

ANALYSIS OF WEBSITE SECURITY STATUS BASED ON PERFORMANCE METRICS

D.T. Dim¹, *Postgraduate Student, dim.dike@yahoo.com*

V.N. Bogatkov¹, *Dr.Sc. (Engineering), Professor, vnbgtk@mail.ru*

¹*Tver State Technical University, Nikitin Quay 22, Tver, 170026, Russian Federation*

Abstract. The paper is relevant due to the constant evolution of Internet infrastructure including conditions for optimal operation of web systems, information sources, e-commerce and various types of services. During the past twenty years, general information in the World Wide Web (WWW) and the number of users has increased almost two hundredfold. With this growth, requirements for technology has become even more critical. However, current research is more about a commercial aspect than the technical one.

The paper pays special attention to assessing website performance quality regarding information technology indicators. It also considers a fuzzy Markov chain to define specific website productivity states.

Keywords: website, metrics, website analysis, website testing, Markov chain.

In design, development and operation of a website as a resource, the main criteria and indicators (so-called metrics) are used in assessing its quality and effectiveness [1, 2]. The issues related to cross-platform, presentation and usability of an interface are not considered.

The following metrics were selected to assess website performance:

- time to the first byte (TTFB, ms);
- domain name server (DNS) lookup time (ms);
- universal resource locator (URL) redirection;
- number of HTTP requests;
- page size (kb);
- connection time (ms).

TTFB. In web applications and websites, there are different sub-categories of latency. TTFB is the time required to receive the first byte from the server after sending HTTP GET request.

DNS lookup time. Like TTFB, DNS lookup refers to a sub-category of latency and reflects the time required to look up the IP address of the corresponding domain.

URL redirection. Automatic transition by a website from one page to another after user's GET request. For example, visiting google.com on a device in the Russian Federation without tunneling defaults to google.ru.

The number of HTTP requests. This is a number of objects required when loading websites. Naturally, higher number of requests results in longer load times.

Page size. The cumulative page size with images, animations, drawings, style sheets, scripts and html/htm code.

Connection time (ms). Another component of latency, which reflects the time required to establish a TCP connection.

Managing these metrics will lead to better site performance, since they belong to a server and network environment of websites. To analyze performance of websites, first, it is necessary to collect data.

There are two main types of data collection tools [3].

1. **Counters.** A small code embedded in webpages loaded by a browser. Most prominent solutions include OpenStat, GoogleAnalytics, Yandex.metrics, LiveInternet and others.

2. **Log analyzers.** They provide the ability to collect statistical data and compose their own special reports through installation on a server.

Counters [4–8] were primarily used to collect data for selected metrics, which further characterizes different website states.

Analysis of website states

The analysis of website security states was based on the following assumptions:

- absence of change in core content;
- the probability of a transition from one state to another depends solely on the current state, and not the previous one;
- the probability of a transition depends on a change in parameter or a set of parameters.

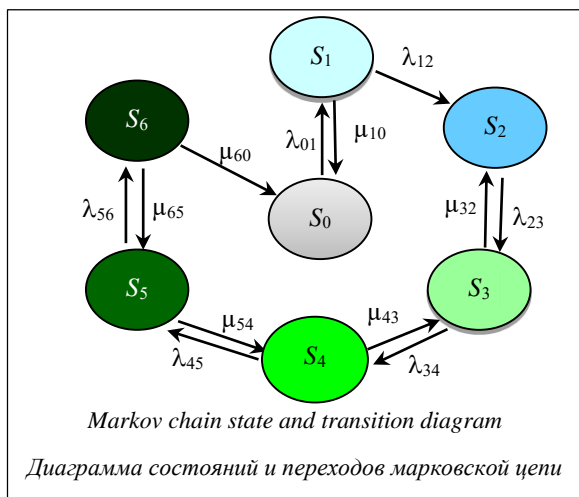
The state of the system at any given time is described by specifying its coordinates. Knowing these values at a given time t , we can determine evolution of the system under the influence of internal and external factors in subsequent time periods.

Each website productivity state is denoted by S_i , where $i = 0 \sim 6$.

$S_0 \sim S_6$. System states determined based on selected metrics. The transition between states occurs due to change (improvement or deterioration) in hardware, software, network, their interaction with each other and the environment (as shown in a figure). The most important part in their interaction is how they directly affect website performance.

Characteristics of different website states are shown in table 1:

S_0 – Inoperative state. Characterized by high DNS lookup time exceeding 1 200 ms, page size >7 168 kb, redirects >7, TTFB exceeding 3 000 ms, number of HTTP requests >180 and the connection time exceeding 5 500 ms. This state is usually accompanied by



HTTP error codes (including but not limited to 404, 502), unavailability of the requested link or page, and expiration of waiting time.

S_1 – first improvement stage. This state corresponds to S_0 with exceptions to connection time and DNS lookup time, which are $1\ 500\text{ ms} < T_{CON} < 5\ 500\text{ ms}$ and $880 < T_{DNS} < 1\ 200\text{ ms}$ respectively.

S_2 – second improvement stage. This state adheres to S_1 except for TTFB and the number of HTTP requests, which are $780 < D < 900\text{ ms}$ and $62 < R_q < 69$ respectively.

S_3 – third improvement stage. This state corresponds to S_2 except for DNS lookup time, which is $160 < T_{DNS} < 380\text{ ms}$.

S_4 – satisfactory functional stage. This state corresponds to S_3 except for the following parameters: DNS lookup time $< 380\text{ ms}$; page size $350 < P < 700\text{ kb}$; TTFB delay is below 280 ms.

S_5 – operational state. This state corresponds to S_4 except for the following parameters: page size $180 < P < 350\text{ kb}$, redirects < 7 and connection below 286 ms.

S_6 – optimal working state of websites. This state corresponds to S_5 with exceptions of the following metrics: DNS lookup time is below 200 ms, redirects < 3 , page size $< 160\text{ kb}$, TTFB delay is below 180 ms and the connection time $< 250\text{ ms}$.

It should be noted that the majority of existing websites on the WWW are located in states $S_3 \sim S_5$ [1].

Consequently, the states are characterized as showed in table 2. Table 3 shows the indicators of existing websites.

Measures for managing website security

The main problem in ensuring proper functioning of websites is in the implementation of a set of measures aimed at maintaining operability, sustainability and development potential of websites. One of the

Identification of website states

Table 1

Идентификация состояния web-сайтов

Таблица 1

| S_i | Time to First Byte TTFB (D) | DNS Lookup (T_{DNS}) | Page size (P) | Redirects (R_d) | HTTP-requests (R_q) | Connection time (T_{CON}) |
|-------|-----------------------------|--------------------------|---------------|---------------------|-------------------------|-------------------------------|
| S_0 | D_1 | T_{DNS1} | P_1 | R_{d1} | R_{q1} | T_{CON1} |
| | D_2 | T_{DNS2} | P_2 | R_{d2} | R_{q2} | T_{CON2} |
| | D_3 | T_{DNS3} | P_3 | R_{d3} | R_{q3} | T_{CON3} |
| S_1 | D_1 | T_{DNS1} | P_1 | R_{d1} | R_{q1} | T_{CON1} |
| | D_2 | T_{DNS2} | P_2 | R_{d2} | R_{q2} | T_{CON2} |
| | D_3 | T_{DNS3} | P_3 | R_{d3} | R_{q3} | T_{CON3} |
| S_2 | D_1 | T_{DNS1} | P_1 | R_{d1} | R_{q1} | T_{CON1} |
| | D_2 | T_{DNS2} | P_2 | R_{d2} | R_{q2} | T_{CON2} |
| | D_3 | T_{DNS3} | P_3 | R_{d3} | R_{q3} | T_{CON3} |
| S_3 | D_1 | T_{DNS1} | P_1 | R_{d1} | R_{q1} | T_{CON1} |
| | D_2 | T_{DNS2} | P_2 | R_{d2} | R_{q2} | T_{CON2} |
| | D_3 | T_{DNS3} | P_3 | R_{d3} | R_{q3} | T_{CON3} |
| S_4 | D_1 | T_{DNS1} | P_1 | R_{d1} | R_{q1} | T_{CON1} |
| | D_2 | T_{DNS2} | P_2 | R_{d2} | R_{q2} | T_{CON2} |
| | D_3 | T_{DNS3} | P_3 | R_{d3} | R_{q3} | T_{CON3} |
| S_5 | D_1 | T_{DNS1} | P_1 | R_{d1} | R_{q1} | T_{CON1} |
| | D_2 | T_{DNS2} | P_2 | R_{d2} | R_{q2} | T_{CON2} |
| | D_3 | T_{DNS3} | P_3 | R_{d3} | R_{q3} | T_{CON3} |
| S_6 | D_1 | T_{DNS1} | P_1 | R_{d1} | R_{q1} | T_{CON1} |
| | D_2 | T_{DNS2} | P_2 | R_{d2} | R_{q2} | T_{CON2} |
| | D_3 | T_{DNS3} | P_3 | R_{d3} | R_{q3} | T_{CON3} |

Where D_1, D_2, D_3 are small, medium, large TTFB; $T_{DNS1}, T_{DNS2}, T_{DNS3}$ – small, medium, large DNS lookup time; P_1, P_2, P_3 – small, medium, large page sizes; R_{d1}, R_{d2}, R_{d3} – small, medium, large number of redirects; R_{q1}, R_{q2}, R_{q3} – small, medium, large number of HTTP-request; $T_{CON1}, T_{CON2}, T_{CON3}$ – small, medium, large connection times

Table 2
Classification of website states

Таблица 2
Классификация состояния web-сайтов

| S_i | Tuple of website states | States |
|-------|--|--------------|
| S_0 | $\langle D_3, T_{DNS3}, P_3, R_{d3}, R_{q3}, T_{CON3} \rangle$ | Inoperable |
| S_1 | $\langle D_3, T_{DNS2}, P_3, R_{d3}, R_{q3}, T_{CON2} \rangle$ | |
| S_2 | $\langle D_2, T_{DNS2}, P_3, R_{d3}, R_{q2}, T_{CON2} \rangle$ | Average |
| S_3 | $\langle D_2, T_{DNS2}, P_3, R_{d3}, R_{q2}, T_{CON2} \rangle$ | |
| S_4 | $\langle D_1, T_{DNS1}, P_2, R_{d3}, R_{q2}, T_{CON2} \rangle$ | Satisfactory |
| S_5 | $\langle D_1, T_{DNS1}, P_1, R_{d2}, R_{q2}, T_{CON1} \rangle$ | |
| S_6 | $\langle D_1, T_{DNS1}, P_1, R_{d1}, R_{q1}, T_{CON1} \rangle$ | Optimal |

Table 3
Indicators of existing websites

Таблица 3
Показатели существующих web-сайтов

| Websites | $D(ms)^*$ | $T_{CON}(ms)$ | $P(kb)$ | Rd | Rq | $T_{DNS}(ms)$ | Measuring tools |
|---------------|------------|---------------|---------|------|------|---------------|---|
| facebook.com | 26 (2.9) | 121 | 377 | 0 | 27 | 8 | GtMetrix https://gtmetrix.com Webpage test http://www.webpagetest.org Google Developer tools (Chrome) Ultra-tools https://www.ultratools.com/ Pingdom https://tools.pingdom.com/ |
| wordpress.org | 229 (2) | 184 | 3072 | 0 | 46 | 1 | |
| wikipedia.org | 48 (0.9) | 23 | 108 | 0 | 9 | 35 | |
| mail.ru | 215 (6.2) | 70 | 622 | 3 | 146 | 133 | |
| narod.ru | 18 (6.3) | 42 | 895 | 0 | 43 | 2 | |
| yandex.ru | 53 (5.4) | 30 | 467 | 0 | 35 | 85 | |
| rakuten.co.jp | 275 (21.3) | 262 | 7520 | 0 | 803 | 175 | |
| dangote.com | 1 (15) | 141 | 1024 | 0 | 39 | 2 | |
| tehetem.com | 2 (14.5) | 30 | 352 | 0 | 16 | 15 | |

* – values in brackets denotes complete load times in seconds as compared to the TTFB in milliseconds.

most important measures is monitoring, which enables not only the ability to keep track of all processes on websites, but also to prevent various safety and efficiency threats in a timely manner. Amongst the probable criteria to monitor, the following might be identified as key areas [9]:

- DNS lookup time;
- response time;
- scheduled task execution;
- wait times for static files;
- databases and their connections.

In this case, each area contains a specific set of metrics with each composition significantly different from the other.

Conclusion

Monitoring and forecasting are the most important parts of managing a complex security process. The technological, technical and economic consequences of information technology development, maximum use of computational and financial resources depends significantly on them. In order to control a process effectively, it is necessary to detect and prevent crisis situa-

tions, which thereby ensures effective safety management. This requires the existence of unified identifiable metrics and states of an information system, as well as an accurate definition of the main threats and the subsequent development of measures to eliminate them.

References

1. Souders S. *High performance web sites: essential knowledge for front-end engineers*. O'Reilly Media Publ., 2007, 170 p.
2. King A.B. *Website Optimization: Speed, Search Engine & Conversion Rate Secrets*. O'Reilly Media Publ., 2008, 398 p.
3. Kaushik A. *Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity*. Sybex Publ., 2009 (Russ. ed.: Vilyams Publ., 2014, 528 p.).
4. *GtMetrix*. Available at: <https://gtmetrix.com> (accessed May 18, 2017).
5. *Webpage test*. Available at: <http://www.webpagetest.org> (accessed May 18, 2017).
6. *Ultra-tools*. Available at: <https://www.ultratools.com/> (accessed May 18, 2017).
7. *Pingdom*. Available at: <https://tools.pingdom.com/> (accessed May 18, 2017).
8. *Google Developer tools in chrome browser* (accessed May 18, 2017).
9. Croll A., Power S. *Complete web monitoring*. O'Reilly Media Publ., 2009, 672 p.
10. *Internet growth statistics*. Available at: <http://www.internetworldstats.com/emarketing.htm> (accessed May 2, 2017).

УДК 004.56
DOI: 10.15827/0236-235X.120.654-657

Дата подачи статьи: 15.06.17
2017. Т. 30. № 4. С. 654–657

АНАЛИЗ СОСТОЯНИЯ БЕЗОПАСНОСТИ WEB-САЙТОВ НА ОСНОВЕ МЕТРИК

Д.Т. Дим, аспирант, dim.dike@yahoo.com;

В.Н. Богатиков, д.т.н., профессор, vnbgtk@mail.ru

(Тверской государственный технический университет, наб. Аф. Никитина, 22, г. Тверь, 170026, Россия)

Актуальность работы заключается в анализе современных трендов развития web-сайтов и интернет-инфраструктуры, а также в отражении условий для оптимальной работы систем, представляющих информацию, выполняющих функции электронной коммерции и предоставляющих различные виды услуг. В последние два десятилетия общая информация во всемирной паутине WWW и количество пользователей увеличились почти в двести раз. С их ростом требования к технике стали еще критичнее, но оценке с коммерческой точки зрения уделяется больше внимания, чем с технической.

В работе особое внимание уделяется оценке качества работы web-сайтов с точки зрения информационно-технологических показателей, рассмотрена нечеткая цепь Маркова с выделением конкретных состояний производительности web-сайтов.

Ключевые слова: *web-сайт, метрика, характеристики сайта, анализ сайта, цепь Маркова.*

Литература

1. Souders S. High performance web sites: essential knowledge for front-end engineers. O'Reilly Media Publ., 2007, 170 p.
2. King A.B. Website optimization: speed, search engine & conversion rate secrets. O'Reilly Media Publ., 2008, 398 p.
3. Кошик А. Веб-аналитика 2.0 на практике. Тонкости и лучшие методики. М.: Вильямс, 2014. 528 с.
4. GtMetrix. URL: <https://gtmetrix.com> (дата обращения: 18.05.2017).
5. Webpage test. URL: <http://www.webpagetest.org> (дата обращения: 18.05.2017).
6. Ultra-tools. URL: <https://www.ultratools.com/> (дата обращения: 18.05.2017).
7. Pingdom. URL: <https://tools.pingdom.com/> (дата обращения: 18.05.2017).
8. Google Developer tools в браузере Chrome (дата обращения: 18.05.2017).
9. Croll A., Power S. Complete Web Monitoring. 2009, 672 p.
10. Internet growth statistics. URL: <http://www.internetworldstats.com/emarketing.htm> (дата обращения: 05.02.2017).

Примеры библиографического описания статьи

1. Dim D.T., Bogatikov V.N. Analysis of website security status based on performance metrics // Программные продукты и системы. 2017. Т. 30. № 4. С. 654–657 (англ.). DOI: 10.15827/0236-235X.120.654-657.
2. Dim D.T., Bogatikov V.N. Analysis of website security status based on performance metrics. *Программные продукты и системы* [Software & Systems]. 2017, vol. 30, no. 4, pp. 654–657. DOI: 10.15827/0236-235X.120.654-657.