Инфокоммуникационные технологии и системы

УДК 004.056.5 **DOI: 10.14529/ctcr200206**

МОДЕЛЬ ОБНАРУЖЕНИЯ ФИШИНГОВЫХ АТАК НА ОСНОВЕ ГИБРИДНОГО ПОДХОДА ДЛЯ ЗАЩИТЫ АВТОМАТИЗИРОВАННЫХ СИСТЕМ УПРАВЛЕНИЯ ПРОИЗВОДСТВОМ

Е.А. Митюков[1], А.В. Затонский[2]

[1] Пермский национальный исследовательский политехнический университет, г. Пермь, Россия,
[2] Пермский национальный исследовательский политехнический университет, Березниковский филиал, г. Березники, Россия

Введение. В связи с ежегодным развитием фишинговых техник злоумышленников, которые направленны на автоматизированные системы управления производством с целью компрометации конфиденциальной информации, актуальной задачей является разработка новых методов определения фишинговых атак, направленных на промышленный сектор. **Цель исследования:** разработка метода защиты от фишинговых атак на пользователей и сервисы автоматизированных систем управления производством. **Материалы и методы.** Для анализа предметной области проанализированы возможные источники литературы. Основываясь на собранной информации из предыдущих исследований, продолжена работа над улучшением архитектуры системы защиты от фишинга. В архитектуру системы добавлены восемь эвристик, направленных на улучшение точности детектирования фишинговых URL (Uniform Resource Locator). Ряд эвристик направлен на семантическую проверку URL в части использования специальных символов, точек, слешей, порта, протокола URL и в том числе длины самого URL. Другие же проверяют валидность SSL/TLS (Secure Sockets Layer/Transport Layer Security) сертификата, ищут фишинговые ключевые слова в URL и сравнивают страну хостинг-провайдера со страной домена верхнего уровня. **Результаты**. Проведены практические исследования новой архитектуры с различными комбинациями эвристик. Приводятся количественные данные, показывающие улучшение ключевых показателей детектирования фишинговых ресурсов системой, которые, в свою очередь, помогают офицеру безопасности принимать решение о фишинговости или легитимности URL. **Заключение**. Представленная система показывает следующие показатели: TPR (True Positive Rate) – 97,85 % и FPR (False positive Rate) – 2,09 %. Также улучшена точность метода до 98,16 %.

Ключевые слова: кибербезопасность, кибератака, автоматизированные системы управления, анти-фишинг, свойства URL.

Введение

На сегодняшний день фишинг – одна из серьезных угроз [1], которая потенциально опасна для каждого пользователя сети Интернет. Фишинг (англ. *phishing*, от *fishing* – рыбная ловля, выуживание и password – пароль) – вид интернет-мошенничества, цель которого – получить идентификационные данные пользователей. Сюда относятся кражи паролей, номеров кредитных карт, банковских счетов и другой конфиденциальной информации [2]. Как правило, в типовой фишинг-атаке злоумышленники пытаются обмануть пользователей сети Интернет, отправляя им специальные электронные письма, содержащие в себе ссылку на нелегитимный веб-сайт. Этот нелегитимный веб-сайт представляется полной копией легитимного веб-сайта, с одной лишь оговоркой: все логины и пароли которые вводит пользователь при попытке авторизации, злоумышленник сохраняет для дальнейшего использования в личных целях [3]. Очевидно, что одним из основных

методов распространения фишинговых веб-ссылок является электронная почта. Но поскольку атаки развиваются, злоумышленники используют все более изощренные подходы для распространения и сокрытия фишинговых веб-сайтов. Увеличение сложности определения фишинговых атак затрудняет возможность обеспечения должного уровня защищенности пользователей [4, 5].

Согласно ежегодному отчету «Лаборатории Касперского» о спаме и фишинге количество обнаруженных фишинговых атак за 2018 г. в два раза больше, чем за 2017 г. Интересный факт, что одним из трендов 2018 г. стали фишинговые атаки, направленные на веб-сайты, имитирующие криптовалютные кошельки, платформы, биржи, для получения логинов и паролей потенциальных жертв. Также эксперты лаборатории поделились информацией о том, что злоумышленники следят за периодами распродаж на крупных торговых платформах и готовятся к таким событиям заранее [6]. На момент публикации и старта рекламных кампаний о распродаже злоумышленники начинают активно распространять подготовленные фишинговые веб-сайты. Статистика за 2018 г. по переходам на веб-сайты фишинговых торговых площадок представлена на рис. 1.

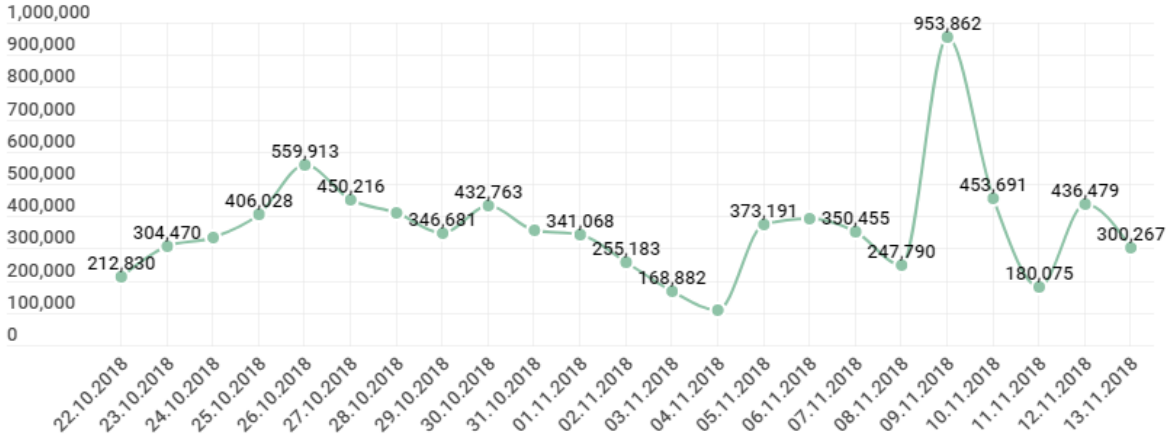

Рис. 1. Статистика за 2018 г. по переходам на фишинговые торговые площадки
Fig. 1. Statistics for 2018 on transitions to phishing trading sites

Фишинговые атаки не обошли и корпоративный сектор в 2018 г. Многие сотрудники крупных компаний получали письма, содержимое которых было представлено, как переписка с партнерами, либо письма с нетипичными расширениями файлов во вложении, дабы обойти имеющиеся средства защиты. В целом в письмах сохранялась тенденция таргетированных фишинговых атак с максимальной проработкой писем под цель [6]. Фишинговые атаки на крупные корпорации могут привести не только к финансовым потерям, но и к более серьезным последствиям. Например, производитель Radiflow в области промышленной кибербезопасности опубликовал информацию о том, что майнер monero был обнаружен на 5 серверах одной из водоснабжающих компаний в Европе [7]. АСУП (автоматизированная система управления производством) компании была опубликована в сети Интернет для передачи информации в систему мониторинга. В случае с майнером последствия не так страшны, но если бы вредонос был более серьезным, результатом атаки могла быть остановка внутренних сервисов компании или потенциально большие убытки.

Несмотря на то, что на сегодняшний день существует множество методов борьбы с фишинговыми атаками, злоумышленники находят новые возможности и техники для реализации фишинговых атак. Цель же нашей работы – это разработка метода, который бы защищал АСУП от фишинговых атак. Для достижения цели были сформированы следующие задачи.

1. Продолжить развитие общей архитектуры системы, которая представлена в предыдущих работах [8].

2. Провести обзор литературы и выбрать существующие методы и алгоритмы, применимые для защиты АСУП.

3. Доработать существующие либо разработать новые методы в соответствии с особенностями АСУП.

4. Провести практическое исследование.

В одной из предыдущих работ мы предложили гибридный подход, адаптированный для АСУП, с целью борьбы с фишингом. Подход состоял из четырех основных модулей:

первый модуль: позволяет получить URL, который запрошен пользователем;

второй модуль: проверяет наличие URL в персональном белом списке пользователя;

третий модуль: проверяет наличие форм авторизации на странице;

четвертый модуль: включает в себя ряд эвристик, уменьшающих частоту ложных срабатываний системы.

Остальная часть статьи организована следующим образом: в разделе 1 представлен литературный обзор. Раздел 2 отражает общую архитектуру системы. Раздел 3 содержит эвристики. В разделе 4 представлены метрики оценки системы. В разделе 5 обсуждаются экспериментальные результаты. В конце статьи представлены выводы.

1. Литературный обзор

Сегодня обнаружение фишинговых атак и противостояние им – одна из наиболее исследуемых проблем. Существует немалое количество работ, связанных с обнаружением фишинговых атак, в этом разделе проведен краткий обзор этих работ.

Одни из исследователей разработали антифишинговый тулбар под названием «Phishark». Ими были проанализированы и изучены характеристики фишинговых атак [9]. В состав тулбара были включены 20 эвристик для обнаружения фишинговых веб-страниц. Каждая из эвристик проверялась на эффективность на наборах данных. В нашем исследовании мы используем эвристику «Проверка подлинности кода страны», входящую в состав тулбара с некоторыми изменениями.

В другом исследовании авторы рассмотрели особенности URL-адреса: дополнительные символы и ключевые слова в составе URL; отсутствие домена верхнего уровня; безопасность транспортного уровня и др. [10]. Также авторы дополнительно предложили правило для определения фишинга на основе ассоциаций. Кроме того, мы считаем, что двойные слеши (косая черта) и точки в содержимом URL – одни из ключевых факторов, на которые стоит обратить внимание при определении легитимности URL.

Модель «PhishDMA» была разработана авторами для людей с нарушением зрения [11]. Это редкая работа, которая направлена на увеличение осведомленности людей с нарушением зрения в вопросах определения фишинговых веб-сайтов. Авторы использовали каскад фильтров, представленный в виде модели, и различные функции для определения фишинга. Функция «определения оценки доступности» из модели «PhishDMA» была нами доработана и использована в нашем исследовании. Интерактивная обучающая игра «Anti-phishing Phil» со схожей целью – повышение осведомленности пользователей в вопросах определения фишинговых ресурсов – была разработана группой исследователей [12]. Каждая игровая сессия длится 10 минут, за которые «игрок» обучается выявлять мошеннические и вредоносные URL-адреса; если «игрок» ошибается, игра предоставляет дополнительную обучающую информацию. В игре есть система управления обучением (LMS), которая позволяет повышать уровень знаний «игрока», основываясь на том, как быстро он обучается.

В одном из предыдущих исследований мы предложили эвристику «Сопоставление контента с доменным именем». Основная идея эвристики была направлена на поиск ключевых слов на исследуемой веб-странице и сопоставление их с доменным именем. В этом исследовании мы добавили дополнительную эвристику, основанную на этом подходе, но направленную на поиск фишинговых ключевых слов на веб-странице.

2. Архитектура системы

Мы разработали следующую архитектуру системы [13] (рис. 2). Основываясь на архитектуре из одной нашей предыдущей работы [8], мы добавили восемь новых эвристик для улучшения точности определения фишинговых ресурсов и уменьшения ложных срабатываний. Цель системы – обеспечение безопасности пользователей АСУП в части проверки посещаемых URL на наличие фишинговых признаков, при посещении ресурсов сети Интернет. На основе архитектуры предлагается следующий алгоритм работы системы.

1. Получить URL, который хочет посетить пользователь в текущий момент.

2. Передать URL серверу фильтрации.

3. Проверить URL по предварительному модулю «белый список». Если URL не обнаружен в белом списке, то URL отправляется дальше для проверки, в противном случае проверки прекращаются.

4. Проверить URL по предварительному модулю «наличие формы логина». Если обнаружена форма логина, то URL отправляется дальше для проверки, в противном случае проверки прекращаются.

5. Проверить URL по всем эвристикам.

6. На основе проверки URL формируется результат проверки, который направляется офицеру безопасности.

7. Офицер безопасности принимает решение: если URL фишинговый, он блокируется офицером безопасности при помощи механизмов DPI (deep packet inspection) решения, в противном случае изменения не производятся.

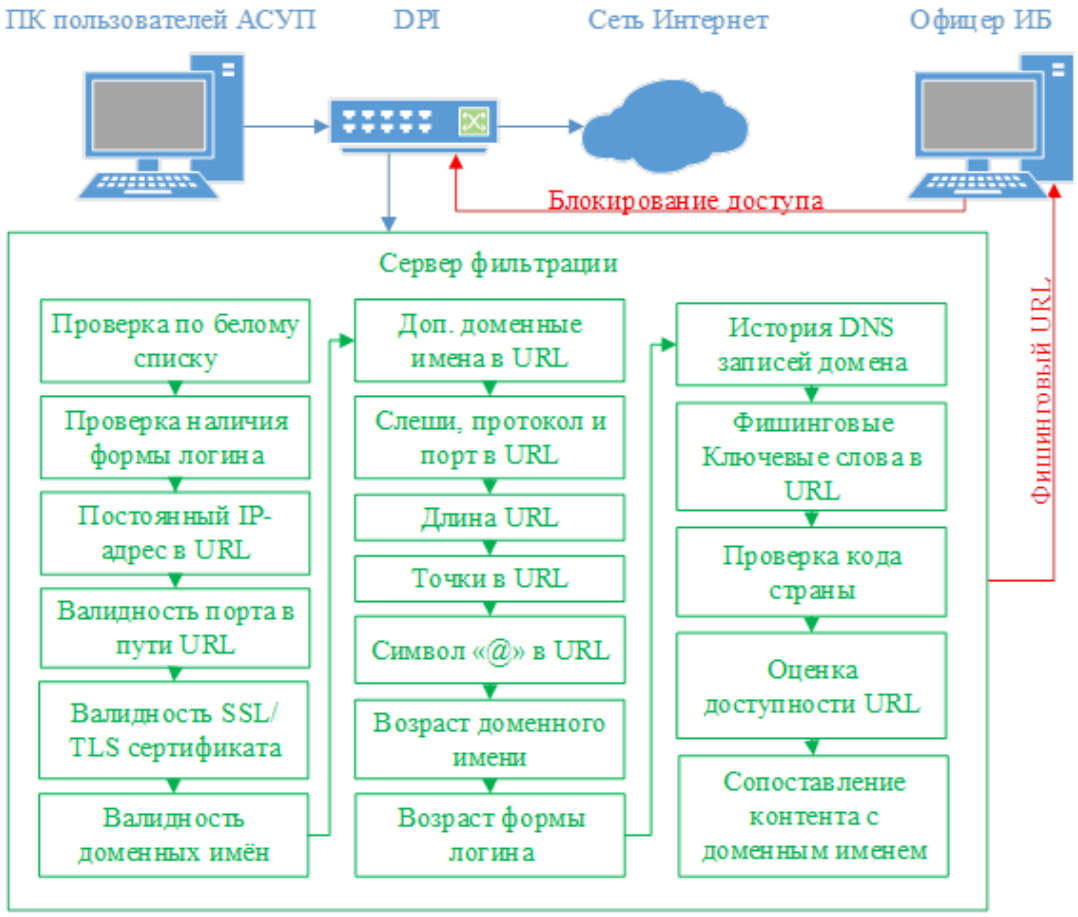

Рис. 2. Общая архитектура системы
Fig. 2. General architecture of a system

Ранее мы не использовали автоматическую блокировку фишинговых URL ввиду особенностей производств. Именно по этой причине предлагаемая система работает в режиме детектирования. Дальнейшим развитием системы будет реализация дополнительного функционала для работы в превентивном режиме с учетом возможных особенностей производств. Это позволит уменьшить скорость реакции блокировки фишинговых URL при их обнаружении.

3. Эвристики

Эвристика 1 – проверка кода страны. Цель этой эвристики – сверить для каждого ресурса код страны домена и код страны хостинг-провайдера, где размещен исследуемый ресурс. Алгоритм работы эвристики: извлечь верхний уровень домена из URL-адреса; разрезолвить доменное имя в

IP-адрес, сделать запрос в соответствующий региональный интернет-регистратор для получения сведений о стране, которой выдан этот IP-адрес. Сегодня существует 5 региональных интернет-регистраторов:

- для Северной Америки – это *American Registry for Internet Numbers*;
- для Европы, Ближнего Востока и Центральной Азии – это *RIPE Network Coordination Centre*;
- для Азии и Тихоокеанского региона – это *Asia-Pacific Network Information Centre*;
- для Латинской Америки и Карибского региона – это *Latin American and Caribbean Network Information Centre*;
- для Африки и региона Индийского океана это *African Network Information Centre*.

Если код страны домена отличается от кода страны хостинга, где размещен ресурс, то ресурс считается подозрительным, в противном случае – ресурс легитимный. При практическом исследовании мы выяснили, что сведения региональных интернет-регистраторов не всегда точны. Поэтому мы реализовали дополнительную возможность использовать запросы к коммерческим и бесплатным базам данных геолокаций.

Эвристика 2 – фишинговые ключевые слова в URL. Как показала практика, фишинговые вебсайты часто содержат в своем контенте и URL-адресе фишинговые ключевые слова. Мы сформировали список фишинговых ключевых слов из публичных источников [14] и добавили дополнительно свои. Соответственно, мы проверяем наличие в контенте и URL-адресе наличие этих слов. Если слова обнаружены, URL считается фишинговым, в противном случае – легитимным.

Эвристика 3 – валидность SSL/TLS-сертификата. Для того чтобы защитить трафик любого веб-сайта, будь то фишинговый или легитимный, используются два метода.

1. Первый метод – использование сертификата SSL/TLS, выданного общедоверенным центром сертификации. В данном случае ресурсу будут доверять все браузеры и операционные системы, в которых добавлен в доверенные корневые центры сертификации тот центр, который выдал данный SSL/TLS-сертификат.

2. Второй метод – использование самоподписанного SSL/TLS-сертификата. Здесь же доверия к ресурсу не будет, если принудительно не разрешить переход по ссылке внутри браузера пользователя.

Оба метода применяются злоумышленниками для того, чтобы шифровать трафик между фишинговым ресурсом и пользовательским ПК. Для противодействия таким атакам мы предварительно проверяем протокол URL-адреса, если это 443 порт – https протокол, тогда URL проверяется данной эвристикой. Мы проверяем всю цепочку в пути сертификата: корневой (root), промежуточные (intermediate), основной сертификаты. Алгоритм работы эвристики:

1. Проверяем протокол страницы и ее контент; если обнаружен порт 443, начинаем проверять URL.

2. Выгружаем цепочку из пути сертификата.

3. Проверяем все сертификаты в цепочке на валидность; если в сертификат выпущен не доверенным центром сертификации, ресурс считается подозрительным.

4. Проверяем поля DN (Distinguished Name), SAN (Subject Alternative Name) и Subject-сертификата на наличие домена из исследуемого URL-адреса в этом поле. Если домен отсутствует, значит сертификат выпущен на другое/другие доменные имена, следовательно, домен подозрительный.

Эвристика 4 – длина URL. Подсчет количества символов в URL-адресе – важная характеристика для детектирования фишинговых ресурсов. Злоумышленники используют длинные URL-адреса для того, чтобы скрыть фишинговую часть адреса в адресной строке. Соответственно, видимая часть адресной строки содержит легитимный URL-адрес, чтобы обмануть пользователя, что он взаимодействует якобы с легитимным ресурсом. Если длина URL-адреса больше допустимого значения 60 символов, то ресурс считается подозрительным, в противном случае – ресурс легитимный.

Эвристика 5 – символ «@» в URL. Цель этой эвристики – поиск символа «@» в адресной строке. В строке вида root@domain.com: первая часть – «root» до символа «@» является логином, а вторая – «domain.com» после символа «@» – доменное имя ресурса. Как правило, сегодня подобный синтаксис практически не используется. Следовательно, если в адресной строке обнаружен символ «@», URL-адрес считается подозрительным, в противном случае – легитимным.

Дополнительно мы разработали механизм очистки запрошенного URL-адреса, отбрасывая левую часть и символ «@», оставляя только доменное имя, тем самым минимизируя риски передачи конфиденциальной информации.

Эвристика 6 – точки в URL. Основная цель этой эвристики – подсчет количества точек в доменном имени URL. Мы провели ряд экспериментов на основе фишинговых URL из публичных ресурсов. Результат показал, что большинство фишинговых ресурсов содержат пять и более точек в доменном имени. Этот результат совпадает с тем, что получили авторы другого исследования [15]. В результате экспериментов в этой эвристике пороговое значение количества точек стало пять. Исходя из этого, эвристика считает количество точек в доменном имени URL: если количество больше порогового значения, URL считается подозрительным, в противном случае – легитимным.

Эвристика 7 – слеши, протокол и порт в URL. В этой эвристике мы проверяем наличие двух слешей после протокольной части, более одной протокольной части и более одного порта в URL-адресе. При исследовании мы обнаружили, что злоумышленники часто используют двойные слеши для сокрытия фишинговой части URL. Например, для URL адреса: https://click.sender.yandex.ru/l/7885/8297/2/L/RERVRklFREU4TnhjckdnWmNKaEExTTE0aEFrWTRLelFQSGtzYlNVWnpXMXg0UWdjSlJnPT06MzUxNjow/*https://phi.ru/app/?from=email_pushapp фишинговой частью является https://phi.ru/app/?from=email_pushapp, где есть и второй протокол и двойные слеши. Алгоритм этой эвристики представляется следующим образом.

1. Проверяется URL на наличие более двух двойных слешей.
2. Проверяется URL на наличие более одной протокольной части.
3. Проверяется URL на наличие более одного порта.
4. Если хотя бы одна из проверок показала положительный результат, URL считается подозрительным, в противном случае – легитимным.

Эвристика 8 – оценка доступности URL. Подразделение WAI (Web Accessibility Initiative) группы W3C (world wide web consortsium) уже с 1999 г. разрабатывает стандарты доступности веб-контента, которые называются WCAG (Web Content Accessibility Guidelines). Основная цель стандартов – обеспечить доступность содержимого сети Интернет для всех категорий интернет-пользователей. Поскольку до сегодняшнего момента было выпущено 3 релиза данных стандартов – WCAG 1.0 была выпущена 05.05.1995; WCGA 2.0 была выпущена 11.12.2008; WCGA 2.1 был выпущен 05.06.2018, – можно смело предположить, что можно брать за основу данные стандарты, поскольку WAI периодически вносят изменения в них и следят за их актуальностью. Обеспечение доступности включает в себя четыре основных принципа: надежность, управляемость, воспринимаемость и понятность. Под надежностью понимается, что пользователи могут получить доступ к контенту по мере появления новых технологий. Управляемость же подразумевает, что пользователи должны иметь возможность управления интерфейсом. В свою очередь воспринимаемость говорит о том, что пользователи должны быть в состоянии воспринимать контент (он должен быть видимым). Понятность говорит о том, что контент и интерфейс должны быть понятны пользователю.

Нами было принято решение использовать WCAG-стандарты для разработки эвристики оценки доступности URL: на первом этапе оценивается доступность исследуемого URL; второй этап – получить URL из эвристики, связанной с системой *Google PageRank*, описанной в одной из наших предыдущих работ, и сделать его оценку доступности; третий этап – сравнение оценки доступности первого и второго URL и формирование предположения о подозрительности данного URL.

Сегодня разработано множество инструментов, в том числе онлайн-сервисов. Для реализации первого этапа, рассмотрев возможные варианты, мы решили использовать несколько онлайн-сервисов, таких как WAVE, Achecker и др. Множественный выбор инструментов обусловлен тем, что результаты проверок одного и того же URL разнились от инструмента к инструменту. Это связано с тем, что не все инструменты поддерживали Javascript и каждый из инструментов имел ряд уникальных проверок.

В основе второго этапа лежит эвристика «Валидность доменных имен», основная идея которой – сделать запрос с адресом исследуемого ресурса к *Google PageRank* и выбрать URL из результата с наилучшим рейтингом. В результате мы получим, с большей долей вероятности,

валидный URL-адрес, оценку доступности которого проводим по тем же сервисам, что на первом этапе.

На третьем этапе сравниваются результаты оценок доступности двух URL, где эталонным считается URL, полученный из второго этапа. Каждой метрике отчета мы присвоили свой вес, в результате чем больше отличий в отчете, тем больше вероятность, что исследуемый URL подозрительный. Проведя практическое исследование, мы столкнулись с проблемой, что система ранжирования возвращает в ряде случаев URL на главную страницу ресурса, а текущий исследуемый URL мог вести на страницу, которая никак не связана с главной (например, на форум). Это связано с тем, что не все страницы ресурсов индексируются. Поэтому мы приняли решение проверять только главную страницу и дополнительно на первом этапе обрезаем URL до домена верхнего уровня, удаляя все, что после. Это увеличило точность эвристики и уменьшило ложные срабатывания. Как только мы решили проблему с ложными срабатываниями, нам удалось увеличить процент соответствия с 60 до 80 %, в результате если оценка доступности URL из первого этапа соответствует оценке URL из второго этапа менее чем на 80 %, то URL считается подозрительным, в противном случае – легитимным.

4. Метрики оценки системы

Для того чтобы оценить работу системы, мы использовали следующие метрики: TPR (True Positive Rate) и FPR (False Positive Rate), как в предыдущей работе; также мы использовали TNR (True Negative Rate), FNR (False Negative Rate), точность и Recall. Соответственно TPR и FPR использовались для сравнения показателей текущей модели с предыдущей версией.

Итак, точность (Accuracy) является одной из общеизвестных мер, применяемых для оценки. Точность – это отношение правильно определенных фишинговых и легитимных ресурсов к общему количеству ресурсов с учетом ошибок классификации, как показано в формуле

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

где TN (True Negative) – количество правильно классифицированных легитимных ресурсов; TP (True Positive) – количество правильно классифицированных фишинговых ресурсов; FP (False Positive) – количество неправильно классифицированных фишинговых ресурсов, которые классифицированы как легитимные; FN (False Negative) – количество неправильно классифицированных легитимных ресурсов, которые классифицированы как фишинговые.

Recall – следующая метрика для оценки работы системы. Recall, в свою очередь, – это отношение TP к сумме TP и FN. Вычисляется Recall по следующей формуле

$$\text{Recall} = \frac{TP}{(TP + FN)}. \quad (2)$$

TNR является третьей метрикой оценки системы. TNR – это отношение TN к сумме TN и FP, метрика представлена формулой

$$\text{TNR} = \frac{TN}{(TN + FP)}. \quad (3)$$

Четвертая метрика, FNR – это отношение FN к сумме FN и TP, метрика вычисляется по формуле

$$\text{FNR} = \frac{FN}{(FN + TP)}. \quad (4)$$

5. Экспериментальные результаты

Основываясь на результатах предыдущих работ [8], вместе с новой архитектурой системы мы провели оценку эффективности системы. Оценка эффективности проводилась по факту проведения двух экспериментов на одном и том же наборе данных, что был собран ранее в предыдущих работах: 1718 фишинговых и 1103 легитимных ресурса. Проведение экспериментов мы разделили на два этапа: первый включал в себя эвристики из прошлых работ плюс четыре эвристики из этой работы; второй этап включал в себя все эвристики.

Итак, для первого эксперимента мы использовали два фильтра и восемь эвристик из предыдущей работы и добавили следующие эвристики: длина URL; символ @ в URL; точки в URL и слеши, протокол и порт в URL. Поскольку исследование URL – важная часть при определении фишинга, мы продвинулись в достижении лучших результатов для разрабатываемой архитектуры. Результаты эксперимента приведены в таблице.

Во втором эксперименте мы оценивали работу системы в целом, со всеми эвристиками. Мы проводили второй эксперимент несколько раз, поскольку не все эвристики работали с необходимой точностью. На основании полученных результатов мы дорабатывали эвристики. Так, для «проверки кода страны» мы добавили возможность использования различных баз геолокации и др. Результаты эксперимента приведены в таблице.

Результаты экспериментов
Results of experiments

Метрика	Архитектура на основе двух фильтров и 8 эвристик	Архитектура на основе двух фильтров и 12 эвристик	Архитектура на основе двух фильтров и 16 эвристик
TPR, %	93,2	96,74	97,85
FPR, %	7,25	3,63	2,09
TNR, %	92,74	96,37	97,91
FNR, %	6,69	3,26	2,15
Accuracy, %	93,08	97,00	98,16
Recall, %	93,31	86,74	97,85

Выводы

Ввиду острой ситуации, связанной с фишинговыми атаками в мире, нами было принято решение разрабатывать антифишинговый метод. Мы провели обзор работ, связанных с разработкой антифишинговых решений. Основываясь на собранных сведениях, мы доработали ряд существующих эвристик и разработали новые. Затем мы продолжили разрабатывать архитектуру метода борьбы против фишинга в АСУП. Метод основан на получении URL-адреса, который запрошен пользователем при помощи DPI. Затем данный URL исследуется по фильтру, основанному на персональном белом списке и модуле поиска формы логина. Следующим этапом URL проходит проверку по ряду различных эвристик, которые возвращают три возможных значения: фишинговый URL, легитимный URL или подозрительный. По факту проверки, если URL оценен как фишинговый или подозрительный, уходит отчет для офицера безопасности, который принимает решение о блокировке этого URL. Нами было проведено два практических исследования разработанного метода. В результате первого, где мы использовали только ряд эвристик, нам удалось достигнуть 96,74 % TPR и 3,63 % FPR. Эти показатели улучшились, по сравнению с предыдущей архитектурой. Проведя второй эксперимент, основываясь на работе всех эвристик системы, в сравнении с 1-м экспериментом мы улучшили показатели TPR и FPR на 1,11 и 1,54 % соответственно. Также мы улучшили точность метода до 98,16 %. Следующим этапом будет разработка системы поддержки принятия решений для автоматической блокировки фишинговых URL.

Литература

1. Список угроз ФСТЭК. – http://bdu.fstec.ru/threat (дата обращения: 15.02.2020).

2. Что такое «фишинг» (2019). – https://encyclopedia.kaspersky.ru/knowledge/what-is-phishing/ (дата обращения: 15.12.2019).

3. Митюков, Е.А. Жизненный цикл фишинговых атак и техники их реализации / Е.А. Митюков // Решение. – 2019. – Т. 1. – С. 140–142.

4. Митюков, Е.А. Аудит безопасности SCADA-систем / Е.А. Митюков, А.В. Затонский, П.В. Плехов // Защита информации. Инсайд. – 2016. – № 4. – С. 72–77.

5. Митюков, Е.А. Уязвимости MS SQL SERVER, или использование хранимых процедур в своих целях / Е.А. Митюков // Защита информации. Инсайд. – 2017. – № 6 (78). – С. 44–47.

6. Спам и фишинг (2018). – https://securelist.ru/spam-and-phishing-in-2018/93453/ (дата обращения: 30.12.2018).

7. Radiflow Reveals First Documented Cryptocurrency Malware Attack on a SCADA Network (2020). – https://radiflow.com/news/radiflow-reveals-first-documented-cryptocurrency-malware-attack-on-a-scada-network/ (дата обращения: 20.01.2020).

8. Mityukov, E.A. Phishing detection model using the hybrid approach to data protection in industrial control system / E.A. Mityukov, A.V. Zatonsky, P.V. Plekhov, N.V. Bilfeld // IOP Conference Series: Materials Science and Engineering. – 2019. – Vol. 537. DOI: 10.1088/1757-899X/537/5/052014

9. Gastellier-Prevost Sophie. Decisive heuristics to differentiate legitimate from phishing sites. La Rochelle, France / Gastellier-Prevost Sophie, Granadillo Gustavo Gonzalez, Laurent Maryline // Proc. of conference on network and information systems security (SAR-SSI). – May 2011. – P. 1–9. DOI: 10.1109/SAR-SSI.2011.5931389

10. Jeeva, S.C. Intelligent phishing URL detection using association rule mining / S.C. Jeeva, E.B. Rajsingh // Human-Centric Comput. Inf. Sci. – 2016. – 6, 10. DOI: 10.1186/s13673-016-0064-3

11. Sonowal, G. PhiDMA – A phishing detection model with multi-filter approach / G. Sonowal, K.S. Kuppusamy // Journal of King Saud University – Computer and Information Sciences. – Jule 2017. – Vol. 32 (1). – P. 99–112.

12. Teaching Johnny not to fall for phish / Kumaraguru Ponnurangam, Sheng Steve, Acquisti Alessandro et al. // Article 7 ACM Transactions on Internet Technology. – May 2010. – 10 (2): 31. DOI: 10.1145/1754393.1754396

13. Затонский, А.В. Информационные технологии: Разработка информационных моделей и систем / А.В. Затонский. – М.: ИЦ Риор, 2014. – 344 с.

14. A framework for detection and measurement of phishing attacks / S. Garera, N. Provos, M. Chew, A.D. Rubin // Proceedings of the 2007 ACM Workshop On Recurring Malcode. – Alexandria, Virginia, USA, ACM, 2007. – P. 1–8. DOI: 10.1145/1314389.1314391

15. Zhang Jian. Highly predictive blacklisting / Zhang Jian, Porras Phillip, Ullrich Johannes // Proc. of the 17th conference on security symposium. – CA, USA: USENIX Association Berkeley; 2008. – P. 107–122.

Митюков Евгений Алексеевич, аспирант кафедры информационных технологий и автоматизированных систем, Пермский национальный исследовательский политехнический университет, г. Пермь; the_nomad@mail.ru.

Затонский Андрей Владимирович, д-р техн. наук, профессор, заведующий кафедрой автоматизации технологических процессов, Пермский национальный исследовательский политехнический университет, Березниковский филиал, г. Березники; avz@bfpstu.ru.

Поступила в редакцию 5 февраля 2020 г.

DOI: 10.14529/ctcr200206

PHISHING ATTACK DETECTION MODEL BASED THE HYBRID APPROACH TO DATA PROTECTION IN INDUSTRIAL CONTROL SYSTEM

E.A. Mityukov*[1]*, the_nomad@mail.ru,
A.V. Zatonsky*[2]*, avz@bfpstu.ru

[1] Perm National Research Polytechnic University, Perm, Russian Federation,
[2]Perm National Research Polytechnic University, Berezniki Branch, Berezniki, Russian Federation

Introduction. Today there is an annual development of phishing techniques cybercriminals who are aimed at industrial control systems in order to compromise sensitive information, the task is to develop new methods for determining phishing attacks aimed at the industrial sector is extremely

important. **Aim**. The article discusses developed method to protection against phishing attacks on users and services of industrial control systems. **Materials and methods**. The possible literature sources of the subject area are analyzed. Based on the information gathered from previous studies, work on improving the architecture of the phishing protection system are continued. With aimed at improving the accuracy of detection of phishing URLs (Uniform Resource Locator), eight heuristics in the system architecture are added. Most heuristics are aimed at semantic verification of URLs, in terms of the use of special characters, periods, slashes, URL protocols and ports, including the length of the URL itself. Additionally, the validity of the SSL/TLS (Secure Sockets Layer/Transport Layer Security) certificate, phishing keywords in the URL and difference the hosting country of the provider with the country of the top-level domain is checked. **Results**. The practical research of the new architecture system with various combinations of heuristics are carried out. Quantitative data showing the improvement of key indicators to detecting phishing URLs by the system are presented. Security Officer decides on phishing or legitimate URL by new architecture of system are helped. **Conclusion**. The presented system shows the following indicators: TPR (True Positive Rate) – 97.85 % and FPR (False positive Rate) – 2.09 %. Also, the accuracy of the method to 98.16 % is improved.

Keywords: cybersecurity, cyberattacks, industrial control system, anti-phishing, URL features.

References

1. *Spisok ugroz FSTEK* [List of Threats of FTECS]. Available at: http://bdu.fstec.ru/threat (accessed 15.02.2020).

2. *Schto takoe "fishing"* [What is "Phishing"] (2019). Available at: https://encyclopedia.kaspersky.ru/knowledge/what-is-phishing/ (accessed 15.12.2019).

3. Mityukov E.A. [Life Cycle of Phishing Attacks and Techniques of their Implementation]. *Reshenie*, 2019, vol. 1, pp. 140–142. (in Russ.)

4. Mityukov E.A., Zatonskiy A.V., Plehov P.V. [SCADA Security Audit]. *Information Security. Insider*, 2016, no. 4, pp. 72–77. (in Russ.)

5. Mityukov E.A. [MS SQL SERVER Vulnerabilities, or the Use of Stored Procedures for their Own Purposes]. *Protect Information. Insider*, 2017, no. 6 (78), pp. 44–47. (in Russ.)

6. *Spam i fishing* [Spam and Phishing]. Available at: https://securelist.ru/spam-and-phishing-in-2018/93453/ (accessed 30.12.2018).

7. Radiflow Reveals First Documented Cryptocurrency Malware Attack on a SCADA Network (2020). Available at: https://radiflow.com/news/radiflow-reveals-first-documented-cryptocurrency-malware-attack-on-a-scada-network/ (accessed 20.01.2020).

8. Mityukov E.A, Zatonsky A.V, Plekhov P.V, Bilfeld N.V. Phishing detection model using the hybrid approach to data protection in industrial control system. *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 537. DOI: 10.1088/1757-899X/537/5/052014

9. Gastellier-Prevost Sophie, Granadillo Gustavo Gonzalez, Laurent Maryline. Decisive heuristics to differentiate legitimate from phishing sites. La Rochelle, France. *Proc. of conference on network and information systems security (SAR-SSI)*, May 2011, pp. 1–9. DOI: 10.1109/SAR-SSI.2011.5931389

10. Jeeva S.C., Rajsingh E.B. Intelligent phishing URL detection using association rule mining. *Human-Centric Comput. Inf. Sci.*, 2016, 6, 10. DOI: 10.1186/s13673-016-0064-3

11. Sonowal Gunikhan, Kuppusamy K.S. PhiDMA – A phishing detection model with multi-filter approach. *Journal of King Saud University – Computer and Information Sciences*, Jule 2017, vol. 32 (1), pp. 99–112.

12. Kumaraguru Ponnurangam, Sheng Steve, Acquisti Alessandro, Cranor Lorrie Faith, Hong Jason. Teaching Johnny not to fall for phish. *Article 7 ACM Transactions on Internet Technology*, May 2010, 10 (2): 31. DOI: 10.1145/1754393.1754396

13. Zatonskiy A.V. *Informacionnye tehnologii: Razrabotka informacionnyh modelej i sistem* [Information Technology: Development of Information Models and Systems]. Moscow, Rior Publ., 2014, 344 p.

14. Garera S, Provos N, Chew M, Rubin AD. A framework for detection and measurement of phi-

shing attacks. *Proceedings of the 2007 ACM Workshop on Recurring Malcode*. Alexandria, Virginia, USA, ACM, 2007. – pp. 1–8. DOI: 10.1145/1314389.1314391

15. Zhang Jian, Porras Phillip, Ullrich Johannes. Highly predictive blacklisting. *Proc. of the 17th conference on security symposium*. CA, USA: USENIX Association Berkeley, 2008, pp. 107–122.

Received 5 February 2020

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Митюков, Е.А. Модель обнаружения фишинговых атак на основе гибридного подхода для защиты автоматизированных систем управления производством / Е.А. Митюков, А.В. Затонский // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». – 2020. – Т. 20, № 2. – С. 56–66. DOI: 10.14529/ctcr200206

FOR CITATION

Mityukov E.A., Zatonsky A.V. Phishing Attack Detection Model Based the Hybrid Approach to Data Protection in Industrial Control System. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*, 2020, vol. 20, no. 2, pp. 56–66. (in Russ.) DOI: 10.14529/ctcr200206