DOI 10.37882/2223–2966.2021.04.05

АНАЛИЗ МЕТОДОВ ОБХОДА ЗАЩИТЫ ВЕБ-САЙТОВ ОТ ИЗВЛЕЧЕНИЯ ДАННЫХ

ANALYSIS OF METHODS TO BYPASS DATA EXTRACTION PROTECTION OF WEBSITES

P. Babaritsky

Summary. This article is devoted to the study and analysis of methods for bypassing the protection of sites based on web technologies from data extraction. At the same time, the main goal of the article is to analyze methods for bypassing the protection of web resources from unauthorized access from the point of view of mathematical models and their algorithmic representation. As part of the study, we propose our own system for analyzing technologies for bypassing site protection. The developed system is based on a set of criteria-signs, on the basis of the analysis of which it becomes possible to comprehensively evaluate algorithms for bypassing the protection of web resources in terms of efficiency, danger level and other characteristic parameters. The scientific novelty of the work is determined by considering the problem of analyzing methods for bypassing the protection of sites precisely at the model level, and not at the technology level. Practical usefulness lies in the effectiveness of the proposed analysis system, as well as in the results of the study, which can later be used in practice to improve methods of protecting websites from unauthorized data extraction.

Keywords: data extraction, websites, the bypass protection, analysis, vulnerabilities.

Бабарицкий Павел Александрович
Аспирант, ФГАОУ ВО «Национальный исследовательский университет ИТМО»
redbear95@gmail.com

Аннотация. Данная статья посвящена исследованию и анализу методов обхода защиты сайтов на основе веб-технологий от извлечения данных. При этом основной целью статьи является анализ методик обхода защиты веб-ресурсов от несанкционированного доступа с точки зрения математических моделей и их алгоритмического представления. В рамках проведенного исследования предлагается собственная система анализа технологий обхода защиты сайтов. Разработанная система основана на совокупности критериев-признаков, на основе анализа которых становится возможной комплексная оценка алгоритмов обхода защиты web-ресурсов с точки зрения эффективности, уровня опасности и прочих характеристических параметров. Научная новизна работы определяется рассмотрением проблемы анализа методов обхода защиты сайтов именно на уровне модели, а не на уровне технологий. Практическая полезность заключается в эффективности предложенной системы анализа, а также в полученных результатах исследования, которые в дальнейшем могут быть использованы на практике для усовершенствования методов защиты веб-сайтов от несанкционированного извлечения данных.

Ключевые слова: извлечение данных, веб-сайты, обход защиты, анализ, уязвимости.

Введение

Сохранение конфиденциальности данных в сети является определяющим требованием качества системы безопасности веб-ресурса. Можно привести большое количество примеров операций на сайтах в сети Интернет, в процессе выполнения которых решение вопроса защиты данных от несанкционированного извлечения становится неотъемлемой задачей (онлайн-покупки, банковские переводы, хранение конфиденциальных данных сотрудников, хранение корпоративных данных организации и прочие).

В данной статье выполняется комплексный анализ методов обхода защиты веб-сайтов от несанкционированного извлечения конфиденциальных данных. Сегодня такие методы активно совершенствуются злоумышленниками, следовательно, задачу совершенствования технологий противодействия подобным методикам следует охарактеризовать как весьма значимую и актуальную. Важной информацией для задач усовершенствования методов защиты сайтов будут являться результаты исчерпывающего анализа методов обхода защиты сайтов. Для наибольшей информативности, эффективности и практической полезности полученных результатов, предлагается провести исследование не на уровне внешнего исполнения (уровень технологии), а на уровне математических закономерностей и принципов работы алгоритмов (уровень модели) злонамеренных методов несанкционированного извлечения данных.

Разработка системы анализа методов обхода защиты веб-сайтов от извлечения данных

В рамках задачи разработки системы анализа методов обхода защиты веб-ресурсов было принято во вни-

Рис. 1. Схема целевых критериев анализа с описанием

мание требование к проработке исследования методов с точки зрения моделей, а не технологий. Таким образом, предложенная система анализа должна быть ориентирована специальным образом не на оценку результатов реализации той или иной технологии обхода защиты (задействованный инструментарий, выполняемые злоумышленником операции и т.д.), на оценку тех характеристических критериев, которые могут быть справедливы для внутренних механизмов метода (математические концепции, модели, алгоритмы и т.д.).

Также в процессе разработки системы анализа была принята в различность технического воплощения методов обхода защиты. Так, например, метод обхода аутентификации на веб-ресурсе и метод обхода защиты на основе визуального ввода (так называемая, CAPTCHA) будут иметь отличающуюся техническую специфику: первый случай будет из области криптографии и шифрования данных, когда заданный пароль может быть установлен злоумышленником с помощью стандартной операции автоматического подбора (перебора); вторая ситуация с обходом CAPTCHA будет относиться к области распознавания образов — в этом случае недоброжелатель может использовать, например, нейронную сеть, которая ориентирована на распознавание символов (ныне существует множество разнообразных готовых библиотек, натренированных на большой выборке данных).

В связи с вышеперечисленными факторами для анализа методов обхода защиты не могут быть избраны стандартные для оценки алгоритмов целевые критерии, поскольку, в этом случае различность технических моделей методов снизит объективность результатов анализа. Необходимо подобрать целевые признаки для анализа таким образом, чтобы была возможность оценить модели разной технической организации из различных предметных областей с минимальным влиянием на объективность полученных результатов исследования. Иллюстрация 1 демонстрирует схему целевых критериев, избранных для анализа методов обхода защиты веб-ресурсов от несанкционированного извлечения данных, с их детальным описанием.

Предполагается, что для каждого метода по каждой целевой характеристике будет в процессе непосредственного исследования экспертом будет выставлена соответствующая оценка от одного до пяти баллов,

где «1» — критерий абсолютно не удовлетворяется, «5» — критерий полностью удовлетворяется. Целью анализа является выявление наиболее мощного и лучшего метода обхода защиты сайтов с точки зрения злоумышленника. Таким образом, наилучший метод для недоброжелателя будет для специалиста по информационной безопасности тем методом, которого следует опасаться более остальных.

В рамках дальнейших мероприятий по анализу методов обхода защиты веб-сайтов от несанкционированного доступа планируется:

- дать подробное описание каждому из рассматриваемых методов;
- провести анализ с выставлением экспертных оценок по целевым характеристикам (см. рисунок 1);
- задокументировать выводы по результатам анализа в виде сравнительной таблицы, дать экспертное заключение.

Описание целевых методов и проведение анализа

Очевидно, что методы обхода имеют непосредственную связь с технологиями, посредством которых организована защита веб-сайта. Глобально рассматриваемые методы обхода могут разделены на три группы:

1. Методы обхода аутентификации (authentication bypass) — критическая уязвимость, характерная максимальным уровнем влияния с точки зрения информационной безопасности.
2. Методы обхода SSL — данная группа методов ориентирована на перехват конфиденциального трафика, передача которого организована на основе криптографического протокола SSL (Secure Socket Layer) для обеспечения безопасного соединения.
3. Методы обхода CAPTCHA — группа методов, направленных на взлом защиты, организованной посредством сокрытия содержимого в нетекстовых элементах. Примером CAPTCHA может быть принуждение пользователя к распознаванию изображений с определенными объектами, либо решение математического примера, представленного на картинке.

Совокупность методов каждой группы может быть представлена с точки зрения механизмов исполнения и алгоритмов функционирования.

Далее выполняется анализ условно определенных моделей обхода защиты веб-ресурсов. Относительно каждой из моделей приводится описание сильных и слабых сторон с учетом целевых критериев, определенных в рамках мероприятий по разработке системы анализа.

Встраивание и выполнение произвольного кода в исходный контекст

Практическими примерами данной модели могут быть SQL-инъекции и XPath-инъекции, реализация которых заканчивается успешным взломом веб-ресурса в 99% случаев [1]. Об эффективности модели свидетельствует статус критической уязвимости с максимальным уровнем угрозы, присвоенный модели специалистами по организации информационной безопасности (максимальная угроза класса A1 по классификации OWASP [2]). Для практической реализации модели достаточно знать логику выполнения запросов или специфику структуры XML-файлов в случае с XPath-инъекцией.

При большой эффективности практическое воплощение модели не требует каких-либо финансовых затрат или использование нагруженной программно-аппаратной архитектуры. При наличии информации об используемой базе данных и хороших познаний в SQL/XML инъекция может быть реализована в относительно сжатые сроки.

Использование готового специализированного программного обеспечения

Для достижения злонамеренных целей — примерами специализированного ПО могут быть реверс-прокси «Modlishka» [3] для обхода двухфакторной аутентификации с генерацией собственного подменного сертификата, а также фреймворк «Frida», позволяющий организовать перехват мобильного SSL-трафика [4]. Использование готового программного обеспечения в злонамеренных целях существенно ускоряет реализацию мероприятий по взлому за счет готовых компонентов и возможности их быстрой настройки (например, злоумышленнику не нужно писать прокси и с нуля настраивать поддельный сертификат, если используется «Modlishka»). При использовании готового ПО следует учитывать требования к программно-аппаратной архитектуре такого специализированного ПО, а также условия использования (большинство подобных продуктов являются платными).

Реализация атак на основе злонамеренных криптоалгоритмов

В эту категорию относятся криптографические модели обхода безопасного соединения на основе SSL. В качестве примеров воплощения таких атак следует

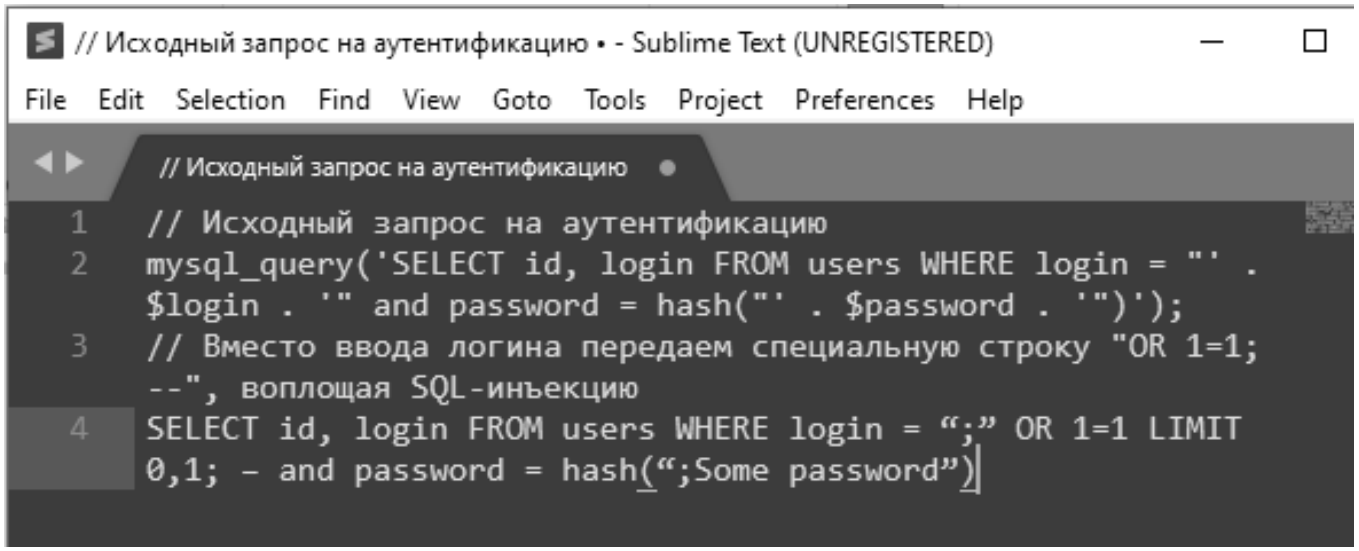

Рис. 2. Пример реализации SQL-инъекции

отметить «BEAST» [5] и «RC4» [6], которые в основе своей криптоаналитической модели используют уязвимости SSL и методики перебора для дешифровки трафика, паролей и прочей конфиденциальной информации. Модели на основе криптоалгоритмов предъявляют жесткие требования к аппаратной архитектуре, ввиду необходимости выполнения большого количества операций дешифровки символов. Так, например, прогон колоссального объема данных через шифр для инициализации атаки с помощью «RC4» может занимать более 32 часов [6]. Также следует отметить высокую сложность практической реализации с точки зрения криптоаналитики.

Написание собственных вредоносных программных модулей и инструкций

В эту категорию моделей можно отнести веб-роботов для автоматизированного сбора данных пользовательского ввода с целью дальнейшей организации базы знаний для взлома защиты на основе CAPTCHA. Также следует отметить алгоритмы на основе наиболее продвинутых технологий, например, модули распознавания символов на основе нейросетей глубокого обучения [7, 8]. Для данных моделей следует отметить высокую эффективность и возможность быстрого получения первых результатов за счет наличия готовых библиотек и наборов обучения нейросетей (так называемых датасетов) и многочисленных обучающих материалов. Более высокая точность и эффективность конечных результатов потребует наиболее глубокого обучения сети для распознавания символов, а также мощную аппаратно-вычислительную инфраструктуру [9].

Документирование и анализ результатов исследования

В процессе выполнения практической части исследования были реализованы два алгоритма, имитирующих атаку с целью извлечения данных, а также два алгоритма защиты противодействующих двум атакующим алгоритмам.

В некоторых ситуациях извлечение данных может выполняться с целью доступа к конфиденциальным данным web-сервиса. Примером таких данных могут быть логины и пароли для учетных записей зарегистрированных пользователей того или иного web-ресурса. Подавляющее большинство сайтов для обработки, хранения и прочих операций с пользовательскими данными использует возможности СУБД. Наиболее распространенным вариантом являются СУБД на основе реляционной модели, поддерживающей язык SQL-запросов. В связи с этим обстоятельством было принято решение реализовать алгоритмы двух альтернативных моделей извлечения данных: SQL-инъекцию и XPath-инъекцию, где, вместо запросов для хранения данных используется синтаксис на основе языка разметки XML [10, 11].

На рисунках 2 и 3 представлены реализованные примеры SQL-инъекций для извлечения конфиденциальных данных с сайта. В инструменте реализации использован наиболее часто встречающийся язык PHP.

В данном примере в исходном запросе отсутствует проверка на ввод некорректных данных. Таким образом, значения передаются из полей ввода формы не-

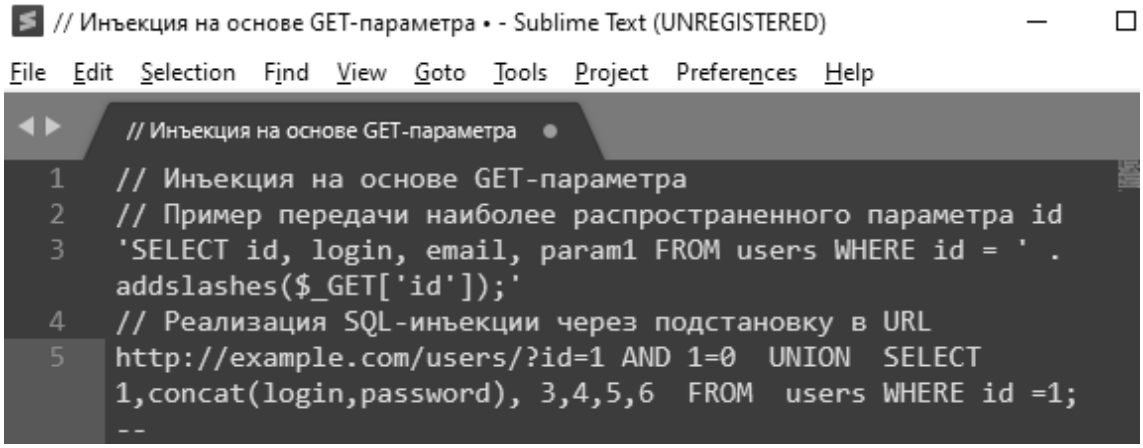

Рис. 3. Пример реализации SQL-инъекции

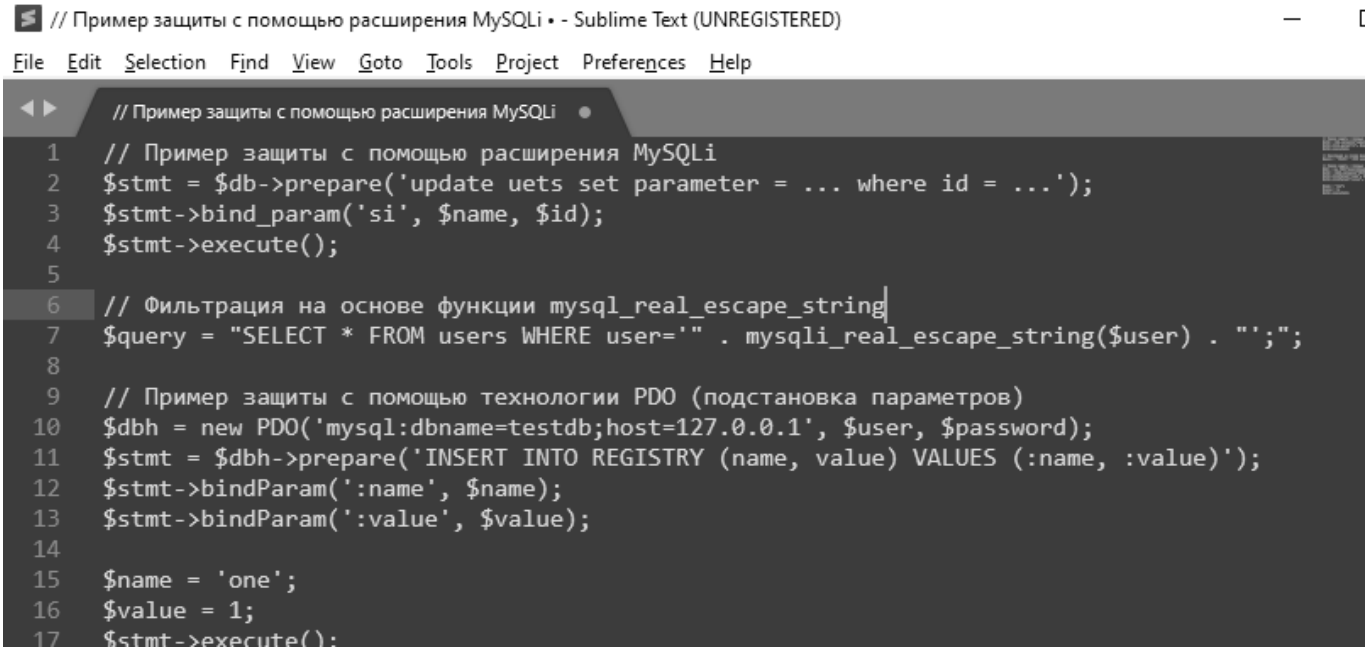

Рис. 4. Пример реализации методик защиты от SQL-инъекций

посредственно в SQL-запрос. В данном случае встраивание строки «OR1=1; —» будет инициировать попытку входа в качестве первого пользователя базы данных, это, как правило, запись администратора. При благополучном воплощении инъекции запрос вернет данные первого пользователя базы данных и, даже, в некоторых случаях, сразу войдет в систему. Так, незарегистрированный пользователь может получить привилегии и доступ к данным в качестве легитимного пользователя web-ресурса и даже администратора.

Приведенная инъекция основана на злоупотреблении потребностью в передаче наиболее частого параметра «id». Сильной стороной данного типа инъекции является возможность использования вне зависимости от вида URL. Благополучная реализация такого запроса-инъекции вернет в ответе набор типа «prepared data», в котором будет id, логин, хеш пароля и ряд других параметров. С помощью специализированных программ хеш пароля может быть дешифрован и использован для несанкционированной аутентификации на web-ресурсе с целью извлечения данных.

На рисунке 4 представлены примеры реализации защиты от приведенных SQL-инъекций. В частности,

```
String username = req.getParameter("username');
String username = req.getParameter("username');

String password = req.getParameter("password');
XPathFactory factory = XPathFactory.newInstance();

Xpath xpath = factory.newXPath();
File file = new File("/usr/webappdata/users.xml');

InputSource src = new InputSource(new FileInputStream(file));
XPathExpression expr = xpath.compile("//users[username/text()=' " +
username + " ' and password/text()=' " + password + ' ']/id/text()');

String id = expr.evaluate(src);
```

Рис. 5. Запрос аутентификации

```
// Запрос получения ID, привязанного к логину и паролю
users[username/text()='admin' and  password/text()='admpass'] /id/text()

// Реализация XPath-инъекции
users[username/text()='admin' and  password/text()='' or '1'='1'
]/id/text()
```

Рис. 6. Исходный запрос и основанная на нем XPath-инъекция

это использование расширения MySQLi, в случае, когда работа с базой данных ведется посредством СУБД MySQL [12]. Настоящее расширение обеспечивает возможность работы со связанными параметрами, за счет чего исходный запрос становится единой конструкцией и исключает возможность встраивания стороннего контекста в виде инъекции.

Другой вариант защиты предполагает использование вместо уязвимой функции «addslashes()», которая не предусматривает большинство методик взлома, альтернативной более надежной функции «mysql_real_escape_string», обеспечивающей фильтрацию строковых параметров.

Третья тактика защиты предполагает использование технологии «PHP Data Objects (PDO) — специализированной прослойки для работы с объектами, делающей возможной подстановку параметров [13].

Все три представленные методики в комплексе могут быть использованы как полноценный алгоритм защиты от SQL-инъекций.

В случае XPath-инъекций код для извлечения данных встраивается в базу данных на языке XML [14, 15]. В рамках демонстрации алгоритма атаки на основе XPath-инъекции была реализована небольшая база данных на XML, содержащая в себе пользователей <user> с их <id>, <username> и <passowrd>.

Для этой базы данных запрос, реализующий аутентификацию, может иметь вид, приведенный в листинге на рисунке 5.

Вид исходного запроса на получение ID легитимного пользователя и реализация XPath-инъекции на основе этого запроса приведены на рисунке 6.

Отсутствие проверки на корректность введенных данных позволяет применить встраиваемую конструкцию. В результате запрос вернет ID пользователя «admin» с пустым паролем ввиду истинности выражения «1=1».

В качестве практической реализации алгоритма защиты от XPath-инъекций предлагается комбинация методов, реализованных в листинге на рисунке 7.

Таблица 1. Результаты анализа алгоритмов извлечения данных web-ресурсов и алгоритмов защиты

Алгоритм /Критерий	SQL-инъекция	XPath-инъекция	Защита от SQL-инъекции	Защита от XPath-инъекции
Простота реализации	5	5	5	5
Скорость реализации	5	5	5	5
Эффективность	5	4	3	3
Сложность архитектуры	4	3	4	3
Стоимость реализации	5	5	5	5
Скорость работы алгоритма	4	4	3	4
Итог	28	26	25	25

```
// Экранирование контекста с ненадежными данными
String FindUserXPath;
FindUserXPath = "//Employee[UserName/text()='" + Request("Username").Replace("'",
"&apos;") + "' And
        Password/text()='" + Request("Password").Replace("'", "&apos;") + "']";

// Использование параметризированного запроса
"users[LoginID/text()= $LoginID and passwd/text()= $password]"
```

Рис. 7. Методы защиты от XPath-инъекций

В случае, когда для завершения ненадежного ввода в динамическом запросе XPath используются кавычки, предлагается экранировать таковое содержимое ненадежного ввода для предотвращения «утечки» ненадежных данных из цитируемого контекста.

Также предлагается использовать параметризированные конструкции, в рамках которых запросы предварительно скомпилированы и вводимые пользователем данные передаются как параметры, а не выражения.

По результатам анализа выполненных разработок алгоритмов атаки и защиты была сформирована итоговая таблица с оценками по каждому целевому критерию анализа

Заключение

По результатам проведенного исследования в полной мере были решены следующие задачи:

1. Разработка и описание собственной системы анализа методов обхода защиты веб-ресурсов.
2. Краткое описание и анализ целевых моделей обхода защиты веб-сайтов от несанкционированного извлечения данных.
3. Документирование результатов анализа методов обхода защиты сайтов.

Полученные результаты анализа позволяют оценить извлечение данных на основе SQL-инъекции как наиболее эффективный и опасный алгоритм атаки. XPath-инъекция также отмечается высоким уровнем эффективности, однако, немного уступает по общей эффективности и сложности архитектуры, поскольку архитектуры СУБД на основе SQL наиболее распространены и понятны. Алгоритмы защиты от реализованных алгоритмов атак показали аналогичные результаты по всем целевым критериям анализа.

Научная новизна работы определяется рассмотрением проблемы анализа методов обхода защиты сайтов именно на уровне модели, а не на уровне технологий.

Практическая полезность заключается в эффективности предложенной системы анализа, а также в полученных результатах исследования, которые в дальнейшем могут быть использованы на практике для усовершенствования методов защиты веб-сайтов от несанкционированного извлечения данных.

ЛИТЕРАТУРА

1. Shoshitaishvili Yan, Ruoyu Wang, Christophe Hauser, Christopher Kruegel, and Giovanni Vigna. "Firmalice-automatic detection of authentication bypass vulnerabilities in binary firmware." In NDSS. 2015.
2. Poston Howard. "Mapping the OWASP Top Ten to Blockchain." Procedia Computer Science 177 (2020): 613–617.
3. Pillay Rishalin. Learn Penetration Testing: Understand the art of penetration testing and develop your white hat hacker skills. Packt Publishing Ltd, 2019.
4. D'Orazio Christian J. and Kim-Kwang Raymond Choo. "A technique to circumvent SSL/TLS validations on iOS devices." Future Generation Computer Systems 74 (2017): 366–374.
5. Meyer Christopher and Jörg Schwenk. "Lessons Learned From Previous SSL/TLS Attacks-A Brief Chronology Of Attacks And Weaknesses." IACR Cryptol. ePrint Arch. 2013 (2013): 49.
6. Weerasinghe T.D.B. and Chamara Disanayake. "Usage of RC4 cipher in SSL configurations in web portals of Sri Lankan banking/non-banking financial institutes and Awareness levels of relevant staff about it." In 2018 National Information Technology Conference (NITC), pp. 1–6. IEEE, 2018.
7. Brodić Darko, Alessia Amelio and Ivo R. Draganov. "Statistical analysis of dice captcha usability." arXiv preprint arXiv:1706.10177 (2017).
8. Yan Jeff and Ahmad Salah El Ahmad. "Breaking visual captchas with naive pattern recognition algorithms." In Twenty-Third Annual Computer Security Applications Conference (ACSAC2007), pp. 279–291. IEEE, 2007.
9. Ursu Cristian. "Techniques for securing web content." Journal of Mobile, Embedded and Distributed Systems 4, no. 2 (2012): 63–79.
10. Kieyzun Adam, Philip J. Guo, Karthick Jayaraman, and Michael D. Ernst. "Automatic creation of SQL injection and cross-site scripting attacks." In 2009 IEEE31st international conference on software engineering, pp. 199–209. IEEE, 2009.
11. Clarke-Salt Justin. SQL injection attacks and defense. Elsevier, 2009.
12. Subagia A., 2016. Membuat web dengan PHP 7 dan Database PDO MySQLi. Elex Media Komputindo.
13. Tatroe Kevin and Peter MacIntyre. Programming PHP: Creating Dynamic Web Pages. O'Reilly Media, 2020.
14. Clincy Victor and Hossain Shahriar. "Web service injection attack detection." In 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST), pp. 173–178. IEEE, 2017.
15. Moul, Varsha R. and K.P. Jevitha. "Web services attacks and security-a systematic literature review." Procedia Computer Science 93 (2016): 870–877.

Журнал «Современная наука: актуальные проблемы теории и практики»