ISSN 1816-0301 (Print)
ISSN 2617-6963 (Online)

УДК 004.738.52
https://doi.org/10.37661/1816-0301-2020-17-3-72-77

Поступила в редакцию 01.04.2020
Received 01.04.2020

Принята к публикации 29.04.2020
Accepted 29.04.2020

Анализ защищенности веб-ресурсов на основе метрики CVSS

Ш. Р. Давлатов[1]✉, П. В. Кучинский[2]

[1]Белорусский государственный университет информатики и радиоэлектроники, Минск, Беларусь
✉E-mail: shohrukh.92@gmail.com

[2]Научно-исследовательское учреждение «Институт прикладных физических проблем имени А. Н. Севченко» Белорусского государственного университета, Минск, Беларусь

Аннотация. На основе анализа данных об уязвимостях веб-ресурсов и метрики CVSS (Common Vulnerability Scoring System) изучено распределение усредненной величины оценки по стандарту CVSS для расчета числового показателя уязвимости по десятибалльной шкале для сайтов Республики Беларусь. Проведена проверка гипотезы о распределении оценки уязвимостей CVSS по закону Пуассона методом критерия хи-квадрат. Установлено, что около 10 % веб-ресурсов из исходной генеральной выборки размером 19 000 имеют критическую усредненную оценку уязвимости. В рамках проведенного исследования создана универсальная система для сбора технической информации об активных веб-ресурсах в сети Интернет из общедоступных каталогов и реестров. Разработаны специальные шаблоны поиска с помощью RegExp-выражений языка программирования JavaScript для точного определения версий технологий, которые были использованы для создания веб-сайтов. На базе полученных данных установлены процентные соотношения используемых технологий, доменов верхнего уровня и географическое расположение серверов, которые обслуживают веб-ресурсы. Предлагаемая система может быть адаптирована под любые уникальные требования, необходимые специалистам по защите информации для проведения аудита безопасности веб-ресурсов.

Ключевые слова: информационная безопасность, оценка защищенности, веб-сайт, веб-сервер, метрика CVSS, язык программирования JavaScript

Для цитирования. Давлатов, Ш. Р. Анализ защищенности веб-ресурсов на основе метрики CVSS / Ш. Р. Давлатов, П. В. Кучинский // Информатика. – 2020. − Т. 17, № 3. – С. 72–77. https://doi.org/10.37661/1816-0301-2020-17-3-72-77

Web resource security analysis based on CVSS metrics

Shohrukh R. Davlatov[1]✉, Pyotr V. Kuchinsky[2]

[1]Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus
✉E-mail: shohrukh.92@gmail.com

[2]A. N. Sevchenko Institute of Applied Physical Problems of Belarusian State University, Minsk, Belarus

Abstract. Based on the analysis of vulnerability data for web resources and the CVSS metric, the distribution of the average CVSS (Common Vulnerability Scoring System standard for calculating a numerical vulnerability score on a ten-point scale) score for the websites of the Republic of Belarus was studied. The hypothesis on the distribution of the CVSS vulnerability assessment according to Poisson's law was tested by chi-square criteria. It was found that about 10% of web resources from the original general of samples of 19000 size have a critical averaged assessment level of vulnerability. As part of this work an universal system for collecting technical information about active web resources on the Internet from public directories and registries has been developed. Specific search templates have been developed using RegExp JavaScript expressions to detect the versions of

technologies that were used to create websites. Based on this data the percentage distribution of used technologies, top-level domains and the geographical location of the servers were calculated. Proposed system can be adapted to any unique conditions required by information security specialists to conduct a security audit of web resources.

Keywords: information security, security assessment, website, web server, CVSS metric, programming language JavaScript

For citation. Davlatov Sh. R., Kuchinsky P. V. Web resource security analysis based on CVSS metrics. *Informatics*, 2020, vol. 17, no. 3, pp. 72−77 (in Russian). https://doi.org/10.37661/1816-0301-2020-17-3-72-77

Введение. С развитием веб-технологий растет и число потенциальных уязвимостей в онлайн-ресурсах. Наличие большого набора инструментов для реализации угроз безопасности информации в сети Интернет определяет актуальность использования систем для анализа безопасности веб-ресурсов. Специалисты по защите информации широко применяют объективные количественные показатели защищенности, которые вычисляются на основе метрик открытой системы оценки CVSS [1]. Существуют открытые базы данных, в которых хранится информация об уязвимостях определенных версий технологий в формате *<название технологии, версия, оценка CVSS>*. Метрика CVSS предлагает простой инструментарий для расчета числового показателя по десятибалльной шкале, который позволяет оперативно принимать решения о том, как реагировать на ту или иную уязвимость [2]. Чем выше значение метрики, тем более оперативная реакция требуется. Например, язык программирования (ЯП) PHP версии 5.4.0 имеет 67 опубликованных уязвимостей в базе https://www.cvedetails.com со средней оценкой CVSS порядка семи.

В рамках настоящей работы была создана универсальная система для автоматического сбора из общедоступных каталогов и реестров технической информации об активных интернет-ресурсах Беларуси, такой как канонические имена доменов, IP-адреса, открытые порты, географическое расположение серверов, уязвимости технологий, версии операционных систем серверов и других программных обеспечений, которые были использованы для разработки и обслуживания веб-ресурсов. На основе полученных данных из открытых источников построено эмпирическое распределение оценок уязвимостей сайтов, предложена новая методика вычисления усредненной оценки уязвимостей веб-ресурсов на базе метрики CVSS, с помощью критерия хи-квадрат проверена гипотеза о том, что случайная величина X (усредненная оценка уязвимостей веб-ресурсов) распределена по закону Пуассона.

Сбор данных и методы исследования. Архитектура разработанной системы основана на шаблоне проектирования «цепочка обязанностей». Данный паттерн позволяет структурировать модули системы таким образом, чтобы запросы передавались последовательно по цепочке обработчиков. Каждый последующий модуль определяет возможность самостоятельной обработки запроса и необходимость ее последующей передачи по цепи. Первое звено цепочки состоит из модуля для получения и первичной обработки списка исходных доменов. С помощью разработанных NodeJS-скриптов автоматически была собрана информация о веб-ресурсах из открытых источников в сети Интернет (URL: https://www.shodan.io и https://censys.io). В результате процесса сканирования удалось собрать данные более 19 000 наиболее популярных веб-ресурсов Беларуси, которые были распределены по следующим категориям:

– работа, бизнес, компании, маркетинг, промышленность, банки и финансы;
– новости, веб-сервисы, онлайн-магазины, торговые площадки, объявления;
– семья, спорт, медицина, здоровье, отдых, досуг, культура и искусство;
– компьютерная техника, электроника, оргтехника, бытовые товары, авто- и мототехника;
– общество, политика, городские сайты, каталоги, образование и наука.

Для каждого отдельного домена параллельно запускается процесс для отправки GET-запроса и получения дополнительной технической информации: IP-адресов, открытых портов, географического расположения веб-серверов и заголовков ответов HTTP. Далее с помощью RegExp-выражений ЯП JavaScript определяются версии технологий, на базе которых были созданы веб-ресурсы. Для решения поставленной задачи достаточно получить исходный код страницы веб-сайта в формате HTML и заголовки ответов сервера [3]. С помощью Fetch API-интерфейса JavaScript были получены ответы на HTTP-запросы всех доменов исходной

выборки. Результаты запросов и заголовки ответов сервера были сохранены в локальной базе данных для последующих процессов обработки и анализа данных. Разработанный скрипт также может обнаруживать типы систем управления контентом, платформы электронной коммерции, версии веб-фреймворков, серверное программное обеспечение и аналитические инструменты. Последним этапом является проверка безопасности каждого веб-ресурса по отдельности на базе публичных программных интерфейсов API [4, 5]. В исследовании в качестве открытых баз данных об уязвимостях технологий были использованы общедоступные сервисы https://vulners.com и https://www.cvedetails.com, которые интегрировались в разрабатываемую систему с помощью API-интерфейса.

Результаты исследования и их обсуждение. На основе полученных данных о версиях технологий была создана диаграмма, показывающая процентное соотношение используемых для разработки веб-ресурсов технологий в семи категориях (рис. 1). Полученные результаты выявили, что наиболее популярными технологиями в своих категориях являются Windows Server – линейка серверных операционных систем от компании Microsoft (37 %), Nginx – веб-сервер и почтовый прокси-сервер (80 %), MooTools – модульный объектно-ориентированный JavaScript-фреймворк для разработки кроссбраузерных веб-приложений (46 %), jQuery – JavaScript-библиотека, фокусирующаяся на взаимодействии JavaScript и HTML (70 %), Google Analytics – сервис для создания детальной статистики посетителей веб-сайтов (44 %), Google AdSense – сервис контекстной рекламы (53 %), WordPress – система управления содержимым сайта с открытым исходным кодом, написанная на ЯП PHP (44 %).

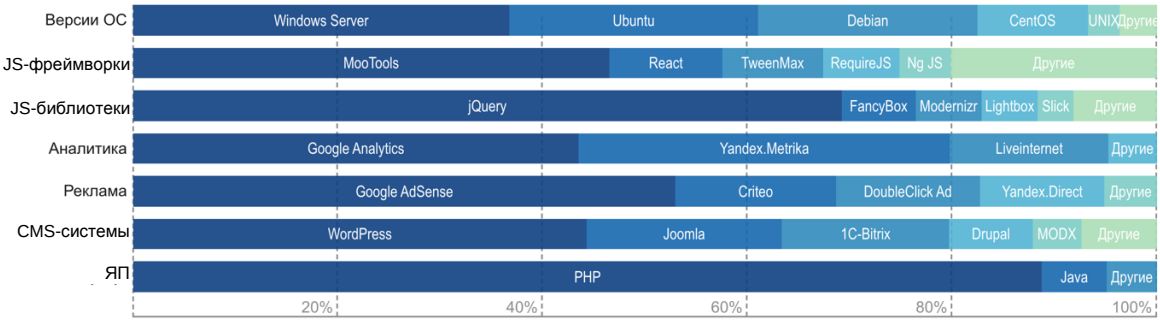

Рис. 1. Процентное соотношение используемых технологий

Данные о географических расположениях серверов, обслуживающих веб-ресурсы из исходной выборки, представлены в виде диаграммы (рис. 2). На горизонтальной оси расположены названия стран, где сосредоточено наибольшее количество веб-серверов: Беларусь, Россия, США, Германия и др. (Нидерланды, Польша, Украина и Великобритания). Каждый столбец показывает количество серверов, которые обслуживают веб-ресурсы определенной категории в той или иной стране. Как видно из рис. 2, примерно 68 % серверов находятся на территории Беларуси. Отметим, что категории веб-сайтов «работа», «бизнес», «новости», «онлайн-магазины», «объявления» являются наиболее популярными.

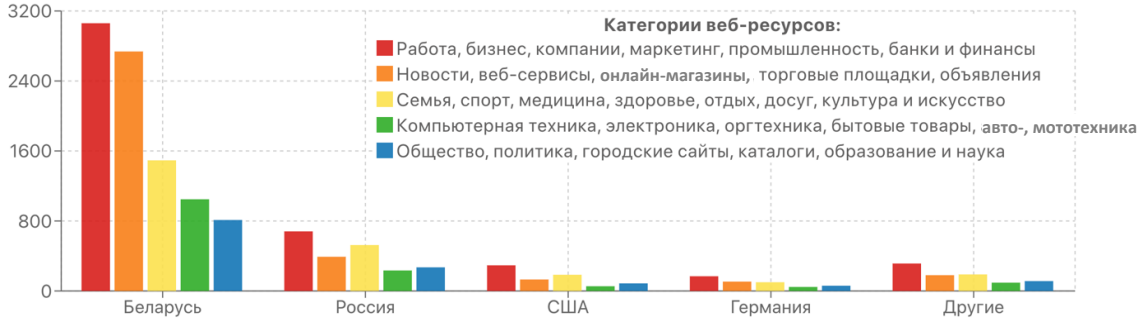

Рис. 2. Распределение веб-ресурсов по категориям и странам расположения серверов

Процентное соотношение доменов верхнего уровня показано на рис. 3 в виде круговой диаграммы. Данные были получены из исходной выборки веб-ресурсов путем приведения доменных имен к «каноническому» виду. Из рисунка видно, что домены BY составляют примерно 75 % из всех имеющихся записей в базе данных, а домены COM и RU – 8,8 и 8,2 % соответственно. Следует отметить, что все остальные домены (NET, БЕЛ, ORG и др.) были объединены в одну категорию с процентной долей 7,5 %.

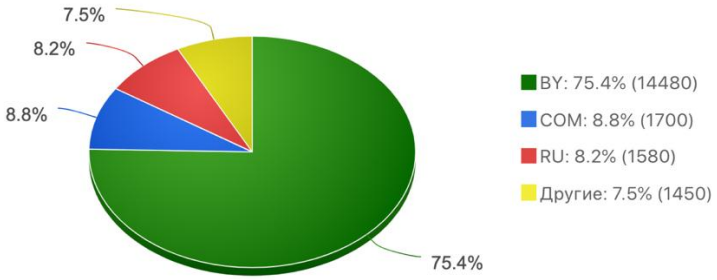

Рис. 3. Процентное соотношение доменов верхнего уровня

Рассмотрим предложенную методику оценки безопасности веб-ресурсов на базе метрики CVSS. Для каждого ресурса из исходной выборки используется усредненная оценка CVSS, рассчитанная по формуле

$$S_i = \frac{\left(\sum_{j=1}^{m} A_j\right)}{m}, 1 \le i \le n,$$

где n – количество элементов в выборке; m – количество распознанных версий технологий и ЯП для заданного веб-ресурса; A_j – оценка уязвимости определенной версии технологии.

Для вычисления значения A_j была выбрана функция максимума по всем известным оценкам CVSS:

$$A_j = \max_{1 \le k \le r} C_k,$$

где r – количество найденных уязвимостей для определенной версии технологии в общедоступной базе https://vulners.com. Целесообразность выбора функции максимума объясняется тем, что злоумышленники, как правило, направляют вектор атаки на наиболее уязвимые части системы. Следовательно, имеет смысл включать в общую усредненную оценку системы максимальные значения метрики CVSS по каждой из версий технологий. Для исследования распределения была создана случайная выборка из исходной базы данных веб-ресурсов, состоящая из N = 2000 элементов (около 10 % записей). На основе этих данных была построена диаграмма эмпирического распределения усредненной оценки уязвимости S_k веб-ресурсов (рис. 4).

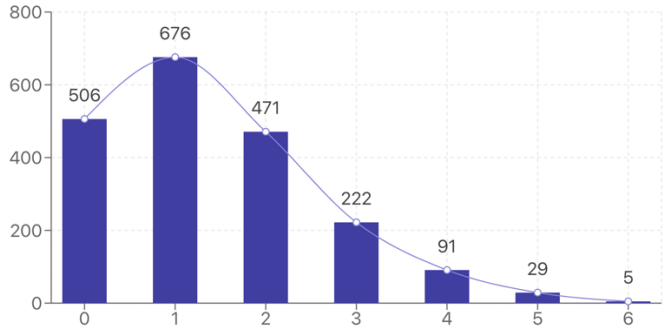

Рис. 4. Диаграмма эмпирического распределения оценок уязвимостей веб-ресурсов

По виду гистограммы распределения можно предположить, что оно подчиняется закону Пуассона. Заметим, что число наблюдений для интервала оценки S_6 равно пяти, поэтому имеет смысл объединить его с предыдущим интервалом S_5. Далее с помощью критерия хи-квадрат при уровне значимости α = 0,05 проверяем гипотезу о том, что случайная величина X (усредненная оценка CVSS уязвимости веб-ресурса) распределена по закону Пуассона. Данное распределение имеет единственный параметр λ, являющийся математическим ожиданием случайной величины X. Заменяем параметр λ наилучшей оценкой по выборке – выборочной средней: $\lambda = \frac{(\sum S_k O_k)}{N} \approx 1{,}41$.

Проверяем нулевую гипотезу H_0 (средняя оценка уязвимости веб-ресурса распределена по закону Пуассона) при уровне значимости α = 0,05, т. е. $p_k = P(X = k) = \frac{(\lambda^k e^{-\lambda})}{k!}$, где $1 \le k \le 6$ (максимальное значение усредненной оценки CVSS в рассматриваемой выборке). Находим теоретические вероятности p_k и теоретические частоты $E_k = p_k \times N = p_k \times 2000$, результаты вычислений заносим в таблицу.

Эмпирические и теоретические частоты

k	Оценки CVSS, S_k	Эмпирические частоты, O_k	Теоретические частоты, E_k	$\frac{(O_k - E_k)^2}{E_k}$
1	0	506	488,287	0,643
2	1	676	688,484	0,226
3	2	471	485,381	0,426
4	3	222	228,129	0,165
5	4	91	80,416	1,393
6	5–6	34	22,677	5,654
Σ		2000	1998,695	χ^2= 8,506

Из расчетной таблицы находим наблюдаемое значение критерия Пирсона: $\chi^2 = \sum \frac{(O_k - E_k)^2}{E_k} \approx$ ≈ 8,5. Так как новое число интервалов (с учетом объединения) m = 6, а закон распределения Пуассона определяется параметром r = 1, то число степеней свободы $q = m - r - 1 = 4$. Соответствующее критическое значение статистики χ^2 для уровня значимости 0,05 при количестве степеней свободы q = 4 равно 9,5. В связи с тем что наблюдаемое значение критерия 8,5 меньше критического значения 9,5 (по таблице критических значений распределения хи-квадрат), следует принять нулевую гипотезу о распределении генеральной совокупности по закону Пуассона.

Отметим, что в исходной генеральной совокупности почти 90 % веб-ресурсов имеют значение усредненной оценки CVSS в интервале [0, 5]. Было выявлено, что остальные 10 % ресурсов имеют высокую критическую оценку уязвимости из-за использования устаревших версий таких технологий, как CMS WordPress, ЯП PHP, веб-сервер nginx и JavaScript-библиотека jQuery.

Заключение. В работе представлена методика оценки безопасности веб-ресурсов на базе метрики CVSS. Установлено, что около 10 % веб-ресурсов Беларуси из исходной генеральной выборки (19 000 доменов) имеют критическую усредненную оценку уязвимости. На основе полученных данных об уязвимостях построено эмпирическое распределение оценки CVSS. С помощью критерия хи-квадрат проверена гипотеза о том, что данное распределение подчиняется закону Пуассона. В рамках проведенного исследования были также разработаны RegExp-выражения на ЯП JavaScript для точного определения версий технологий, которые были использованы для создания веб-сайтов. Установлены процентные соотношения применяемых технологий, доменов верхнего уровня и географическое расположение серверов, которые обслуживают данные веб-ресурсы.

Список использованных источников

1. Дойникова, Е. В. Оценка защищенности компьютерных сетей на основе метрик CVSS / Е. В. Дойникова, А. А. Чечулин, И. В. Котенко // Информационно-управляющие системы. – 2019. – № 6. – С. 76–87. https://doi.org/10.15217/issn1684-8853.2017.6.76
2. Li, H. Study on the distribution of CVSS environmental score / H. Li, R. Xi, L. Zhao // 5th Intern. Conf. on Electronics Information and Emergency Communication, Beijing, China, 14–16 May 2015. – Beijing, 2015. – P. 1–4. https://doi.org/10.1109/ICEIEC.2015.7284502
3. Exploring the intersections of web science and accessibility [Electronic resource] / T. Bostic [et al.] // The MITRE Corporation Scientific Journal. – 2019. – Mode of access: https://arxiv.org/abs/1908.02804. – Date of access: 12.03.2020.
4. Likarish, P. A targeted web crawling for building malicious javascript collection / P. Likarish, E. Jung // Proc. of the ACM First Intern. Workshop on Data-Intensive Software Management and Mining, Hong Kong, China, Nov. 2009. – Hong Kong, 2009. – P. 23–26. https://doi.org/10.1145/1651309.165131
5. A quantitative evaluation model for network security / D. Man [et al.] // Proc. of the 2007 Intern. Conf. on Computational Intelligence and Security, Harbin, China, 15–19 Dec. 2007. – Harbin, 2007. – P. 773–777.

References

1. Dojnikova E. V., Chechulin A. A., Kotenko I. V. Ocenka zashhishhennosti komp'juternyh setej na osnove metrik CVSS [Security assessment of computer networks based on CVSS metrics]. Informacionno-upravljajushhie sistemy [*Management Information Systems*], 2019, no. 6, pp. 76–87 (in Russian). https://doi.org/10.15217/issn1684-8853.2017.6.76
2. Li H., Xi R., Zhao L. Study on the distribution of CVSS environmental score. *5th International Conference on Electronics Information and Emergency Communication, Beijing, China, 14–16 May 2015*. Beijing, 2015, pp. 1–4. https://doi.org/10.1109/ICEIEC.2015.7284502
3. Bostic T., Stanley J., Higgins J., Chudnov D., Montgomery B., Brunell J. Exploring the intersections of web science and accessibility. *The MITRE Corporation Scientific Journal*, 2019. Available at: https://arxiv.org/abs/1908.02804 (accessed 12.03.2020).
4. Likarish P., Jung E. A targeted web crawling for building malicious javascript collection. *Proceedings of the ACM First International Workshop on Data-Intensive Software Management and Mining, Hong Kong, China, November 2009*. Hong Kong, 2009, pp. 23–26. https://doi.org/10.1145/1651309.165131
5. Man D., Yang W., Yang Y., Wang W., Zhang L. A quantitative evaluation model for network security. *Proceedings of the 2007 International Conference on Computational Intelligence and Security, Harbin, China, 15–19 December 2007*. Harbin, 2007, pp. 773–777.

Информация об авторах

Давлатов Шохрух Рустамович, аспирант кафедры защиты информации, Белорусский государственный университет информатики и радиоэлектроники, Минск, Беларусь.
E-mail: shohrukh.92@gmail.com

Кучинский Петр Васильевич, доктор физико-математических наук, директор, научно-исследовательское учреждение «Институт прикладных физических проблем имени А. Н. Севченко» Белорусского государственного университета, Минск, Беларусь.

Information about the authors

Shohrukh R. Davlatov, Postgraduate Student of the Department Information Security, Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus.
E-mail: shohrukh.92@gmail.com

Pyotr V. Kuchinsky, Dr. Sci. (Phys.-Math.), Director, A. N. Sevchenko Institute of Applied Physical Problems of Belarusian State University, Minsk, Belarus.