2019 IEEE 31-я Международная конференция по инструментам с искусственным интеллектом (ICTAI)

Интеграция механизма внимания и глубокой нейронной сети для обнаружения доменных имен DGA

Фангли Рен
Институт информационной инженерии,
Китайская академия наук.
Школа кибербезопасности,
Университет Китайской наук.
Электронная почтarenfangli@iie.ac.cn

Чжэнвэй Цзян
Институт информационной инженерии,
Китайская академия наук.
Электронная
почтajiangzhengwei@iie.ac.cn

Цзянь Лю
Институт информационной инженерии,
Китайская академия наук.
Электронная почтaliujian6@iie.ac.cn

***Аннотация* - Алгоритмы генерации *доменов* (DGA) используются вредоносными программами для генерации доменных имен, с помощью которых они подтверждают точки рандеву со своими командно-контрольными (C2) серверами. Обнаружение доменных имен DGA является одной из важных технологий для обнаружения командно-контрольных коммуникаций. Учитывая случайность доменных имен DGA, в недавних работах по обнаружению DGA для классификации доменных имен использовались методы машинного обучения, основанные на архитектурах извлечения признаков и глубокого обучения. Однако эти методы плохо работают с семействами DGA, основанными на словарных списках, которые генерируют доменные имена путем случайного объединения словарных слов. В данной работе мы предложили модель ATT-CNN-BiLSTM для обнаружения и классификации доменных имен DGA. Во-первых, слой конволюционной нейронной сети (CNN) и двунаправленной длинной кратковременной памяти (BiLSTM) использовался для извлечения особенностей информации о доменных последовательностях; во-вторых, слой внимания использовался для распределения соответствующего веса извлеченной глубинной информации домена. Наконец, сообщения о признаках домена с разными весами поступали в выходной слой для задач обнаружения и классификации. Результаты экспериментов демонстрируют эффективность предложенной модели как на обычных доменных именах DGA, так и на именах, основанных на списках слов. Точнее говоря, мы получили F1 оценку 98,92 % для задачи обнаружения и макросреднюю F1 оценку 81 % для задачи классификации доменных имен DGA.**

Ключевые слова* - *кибербезопасность, алгоритм генерации доменов, механизм внимания, глубокое обучение

I. Введение

Система доменных имен (DNS) - это важная инфраструктура Интернета, которая сопоставляет легко запоминающиеся имена хостов со скучными и трудно запоминающимися IP-адресами. Она предоставляет критически важные вспомогательные услуги для нормальной работы различных доменных веб-приложений, электронной почты и распределенных систем. Благодаря способности DNS-трафика проникать через брандмауэр [1], злоумышленники начинают использовать DNS для проведения различных кибератак. Некоторые крупные атакующие организации даже могут использовать уникальную способность DNS для построения атакующего поведения, что серьезно влияет на общую работу сети. В последние годы различные вспышки вредоносного ПО нанесли значительный ущерб правительству, энергетике, производству и другим ключевым информационным инфраструктурам. Защита от вредоносного ПО имеет решающее значение для безопасности киберпространства.

Современные вредоносные программы, такие как ботнеты, программы-вымогатели и современные постоянные угрозы, часто используют службу DNS для связи с командно-контрольным (C&C) сервером для передачи файлов и обновления программного обеспечения. Для этого на сервере Вредоносная программа должна знать, к какому C&C-серверу подключаться. В прежние времена для этого использовались простые подходы - жесткое кодирование IP-адреса или доменного имени. Но эти методы легко перехватить с помощью черного списка, трафик на определенный IP-адрес может быть тривиально заблокирован, а доменные имена могут быть легко захвачены или вырезаны. Для повышения надежности и устойчивости связи между C&C-сервером и вредоносным ПО, вредоносное ПО обычно использует алгоритмы генерации доменов (DGA) для автоматического создания большого набора псевдослучайных доменных имен, а затем выбирает одно или несколько эффективных доменных имен для разрешения IP-адреса, чтобы реализовать связь с C2-сервером и избежать блокировки методом черного списка. Задача подтверждения того, что доменное имя сгенерировано DGA, является важнейшим этапом защиты от вредоносного ПО.

В последние годы широко ведутся исследования по выявлению доменов DGA. От методов, основанных на ручном подборе признаков [2] на основе машинного обучения, до применения глубокого обучения. Вудбридж [3] использовал простую модель сетей долговременной кратковременной памяти (LSTM) на уровне символов для идентификации доменов DGA, которая имела высокий уровень эффективности, за исключением классов DGA, напоминающих английские слова. Тран [4] предложил новую модель на основе LSTM для обнаружения ботнетов, которые используют DGA для генерации большого количества доменов в качестве командно-контрольного сервера. Сакс [5] предложил модель на основе CNN для обнаружения доменных имен DGA. Однако они не проверяли эффективность модели на основе словесного списка DGA, который является разновидностью DGA, имитирующего состав и методы именования обычных доменных имен, что также делает доменные имена DGA более скрытными и затрудняет их обнаружение. Исследований по обнаружению семейств DGA, созданных на основе английских списков слов, до сих пор не проводилось, что делает проблему обнаружения DGA-доменов на основе списков слов еще более сложной.

В этой статье мы предложили модель глубокой нейронной сети с механизмом внимания (ATT-CNN-BiLSTM) для обнаружения и классификации доменных имен DGA. Основная идея нашей ансамблевой модели заключается в том, что достоверность контекста, присущего доменам, может содержать достаточную информацию, с помощью которой можно отличить доменные имена DGA, особенно те, которые основаны на списках слов.

В целом, данная статья содержит следующие материалы:

- Задача обнаружения доменных имен похожа на задачу обработки естественного языка. Мы строим модель для изучения семантических связей между доменными именами. CNN и BiLSTM могут получить информацию

2375-0197/19/$31.00 ©2019 IEEE DOI
10.1109/ICTAI.2019.00121

из прошлых и будущих состояний. Поскольку этот метод позволяет более полно передать информацию о текстовом контексте, мы применяем его для обнаружения и классификации доменных имен в DGA.

- Встроив в модель механизм внимания, способный улавливать критическую информацию, мы вычисляем важность корреляции между выходами слоя BiLSTM и получаем общую характеристику последовательностей в соответствии со важности. С использованием модели был проведен эксперимент на собранном наборе данных. Результаты показали, что разработанная модель может эффективно решить проблему обнаружения доменных имен DGA и повысить производительность.

- Предложенная модель может работать как с онлайн, так и с офлайн . Результаты экспериментов показывают, что модель может успешно классифицировать доменные имена на основе списков слов, которые другие современные подходы не смогли определить.

Остальная часть статьи организована следующим образом. В разделе 2 представлены общие сведения об обнаружении доменных имен DGA, а также обзор соответствующих работ. В разд. 3 мы описываем ансамблевую модель, разработанную для обнаружения доменных имен DGA. Набор данных и экспериментальная оценка предложенной модели подробно описаны в разд. 4. Наконец, в разделе 5 мы подводим итоги и обсуждаем возможные будущие работы.

II. История вопроса и сопутствующие работы

A. Алгоритмы генерации доменов

За последние несколько лет большинство семейств вредоносных программ стали использовать другой подход для связи со своими удаленными серверами. Вместо использования жестко закодированных адресов серверов некоторые семейства вредоносных программ теперь используют алгоритм генерации доменов (DGA). DGA - это класс алгоритмов, которые принимают на вход семя, выдают строку и добавляют к ней домен верхнего уровня (TLD), например .com, .net [6]. Методы DGA различаются по сложности, для борьбы с обнаружением вредоносных доменных имен на основе признаков некоторые новые DGA имитируют состав и методы именования обычных доменных имен, что называется DGA на основе словесных списков доменов, усложняющих обнаружение. Matsnu [7] извлекает существительные и глаголы из встроенного списка из более чем 1300 слов для формирования доменов, представляющих собой 24-символьные фразы. Вредоносная программа suppobox [8] создает такие домены, как heavenshake.net, heavenshare.net и leadershare.net, объединяя два псевдослучайно выбранных словаря английского языка, например christinepatterson.net.

B. Методы обнаружения ДГА

Исследование, посвященное отличию легитимных доменных имен от DGA, ведется уже много лет. Легитимные доменные имена и имена DGA отличаются как по структуре, так и по поведению. Анализируя поведение и структуру доменного имени, можно определить, ли оно доменным именем DGA или нет. Ядав [9] применил статистическую технику для обнаружения алгоритмически сгенерированных доменных имен, смоделировав временную корреляцию и информационную энтропию характеристик успешных и неудачных имен. Антонакокис [10] использовал подход к кластеризации по длине, уровню случайности и распределению частот символов, включая распределение n-грамм наблюдаемых доменных имен. Затем была использована скрытая марковская модель для определение вероятности того, что доменное имя является DGA.

Дальнейшие исследования в области обнаружения DGA развиваются в направлении все более широкого использования технологий машинного обучения. Занг [11] предложил алгоритм обнаружения с использованием кластерной корреляции, который идентифицирует доменные имена, сгенерированные DGA или его разновидностями. При этом используются такие характеристики, как TTL, распределение IP-адресов, информация whois и историческая информация о доменных именах. Чжан [12] предложил алгоритм обнаружения, который анализирует особенности доменных имен, такие как состав символов и лексическая иерархическая структура, включающая длину доменного имени, частоту символов и двойные буквы, для обнаружения вредоносных доменных имен. Ядав [13] проанализировал расстояние KL, коэффициент Жаккара и расстояние редактирования доменных имен DGA. Бильге [14] предложил систему EXPOSURE, извлекшую 15 признаков доменных имен и использовавшую для классификации дерево решений J48. Рагурам [15] построил модель обнаружения нормальных доменных имен для быстрой идентификации аномальных доменных имен, Грилл [16] изучил метод, использующий только информацию NetFlow о трафике DNS, а не доменные имена, Ванг [17] предложил использовать сегментацию слов для извлечения лексем из доменных имен для обнаружения DGA-доменных имен с помощью таких признаков, как количество символов и цифр. Однако в реальной сети такие признаки трудно извлечь и собрать.

Между тем глубокие нейронные сети продемонстрировали свою способность находить и извлекать релевантные признаки, а также повысили точность классификации по сравнению с традиционными методами машинного обучения. Для обнаружения доменных имен DGA Вудбридж [3] показал, что сети долговременной кратковременной памяти (LSTM) на уровне символов чрезвычайно эффективны при обнаружении доменных имен DGA. В ходе последующих исследований Тран [4] предложил новую модель на основе LSTM для обнаружения ботнетов, которые используют DGA для генерации большого количества доменов в качестве командно-контрольного сервера. Сакс [5] предложил модель на основе CNN для обнаружения доменных имен DGA. Андерсон [18] исследовал использование методов состязательного обучения обнаружения DGA. Шибахара [19] предложил другой алгоритм, использующий RNN на изменениях в сетевом взаимодействии с целью сокращения времени анализа вредоносного ПО. Эти модели, основанные на глубоких нейронных сетях, имели высокую точность обнаружения доменных имен DGA с высокой случайностью, но низкий уровень идентификации семейств DGA, похожих на английские слова, например suppobox, что приводило к высоким ложноположительным результатам для обычных доменных имен и снижению доверия к модели.

В последнее время достигнут определенный прогресс в решении проблемы низкой эффективности обнаружения в доменах ДГА на основе списков слов. Ян Лю [20] предложил классификатор случайного леса, который использовал извлеченные вручную признаки, такие как частота слов, теги части речи и корреляции слов, для классификации ДГА на основе списков слов. Перейра [21] использовал метод WordGraph для обнаружения ДГА на основе словаря, используя теорию графов, и его метод превосходит конволюционные нейронные сети для ДГА на основе словаря. Кох [22] предложил подход, сочетающий предварительно обученную контекстно-чувствительную модель встраивания слов ELMo с простым полносвязным классификатором для классификации доменов на основе информации на уровне слов. Куртин [23] предложил меру smashword score, которая отражает, насколько близко сгенерированные DGA домены похожи на английские слова, и построил модель машинного обучения, состоящую из RNN, используя обобщенный тест отношения правдоподобия

(GLRT), а также включает побочную информацию, такую как WHOIS. Объединенная модель была способна эффективно идентифицировать семейства DGA с высоким показателем smashword, такие как сложные семейства matsnu и suppobox.

C. Механизм внимания

Механизм внимания в глубоком обучении имитирует характеристики внимания человеческого мозга, которые можно понимать как постоянное внимание к более важной информации. Бахданау [24] впервые предложил применить механизм внимания в нейросетевом машинном переводе. Механизм внимания недавно продемонстрировал успех в широком спектре задач, таких как распознавание речи, машинный перевод и распознавание изображений. Он может использоваться самостоятельно или в качестве слоя в других гибридных моделях. Назначение весов внимания в нейронных сетях достигло большого успеха в различных задачах машинного обучения. Луонг [25] разработал два новых типа моделей на основе внимания для машинного перевода. С тех пор все больше и больше исследований интегрируют механизмы внимания в классификацию текстов, классификацию отношений и извлечение абстракций. Янг [26] предложил иерархическую сеть внимания для классификации документов, которая имеет два уровня механизмов внимания, применяемых как на уровне слов, так и на уровне предложений. Ма представил три вида временного внимания на разных временных шагах. Рийке [27] использовал двухуровневый механизм внимания для извлечения резюме. Чжоу [28] предложил основанную на внимании модель двуязычного представления, которая изучает документы на исходном и целевом языках, а также иерархический механизм внимания для двуязычной LSTM-сети. Модель, интегрированная с механизмом внимания, способна большое внимание важному содержанию, словам и предложениям в окружающем контексте заданной входной последовательности. Успешное применение механизма внимания в естественном языке дает толчок для обнаружения доменных имен DGA, в основном на основе списков слов.

III. Методология

Основываясь на вышеизложенном, мы предлагаем модель с механизмом внимания для обнаружения и классификации DGA, которая называется ATT-CNN-BiLSTM модель. В ней механизм внимания к именам домена представляет собой

представила.

Архитектура модели ATT-CNN-BiLSTM с механизмом внимания для обнаружения доменных имен представлена на рис. 1. Она состоит из пяти компонентов: слоя встраивания, слоя CNN, слоя BiLSTM, слоя внимания и выходного слоя. Перед выходным слоем мы обучаем последовательность доменных имен и используем стратегию отсева, чтобы избежать перебора.

A. Слой встраивания

Поскольку входная последовательность, принимаемая нейросетевой моделью, представляет собой вектор фиксированной длины, а доменное имя - строка, необходимо ввести этап векторизации доменного имени для преобразования строки во входной формат нейронной сети. К необработанным доменным именам применяется предварительная обработка. А именно преобразование символов верхнего регистра в нижний, в первую очередь из-за того, что различение символов верхнего и нижнего регистра может привести к проблеме регуляризации. Затем TLD отбрасываются, и в качестве исходных данных для модели остается только основной домен. Последовательность доменов можно обозначить как $C_i = \{c_1, c_2, c_3, \ldots c_n\}$, где n - длина домена. Например, входное доменное имя christinepatterson.net, сгенерированное suppobox, после предварительной обработки будет выражено как *{c, h, r, i, s, t, i, n, e, p, a, , e, r, s, o, n}*. Слой встраивания работает только со строками фиксированной длины m. Если

если длина входной строки больше m, строки, превышающие m, должны быть обрезаны. Если длина входной строки меньше m, строка должна быть дополнена. Встраивающий слой реализует такую функцию, которая проецирует последовательности входных символов из входной области на последовательность векторов. Затем входная строка кодируется как вектор *v* длины d, которая является переменным параметром.

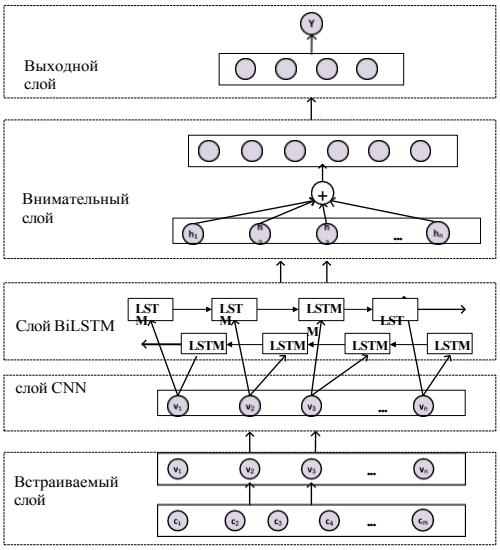

Рис. 1. Архитектура модели ATT-CNN-BiLSTM

B. Слой CNN

После слоя встраивания входная последовательность была встроена в матрицу $m \times d$, следующим шагом является извлечение локально обнаруженных признаков с помощью слоя CNN, который мы используем в качестве компонента извлечения признаков в нашей модели. Фильтры t охватывают всю длину вставки символов d, которая представляет собой скользящую свертку ядер по последовательности вложенных символов. Используя фильтр шириной h и нелинейную функцию, операция свертки может генерировать новый признак $_{oi}$, как показано в уравнении (1).

$$o_i = f(\mathrm{W} \cdot v_i + \mathrm{b}) \quad (1)$$

Где *W* - матрица весов, $_{vi}$ - вектор встраивания доменного имени, b - член смещения, а f - нелинейная функция. С помощью слоя CNN обнаруживается схожий последовательный шаблон доменных имен. Слой CNN состоит из сверточного слоя и слоя объединения. Сверточный слой обладает свойствами локальной связи и разделения веса, что позволяет снизить сложность модели, а объединяющий слой - уменьшить размер параметров и предотвратить перебор. Поскольку доменное имя является одномерным, мы выбираем свертку 1D, которая использует фильтр, применяемый к окну символов с для создания новой карты признаков, и операцию объединения 1D, которая представляет собой нелинейную понижающую выборку, используемую для получения наиболее значимых признаков в качестве следующего шага.

C. Слой BiLSTM

Мы используем BiLSTM для получения контекстуальной характеристики каждого персонажа на большом расстоянии. В двунаправленной архитектуре есть два слоя скрытых узлов из двух отдельных LSTM. Эти два LSTM отражают зависимости в разных направлениях. Первые скрытые слои имеют рекуррентные связи от последних слов, в то время как направление рекуррентных связей второго слоя меняется, передавая

активация в последовательности назад. Таким образом, в слое LSTM мы можем получить прямое скрытое состояние из прямой сети LSTM и обратное скрытое состояние из
обратная LSTM-сеть. Двухсоставная сеть передает информацию о композиционной семантике в обоих направлениях последовательностей символов.

BiLSTM состоит из прямого и обратного LSTM, поскольку выходной вектор слоя CNN - это . $V=\{v_1, v_2, v_3, \ldots v_z\}$, прямая LSTM считывает входной вектор сv_1 наv_z , а обратная LSTM считывает входной вектор сv_z наv_1, nam ely_, the rever→s(.)e→o(.)rde→r(), между тем пара скрытые состояния $(h_1, h_2, \ldots, h_z)$ $(h_1, h_2, \ldots, h_z)$ генерируются, мы и
можно получить выход слоя BiLSTM, комбинируя два скрытых состояния, как показано в уравнении (2).

$$h_i = \lceil h_i, h_{(i} \rceil \quad (2)$$

D. Слой внимания

Как было описано выше, некоторые DGA генерируют доменные имена путем объединения слов из словаря, такие семейства DGA всегда следуют фиксированному шаблону в комбинации слов. Распределение отдельных символов в доменных именах больше не является случайным и приблизительно нормальным, но поведение выбора слов из словаря для генерации доменного имени является случайным, что отражается на соединительной части слов. И многие модели обнаружения DGA, основанные на RNN, представляют последовательности слов только с помощью скрытого слоя в конечном узле. В данной работе скрытые состояния во всех позициях рассматриваются с различными весами внимания. Поскольку для обнаружения домена DGA фокусировка на некоторых определенных частях последовательностей будет эффективна для отсеивания нерелевантного шума DGA. Мы применяем механизм внимания для захвата релевантных признаков с выхода слоя BiLSTM, которые важны для предсказания и которые необходимо учитывать при чтении входной последовательности и накоплении ее до представления состоянии ячейки. Чтобы уловить взаимосвязь между текущим скрытым состоянием и всеми предыдущими скрытыми состояниями и использовать информацию из всех предыдущих скрытых состояний, w e_→ принимает_→общее внимание, чтобы уловить взаимосвязь между h_i и h_i , Каждое доменное имя состоит из нескольких слов или символов, которые имеют одинаковый вес. Многие слова несут шум или бесполезную информацию, поэтому мы обучили веса для каждого символа с помощью механизма внимания, чтобы сосредоточиться на ключевых характеристиках. Формулы для вычисления вектора веса внимания следующие:

$$\alpha = v^T \tanh\lceil \lfloor h_i, h_i \rfloor \rightarrow \rightarrow \rceil \quad (3)$$

$$\alpha = \text{softmax}\left(\lceil \alpha_1, \alpha_2, \ldots \alpha_{t(t-1)} \rfloor\right) \quad (4)$$

$[h_1, h_2, \ldots h_t]$ - это входная матрица, которая создается слоем BiLSTM. Затем вектор контекста может быть вычислен основе вектора весов внимания и скрытых состояний как Уравнение (5).

$$c_t = \sum_{i}^{t-1} \alpha_{ti} h_i \quad (5)$$

Скрытое состояние внимания $h(')$ вычисляется по уравнению (6), который основан на текущем скрытом состоянии h_t и векторе контекстаc_t , где W_c - весовая матрица слоя внимания. Весовой вектор может автоматически изучать особенности слов и записывать значимую информацию в домене. Признак домена может быть представлен путем умножения весового вектора. и r - размерность состояния внимания. Кроме того, в плотном слое используется отсев, чтобы избежать чрезмерной подгонки.

$$h' = \tanh(W_c [c_t; h_t]) \quad (6)$$

E. Выходной слой

В выходном слое есть два плотных слоя. Выход первого плотного слоя поступает во второй плотный слой с n скрытыми нейронами, где n - номер класса доменных имен. Затем мы подаем вектор внимания после отсева в функцию softmax для прогнозирования. w_s иb_s - параметры, которые необходимо выучить. В качестве активации выбрана функция softmax, результат расчета которой лежит в диапазоне от 0 до 1.

$$p = \text{softmax}\left(\left[w h'' + b\right]\right) \quad (7)$$

Из приведенной выше модели мы получаем вероятность p, которая будет обучаться непосредственно на основе характеристик, чтобы определить, является ли домен
имя является доброкачественным или DGA на задаче обнаружения, и к какому семейству доменов принадлежит доменное имя на задаче классификации.

IV. Эксперименты

Этот раздел начинается с описания необходимых деталей наборов данных доброкачественных доменных имен и доменных имен DGA, методов сравнения и схемы эксперимента, описанных в статье.

A. Описание набора данных

Набор данных маркированных доменных имен, используемый в данной работе, включает доброкачественные доменные имена и имена DGA. Доброкачественные доменные имена были взяты из Alexa [29]. Мы загрузили 500 000 лучших доменных имен в Alexa. Alexa ранжирует сайты на основе их популярности с точки зрения количества просмотров страниц и числа уникальных посетителей. Доменные имена DGA были собраны из двух взаимодополняющих источников. А именно: John Bambenek [30] и 360netlab [31]. Общее количество доменных имен DGA составляет 300 046, и они соответствуют 19 различным семействам вредоносных программ. В том числе Cryptolocker, locky, suppobox и т. д. Семейства DGA можно разделить на три категории: схема, основанная на арифметике, схема, основанная на списке слов, и схема, основанная на части списка слов. Первая схема обычно вычисляет последовательность, которая имеет прямое ASCII-представление, пригодное для доменного имени. Схема на основе списка слов состоит из конкатенации последовательности слов из одного или нескольких списков слов. Схема, основанная на частичном списке слов, означает, что алгоритм сочетает в себе схему, основанную на списке слов, и схему, основанную на арифметике, например banjori и beebone. Генерируемое доменное имя имеет вид "bnwgbyplay.com", где
первые шесть символов "bnwgby" основаны на арифметике, а последние четыре символа "play" - на списке слов. В наборе данных, 80% данных используется для обучения, а остальные 20% - для тестирования. Распределение данных показано в таблице I.

ТАБЛИЦА I. КРАТКОЕ ОПИСАНИЕ СОБРАННОГО НАБОРА ДАННЫХ

Тип домена	Схема	Образец
Алекса	/	500000
банджори	на основе частичного списка слов	25000
Бибон	на основе частичного списка слов	210
cryptolocker	арифметический	6000
dyre	арифметический	823
emotet	арифметический	20000
геймовер	арифметический	16615
locky	арифметический	8028
мацну	список слов	20694
мурофет	арифметический	26520
некур	арифметический	21843
Пост	арифметический	22000

pykspa_v1	арифметический	23293
qakbot	арифметический	26058
рамнит	арифметический	24500
rovnix	арифметический	25800
suppobox	список слов	10055
тинба	арифметический	19766
urlzone	арифметический	1845
летучие	на основе частичного списка слов	996

B. Экспериментальная установка

В наших экспериментах мы рассматривали TensorFlow в

в сочетании с Keras как программным фреймворком. Для увеличения скорости вычислений градиентного спуска в глубоком обучении

архитектуры, мы используем TensorFlow с поддержкой GPU в одном Nvidia Tesla P100. Постоянно корректируя и оптимизируя параметры в наших экспериментах, наиболее эффективные гиперпараметры были установлены следующим образом:

- Размерность вектора встраивания была установлена на 128 в слое встраивания.
- Нейронные модели обучались на обучающем множестве с размером партии 128.
- Количество узлов скрытого слоя в модели BiLSTM было установлено равным 128.
- Фильтры слоя CNN были установлены на 64.
- В качестве алгоритма оптимизации был использован RMSProp.
- Коэффициент отсева был установлен на уровне 0,2.

C. Сравнение с базовыми методами

Мы установили три базовых метода для сравнения с предложенной моделью ATT-CNN-LSTM. В эксперименте использовались следующие модели.

1) LR-модель

Сначала извлекаются признаки с помощью частоты терминов и обратной частоты документов (TF-IDF), затем мы используем логистическую регрессию (LR), которая является классической моделью машинного обучения, чтобы классифицировать доменные имена.

2) LSTM-модель

Модель LSTM, предложенная в [5], была принята в качестве сравнительной модели в нашей работе, которая использовала только слой LSTM для извлечения признаков и классификации доменных имен.

3) CNN-BiLSTM

CNN-BiLSTM - это модель, в которой механизм внимания не включен, а остальные слои и настройки параметров такие же, как в модели AB-BiLSTM.

4) Модель ATT-CNN-BiLSTM

ATT-CNN-BiLSTM - это модель, которую мы предложили для обнаружения доменных имен DGA в данной статье.

D. Дизайн эксперимента

Чтобы оценить эффективность моделей, мы использовали стратегию 10-кратной кросс-валидации на наборе данных. Сначала набор данных разбивается на 10 складок, затем девять складок обучаются, а оставшаяся используется для проверки. Этот процесс использует каждую складку для проверки один раз и повторяется десять раз. Наконец, все метрики по проверенным складкам усредняются, и получается более точная оценка производительности.

Для оценки классификации мы используем стандартные метрики accuracy, precision, recall и F1-score.

эффективность обнаружения вредоносных доменных имен. Формула метрики может быть определена как уравнения (8)-(11).

$$точность = \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

$$точност = \frac{TP}{TP+FP} \tag{9}$$

$$отзыв = \frac{TP}{TP+FN} \tag{10}$$

$$= \frac{2* точность* отзыв}{точность+ отзыв} \tag{11}$$

где TP - истинные положительные результаты, TN - истинные отрицательные результаты, FN - истинные отрицательные результаты. False Negatives и FP - False Positives соответственно. Мы определяем вредоносные доменные имена как положительные, а доброкачественные - как отрицательные. Поскольку классы в наборе данных несбалансированы, мы также использовали кривую операционной характеристики приемника (ROC) для оценки статистики площади под кривой (AUC).

V. Экспериментальные результаты

В этом разделе мы приводим результаты оценки эксперимента и анализируем результат каждой модели.

A. Задача обнаружения и классификации

- Задача обнаружения

На рис. 2 показаны ROC-кривые каждой модели. Значение AUC для ATT-CNN-BiLSTM составляет 0,9990, результаты показывают, что модель, предложенная нами в данной статье, показала лучшие результаты в распознавании доменных имен как доброкачественных или вредоносных по сравнению с моделями LR, LSTM и CNN-BiLSTM.

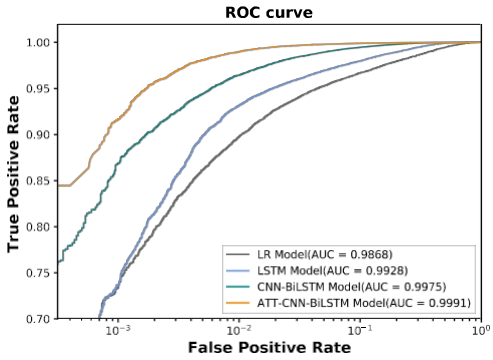

Рис. 2. ROC-кривая для обнаружения доменных имен DGA

Сравнение точности, прецизионности, запоминания и F1 score методов представлено на рис. 3. Модель, основанная на механизме внимания, превосходит три другие модели по точности, прецизионности, запоминанию и F1 score. Модель ATT-CNN-BiLSTM имеет лучший результат обнаружения по F1-баллу 0,9872, чем традиционная LR-модель, использующая статистические характеристики, что указывает на то, что новое доменное имя DGA уклоняется от традиционных статистических характеристик, и LR-модель не подходит для обнаружения таких доменных имен. Показатели recall и precision модели ATT-CNN- BiLSTM выше, чем у моделей LSTM и CNN-BiLSTM, показавших наилучшие результаты в традиционном методе, соответственно, что говорит о том, что механизм внимания позволяет лучше обнаруживать доменные имена DGA и повышает общий эффект обнаружения.

По показателям точности и прецизионности модель ATT-CNN-BiLSTM в среднем на 3 % превосходит модель на основе признаков с LR, а показатель recall почти на 5 % выше, чем у модели LR. С одной стороны, модель на основе признаков опирается на другие инструменты обработки естественного языка, накопленные ошибки оказывают большое влияние на производительность, и эффективное извлечение признаков оказывается недостаточным. С другой стороны, внешняя семантика, такая как частота слов, оказывает ограниченное влияние на задачу обнаружения доменного имени DGA, особенно на основе списка слов доменного DGA, в то время как нейронные сети могут кодировать семантическую информацию в высокоразмерное скрытое пространство признаков и извлекать больше признаков.

По сравнению с моделью LSTM, значение F1 модели ATT-CNN-BiLSTM на 2% выше. Модель CNN-BiLSTM

Модель также имеет более высокие показатели точности и F1, чем модель LSTM. Это объясняется тем, что модели CNN-BiLSTM и ATT- CNN-BiLSTM сочетают в себе преимущества CNN и BiLSTM, которые позволяют улавливать локальные особенности и информацию о зависимости на больших расстояниях в последовательностях доменных имен.

По сравнению с моделью CNN-BiLSTM, значение F1 модели ATT-CNN-BiLSTM на 1 процент выше. Так как обе модели отличаются только слоем внимания, который может придавать относительно большой вес таким важным признакам, как совместная часть нескольких слов доменных имен. Это говорит о том, механизм внимания эффективен для задачи обнаружения доменных имен DGA и может улучшить общий эффект обнаружения.

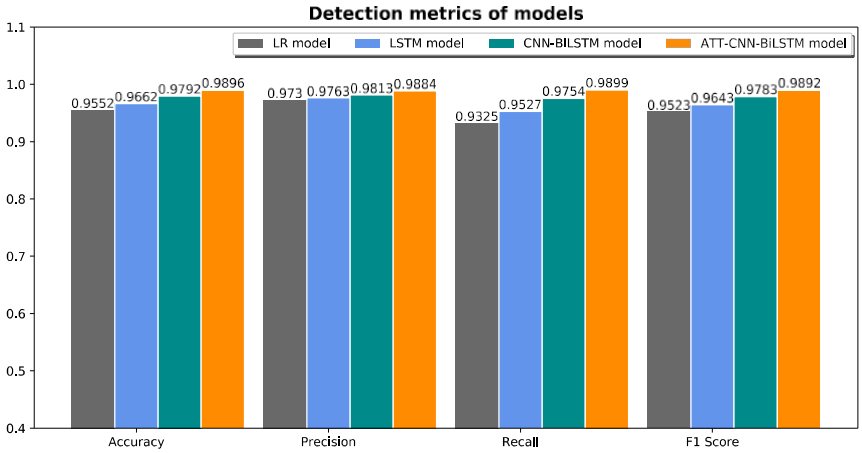

Рис. 3. Результаты оценки обнаружения доменных имен DGA

- Задача классификации

На этом же наборе данных был проведен эксперимент по многоклассовой классификации. Для задачи классификации в данной работе мы использовали такие метрики, как точность, отзыв и F1. Поскольку показатели precision, recall и F1 являются специфическими для каждого класса показателями, их необходимо усреднить, чтобы получить общий результат. Микросредние суммируют индивидуальные показатели TP, FP и FN для всех классов, а макросредние берут среднее значение индивидуальных показателей для всех классов. Другими словами, макро-средние рассматривают все семейства DGA одинаково, в то время как микро-средние отдают предпочтение классу с большим количеством образцов. Как показано в таблице II, модель ATT-CNN-BiLSTM показала лучшие показатели как на микро-, так и на макро-средних (микро-среднее значение F1-score составило 0,89, а макро-среднее значение F1-score - 0,81).

B. Анализ и обсуждения

Мы можем уточнить анализ результатов эксперимента, рассмотрев результаты обнаружения и классификации для каждого семейства вредоносных программ. По сравнению с LR- и LSTM-моделями, F1 score модели CNN-BiLSTM достиг 0,9887, а средний макрос - 0,72, что доказывает, что с помощью слоя CNN мы можем эффективно извлекать локальные признаки. Модель ATT-CNN-BiLSTM получила лучшие показатели точности, прецизионности, запоминания и F1 среди всех моделей, что доказывает, что хорошая способность извлекать общие закономерности в домене названия. В схемах DGA на основе арифметики, например, для вредоносных программ dyre и urlzone, результат F1 составил более 92 % при малом размере менее 2000. В DGA на основе частичных списков, поскольку в них также применялась случайная стратегия, модель ATT-CNN-BiLSTM показала лучшую эффективность обнаружения таких схем, как banjori и volatile malware, которая достигла 0,99 балла F1.

Что касается семейств DGA на основе списков слов, то распределение отдельных символов в доменных именах перестает быть случайным и становится приблизительно нормальным. Поведение выбора слов из словаря для генерации доменного имени является случайным, отражается на соединительной части слов. Предложенная модель ATT- CNN-BiLSTM может отразить такую случайность с помощью встроенного механизма внимания. В случае с suppobox оценка F1 составила 0,91 по сравнению с 0,46 для биграммной модели и
0,33 для модели CNN-BiLSTM, а для вредоносной программы matsnu мы получили оценку F1 0,81, что было лучше, чем у трех других моделей. После интеграции механизма внимания в нейронную модель способность к обучению шаблону сочетания списков слов, используемых DGA, повысилась до определенного уровня.

ТАБЛИЦА II. ТОЧНОСТЬ, ОТЗЫВ И ОЦЕНКА F1 ДЛЯ КЛАССИФИКАЦИИ

Тип домена	LR			LSTM			CNN-BiLSTM			ATT-CNN-BiLSTM			Поддержка
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	
доброкачественный	0.94	0.99	0.96	0.95	0.98	0.97	0.94	0.99	0.96	0.99	0.98	0.99	99999

банджори	0.97	0.98	0.98	0.99	1.00	0.99	1.00	0.99	0.99	1.00	1.00	1.00	5012
Бибон	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.98	1.00	0.99	42
Cryptolocker	0.00	0.00	0.00	0.26	0.24	0.25	0.21	0.02	0.03	0.28	0.27	0.28	1200
dyre	0.95	0.99	0.97	0.99	1.00	0.99	0.99	0.98	0.99	0.99	1.00	1.00	167
emotet	0.65	0.81	0.72	0.66	1.00	0.79	0.66	1.00	0.79	0.65	0.81	0.72	3985
геймовер	0.34	0.11	0.16	0.90	0.00	0.01	0.53	0.01	0.02	0.35	0.29	0.31	3323
locky	0.26	0.03	0.06	0.67	0.00	0.00	0.14	0.00	0.00	0.43	0.26	0.32	1610
мацну	0.27	0.26	0.27	0.69	0.86	0.77	0.76	0.78	0.76	0.75	0.87	0.81	4133
мурофет	0.96	0.99	0.98	0.95	1.00	0.97	0.97	0.99	0.98	0.98	1.00	0.99	5304
Пост	0.59	0.73	0.66	0.78	0.73	0.76	0.68	0.87	0.76	0.83	0.76	0.79	4373
некур	0.24	0.22	0.23	0.52	0.32	0.40	0.56	0.19	0.29	0.61	0.66	0.63	4400
pykspa_v1	0.9	0.88	0.89	0.92	0.96	0.94	0.98	0.93	0.96	0.98	0.98	0.98	4662
qakbot	0.64	0.41	0.50	0.71	0.64	0.67	0.62	0.72	0.67	0.75	0.65	0.70	5217
рамнит	0.38	0.49	0.43	0.60	0.71	0.65	0.62	0.71	0.66	0.60	0.75	0.67	4900
rovnix	0.85	0.90	0.87	1.00	0.99	1.00	1.00	0.99	0.99	1.00	0.99	1.00	5160
suppobox	0.75	0.33	0.46	0.87	0.20	0.33	0.99	0.09	0.17	0.87	0.94	0.91	2011
тинба	0.59	0.80	0.68	0.67	0.96	0.79	0.77	0.98	0.86	0.91	0.98	0.94	3955
urlzone	0.86	0.75	0.80	0.96	0.91	0.93	0.98	0.91	0.94	0.98	0.92	0.95	369
летучие	0.98	1.00	0.99	0.97	0.69	0.81	1.00	0.36	0.53	0.99	0.99	0.99	185
микроавт.	0.80	0.80	0.80	0.86	0.86	0.86	0.87	0.87	0.87	0.89	0.89	**0.89**	120007
среднее макросостояние	0.66	0.63	0.63	0.69	0.65	0.66	0.72	0.73	0.72	0.81	0.81	**0.81**	120007

Но есть и исключения: в таблице II показано, что вредоносные программы Cryptolocker, gameover и locky не были правильно классифицированы. В задаче обнаружения модель ATT-CNN-BiLSTM может отличить доменные имена, сгенерированные Cryptolocker gameover и locky, от обычных, но в задаче классификации все модели имеют низкую производительность для этих классов вредоносных программ, хотя ATT-CNN- BiLSTM получает самую высокую точность классификации. Причина в том, что эти вредоносные программы используют серию умножений, делений и модулей на основе одного семени для генерации доменных имен DGA. Распределение униграмм Cryptolocker и Locky показано на рис. 4, которое выглядит совершенно одинаково, а одинаковое распределение символов в доменных именах приводит к ошибочной классификации по двум схожим классам. Большая часть неверно классифицированных DGA Cryptolocker была отнесена Locky. Между тем, как показано на рис. 5, вредоносные программы Gameover и Post имеют схожую ситуацию, что приводит к плохому эффекту классификации для вредоносных программ Gameover.

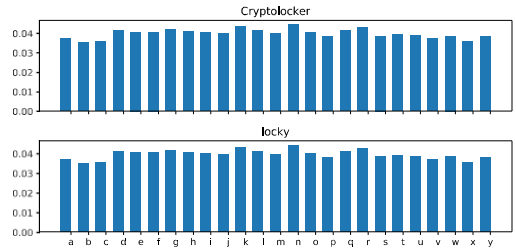

Рис. 4. Распределения униграмм для Cryptolocker и locky

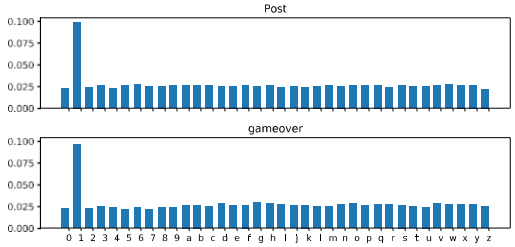

Рис. 5. Распределения униграмм для gameover и Post

VI. Выводы

В этой статье мы рассмотрели проблему обнаружения DGA-доменов, поскольку многие вредоносные программы имитируют шаблон обычных доменных имен, объединяя псевдослучайно выбранные слова из английского словаря для создания доменных имен и достижения эффекта сокрытия и противостояния. Мы предложили эффективную модель ATT-CNN-BiLSTM, которая объединила механизм внимания и глубокую нейронную сеть для обнаружения и классификации доменных имен. По сравнению со статистическими характеристиками и наиболее широко используемой моделью LSTM, модель ATT-CNN-BiLSTM может повысить точность задачи обнаружения без ручного извлечения признаков и достигла лучшей производительности на основе арифметики, части слов и списка слов DGA, таких как семейства matsnu и suppobox. в задаче классификации. Результаты экспериментов показывают эффективность модели, с помощью которой мы получаем F1 оценку в среднем 98,92% для обнаружения и F1 оценку в среднем 81% для классификации доменных имен. Полученные результаты доказывают, что механизм внимания может повысить достоверность идентификации доменных имен DGA на основе списков слов.

Будущая работа будет направлена на интеграцию этой нейронной модели в более крупную архитектуру машинного обучения для обнаружения киберугроз в реальных данных трафика. Мы также намерены добавить побочную информацию для повышения точности классификации и разграничения классов DGA, генерирующих похожие доменные имена.

Благодарность

Работа выполнена при поддержке Стратегической программы приоритетных исследований Китайской академии наук (грант № XDC02030200) и Национального фонда естественных наук Китая (грант № 61572481).

Ссылки

[1] Эндрюс, Марк. "Отрицательное кэширование DNS-запросов (DNS NCACHE)". 1998.

[2] Y. Жауниярович, И. Халил, Т. Ю. и М. Дасье, "Обзор обнаружения вредоносных доменов с помощью анализа данных DNS", ACM Computing Surveys, vol. 51, no. 4, 67:1-67:36, 2018.

[3] Дж. Вудбридж, Х.С. Андерсон, А. Ахуджа и Д. Грант, "Предсказание алгоритмов генерации доменов с помощью сетей долговременной кратковременной памяти", arXiv:1611.00791 [cs.CR], 2016.

[4] Хьеу Мак, Дук Тран, Ван Тонг, Линь Гианг Нгуен и Хай Ань Тран. "Обнаружение ботнета DGA с помощью методов контролируемого обучения".

В материалах Восьмого международного симпозиума по информационным и коммуникационным технологиям, SoICT 2017, стр. 211-218.

[5] J. Сакс, К. Берлин, "eXpose: конволюционная нейронная сеть с вкраплениями на уровне символов для обнаружения вредоносных URL-адресов, путей к файлам и ключей реестра", 2017.

[6] D. Плохманн, К. Якдан, М. Клатт, Дж. Бадер, Э. Герхардс-Падилья, "Всестороннее исследование измерений вредоносных программ, генерирующих домены", Труды 25-го симпозиума по безопасности USENIX (SECURITY), 2016.

[7] Станислав Скуратович. "Технический отчет по Мацну". http://aiweb.techfak.uni-bielefeld.de/content/bworld-robot-control- software/, 2015.

[8] J. Геффнер, "Сквозной анализ алгоритма генерации доменов семейство вредоносных программ". Black Hat USA 2013.

[9] S. Yadav, A. K. K. Reddy, A. N. Reddy, S. Ranjan, "Detecting algorithmically generated domain-flux attacks with DNS traffic analysis", Networking IEEE/ACM Transactions on, vol. 20, no. 5, pp. 1663-1677, 2012.

[10] *М. Антонакокис, Р. Пердиши, Д. Дагон, В. Ли и Н. Фемстер, "Создание динамической системы репутации для DNS", симпозиум по безопасности USENIX, 2010. стр. 273-290.*

[11] X. Zang, J. Gong, and X. Hu, "Detecting malicious domain name based on AGD," Journal on Communications, vol. 39, no. 7, pp. 15-25, 2018.

[12] Y. Чжан, Ю. Чжан и Ж. Сяо, "Обнаружение вредоносных доменных имен на основе DGA", в материалах Международной стандартной конференции по надежным вычислениям и сервисам, Пекин, Китай, ноябрь 2013 г. стр. 130-137.

[13] S. Ядав, А. К. Редди, А. Редди и С. Ранджан, "Обнаружение алгоритмически сгенерированных вредоносных доменных имен", конференция по измерениям в Интернете (IMC), 2010.

[14] L. Бильге, С. Сен, Д. Бальзаротти, Э. Кирда, К. Крюгель, "Exposure: Пассивная служба анализа DNS для обнаружения и сообщения о вредоносных доменах", ACM Trans. Inf. Syst. Secur., vol. 16, no. 4, Apr. 2014.

[15] J. Рагурам, Д. Дж. Миллер, Г. Кесидис, "Неконтролируемое обнаружение аномалий с низкой задержкой алгоритмически сгенерированных доменных имен с помощью генеративного вероятностного моделирования", J. Adv. Res., vol. 5, no. 4, pp. 423- 433, 2014.

[16] M. Гриль, И. Николаев, В. Валерос и М. Рехак, "Обнаружение вредоносных программ DGA с помощью NetFlow", в Proc. IFIP/IEEE Int. Symp. Integr. Netw. Manag. (IM), Ottawa, ON, Canada, 2015, pp. 1304-1309.

[17] Wang W, Shirley K: "Breaking bad: обнаружение вредоносных доменов с помощью сегментации слов". 2015.

[18] H. С. Андерсон, Дж. Вудбридж, Б. Филар, "DeepDGA: Adversarially- tuned domain generation and detection" in Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security, ACM, 2016.

[19] T. Шибахара и др., "Эффективный динамический анализ вредоносного ПО на основе поведения сети с использованием глубокого обучения", Proc. IEEE Global Commun. Conf. (GLOBECOM), pp. 7-18, Feb. 2017.

[20] L. Янг, Г. Лю, Ж. Чжай, Ю. Дай, З. Янь, Ю. Цзоу и В. Хуанг, "Новый метод обнаружения для DGA на основе слов", в Трудах 4-й Международной конференции по облачным вычислениям и безопасности, том 2, 2018. стр. 472-483.

[21] M. Перейра, С. Коулман, Б. Ю, М. Де Кок и А. Насименто, "Извлечение словаря и обнаружение алгоритмически сгенерированных доменных имен в пассивном DNS-трафике", в материалах 21-го Международного симпозиума по исследованиям в области атак, вторжений и защиты, 2018, стр. 295-314.

[22] Koh J J, Rhodes B. "Inline Detection of Domain Generation Algorithms with Context-Sensitive Word Embeddings." in Proceedings of IEEE International Conference on Big Data. "2018. 2966-2971.

[23] R. R. Curtin, A. B. Gardner, S. Grzonkowski, A. Kleymenov, A. Mosquera, "Detecting DGA domains with recurrent neural networks and side information", 2018.

[24] D. Бахданау, К. Чо, Й. Бенгио, "Нейромашинный перевод путем совместного обучения выравниванию и переводу", Международная конференция по изучению представлений, 2015.

[25] T. Луонг, Х. Фам, К. Д. Мэннинг, "Эффективные подходы к нейронному машинному переводу на основе внимания", Proc. Эмпирические методы Natural Lang. Process., pp. 1412-1421, 2015.

[26] Z. Янг, Д. Янг, К. Дайер, Х. Хе, А. Смола и Э. Хови, "Иерархические сети внимания для классификации документов", в NAACL,2017, стр. 1480-1489.

[27] M. D. Rijke, M. D. Rijke, M. D. Rijke, M. D. Rijke, M. D. Rijke, M. D. Rijke, M. D. Rijke, "Leveraging Contextual Sentence Relations for Extractive Summarization Using a Neural Attention Model", Международная конференция ACM SIGIR Conference on Research and Development in Information Retrieva1, pp. 95-104, 2017.

[28] Zhou X, Wan X, Xiao J. "Attention-based LSTM network for cross- lingual sentiment classification." In Proceedings of the 2016 conference on empirical methods in natural language processing. Austin: Association for Computational Linguistics;2016, pp. 247-56.

[29] Alexa. Информационная веб-компания. http://www.alexa.com/, 2007.

[30] Bambenek Consulting. http://osint.bambenekconsulting.com/feeds, 2019.

[31] 360netlab, https://data.netlab.360.com/dga/, 2019.