УДК 004.056.57
Doi: 10.31772/2712-8970-2021-22-3-414-424

Для цитирования: Жуков В. Г., Пигалев Я. В. Обнаружение информационного взаимодействия объектов информационной системы с DGA доменами // Сибирский аэрокосмический журнал. 2021. Т. 22, № 3. С. 414–424. Doi: 10.31772/2712-8970-2021-22-3-414-424.

For citation: Zhukov V. G., Pigalev Y. V. Detection of information system objects interaction with DGA domains. *Siberian Aerospace Journal*. 2021, Vol. 22, No. 3, P. 414–424. Doi: 10.31772/2712-8970-2021-22-3-414-424.

Обнаружение информационного взаимодействия объектов информационной системы с DGA доменами

В. Г. Жуков, Я. В. Пигалев*

Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева
Российская Федерация, 660037, г. Красноярск, просп. им. газ. «Красноярский рабочий», 31
*E-mail: pigalevyan1998@mail.ru

В настоящее время разработчики вредоносного программного обеспечения активно применяют технику генерации доменных имен DGA для установления информационного взаимодействия между вредоносным программным обеспечением и их командными центрами управления. Генерация доменных имен в соответствии с заданным алгоритмом позволяет вредоносному программному обеспечению обходить блокировки средств защиты информации, делая их малоэффективными и устанавливать канал связи для получения команд управления и их параметров, а также для передачи информации из информационной системы на внешние ресурсы, контролируемые злоумышленниками. Таким образом, необходимо разрабатывать новые подходы к решению задачи обнаружения сгенерированных с помощью DGA доменных имен в DNS трафике информационной системы.

В рамках проведенного исследования авторами разработано решение для обнаружения информационного взаимодействия объектов информационной системы с DGA доменами, основанное на применении машинного обучения. Обнаружение информационного взаимодействия происходит в два этапа. На первом этапе методами машинного обучения решается задача классификации для каждого DNS имени из общего потока DNS запросов информационной системы. На втором этапе для каждого DNS имени, классифицированного как DGA, осуществляется обогащение данными из внешних источников и принятие окончательного решения о вредоносном характере запроса на разрешение данного DNS имени с последующим оперативным уведомлением администратора безопасности по каналам электронной почты.

В работе приведено описание процесса разработки классификатора на основе машинного обучения, определены входные характеристические данные DNS имени, необходимые для классификации, представлены результаты обучения классификатора на представительном множестве тестовых данных. Обоснована логика принятия решения о вредоносном характере DNS запросов. Разработанное решение было апробировано в рамках экспериментального стенда. Предложены рекомендации по поддержке корректной работы классификатора на основе машинного обучения.

Применение разработанного решения сделает возможным апостериорное обнаружение информационного взаимодействия вредоносного программного обеспечения со скомпрометированных объектов информационной системы с серверами командных центров управления злоумышленников.

Ключевые слова: информационная безопасность, DNS, Domain Generation Algorithm.

Detection of information system objects interaction with DGA domains

V. G. Zhukov, Y. V. Pigalev*

Reshetnev Siberian State University of Science and Technology
31, Krasnoyarskii rabochii prospekt, Krasnoyarsk, 660037, Russian Federation
*E-mail: pigalevyan1998@mail.ru

Currently, malware developers are actively using domain name generation technique called DGA to establish communication between malware and its command centers. Domain name generation in accordance with a given algorithm allows malicious software to bypass information protection tools blacklists, thus making blacklists ineffective, and establish a communication channel to receive control commands and parameters, as well as to transfer information from the information system to external resources controlled by the attackers. Thus, it is necessary to develop new approaches to DGA generated domain names detection using DNS traffic of an information system.

During the research, the authors have developed a solution for detecting the information system objects interaction with DGA domains based on the use of machine learning. Detection of this interaction occurs in two stages. On the first stage the classification task is being solved for each DNS name from overall information system DNS stream. On the second stage, for each DNS name classified as a DGA, corresponding DNS query is being enriched using data from external sources and a final decision about the malicious nature of the request to resolve this DNS name is being made, followed by notification of the security administrator via e-mail channels.

The paper describes the process of developing a classifier based on machine learning, defines the input data of the DNS name necessary for classification, presents the results of classifier training on a representative set of test data. The logic of making a decision about the malicious nature of DNS requests has been substantiated. The developed solution was tested using experimental stand. Recommendations for correct classifier operation support are proposed.

Application of the developed solution will make a posteriori detection of information interaction of malicious software working on compromised objects of the information system with the servers of the attackers command and control centers possible.

Keywords: information security, DNS, Domain Generation Algorithm.

Введение

Протокол DNS является инфраструктурообразующим и, как правило, по умолчанию разрешен в информационных системах (ИС) организаций вне зависимости от сферы их деятельности. Информационные потоки DNS трафика, в общем случае, либо недостаточно контролируются, либо не контролируется вообще. Именно по этой причине современное вредоносное программное обеспечение (ВПО) очень часто использует протокол DNS для связи с серверами управления (C&C, Command and Control Server), что подтверждается многочисленными исследованиями, например, Spamhaus за 2019 г. [1].

Средства защиты информации препятствуют такому взаимодействию путем выявления и блокирования DNS запросов на разрешение доменных имен C&C центров, например, с помощью механизма черных списков. Чтобы обойти эти ограничения, злоумышленники используют специальное программное обеспечение (ПО) для генерации доменных имен в соответствии

с заданным алгоритмом – Domain Generation Algorithm (DGA). Применение DGA позволяет злоумышленникам уйти от статического списка доменных имен C&C и сделать черные списки, применяемые средствами защиты, малоэффективными – DGA позволяет генерировать произвольное количество вредоносных доменов, добавить их всех в черный список невозможно [2]. Таким образом, традиционные средства защиты информации с использованием черных списков не являются эффективными и для решения задачи обнаружения DGA доменов необходим другой подход, потому что сам факт исходящего DNS запроса на разрешение DGA имени C&C свидетельствует о скомпрометированном узле внутри защищаемой инфраструктуры или попытке такой компрометации. Одним из перспективных решений является применение методов машинного обучения для автоматизированного обнаружения информационного взаимодействия объектов информационной системы с DGA доменами. В рамках исследования авторами было разработан алгоритм и ПО, позволяющее обнаруживать факты такого информационного взаимодействия.

Основные этапы работы средства обнаружения DGA

В качестве инфраструктуры ИС рассматривается доменная сеть под управлением Microsoft Windows, под объектом ИС понимается любой активный сетевой узел, который может генерировать DNS запросы. Взаимодействие объекта ИС с DGA доменом заключается, как минимум, в инициировании объектом ИС DNS запроса на разрешение DGA доменного имени.

Обнаружить взаимодействие объектов ИС с DGA доменами возможно путем апостериорного анализа записей журнала DNS запросов.

Локальные записи о DNS запросах пересылаются на контроллер домена средствами Windows Log Forwarding, где и происходит их последующая обработка, направленная на обнаружение DGA доменов. Концептуальная схема обнаружения информационного взаимодействия объектов ИС с DGA доменами представлена на рис. 1.

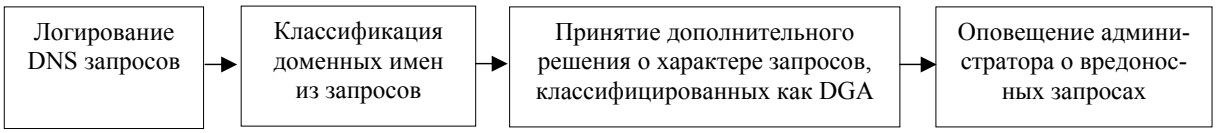

Рис. 1. Схема обнаружения DGA

Fig. 1. DGA detection scheme

Таким образом, процесс обнаружения разделен на два основных этапа:

1) классификация DNS запросов на основе машинного обучения;

2) дополнительная обработка доменных имен, классифицированных как DGA, с принятием окончательного решения о вредоносном характере запроса.

Рассмотрим боле подробно перечисленные этапы работы.

Этап 1: классификация. Все DNS запросы обрабатываются и хранятся в виде записей таблицы базы данных SQLite на контроллере домена. Структура таблицы записей представлена в табл. 1.

На первом этапе доменное имя из каждой записи таблицы классифицируются с помощью машинного обучения на основе атрибутов его доменного имени. Классификация на основе атрибутов доменных имен была выбрана в первую очередь из-за независимости от изменений, вносимых злоумышленниками, в алгоритм работы DGA ВПО [3].

Таблица 1

Описание полей записей таблицы *dns*

Поле записи	Описание
ID	Идентификатор записи
query	Доменное имя
answer	Ответ на запрос (IP адрес)
time	Время и дата запроса
hostname	Имя сетевого узла
status	Статус запроса
image	Приложение, сделавшее запрос
class	Класс имени, устанавливается после обработки классификатором в значение «DGA» или «REAL»

На первом этапе доменное имя из каждой записи таблицы классифицируются с помощью машинного обучения на основе атрибутов его доменного имени. Классификация на основе атрибутов доменных имен была выбрана в первую очередь из-за независимости от изменений, вносимых злоумышленниками, в алгоритм работы DGA ВПО [3].

Разработка классификатора на основе машинного обучения

Задачей классификатора является отнесение каждого доменного имени к одному из двух классов:

1) DGA – DNS запрос на разрешение такого имени рассматривается как вредоносный;

2) реальное доменное имя – DNS запрос на разрешение такого имени рассматривается как легитимный.

Классификатор разработан на языке Python. Для машинного обучения, обработки данных и оценки классификатора использовался набор библиотек scikit-learn [4].

По результатам анализа исследований [5–9] в качестве атрибутов доменного имени, на основании которых будет производиться классификация, были выбраны следующие атрибуты:

1) длина доменного имени;

2) отношение суммы длин всех осмысленных слов (слов, встречающихся в словарях человеческого языка) в доменном имени к общей длине имени;

3) отношение длины самого длинного осмысленного слова в доменном имени к общей длине имени;

4) отношение количества цифр в доменном имени к его общей длине, вычисляется по формуле;

5) расстояние Левенштейна между текущим и предыдущим доменным именем – минимальное количество символов, которые нужно добавить, удалить или изменить, чтобы из предыдущего доменного имени получилось текущее (например, расстояние Левенштейна между test.ru и 1t3st.su равняется 3), именно эта метрика является наиболее подходящей, так как в отличие от, например, расстояния Хэмминга, она не требует одинаковой длины двух строк, к тому же в аналогичных исследованиях DGA используется именно эта метрика [5];

6) информационная энтропия по определению Шеннона;

7) отношение количества гласных к количеству согласных доменного имени;

Так как задача классификации носит бинарный характер, необходима выборка DGA сгенерированных и реальных доменных имен. Выборка имен использовалась для обучения и тестирования классификатора, а также при его финальной оценке.

Реальные доменные имена были взяты из списка самых популярных доменных имен, составленного DomCop [10], источником DGA доменов стал Bambenek Consulting [11] – эти источники уже находили применение при разработке средств по выявлению DGA доменов [5; 6].

Размеры обоих выборок составили 25000 доменных имен, что в сумме дало 50000 доменных имен. Суммарная выборка из 50000 доменных имен была разбита на две: 80 процентов составила обучающая выборка, оставшиеся 20 – тестовая.

В качестве ядра для классификатора был выбран алгоритм случайного леса (Random Forest), положительно зарекомендовавший себя при решении аналогичных задач [3; 12].

Перед обучением классификатора на основе алгоритма Random Forest, было проведено предварительное тестирование этого алгоритма с помощью стратифицированного варианта кросс-валидации с 10-ю блоками на обучающей выборке. Применяя метод кросс-валидации, обучающая выборка доменных имен разбивается случайным образом на десять блоков одинакового размера, затем поочередно каждый блок рассматривается как тестовая выборка, а остальные девять блоков – как обучающая выборка. Для каждого такого блока рассчитывается таблица сопряженности. Далее подсчитывается итоговая таблица сопряженности средняя по 10 блокам.

В итоговой таблице сопряженности (табл. 2) представлены значения среднего количества правильно определенных имен, ошибок первого, второго рода в виде процентного отношения к количеству доменных имен в одном блоке, состоящем из 4000 доменных имен.

Таблица 2

Средняя таблица сопряженности для Random Forest при тестировании с помощью кросс-валидации

		Классифицировано, %		Итого, %
		DGA	Реальное	
Фактически, %	DGA	48,6	1,26	49,86
	Реальное	0,77	49,37	50,14
Итого, %		49,37	50,63	100

Показатели ошибок для алгоритма Random Forest при тестировании с помощью кросс-валидации на обучающей выборке удовлетворительны.

На основе итоговой таблицы сопряженности, была подсчитана точность классификации (*accuracy*) по формуле

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

где TP – количество истинно положительных случаев; TN – количество истинно отрицательных случаев; FP – количество ложно положительных случаев; FN – количество ложно отрицательных случаев.

Random Forest показал высокую точность при тестировании на обучающей выборке с помощью кросс-валидации – 98 %.

Для финальной оценки обученного классификатора вычисляется таблица сопряженности, получаемая с помощью классификации имен из тестовой выборки (10000 доменных имен), и подсчитывается точность.

Таблица сопряженности представлена в табл. 3.

Таблица 3

Таблица сопряженности Random Forest на тестовом множестве

		Классифицировано, %		Итого, %
		DGA	Реальное	
Фактически, %	DGA	49,42	1,14	50,56
	Реальное	0,77	48,67	49,44
Итого, %		50,19	49,81	100

Random Forest на тестовом множестве имеет точность, равную 98,09 %.

Полученные результаты (точность, количество ошибок первого и второго рода) позволяют перейти ко второму этапу исследования.

Этап 2: обогащение результатов классификации и принятие решения

Для снижения возможных ошибок классификации все записи в таблице DNS базы данных, в которых доменные имена классифицированы как «DGA», отбираются для дальнейшего обогащения и принятия решения о их вредоносном характере.

Принятие решения о вредоносном характере запроса производится на основе расчета коэффициента угрозы *Threat*, который вычисляется на основе параметров соответствующего DNS запроса.

Параметры вычисляются по результатам обогащения запроса из внешних источников информации и отражают такие отличительные свойства DNS запросов на разрешение DGA имен, как:

1) одно доменное имя может разрешаться в несколько IP-адресов и в соответствии с исследованием EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis [13] вредоносные домены одного семейства ВПО, как правило, разрешаются в IP-адреса разных стран;

2) как правило, DGA домены сгенерированы за час до атаки и действуют в пределах 24-х часов [12; 14];

3) DGA имена мало документированы: нельзя получить информацию об организации-владельце DGA имени, администраторе домена;

4) во время работы ВПО с использованием DGA, ВПО на объекте информационной системы перебирает набор сгенерированных имен в поиске доступного путем запроса на каждое из них, большая часть завершается с сообщением об ошибке NXDOMAIN, что говорит о том, что доменное имя не найдено – несуществующий домен [12].

Формула вычисления коэффициента *Threat* представлена далее:

$$Threat = \sum_{i=1}^{7} x_i, \tag{2}$$

где x_1 – устанавливается в 1, если количество стран – владельцев IP-адресов в ответах на DNS запрос больше 2; x_2 – устанавливается в 1, если в whois-ответе на доменное имя нет имени организации владельца доменного имени; x_3 – устанавливается в 1, если в whois-ответе нет имени администратора; x_4 – устанавливается в 1, если было установлено, что за текущий день количество DNS запросов, результирующих ошибкой NXDOMAIN, больше порогового значения; x_5 – устанавливается в 1 в случае, если разница между датой регистрации домена и временем DNS запроса меньше 1-го часа; x_6 – устанавливается в 1, если разница между датой окончания срока регистрации домена и временем запроса меньше 1-го дня; x_7 – устанавливается в 1, если в whois-ответе нет даты регистрации домена и даты окончания регистрации домена.

Параметры носят бинарный характер, по умолчанию каждый параметр равен 0.

В случае, если коэффициент угрозы *Threat* превышает значение 3, то принимается решение о том, что соответствующий DNS запрос действительно вредоносен. В противном случае принимается решение о легитимности DNS запроса: произошла ошибка классификации.

После принятия решения запрос, соответствующее решение и данные, полученные путем обогащения, записываются в таблицу suspicious базы данных для анализа работы средства.

Структура записей таблицы suspicious представлена в табл. 4.

Таблица 4

Описание дополнительных полей записей suspicious

Поле записи	Описание
country_number	Количество стран-владельцев IP-адресов в ответе на запрос
registrar	Имя администратора домена
creation_date	Дата регистрации домена
expiration_date	Дата окончания срока регистрации
organisation	Имя организации – владельца доменного имени
NXDOMAIN_query_count	Количество ответов-ошибок NXDOMAIN для комбинации приложения и компьютера
domainStatus	Статус доменного имени, «Up» – доступен, «Down» – недоступен
queryType	Принятое на основе вычисленных параметров решение о вредоносном характере DNS запроса, «Malicious» – вредоносный, «Benign» – легитимный
parentRecord	Ссылка на соответствующую запись в таблице DNS

О DNS запросах, классифицированных и подтвержденных как DGA, по электронной почте администратору высылаются оповещения. Оповещение содержит основную информацию о соответствующем вредоносном DNS запросе.

Тестирование работы средства обнаружения взаимодействия информационных объектов информационной системы с DGA доменами

Тестирование проводилось в тестовой доменной сети Windows, состоящей из двух компьютеров, совершающих DNS запросы, и контроллера домена, на котором установлено средство обнаружения DGA.

Все DNS запросы компьютеров в доменной сети записывались в журнал контроллера домена. Для тестирования с компьютеров тестовой сети предварительно были произведены DNS запросы на разрешение набора реальных имен и три на DGA имена.

Запросы на DGA имена симулировали перебор вредоносным программным обеспечением на компьютере набора DGA имен в поиске действующего. По этой причине первые два запроса вернули ошибку NXDOMAIN (ijoratsdxgwubk.ru и bsqncknwntpill.ru), последний вернул IP-адрес C&C сервера (oqunedkxrrrd.ru).

Запросы были записаны средством на контроллере домена из журнала в таблицу DNS базы данных. Фрагмент запросов представлен на рис. 2, запросы на DGA имена выделены.

ID	query	answer	time	hostname	status	image	class
...	Фильтр	Фильтр	Фильтр	Фильтр	Ф...	Фильтр	Фи...
517	oqunedkxrrrd.ru	::ffff:23.61.215.146	2020-06-13 11:56:03.2...	win8hostClone...	0	C:\virusexample.exe	NULL
518	bsqncknwntpill.ru		2020-06-13 11:55:49.5...	win8hostClone...	9003	C:\virusexample.exe	NULL
519	ijoratsdxgwubk.ru		2020-06-13 11:55:18.3...	win8hostClone...	9003	C:\virusexample.exe	NULL
534	ndj6iayz7u2mbga4pqu4z...	::ffff:185.15.175.157	2020-06-13 11:54:23.4...	win8hostClone...	0	C:\Program Files (x86)\Internet ...	NULL
535	ndj6iayz7u2mbga4pqu4z...	::ffff:185.15.175.158	2020-06-13 11:54:23.4...	win8hostClone...	0	C:\Program Files (x86)\Internet ...	NULL

Рис. 2. Выгруженные DNS запросы

Fig. 2. Stored DNS quieries

После окончания записи запросов начался этап классификации: все доменные имена из базы данных были классифицированы на основе машинного обучения. Далее начался второй этап: все записи из таблицы, имеющие результатом классификации DGA, были отобраны для обогащения и определения характера запроса.

Запросы были обогащены с помощью whois запросов на имя домена, геолокационной проверки IP-адреса, подсчитывания количества ответов NXDOMAIN, затем параметры были вычислены на основе результатов обогащения.

Для соответствующей этим тестовым DGA запросам комбинации приложения и компьютера («win8hostClone» и «C:\virusexample.exe») в базе данных существовали 2 доменных запроса с ошибкой NXDOMAIN, из-за чего параметр x_4 установился в 1 для каждой записи с соответствующей комбинацией приложения и названия компьютера, все три домена недоступны (таким образом, нет данных об организации, администраторе домена, датах начала и конца регистрации, что установило соответствующие параметры x_2, x_3, x_7 в 1). Значения остальных параметров остались по умолчанию.

Следовательно, по формуле (2), коэффициент угрозы для тестовых DGA запросов был равен:

$$Threat = 0 + 1 + 1 + 1 + 0 + 0 + 1 = 4. \quad (3)$$

Значение коэффициента угрозы равно 4, на основе чего было принято решение, что запросы действительно вредоносные.

После классификации, обогащения, определения характера отобранные записи с дополнительными сведениями были записаны в таблицу suspicious базы данных.

Фрагмент таблицы suspicious с данными, полученными при обогащении, для запросов, представленных на рис. 2, приведен на рис. 3. Для всех 3-х вредоносных тестовых запросов было принято решение об их вредоносном характере (на это указывает значение «Malicious» в поле «queryType»), для реальных доменных запросов было принято решение об их легитимности (на это указывает значение «Benign» в поле «queryType»).

ID	query	answer	time	hostname	image	country_number	registrar	queryType
119	oqunedkxrrrd.ru	::ffff:23.61.215.1...	2020-06-13 ...	win8hostClone.d...	C:\virusexample.exe	1	Error, domain is unavaible	Malicious
120	bsqnckrwntpill.ru		2020-06-13 ...	win8hostClone.d...	C:\virusexample.exe	0	Error, domain is unavaible	Malicious
121	ijoratsdxgwubk.ru		2020-06-13 ...	win8hostClone.d...	C:\virusexample.exe	0	Error, domain is unavaible	Malicious
122	ndj6iayz7u2mbg...	::ffff:185.15.175....	2020-06-13 ...	win8hostClone.d...	C:\Program Files (x...	1	RU-CENTER-RU	Benign
123	ndj6iayz7u2mbg...	::ffff:185.15.175....	2020-06-13 ...	win8hostClone.d...	C:\Program Files (x...	1	RU-CENTER-RU	Benign
124	t3848077496801...	::ffff:50.116.239....	2020-06-13 ...	win8hostClone.d...	C:\Program Files (x...	1	Amazon Registrar, Inc.	Benign
125	ndj6iayz7u2mbg...	::ffff:185.15.175....	2020-06-13 ...	win8hostClone.d...	C:\Program Files (x...	1	RU-CENTER-RU	Benign
126	ndj6iayz7u2mbg...	::ffff:185.15.175....	2020-06-13 ...	win8hostClone.d...	C:\Program Files (x...	1	RU-CENTER-RU	Benign

Рис. 3. Результат работы средства

Fig. 3. Results

Администратор был оповещен о каждом вредоносном запросе на DGA имя. Тестовое оповещение об одном из вредоносных запросов представлено на рис. 4.

Subject: Security Notification 13:19 (2 часа назад)

кому:

To: admtestdns123@gmail.com'
From: notificationdnstest@gmail.com

Malicious DGA query oqunedkxrrrd.ru detected at win8hostClone.dnstest.ru from C:\Program Files (x86)\Internet Explorer\iexplore.exe . Domain resolved in ::ffff:23.61.215.146 , time of query is 2020-06-13 11:56:03.2279401 . Domain is Down

Рис. 4. Оповещение об обнаружении вредоносного запроса

Fig. 4. Notification about detected malicious quiery

Таким образом, в ходе тестирования было обнаружено взаимодействие объектов ИС с DGA доменами. Данные о вредоносных запросах, хранящиеся в базе данных и отсылаемые по электронной почте администратору, позволяют определить факт и обстоятельства компрометации объекта ИС.

Поддержка корректной работы классификатора на основе машинного обучения

Классификаторы на основе алгоритмов машинного обучения деградируют со временем использования, к тому же, классификаторы имеют тенденцию работать в практических условиях хуже, чем они работали при тестировании [15].

Предполагается, что, кроме непредсказуемых изменений, точность классификатора со временем может снижаться из-за изменения общего алгоритма DGA генерации вредоносных доменов (т. е. структура имени изменится).

Таким образом, для поддержки точности классификатора, необходимо следить за его работой и, в случае необходимости, модифицировать его и/или переобучить: изменить набор атрибутов доменного имени, обучить классификатор на другом, более новом, наборе доменных имен.

Заключение

По результатам проведенных исследований авторами был разработан и программно реализован двухэтапный алгоритм обнаружения DGA: классификация с помощью машинного обучения на базе алгоритма Random Forest и принятие решения о характере запросов на основе результатов обогащения.

Применение разработанного программного обеспечения позволяет апостериорно обнаружить взаимодействие объектов ИС с DGA доменами. Таким образом появляется возможность обнаружить факт скомпрометированности объекта ИС и повысить его защищенность, путем совместного использования разработанного средства с другими СЗИ.

Средство обнаружения разработано для анализа DNS запросов в доменной сети Microsoft Windows, но его ядро – классификатор на основе машинного обучения и логика принятия решения о вредоносном характере – применимо и для других операционных систем и оборудования.

Библиографические ссылки

1. Spamhaus Botnet Threat Report 2019 [Электронный ресурс]. URL: https://www. spamhaus.org/news/article/793/spamhaus-botnet-threat-report-2019 (дата обращения 02.02.2020).

2. Threat Brief: Understanding Domain Generation Algorithms (DGA) [Электронный ресурс]. URL: https://unit42.paloaltonetworks.com/threat-brief-understanding-domain-generation-algorithms-dga/ (дата обращения 05.08.2020).

3. Sivaguru R., Choudhary C. An Evaluation of DGA Classifiers // IEEE International conference on Big Data, Seattle, USA, 2018. P. 5058–5067.

4. Scikit-learn: machine learning in Python [Электронный ресурс]. URL: https://scikit-learn.org/stable (дата обращения 03.01.2020).

5. Li Y., Xiong K. Machine Learning Framework for Domain Generation Algorithm-Based Malware Detection. IEEE Access, 2019. P. 32765–32782.

6. Anderson H. S., Woodbridge J. DeepDGA: Adversarially – Tuned Domain Generation and Detection. Proceedings of the 2016 ACM Workshop and Artificial Intelligence and Security. 2016. P. 13–21.

7. Anderson H. S., Woodbridge J. Predicting Domain Generation Algorithms with Long Short-Term Memory Networks. Endgame, Inc, 2016. 13 p.

8. Gupta B., Sheng M. Machine Learning for Computer and Cyber Security: Principles, Algorithms, and Practices. Taylor and Francis Group, 2019. 364 p.

9. Alazab M., Tang M. Deep Learning Applications for Cyber Security. Springer Nature Switzerland, 2019. 246 p.

10. Top 10 million Websites based on Open data from Common Crawl & Common Search [Электронный ресурс]. URL: https://www.domcop.com/top-10-million-websites (дата обращения 03.02.2020).

11. Bambenek Consulting [Электронный ресурс]. URL: http://osint.bambenekconsulting.com/feeds/dga-feed.txt (дата обращения 16.01.2020).

12. Wang Z., Jia Z. A Detection Scheme for DGA Domain Names. SVM Proceedings of the 2018 International Conference on Mathematics, Modelling, Simulation and Algorithms. New York, USA, 2018. P. 257–263.

13. Bilge L., Kirda E. EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis. Proceedings of the Network and Distributed System Security Symposium, San Diego, USA, 2011. 17 p.

14. Plohmann D., Yakdan K. A Comprehensive Measurement Study of Domain Generating Malware. Proceedings of the 25th USENIX Security Symposium, Austin, USA, 2016. P. 263–278.

15. Why Machine Learning Models Degrade in Production [Электронный ресурс]. URL: https://towardsdatascience.com/why-machine-learning-models-degrade-in-production-d0f2108e9214 (дата обращения 25.05.2020).

References

1. Spamhaus Botnet Threat Report 2019. Available at: https://www.spamhaus.org/news/article/793/spamhaus-botnet-threat-report-2019 (accessed: 02.02.2020).

2. Threat Brief: Understanding Domain Generation Algorithms (DGA). Available at: https://unit42.paloaltonetworks.com/threat-brief-understanding-domain-generation-algorithms-dga/ (accessed: 05.08.2020).

3. Sivaguru R., Choudhary C. An Evaluation of DGA Classifiers. IEEE International conference on Big Data, Seattle, USA, 2018, P. 5058–5067.

4. Scikit-learn: machine learning in Python. Available at: https://scikit-learn.org/stable (accessed: 03.01.2020).

5. Li Y., Xiong K. Machine Learning Framework for Domain Generation Algorithm-Based Malware Detection. IEEE Access, 2019, P. 32765–32782.

6. Anderson H. S., Woodbridge J. DeepDGA: Adversarially – Tuned Domain Generation and Detection. Proceedings of the 2016 ACM Workshop and Artificial Intelligence and Security, 2016, P. 13–21.

7. Anderson H. S., Woodbridge J. Predicting Domain Generation Algorithms with Long Short-Term Memory Networks. Endgame, Inc, 2016, 13 p.

8. Gupta B., Sheng M. Machine Learning for Computer and Cyber Security: Principles, Algorithms, and Practices. Taylor and Francis Group, 2019, 364 p.

9. Alazab M., Tang M. Deep Learning Applications for Cyber Security. Springer Nature Switzerland, 2019, 246 p.

10. Top 10 million Websites based on Open data from Common Crawl & Common Search. Available at: https://www.domcop.com/top-10-million-websites (accessed 03.02.2020).

11. Bambenek Consulting. Available at: http://osint.bambenekconsulting.com/feeds/dga-feed.txt (accessed 16.01.2020).

12. Wang Z., Jia Z. A Detection Scheme for DGA Domain Names. SVM Proceedings of the 2018 International Conference on Mathematics, Modelling, Simulation and Algorithms, New York, USA, 2018, P. 257–263.

13. Bilge L., Kirda E. EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis. Proceedings of the Network and Distributed System Security Symposium, San Diego, USA, 2011, 17 p.

14. Plohmann D., Yakdan K. A Comprehensive Measurement Study of Domain Generating Malware. Proceedings of the 25th USENIX Security Symposium, Austin, USA, 2016, P. 263–278.

15. Why Machine Learning Models Degrade in Production. Available at: https://towardsdatascience.com/why-machine-learning-models-degrade-in-production-d0f2108e9214 (accessed 25.05.2020).

Жуков Вадим Геннадьевич – кандидат технических наук, доцент кафедры безопасности информационных технологий; Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева. E-mai: zhukov@mail.sibsau.ru.

Пигалев Ян Вячеславович – магистрант; Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева. E-mail: pigalevyan1998@mail.ru.

Zhukov Vadim Gennadevich – Cand. Sc., Associate Professor at the Department of Information Technology Security, Reshetnev Siberian State University of Science and Technology. E-mail: zhukov.sibsau@gmail.com.

Pigalev Yan Vyacheslavovich – Master's Degree Student; Reshetnev Siberian State University of Science and Technology. E-mail: pigalevyan1998@mail.ru.