УДК 004.774 Вестник УрФО № 4(54) / 2024, с. 30–36

Жусов Д. Л., Макеев С. М., Соколов А. Н. DOI: 10.14529/secur240404

АНАЛИЗ СТРУКТУРЫ WEB-САЙТОВ ДЛЯ ИДЕНТИФИКАЦИИ ОБЪЕКТОВ ЭКСПЛУАТАЦИИ УГРОЗ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

В работе представлен вариант решения задачи определения потенциальных угроз информационной безопасности web-сайтов через анализ их структуры и сбор статистических данных о посещаемости отдельных информационных ресурсов. Представлено алгоритмическое обеспечение рассмотренной процедуры. Описана структура данных, обобщающих сведения для идентификации угроз web-сайтов.

Ключевые слова: *web-сайт, гиперссылка, гипертекстовый переход, структура web-сайта, информационная безопасность*

Zhusov D. L., Makeev S. M., Sokolov A. N.

ANALYSIS OF THE STRUCTURE OF WEB SITES TO IDENTIFY OBJECTS OF EXPLOITATION OF INFORMATION SECURITY THREATS

The paper presents a solution to the problem of identifying potential threats to the information security of web sites through the analysis of their structure and the collection of statistical data on the attendance of individual information resources. The algorithmic support of the considered procedure is presented. The structure of data summarizing information for the identification of threats to web sites is described.

Keywords: *website, hyperlink, hypertext navigation, website structure, information security*

Введение

Совершенствование механизмов государственного управления обусловило широкое применение современных информационных технологий для организации взаимодействия государства и общества. Важная роль в структуре инструментов совершенствования информационного общества страны отводится web-сайтам, функционирующих в сети Интернет и обеспечивающих доступ к информации для широкого круга пользователей [1]. Вопросы эффективного функционирования таких инструментов с функциональной точки зрения нормативно определены в [1, 2], а с точки зрения информационной безопасности – в [3].

Компоненты web-сайта в форме исполняемых модулей (сценариев), предназначенных для динамического формирования информационных ресурсов, могут выступать в качестве источников реализации различных классов компьютерных атак [4]. Статистика Positive Technologies [5] свидетельствует о росте числа компьютерных атак на web-приложения. Применение межсетевых экранов web-приложений направлено на обнаружение и блокировку сетевых компьютерных атак, однако не предоставляет возможности анализа инструментария их реализации – конкретного скрипта (приложения), в составе которого присутствуют фрагменты кода – угроз информационной безопасности. Таким образом, можно предположить, что обеспечение информационной безопасности web-сайтов может быть напрямую связано с идентификацией их компонентов, которые потенциально могут быть использованы нарушителем для реализации компьютерных атак. Учитывая, что в качестве таких компонентов web-сайтов выступают информационные ресурсы (web-приложения), то актуальной является задача определения их структуры и последующей оценки количества обращений к ним с привязкой посещаемости к потенциальной эксплуатации уязвимости. Следовательно, анализ структуры web-сайтов может способствовать идентификации источников реализации угроз информационной безопасности и будет в конечном итоге определять повышение защищенности web-сайтов.

Результаты мониторинга [6] свидетельствуют об отсутствии единого подхода к формированию структуры web-сайтов, а существующие способы проверки web-сайтов зачастую слабо автоматизированы, что делает процесс анализа достаточно трудоемким и затратным по времени.

Исследования [7] свидетельствуют о возможности представления структуры web-сайта (рис. 1) графом $G=(V,E)$, где V – множество вершин графа (web-страниц), а E – множество ребер графа (гипертекстовых переходов по сайту – гиперссылок). Для расчета глубины гипертекстовых переходов могут быть использованы известные алгоритмы обхода вершин графа – «*поиск в глубину*» и «*поиск в ширину*» [7].

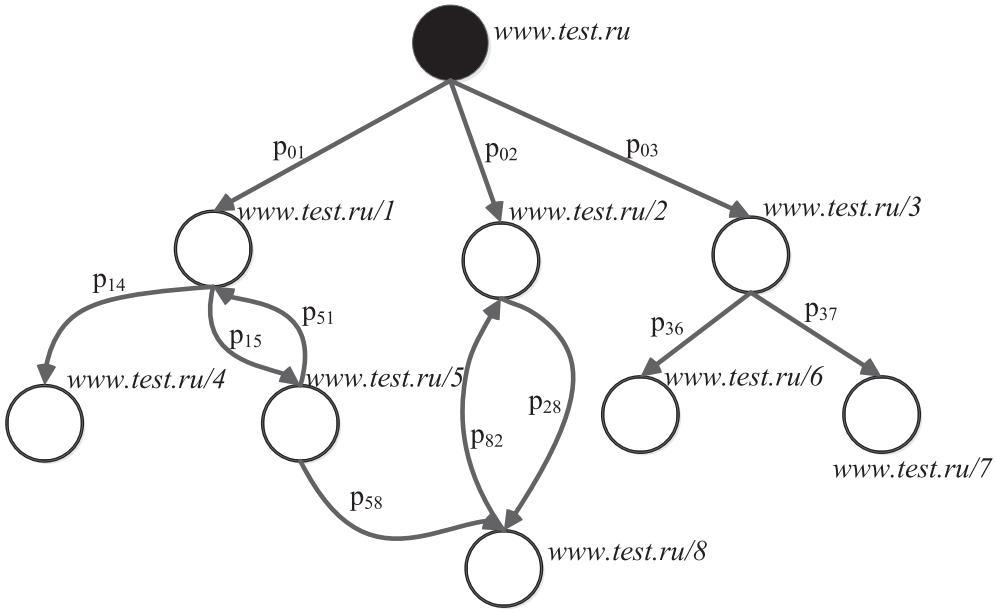

Рис. 1. Графовое представление web-сайта

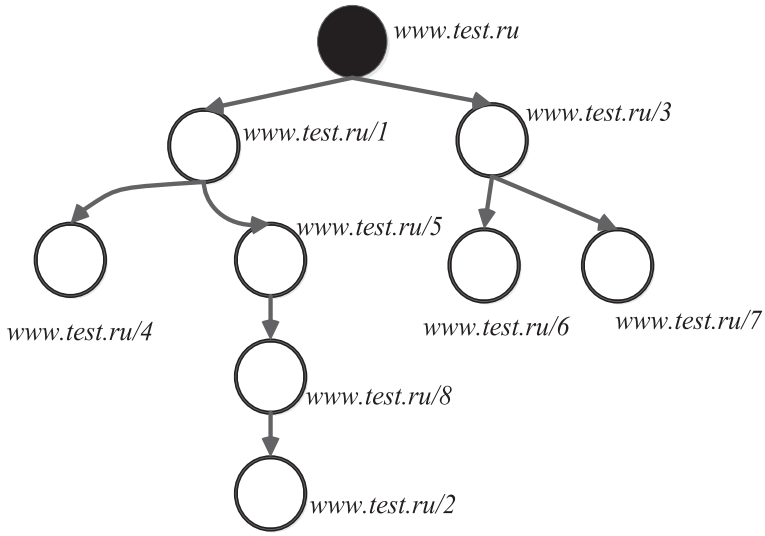

Рис. 2. Дерево web-сайта, построенное алгоритмом поиска «в глубину»

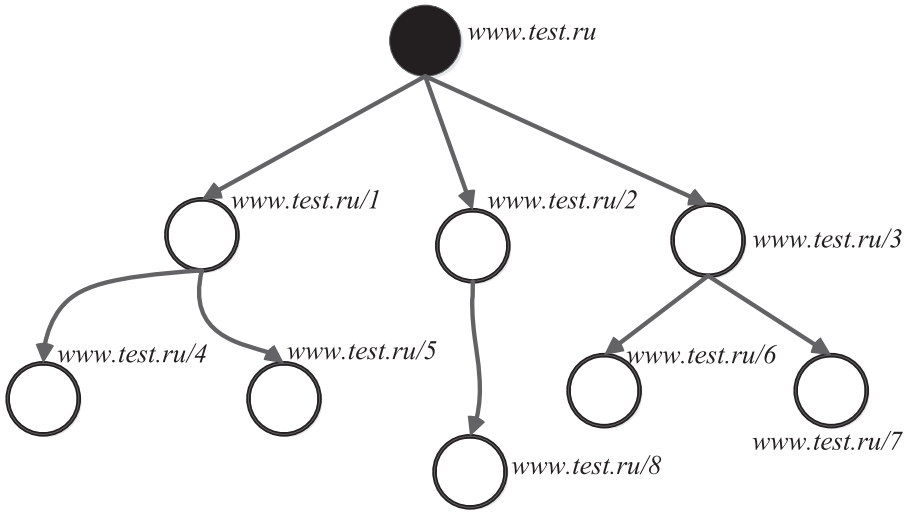

Рис. 3. Дерево web-сайта, построенное алгоритмом поиска «в ширину»

Дерево web-сайта, построенное по алгоритму «в глубину» (рис. 2), и рассчитанное значение максимальной глубины гипертекстовых переходов (равное 4-м) не соответствует действительности, что позволяет говорить о невозможности применения данного алгоритма. В то же время, дерево web-сайта, построенное по алгоритму «в ширину» (рис. 3), и значение максимальной глубины гипертекстовых переходов (равное 3-м) соответствует истинной структуре web-сайта (рис. 1).

Рассчитанная в результате поиска «в ширину» максимальная глубина гипертекстовых переходов равна 3, что соответствует фактическому значению (рис. 1) и свидетельствует о целесообразности применения алгоритма поиска «в ширину» для решения поставленной задачи.

Проведенные исследования позволили разработать алгоритм анализа структуры web-сайтов (рис. 4).

Основу разработанного алгоритма составляет набор рекурсивных процедур, представленный блоками 3 – 15. Сложность его определяется поиском в коде загруженной web-страницы тегов $<a>$ с атрибутом *href* и количеством страниц анализируемого web-сайта и вычисляется как

$$O(\eta) \cdot O(V + E),$$

где η – количество символов в коде web-страниц.

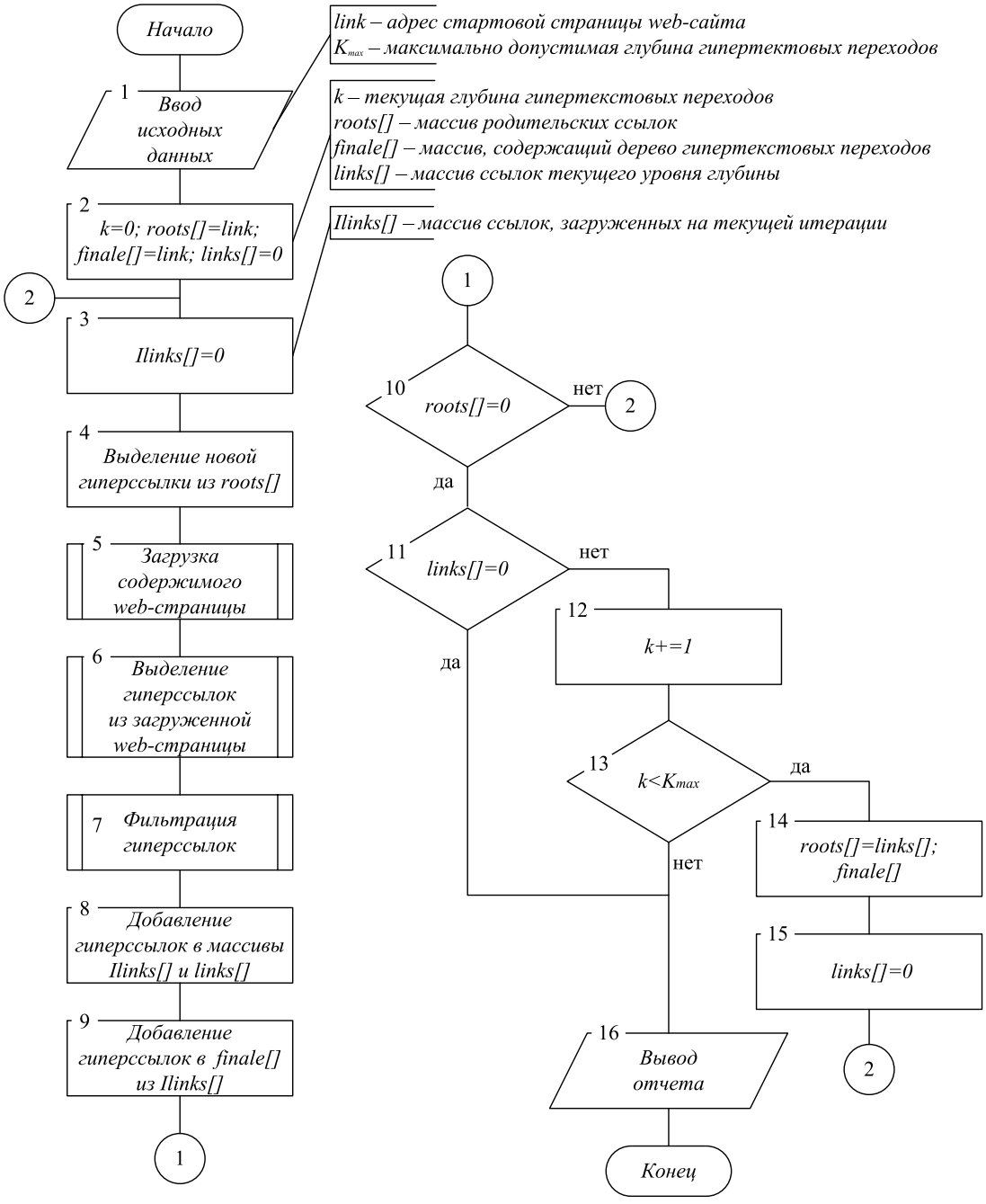

Рис. 4. Алгоритм анализа структуры web-сайта

Каждому узлу сформированного дерева web-сайта ставится в соответствие адрес доступа к нему. Это позволяет сформировать в табличной форме для каждого модуля обработки (сценария, см. табл. 1) не только взаимосвязи с другими информационными ресурсами, но и с полученными данными интегрированной в web-сайт системы статистики.

Статистический анализ данных количества посещений web-сайта позволит определить либо значимые события, вызвавшие интерес со стороны пользователей web-сайта, либо необходимость более углубленного анализа данных при отсутствии публикаций за конкретный период.

Таблица 1

Фрагмент данных анализа структуры web-сайта

Адрес ресурса	Количество взаимосвязей	Глубина переходов	Количество посещений
http://test.ru/index.html	15	0	100
http://test.ru/1/scr1.php	4	1	250
http://test.ru/1/1/scr2.php	2	2	3500
http://test.ru/1/2/scr3.php	2	2	100

Рассмотрим потенциальную ситуацию на примере (табл. 1).

В представленном примере (табл. 1) иллюстрируется ситуация, при которой количество посещений стартовой страницы (100) равно количеству обращений к сценарию *scr3*(100), однако это существенно меньше числа обращений к другим компонентам web-сайта – сценариям *scr1*(250) и *scr2* (3500). Такие сценарии являются серверным расширением функционала web-сайта и могут являться потенциальными объектами эксплуатации угроз информационной безопасности.

С точки зрения реализации угроз функционированию web-сайта (в первую очередь контроля целостности и доступности web-сайта) это означает, что серверные сценарии эксплуатируются (выполняются) существенно чаще числа обращений к стартовой странице. Для исполнения сценариев необходимы значения, которые вводятся пользователем в соответствующих формах на страницах web-сайта, которые являются значениями параметров, передаваемых сценариям. Такой набор статистических данных при известной структуре web-сайта может являться потенциально опасной ситуацией. Это требует более детального анализа сетевого трафика к защищаемым ресурсам или анализа исходного кода серверных сценариев на предмет эксплуатации уязвимостей межсайтового выполнения сценариев, SQL-инъекций и т.п. [8].

Возможным направлением дальнейших исследований может являться углубленный статистический анализ данных о посещаемости информационных ресурсов web-сайтов (обращении к ним) в зависимости от источников и пролонгированных во времени наборах обращений к web-сайтам с применением интеллектуальных методов анализа.

Заключение

Таким образом, в результате исследований представлен вариант алгоритмического обеспечения задачи построения структуры web-сайтов. Во взаимосвязи с описанием возможных сценариев компьютерных атак представленный подход способствует идентификации объектов эксплуатации угроз информационной безопасности в структуре web-сайтов, что в целом положительно влияет на их защищенность. Это, в свою очередь, положительно сказывается на стабильности функционирования, в частности web-сайтов государственных органов.

Литература

1. Федеральный закон от 09.02.2009 № 8-ФЗ «Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления».

2. Приказ Минэкономразвития России от 15.11.2022 № 624 «Об утверждении Требований к технологическим, программным и лингвистическим средствам обеспечения пользования официальными сайтами федеральных органов исполнительной власти и подведомственных им организаций».

3. Указ Президента РФ от 05.12.2016 № 646 «Об утверждении Доктрины информационной безопасности Российской Федерации».

4. Синев С.Г., Козачок В.И., Комашинский В.В., Жусов Д.Л. Модель фильтрации потока запросов к web-серверу // Безопасность информационных технологий, № 3, 2007. – С. 75 – 80.

5. Актуальные киберугрозы: IV квартал 2023 года – 2024. URL: https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2023-q4/ (дата обращения: 01.12.2024).

6. Результаты мониторинга официальных сайтов федеральных органов исполнительной власти – 2013. URL: http://svobodainfo.org/ru/ node/2527 (дата обращения: 01.12.2024).

7. Томас Х. Кормен, Чарльз И. Лейзерсон, Рональд П. Ривест, Клиффорд Штайн Алгоритмы: построение и анализ – 2-е изд. – М.: Вильямс, 2006. – 1296 с.

8. Низамутдинов М.Ф. Тактика защиты и нападения на web-приложения. – СПб.: БХВ-Петербург, 2005. – 432 с.: ил.

References

1. Federal'nyy zakon ot 09.02.2009 № 8-FZ «Ob obespechenii dostupa k informatsii o deyatel'nosti gosudarstvennykh organov i organov mestnogo samoupravleniya».

2. Prikaz Minekonomrazvitiya Rossii ot 15.11.2022 № 624 «Ob utverzhdenii Trebovaniy k tekhnologicheskim, programmnym i lingvisticheskim sredstvam obespecheniya pol'zovaniya ofitsial'nymi saytami federal'nykh organov ispolnitel'noy vlasti i podvedomstvennykh im organizatsiy».

3. Ukaz Prezidenta RF ot 05.12.2016 № 646 «Ob utverzhdenii Doktriny informatsionnoy bezopasnosti Rossiyskoy Federatsii».

4. Sinev S.G., Kozachok V.I., Komashinskiy V.V., Zhusov D.L. Model' fil'tratsii potoka zaprosov k web-serveru // Bezopasnost' informatsionnykh tekhnologiy, № 3, 2007. – S. 75 – 80.

5. Aktual'nyye kiberugrozy: IV kvartal 2023 goda – 2024. URL: https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2023-q4/ (data obrashcheniya: 01.12.2024).

6. Rezul'taty monitoringa ofitsial'nykh saytov federal'nykh organov ispolnitel'noy vlasti – 2013. URL: http://svobodainfo.org/ru/ node/2527 (data obrashcheniya: 01.12.2024).

7. Tomas KH. Kormen, Charl'z I. Leyzerson, Ronal'd P. Rivest, Klifford Shtayn Algoritmy: postroyeniye i analiz – 2-ye izd. – M.: Vil'yams, 2006. – 1296 s. 8. Nizamutdinov M.F. Taktika zashchity i napadeniya na web-prilozheniya. – SPb.: BKHV-Peterburg, 2005. – 432 s.: il.

Жусов Дмитрий Леонидович, кандидат технических наук, доцент, сотрудник федерального государственного казённого военного образовательного учреждения высшего образования «Академия Федеральной службы охраны Российской Федерации». 302020, г. Орёл, ул. Приборостроительная, д. 35. E-mail: d.zhusov@mail.ru.

Макеев Сергей Михайлович, кандидат технических наук, доцент кафедры «Защита информации ФГАОУ ВО «Южно-Уральский государственный университет (национальный исследовательский университет)». 454080, г. Челябинск, проспект Ленина, д. 76. E-mail: makeevsm@susu.ru.

Соколов Александр Николаевич, кандидат технических наук, доцент, заведующий кафедрой защиты информации ФГАОУ ВО «Южно-Уральский государственный университет (национальный исследовательский университет)». 454080, г. Челябинск, проспект Ленина, д. 76. E-mail: sokolovan@susu.ru

Zhusov Dmitry Leonidovich, Candidate of Technical Sciences, Associate Professor, employee of the Academy of the Federal Guard Service of the Russian Federation. 302020, Orel, Priborostroitelnaya St., 35. E-mail: d.zhusov@mail.ru.

Makeev Sergey Mikhailovich, Candidate of Technical Sciences, Associate Professor of Information Security Department, Federal State Autonomous Educational Institution of Higher Education "South Ural State University (national research university)". 454080, Chelyabinsk, Lenina avenue, 76. E-mail: makeevsm@susu.ru.

Sokolov Alexander Nikolayevich, Candidate of Technical Sciences, Associate Professor, Head of Department of Information Security, Federal State Autonomous Educational Institution of Higher Education "South Ural State University (national research university)". 454080, Chelyabinsk, Lenina avenue, 76. E-mail: sokolovan@susu.ru.