд. В таблице 1 показаны результаты мониторинга обученной виртуальной машины с использованием обучающего модуля. В таблице 2 показаны результаты одновременного мониторинга трех виртуальных машин на трех серверах.

Таблица 1

Результаты обнаружения различных алгоритмов машинного обучения

Метод	Accuracy, %	FP, %	FN, %	Точность,%	Полнота,%	F1-Score
LR	94,36	0,00	7,85	100,00	92,15	0,9591
SVM Linear Kernel	93,85	1,41	7,92	99,41	92,08	0,9560
SVM RBF Kernel	93,90	2,46	7,51	98,96	92,49	0,9562
SVM Poly Kernel	94,07	3,38	7,23	98,58	92,77	0,9559
Decision Tree	94,24	1,73	7,04	99,29	92,77	0,9559
Naïve Bayes	94,92	0,00	7,07	100,00	92,93	0,9293
Random Forest	94,96	0,81	6,60	99,67	93,40	0,9643
K-means	64,05	22,93	41,07	86,75	58,93	0,7019
Gaussian EM	62,26	95,88	13,39	69,56	86,61	0,7715

Таблица 2

Совместные результаты обнаружения трех виртуальных машин

Метод	Accuracy, %	FP, %	FN, %	Точность, %	Полнота, %	F1-балл
LR	97,77	0,37	3,82	99,68	96,18	0,9790
SVM Linear Kernel	99,73	0,068	0,44	99,94	99,56	0,9975
SVM RBF Kernel	98,15	3,78	0,24	96,93	99,76	0,9832
SVM Poly Kernel	99,13	0,40	1,27	99,66	98,73	0,9920
Decision Tree	99,07	0,061	0,0167	99,95	98,33	0,9913
Naïve Bayes	98,47	3,07	0,27	97,51	99,73	0,9861
Random Forest	99,53	0,00	0,09	100,00	99,12	0,9956
K-means	87,76	0,44	22,05	99,54	77,95	0,8743
Gaussian EM	66,53	13,17	50,37	81,94	49,63	0,6182

Анализ результатов обнаружений. Полученные данные разделены на собранные данные на обучающие выборки (80%) и тестовые выборки (20%). Использована перекрестная проверка для оценки производительности. Также применен многомерный анализ данных результатов и использовано несколько показателей эффективности. Точность показателей указывает на правильность обнаружения в целом по тестируемым образцам. Ложноположительные (FP) и ложноотрицательные (FN) указывают на ложные тревоги и промахи соответственно.

Точность показывает какая часть тревог является истинной. Отзыв показывает долю обнаруженных атак. Оценка F1 является часто используемым критерием для баланса FP и FN. Чем выше показатель F1, тем выше производительность алгоритма.

Для вычисления F1 и точности используются следующие формулы:

$$\text{точность} = \frac{TP}{TP + FP},$$

где TP — True positive, классификатор верен.

$$F1 = 2 * \frac{\text{точность} * \text{полнота}}{\text{точность} + \text{полнота}}.$$

Результаты вычислений данных переменных представлены в таблицах 1 и 2.

В эксперименте по мониторингу одного хост-сервера (таблица 1) все контролируемые алгоритмы достигают точности более 93 % и более 0,95 баллов F_1 (за исключением наивного байесовского алгоритма с F_1 = 0,9293). Среди них Random Forest работает лучше всего с точностью 94,96 % и 0,9643 балла F_1. Кроме того, Random Forest обеспечивает самая высокая полнота, что означает, что он обнаруживает наибольшее количество атак среди всех сравниваемых алгоритмов. Наивный байесовский метод обеспечивает нулевую скорость FP и относительно низкий FN.

В последних двух строках таблицы 2 показаны алгоритмы k-means и гауссовские алгоритмы EM, которые не так эффективны, как другие, так как являются алгоритмами обучения без учителя [5], то есть они обучаются на немаркированных выборках. Чтобы повысить производительность этих

алгоритмов обучения без учителя можно использовать совместные данные из нескольких виртуальных машин для обучения и чаще переобучать модель, поскольку немаркированные данные легче получить, чем помеченные. Хотя есть опасения по поводу неконтролируемого обучения, данные алгоритмы могут быть интегрированы в механизм онлайн-обновления системы путем предварительной кластеризации чистых образцов для онлайн-обучения.

В эксперименте по мониторингу нескольких хостов (таблица 2) все алгоритмы машинного обучения дают лучшие результаты, чем в эксперименте по мониторингу одного хоста. Достигли наивысшего показателя F_1-балл 0,9975 и точности 99,73 %, используя SVM с линейным ядром. Четыре алгоритма (SVM с ядрами Linear и Poly, Decision Tree и Random Forest) обеспечивают точность более 99 %. K-means улучшается на 23 % в этом эксперименте. Алгоритм Gaussian EM улучшается только на 3 %, что указывает на то, что данные не подчиняются многомерному нормальному распределению.

Заключение. Предложена система обнаружения DDoS-атак, основанная на машинном обучении, для предотвращения атак на стороне источника в облачном сервисе. Извлечены статистические характеристики двух DDoS-атак и запущены реальные экспериментальные атаки для оценки.

Предлагаемая система способна обнаруживать атаки с высокой точностью (99,7 %) и низким уровнем ложных срабатываний (< 0,07 %). Обнаружив DDoS-атаки на исходные виртуальные машины в облаке, можно «подавить атаки в зародыше», а также защитить репутацию облачного провайдера.

Библиографический список

1. The dark menace: Characterizing network-based attacks in the cloud / R. Miao, R. Potharaju, M. Yu, N. Jain // In: Proceedings of the 2015 ACM Conference on Internet Measurement Conference. — ACM, 2015. — P. 169–182. 10.1145/2815675.2815707

2. DDos-resilient scheduling to counter application layer attacks under imperfect detection. / S. Ranjan, R. Swaminathan, M. Uysal, E. W. Knightly // In: INFOCOM. Citeseer, 2006. https://doi.org/10.1109/INFOCOM.2006.127

3. Detecting distributed denial of service (DDos) attacks through inductive learning, / S. Noh, C. Lee, K. Choi, G. Jung // In: International Conference on Intelligent Data Engineering and Automated Learning. — Springer, 2003, — P. 286–295. 10.1007/978-3-540-45080-1_38

4. Amor, N. B. Naive bayes vs decision trees in intrusion detection systems / N. B. Amor, S. Benferhat, Z. Elouedi // In: Proceedings of the 2004 ACM symposium on Applied computing. — ACM, 2004 — P. 420–424. 10.1145/967900.967989

5. Proactive detection of distributed denial of service attacks using mib traffic variables-a feasibility study / J. B. Cabrera, L. Lewis, X. Qin, [et al.] // In: Integrated Network Management Proceedings, 2001. — IEEE/IFIP International Symposium on. IEEE, 2001. — P. 609–622. 10.1109/INM.2001.918069

Об авторах:

Сафарьян Ольга Александровна доцент кафедры «Кибербезопасность информационных систем» Донского государственного технического университета (344003, РФ, г. Ростов-на-Дону, пл. Гагарина, 1), кандидат технических наук, доцент, safari_2006@mail.ru

Джабиа Нзи Жан Максим, студент кафедры «Кибербезопасность информационных систем» Донского государственного технического университета (344003, РФ, г. Ростов-на-Дону, пл. Гагарина, 1), jjjmaxime@gmail.com

About the Authors:

Safaryan, Olga A., Associate Professor, Cybersecurity of Information Systems Department, Don State Technical University (1, Gagarin sq., Rostov-on-Don, 344003, RF), Cand.Sci. (Eng.), Associate Professor, safari_2006@mail.ru

Djabia Nzi Jean Maxim, Student, Cybersecurity of Information Systems Department, Don State Technical University (1, Gagarin sq., Rostov-on-Don, 344003, RF), jjjmaxime@gmail.com