

Estimating the state of website security as poorly regulated mechanisms based on fuzzy logic methods

Yasir Ali Matnee

Department of Computer Science, Basic Education College, University of
Diyala, IRAQ
ee22a12@gmail.com

Recived : 3\5\2018

Revised : 23\5\2018

Accepted : 24\5\2018

Available online : 7/8/2018

DOI: 10.29304/jqcm.2018.10.3.413

Abstract

The research is devoted to the analysis of page loading time, which is an important indicator for any web-site. As a rule, the sites are hosted on web-servers with certain characteristics. The users interact with the environment is largely aggressive and uncertain (external threats such as penetration, denial of service, the introduction of code into the language of structured SQL queries, etc.). It should be noted that there are the uncertainties generated by hardware and software are also important. Any site can be sub- is influenced by the influence of the external environment and various subsystems of servicing the functions of web-sites, which leads to occurrence of contingencies and generates uncertainty of his work. The degree of uncertainty is cannot always be assessed on the basis of statistical material alone. This leads to an increase in the number of methods and means of intellectualizing the performance of evaluations on the basis of methods of artificial intelligence, and methods based on the use of fuzzy estimates. The assessment of the security status of Internet sites as poorly formalized objects on the basis of fuzzy logic methods considered in the article is implemented in the form of technology for assessing the states of reliability of a web site. The proposed approach allows a more flexible adaptation to the particular problem and allows diagnosis object already in the step of calculating an integral index of reliability. Flexibility is achieved due to the fact that the state of reliability can be estimated At once on several reliability indicators.

Keywords— fuzzy logic, methods of system analysis, metrics, analysis and features of a web-site.

Introduction

Currently, information systems (IS) are actively developing, including the intellectual ones. An important link in the development of information technologies are web-sites, which number is more than 4 billion [1-3], and their number is growing rapidly. Therefore, the analysis of the quality and reliability of the created and operated programs, the underlying requirements for the work of web-sites are special requirements. Web-sites depending on the location of elements (pages, sections, navigation) and their relationship with each other can have a different structure [4]. The key ones are linear, hierarchical and lattice. In the design, development and operation of a web site as an information resource, the main criteria and indicators evaluation of its quality, reliability and efficiency.

Description of the metrics of websites

The selection of the best characteristics, which are included in the basis of the web application, implies the use of such a thing as "metrics". Basically, the metric represents a numerical characteristic of the system. When these characteristics are formed, various factors

are taken into account [6]. The most common ones are the mean time between failures MTBF (ms), access (%), delay (ms), channel bandwidth (Kb), time to the first byte TTFB (ms), DNS domain server search time (ms), redirection of the universal URL resource pointer, the number of HTTP requests, the size of the main page (Kb), the connection time (ms). There are three types of metrics for assessing web-sites: server, user and network.

Server metrics allow you to determine the system resources used and the possibility of resource conflicts. These metrics are aimed at tracking the resources of the machine level, such as network, memory, processor and disk utilization. They give an idea of the internal conflicts that underlie the computer. Allocate the hardware and software metrics of the server systems (see table 1 and table 2).

Table (1) Hardware metrics of server appliances

Metrics	Calculation / Range	Comment
Latency	-	Waiting time before sending data or the time at the beginning of the transfer. The shorter portion of the transmitted data, the more frequently occurs latency. The transmission speed is characterized by the maximum channel capacity at large portions of the transmitted data, while the costs for latency are reduced
Mean time between failures MTBF	$\frac{Tu}{NE}$, where TU is the total working hours; NE - number of failures	The metric of the work of equipment, set by the manufacturer. Due to the reliable operation of modern computer equipment, this metric is missing from some manufacturers or is given as a lifetime warranty
Network tracking indicator	$(\frac{Td}{T})$ where Td is the operating time; T - total time	Describes the time of the system. Similar to the MTBF metric, but only provides network maintenance problems
Site volume	-	Number in Kb
Number of I / O operations	SD * ND, where SD is the I / O disk speed, ND is the number of disks	
Time to first byte TTFB	TFB	The required wait time before the first byte of the requested resource arrives from the server after sending the HTTPGET request

Table (2) Software metrics of server systems

Metrics	Calculation / Range	Comment
Number of threads	NTh	Multithreading
Replication delay	Tp	Parameter - time in ms

You can track metrics to determine performance and reliability aspects: It is necessary to be aware of the interdependence between the system indicators and the application load. Probably, the system will need additional hardware resources (real or virtual). In the case of constant load, data values increase metrics. This may be due to external causes: background tasks, constantly running tasks, network activity or input / output (I / O) devices.

WebPageTest.org defines TTFB as the browser wait time until the first byte of the requested resource is received, which begins after the DNS lookup time and connection time [7]. Some sources combine DNS time, connection time, and latency in the TTFB metric, since TTFB represents the amount of time required to respond to the server and create a web page.

Figure (1) shows the TTFB value of 124 ms, which is essentially perfect. Typically, the optimal TTFB should be in the range of 5-180 ms. It is considered that the web page is slow if it has a large TTFB, since the start time of the display will be delayed. This is a form of feedback used to evaluate web sites for effectiveness research. It should be noted that (uodiyala.edu) loads most of the components, such as Java Script JS, CSS style sheets, Flash animation, not immediately, so the website's delay is minimal. The delays of the web page are mentioned in different sources [2, 6]. Metrics (Figure 1) are as follows:

- * DNS lookup time- the time to search for the IP address for the corresponding domain;
- * Connection time - the time required to establish a TCP connection;
- * Timeout - waiting time until the first byte is received after the connection is established;
- * Content downloads time - the time it takes to download the entire object.

					Document Complete			Fully Loaded			
Load Time	First Byte	Start Render	User Time	Speed Index	Time	Requests	Bytes In	Time	Requests	Bytes In	Cost
120.050s	0.277s	2.193s	0.526s	29970	-	0	9,169 KB	120.050s	107	9,169 KB	\$\$\$\$

URL: <http://www.uodiyala.edu.iq/>
Host: www.uodiyala.edu.iq
IP: 94.228.39.10
Error/Status Code: 200
Priority: Very High
Client Port: 65115
Request Start: 0.135 s
DNS Lookup: 27 ms
Initial Connection: 107 ms
Time to First Byte: 124 ms
Content Download: 479 ms
Bytes In (downloaded): 47.6 KB
Uncompressed Size: 47.2 KB
Bytes Out (uploaded): 0.4 KB

Figure (1) Key metrics for web page delay

Table (3): Hardware metrics of user systems

Metrics	Calculation / Range	Comment
Requests per second RPS	$\left(\frac{P}{P_0}\right) * T_Z^{-1}$	P - memory, PO - RAM, TZ - preset time
The amount of memory for programs	-	Measured in MB

Table (4): Software metrics of user systems

Metrics	Calculation / Range	Comment
Bandwidth of the channel	<u>RWIN</u> RTT	RTT - Round-Trip -Time, RWIN - TCP receive window
Simultaneous Users	-	Number
Application response time	100~300	Integral performance of IP in terms of user [8]. The amount of time between the appearance of the user's request to the application and the receipt of a response to the request. Depends on the type of user request, from which user and to which server it is accessed, from the current state of the network elements and the settings of the operating systems and DBMS [9]
Download time	1-10	The time it takes to fully load a web page with a browser. Measured in seconds

Stages of loading a web page: request→ forwarding→ searching in the DNS cache→ TCP→

receipt→ processing→ reply→ download.

TTFB can be from 1-5 sec to 100-200 ms, but the page loads much faster and will be ready for use in a shorter time. Many web-sites are seeing a general increase TTFB 5-10 times. There are also some disadvantages to Gzip compression:

- * increases the total server load during compression;
- * Data processing can take a long time, since the first byte is not sent until the compression is completed;
- * A large TTFB often causes the user to re-create the current request to the web server, which increases the overall load and the required resources due to consecutive requests.

Network metrics are associated with the emergence of network problems, which are accompanied by a decrease in productivity. Network delays lead to an increase in the duration of query execution (Table 5).

The application response time, as a rule, is formed from the time.

- * Preparation of user requests;
- * Transfer of requests between the user and the server through network segments and intermediate communication equipment;
- * Processing requests on the server and sending responses to the user;
- * processing the responses received on the user's device.

To determine the optimal performance of IP in order to determine which metrics are determining, a service level agreement must be drawn up between the various services.

Web pages are often compressed in the Gzip format to reduce the size of the downloaded file, which prevents sending the first byte until compression will not be completed, and greatly increases TTFB.

Table (5): Website network metrics

Metric	Formula / Range	Comment
Network Latency	$\sum(tp + tl + Q(t))$, where tp is the packet delay; tl - propagation delay; Qt - delay in queue	The performance of the network is extremely important for cloud applications, as it is a conductor through which all information passes
Bandwidth of the channel	$\frac{1.26 * MSS}{RTT * \sqrt{L}}$, where RTT - Round-Trip-Time; MSS - segment size; L - loss of frames	
Packet Throughput PPS	$\frac{Rwin}{RTT}$,where RTT - Round-Trip-Time; Rwin - TCP Receive Window	Reflects the number of frames transmitted per time unit . It gives an opportunity to assess whether the equipment copes with the load and whether its performance corresponds to the declared
Frame loss	$\frac{8}{3W^2}$ where W is the segment load	
Network availability	$\frac{MTBF}{MTBF + MTTR}$, where MTBF is the mean time between failures; MTTR - Mean Time To Repair	It is used to assess the reliability and stability of the network. Displays the time that the network is functioning without fail or need to reboot for administrative or maintenance purposes
Bandwidth	$(F_{max} - F_{min})$, where F_{max} -Max. frequency; and F_{min} -min. frequency	
Response time	$\frac{B_{max}}{F}$, where B_{max} - the maximum bandwidth; F is the number of flows	The average speed of the full load of the pages of the website. Use a weighted average score for users, servers, and day periods. Time in seconds

Response time is an important indicator, first of all, for any visitor to the site. Therefore, it is important for the site owner to download the main page of the website. Many users do not have enough speed to quickly download large portals. Waiting for the page to load fully should not exceed 5-10 seconds. To date, for example, MS Windows does not have performance counters to measure the latency of individual application requests.

However, there is a "Resource Monitoring" which is an excellent tool for analyzing network traffic on the local machine. "Resource Monitoring" provides information about lost packets and additional information about the delay of current TCP / IP sessions. The information about the lost packets makes it possible to represent the quality of the connection. The delay describes the time it takes to completely traverse a TCP / IP packet.

Approaches to determining the state

Determining the security status of Internet sites is a complex task, the solution of which requires an integrated or systemic approach. Diagnosis is performed based on a fuzzy set. To do this, it is necessary to determine the degree of fuzziness of all terms relative to the center.

Let $X = \{x\}$ be a family of objects denoted by x , then the set A in X is $A = \{x, \mu_A(x)\}$, $x \in X$, where $\mu_A(x)$ is the degree of belonging of x to A . When the sets A and B ($A \cup B$) are joined, $\mu_{A \cup B}(x) = \text{Max}(\mu_A(x), \mu_B(x))$, $x \in X$.

(1)

The intersection of A and B ($A \cap B$) has the relation $\mu_{A \cap B}(x) = \text{min}(\mu_A(x), \mu_B(x))$, $x \in X$.

(2)

The degree of fuzziness is determined from the implication A on B, then we find the inverse implication B on A and compare the obtained implications:

$$A \rightarrow B = \max(\bar{A}, B) = \max(1 - A, B);$$

(3)

$$B \rightarrow A = \max(\bar{B}, A) = \max(A, 1 - B);$$

(4)

$$A \equiv B = \min(A, B) = \min(\max(1 - A, B), \max(A, 1 - B)).$$

(5)

Proposed evaluation criteria based on data collected from publicly available sources

Table (6) : Website working parameter's state

	S_0	S_1	S_2	S_3	S_4	S_5	S_6
D	> 780	600 ~ 700	381 ~ 450	321 ~ 380	250 ~ 310	181 ~ 260	5 ~ 190
T_{DNS}	>1200	880~ 1200	580~ 880	380~ 580	300~ 380	200~ 300	<200
P	> 7168	4096~ 7168	1536~ 4096	700 ~ 1536	350 ~ 700	180 ~ 350	>180
R_d	>8	8	7	6	5	4	<3
R_q	>70	54~67	45 ~ 55	34 ~48	27~ 35	19 ~ 29	<20
T_{CON}	>5500	1500~ 5500	1550~750	340~780	285~ 350	248~ 286	<250

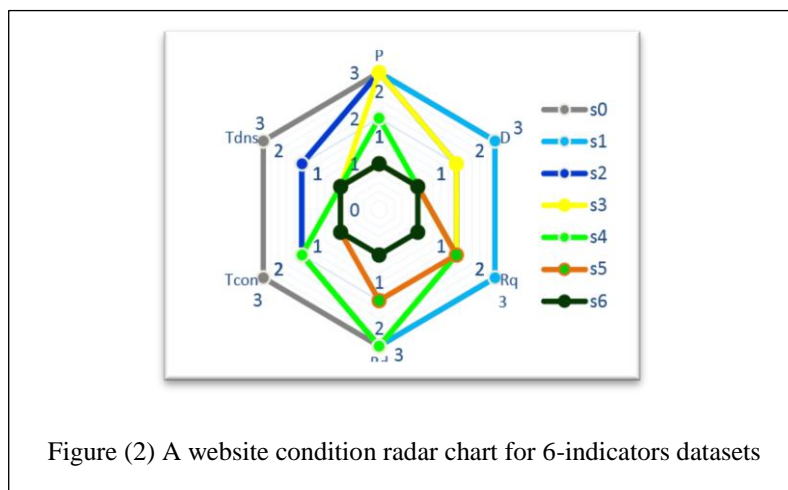


Figure (2) A website condition radar chart for 6-indicators datasets

Based on (Table 6), we construct a diagram (Figure 2), which shows the position of possible states of the web site operation in space. It can be concluded that the closer to the center, the higher the safety of the site.

Conclusions

The assessment of the security status of Internet sites as poorly formalized objects on the basis of fuzzy logic methods considered in the article is implemented in the form of technology for assessing the states of reliability of a web site. The proposed approach allows a more flexible adaptation to the particular problem and allows

diagnosis object already in the step of calculating an integral index of reliability. Flexibility is achieved due to the fact that the state of reliability can be estimated At once on several reliability indicators. Diagnostics at the stage of computing the integral index of security states is achieved due to the fact that the calculations can be divided into different stages. Each stage evaluates any of the individual indicators, which ultimately leads to a conclusion about the state of the corresponding element.

References

- 1- Antal V, Toine B, Maurice K: Estimating search engine index size variability, published online at Springerlink.com ,(2015).
- 2- Simon K. Digital In 2017 Global overview, Hoot suite , (2017).
- 3- Felke-Morris T. Web development and design of the fund with HTML5. Pearsom Publ., 672 p (2012).
- 4- Daniel Cederholm , Web Standards Solutions: The Markup and Style Handbook, Special Edition, amazon (2009).
- 5- Souders S. Even Faster Web Sites: Performance Best Practices for Web Developers , 221 p (2009).
- 6- Rick V., Andy D., Marcel D. : WebPageTest Web Performance Testing , O'Reilly Media,(2015).
- 7- Clifton B. Advanced Web Metrics with Google Analytics. 3rd ed. Sybex, John Wiley & Sons Publ., Indianapolis, Indiana, 600 p (2012) .
- 8- Claypool M., Claypool K. Latency can kill : precision and deadline in online games. Proc. MMSys '10 Proc. 1st annual ACM SIGMM Conf. on Multimedia systems. NY, ACM, pp. 215–222 (2010).
- 9- Shamim N. Pakzad; Gregory L. Fenves; Sukun Kim; and David E. Culler :Design and Implementation of Scalable Wireless Sensor Network for Structural Monitoring, Journal of Infrastructure Systems Vol. 14, Issue 1 (2008).

تقدير حالة أمن موقع الويب كآليات ضعيفة التنظيم تستند إلى أساليب منطقية ضبابية

ياسر علي مطني

قسم الحاسبات، كلية التربية الاساسية ، جامعة ديالى

ee22a12@gmail.com

المستخلص :

تُخصّص البحث لتحليل وقت تحميل الصفحة ، وهو مؤشر مهم لأي موقع ويب. وكقاعدة عامة ، يتم استضافة المواقع على خوادم الويب ذات خصائص معينة. يتفاعل المستخدمون مع البيئة بشكل كبير وغير مؤكد (التحديات الخارجية مثل الاختراق ، الحرمان من الخدمة ، إدخال الشفرة في لغة استعلامات SQL المنظمة ، إلخ). وتجدر الإشارة إلى أن هناك أوجه عدم اليقين التي تولدها الأجهزة والبرمجيات هي أيضا. يمكن أن يتأثر أي موقع بتأثير البيئة الخارجية والأنظمة الفرعية المختلفة لخدمة وظائف مواقع الويب ، مما يؤدي إلى حدوث حالات طارئة ويولد عدم يقين من عمله. لا يمكن دائما تقييم درجة عدم اليقين على أساس المواد الإحصائية وحدها. وهذا يؤدي إلى زيادة في عدد الطرق والوسائل لإضفاء الطابع الذهني على أداء التقييمات على أساس أساليب الذكاء الاصطناعي ، والأساليب القائمة على استخدام التقديرات غير الواضحة. يتم تنفيذ تقييم الوضع الأمني لمواقع الإنترنت كأجسام ضعيفة الشكل على أساس أساليب المنطق الضبابي التي يتم النظر فيها في المقالة في صورة تقنية لتقييم حالات موثوقية موقع الويب. يسمح النهج المقترح بتكيف أكثر مرونة مع المشكلة المعينة ويسمح لكائن التشخيص بالفعل في خطوة حساب مؤشر متكامل للاعتمادية. يتم تحقيق المرونة بسبب حقيقة أن حالة الموثوقية يمكن تقديرها مرة واحدة على عدة مؤشرات موثوقية.

الكلمات المفتاحية - المنطق الضبابي ، طرق تحليل النظام ، المقاييس ، التحليل وخصائص موقع الويب.