UDC 004.94 **doi: 10.20998/2522-9052.2017.1.11**

A. Semenova[1], M. Dubrovskyi[2], V. Savitskyi[2]

[1] National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine
[2] Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine

A GERT MODEL OF AN ALGORITHM FOR ANALYZING SECURITY OF A WEB APPLICATION

The **subject** of the study in the article is the mathematical network GERT model algorithm for analyzing the security of web applications, which allows you to find an arbitrary distribution function and the probability density function for the execution time of security of a Web application analysis algorithm. **Objectives:** The analysis of the problem and formulation of the task, task solution, flow chart of security of a Web application analysis, GERT model of security of a Web application analysis algorithm, probability density function for the execution time of security of a Web application analysis algorithm. The **methods** that are used: Methods of graph theory, security testing algorithms, methods of probability theory and mathematical statistics. The following **results** are obtained. An algorithm for testing the security of web applications is developed. A mathematical model of the algorithm for testing Web application security was developed, the model allowed to find an arbitrary distribution function of the statistical value of the vulnerability testing time. The probability distribution function for testing the security of web applications is found. This will make calculations and identify the most likely case of the law of distribution of the random value of the time of testing Web application security. **Conclusion.** A mathematical model of the algorithm security of a Web application analysis has been developed based on an exponential GERT network that is different from known models through taking into account DOM structure execution or analysis. The model can be used to study processes in automated systems as well as to develop new data security tools and protocols. Using exponential stochastic GERT models makes it possible to employ results obtained in an analytical form (functions, distribution densities) for comparative analysis and studies of more complex computer systems using mathematical methods.

Keywords: security of a Web application, a mathematical model, GERT model.

Introduction

Since the demand for web applications as well as web services is high, criminals have developed a keen interest in their potential vulnerabilities. Being originally targeted against server-side components, the key threats eventually turn into attacks on common users.

An analysis of Open Web Application Security Project (OWASP TOP-10) [1, 6] data has shown cross site scripting (hereinafter referred to as XSS) to be one of the most dangerous attack types (vulnerabilities).

An analysis of related literature has demonstrated that XSS is a user data validation error enabling a JavaScript code to be transmitted to the user's browser for execution. Such attacks are also commonly known as HTML injections as they are essentially similar to SQL injections, though the injected code is executed in the user's browser unlike in SQL injections.

The analysis of the problem and formulation of the task

According to [1-3, 6-8, 13], the term XSS generally refers to immediate [1] and deferred [6] cross site scripting. In immediate XSS, the attacked server returns the malicious code (JavaScript) immediately as a response to an HTTP request. Deferred XSS means that the malicious code is stored in the attacked system and can later be injected to an HTML page of the vulnerable system. It follows from this classification that XSS fundamentally consists in the browser sending the malicious code to the server, after which it is returned either to the browser (immediate XSS) or to any other browser (deferred XSS).

A number of articles on the Internet provide a detailed description of the basic mechanisms of such threats as well as potential ways of quenching them. However, to identify the threats and the possible consequences of their spreading within secure IT project management as well as to develop optimal solutions to the problem requires the process of their initialization and spreading to be formalized mathematically.

Simulation of security of a Web application appears to be of special importance in this respect since DOM XSS is an XSS type where the result of the attack is stored not in the server response and thus not in the HTML code but in the DOM structure of the HTML page. The results of attacks made through such vulnerabilities can only be detected when executed or by a DOM structure analysis. The attack mechanism here is still a JavaScript code injection to the vulnerable segment.

Task solution

In order to mathematically formalize the algorithm of security of a Web application analysis, we will refer to the fundamentals of GERT network modeling as described by [9-12].

Fig. 1 presents a flowchart for security of a Web application analysis.

The key stages according to the algorithm are as follows:

1. All <script> tags are retrieved from the code of the page analyzed and a list of tags to be analyzed is made.

2. A tag content analysis performed. In case the tags contain no code and only contain reference to a remote file, the file is accessed and the code is retrieved from it. The file contents is then analyzed for presence

of potentially dangerous sections of code (sinks) that use client input (source).

The following can serve as source examples:

document.URL;
document.documentURI;
location.href;
location.search;
location;
window.name;
document.referrer;
Sink examples;
document.write;
(element).innerHTML;
eval;
setTimout / setInterval;
execScript.

3. If the code uses the source, an attack is executed with a specific marker that can be traced in the page DOM structure after the code has been executed (e.g. an injection of a text content to the DOM).

4. The DOM content is checked for presence of the marker. If the marker is in the DOM following the attack, it indicates a DOM vulnerability.

5. Steps 2–4 are executed for each script tag on the page.

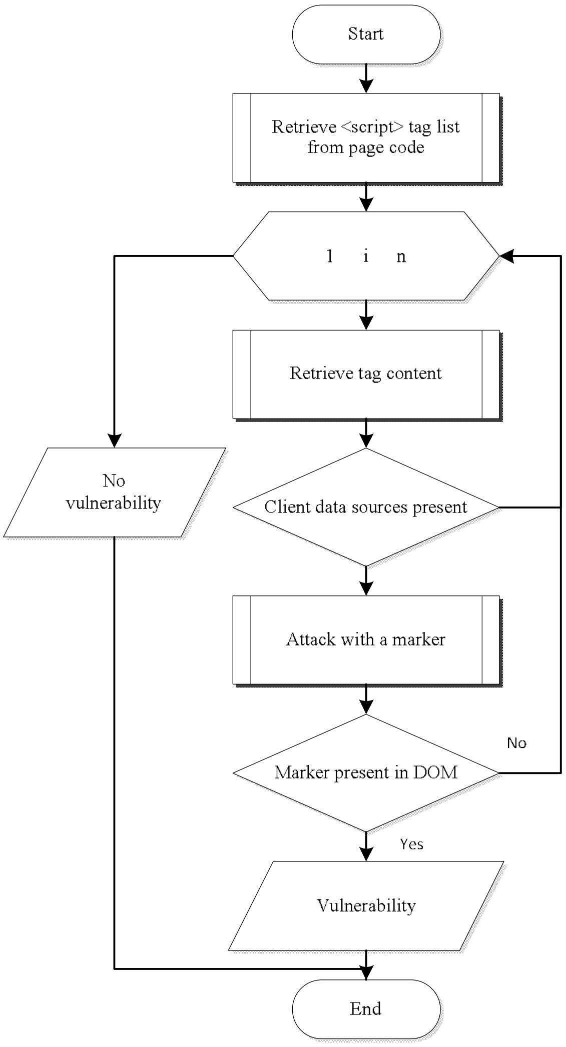

Fig. 1. Flow chart of security of a Web application analysis

Let us build a GERT network model of security of a Web application analysis algorithm according to the description. Fig. 2 provides a graphic presentation of the GERT model.

In the network presented here, graph nodes are interpreted as states of the computer system during DOM structure operation while graph edges are interpreted as probability and timing data for transitions between states. In particular, Edge (1,2) represents the time of tag content retrieval and analysis. Edge (2,3) represents timing characteristics of an attack that is executed in case of a source structure being present in the code. Edge (2,4) sets a random time of accessing the content of the remote file (sink search). Edge (4,2) represents the return to the state of executing the attack. Edge (3,5) describes the continuation of the attack, in particular checking the DOM contents. Edge (5,6) represents the time of decision on vulnerability while Edge (5,1) represents the timing of transition to the next tag.

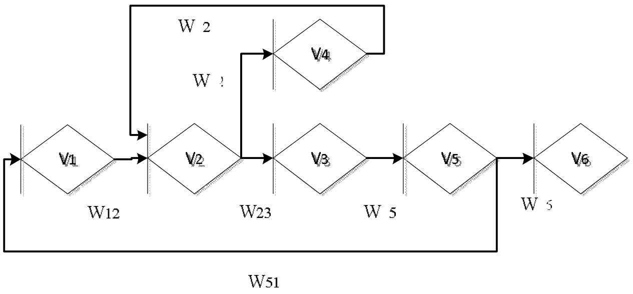

Fig. 2. GERT model of security of a Web application analysis algorithm

Look on a Table 1 for edge characteristics of the model.

The equivalent W function of execution time of security of a Web application analysis algorithm is as follows:

$$W_E(s) = \frac{W_{12}W_{23}W_{35}W_{56} + W_{12}W_{24}W_{42}W_{23}W_{35}W_{56}}{1 - W_{12}W_{23}W_{35}W_{51} - W_{12}W_{24}W_{42}W_{23}W_{35}W_{51}} =$$

$$= \frac{p_1 p_2^2 \lambda_1 \lambda_2^2 \left(p_4 \lambda_4 (\lambda_3 - s)^2 (\lambda_5 - s) + p_3^2 q_1 \lambda_3^2 \lambda_5 (\lambda_4 - s)\right)}{(\lambda_4 - s)\begin{pmatrix} (\lambda_1 - s)(\lambda_2 - s)^2 (\lambda_3 - s)^2 (\lambda_5 - s) - \\ - p_1 \lambda_1 p_2^2 \lambda_2^2 q_1 \lambda_5 (\lambda_3 - s)^2 - \\ - p_1 p_2^2 p_3^2 \lambda_1 \lambda_2^2 q_1 \lambda_3^2 \lambda_5 \end{pmatrix}}, \quad (1)$$

where $1 - p_4 = q_1$.

The point of interest of the process is characterized by high diversity of data analyzed and processed. Feedback can be organized in various ways. Fig. 2 presents the cycles as transitions $W_{12} \to W_{24} \to W_{42}$, $W_{12} \to W_{23} \to W_{35} \to W_{51}$.

No simple methods of finding singular points of function $\Phi_E(z)$ of real variables substitution $(z = -i\varsigma)$, where ς is a real variable, apply to GERT networks with cycles. This is due to the fact that finding singular points requires solving nonlinear equations. The more complex the GERT network structure is, the more complicated is the equation. A substitution is therefore suggested for modelling.

Complex transformation $z=-s$ yields

$$\Phi(\mathrm{z})=\frac{\mathrm{u}z^3+vz^2+bz+k}{(\lambda_4+z)(z^6+cz^5+dz^4+gz^3+hz^2+wz+m)}, \quad (2)$$

where $u=-p_1p_2^2p_4\lambda_1\lambda_2^2\lambda_4$,

$v=p_1p_2^2p_4\lambda_1\lambda_2^2\lambda_4(\lambda_5+2\lambda_3)$, $c=\lambda_1+2\lambda_2+2\lambda_3+\lambda_5$,

$b=-p_1p_2^2p_4\lambda_1\lambda_2^2\lambda_4\lambda_3(2\lambda_5-\lambda_3)$,

$k=-p_1p_2^2\lambda_1\lambda_2^2\lambda_3^2\lambda_4\lambda_5(p_4+p_3^2q_1)$,

$$d=-\begin{pmatrix}2\lambda_3\lambda_5+\lambda_1\lambda_5+2\lambda_2\lambda_5+\lambda_3^2+2\lambda_1\lambda_3+4\lambda_2\lambda_3+\\+2\lambda_1\lambda_2+\lambda_2^2\end{pmatrix},$$

$$g=\begin{pmatrix}\lambda_3^2\lambda_5+4\lambda_1\lambda_2\lambda_5+4\lambda_2\lambda_3\lambda_5+\lambda_2^2+\lambda_3^2\lambda_1+2\lambda_3^2\lambda_2+\\+4\lambda_1\lambda_2\lambda_3+2\lambda_2^2\lambda_3+\lambda_2^2\lambda_1\end{pmatrix},$$

$$h=-\begin{pmatrix}\lambda_1\lambda_3^2\lambda_5+2\lambda_2\lambda_3^2\lambda_5+4\lambda_1\lambda_2\lambda_3\lambda_5+2\lambda_2^2\lambda_3\lambda_5+\\+\lambda_2^2\lambda_3^2+2\lambda_1\lambda_2^2\lambda_3-p_1p_2^2q_1\lambda_1\lambda_2\lambda_5\end{pmatrix},$$

$$w=\begin{pmatrix}\lambda_1\lambda_2\lambda_3^2\lambda_5+\lambda_2^2\lambda_3^2\lambda_5+2\lambda_1\lambda_2^2\lambda_3\lambda_5+\\+\lambda_1\lambda_2^2\lambda_3-2p_1p_2^2q_1\lambda_1\lambda_2\lambda_3\lambda_5\end{pmatrix},$$

$$m=\left(p_1p_2^2q_1\lambda_1\lambda_2\lambda_3^2\lambda_5+p_1p_2^2p_3q_1\lambda_1\lambda_2^2\lambda_3^2\lambda_5-\lambda_1\lambda_2^2\lambda_3\lambda_5\right).$$

Table 1. **Model Edge Characteristics**

No.	Edge	W function	Probability	Moment generating function
1	(1,2)	W_{12}	p_1	$\lambda_1/(\lambda_1-s)$
2	(2,3)	W_{23}	p_2	$\lambda_2/(\lambda_2-s)$
3	(2,4)	W_{24}	p_3	$\lambda_3/(\lambda_3-s)$
4	(3,5)	W_{35}	p_2	$\lambda_2/(\lambda_2-s)$
5	(5,6)	W_{56}	p_4	$\lambda_4/(\lambda_4-s)$
6	(5,1)	W_{51}	$1-p_4$	$\lambda_5/(\lambda_5-s)$
7	(4,2)	W_{42}	p_3	$\lambda_3/(\lambda_3-s)$

Probability density function for the execution time of security of a Web application analysis algorithm is as follows:

$$\varphi(x)=\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}e^{zx}\frac{\mathrm{u}z^3+vz^2+bz+k}{(\lambda_4+z)\begin{pmatrix}z^6+cz^5+dz^4+gz^3+\\+hz^2+wz+m\end{pmatrix}}dz, \quad (3)$$

where integration is carried out with the Bromwich-Wagner integral [4]

The method of integration depends of whether the function $\Phi(\mathrm{z})$ has simple poles only or poles of some order. Where the function $\Phi(\mathrm{z})$ has simple poles only, the expression $e^{zx}\Phi(z)$ can be presented as follows:

$$e^{zx}\Phi(z)=\frac{e^{zx}(\mathrm{u}z^3+vz^2+bz+k)}{\begin{pmatrix}z^7+a_6z^6+a_5z^5+a_4z^4+\\+a_3z^3+a_2z^2+a_1z+a_0\end{pmatrix}}=\frac{\mu(z)}{\psi(z)}, \quad (4)$$

where $a_6=\lambda_4+c$, $a_5=c\lambda_4+d$, $a_4=d\lambda_4+g$, $a_3=g\lambda_4+h$, $a_2=h\lambda_4+w$, $a_1=w\lambda_4+m$, $a_0=m\lambda_4$.

In this case, probability density function for the execution time of security of a Web application analysis algorithm is as follows:

$$\phi(x)=\sum_{k=1}^{6}\mathrm{Re}s\left[e^{zx}\Phi(z)\right]=\sum_{k=1}^{7}\frac{\mu(z_k)}{\psi(z_k)}=$$
$$=\sum_{k=1}^{7}\frac{e^{zx}\left(uz^3+vz^2+bz+k\right)}{\begin{pmatrix}7z_k^6+6a_6z_k^5+5a_5z_k^4+4a_4z_k^3+\\+3a_3z_k^2+2a_2z_k+a_1\end{pmatrix}}. \quad (5)$$

Apart from the solutions determined by the roots of the equation $z^6+cz^5+dz^4+gz^3+hz^2+wz+m=0$, the function $\Phi(z)$ can have a second or third order pole where the value of λ_4 equals that of the roots z_2, z_3, z_4, z_5, z_6, z_7. In these cases distribution density for message transmission time $\varphi(x)$ can be calculated by the formula for finding the residues r_{-1} of the poles z_k of the order n:

$$r_{-1}=\frac{1}{(n-1)!}\lim_{z\to z_k}\frac{d^{n-1}\left((z-z_k)^n e^{zx}\Phi(z)\right)}{dz^{n-1}}. \quad (6)$$

Expression (5) is a fractional rational function of Z with a denominator degree higher than the numerator degree. It therefore meets the conditions of Jordan's lemma [4, 5]. The function $\Phi(z)$ has poles in points $z_1=-\lambda_4$. The polynomial

$$z^6+cz^5+dz^4+gz^3+hz^2+wz+m$$

brings about seven more poles. The equation

$$z^6+cz^5+dz^4+gz^3+hz^2+wz+m=0 \quad (7)$$

can be solved by any method, e.g. Viet's formulas [3, 4]. As the result, singular points z_2, z_3, z_4, z_5, z_6, z_7 are found.

Conclusions

Therefore, a mathematical model of the algorithm security of a Web application analysis has been developed based on an exponential GERT network that is different from known models through taking into account DOM structure execution or analysis.

The model can be used to study processes in automated systems as well as to develop new data security tools and protocols.

Using exponential stochastic GERT models makes it possible to employ results obtained in an analytical form (functions, distribution densities) for comparative analysis and studies of more complex computer systems using mathematical methods.

REFERENCES

1. About The Open Web Application Security Project – OWASP, available at : https://www.owasp.org/index.php/About_The_Open_Web_Application_Security_Project (last accessed December 26, 2016).

2. Babincev, I. and Vuletic, D. (2016), “Web application security analysis using the kali Linux operating system”, *Vojnotehnicki glasnik. Military Technical Courier*, Vol. 64 № 2 available at : https://cyberleninka.ru/article/v/web-application-security-analysis-using-the-kali-linux-operating-system (last accessed December 26, 2016).
3. Baranov, P. and Beybutov E. (2015) “Securing information resources using Web application firewalls”, *Business Informatics*, No. 4 (34). pp. 71-78.
4. Edvards, G. (1980), Poslednyaya teorema Ferma. Geneticheskoye vvedeniye v algebraicheskuyu teoriyu chisel, Moskva : Mir, 486 p.
5. Gmurman, V.Ye. (2003), *Teoriya veroyatnostey i matematicheskaya statistika*, Moskva : Vysshaya shkola, 479 p.
6. Il'yenko, F.V. and Prikhod'ko, T.A. (2013), “Problemy uyazvimosti Web i sredstva dlya analiza bezopasnosti Web-prilozheniy”, *Ínformatsíyní upravlyayuchí sistemi ta komp'yuterniy monitoring*. Materiali III mizhnarodnoí naukovo-tekhnichnoí konferentsií studentiv, aspirantiv ta molodikh vchenikh, Donets'k, DonNTU, available at : http://masters.donntu.org/2013/fknt/ilyenko/library/sredstva_analiza_bezopasnosti_web_ilyenko_prixodko.pdf (last accessed December 26, 2016).
7. Category: OWASP Top Ten Project – OWASP, available at : https://www.owasp.org/index.php/Category:OWASP_Top_ Ten_Project. (last accessed December 26, 2016).
8. Cohen, W., Ravikumar, P. and Fienberg S. A Comparison of String Metrics for Matching Names and Records, available at : https://www.cs.cmu.edu/afs/cs/Web/People/wcohen/postscript/kdd-2003-match-ws.pdf (last accessed December 26, 2016).
9. Pritsker, A.A.B. and Happ, W.W. {1966), GERT : “Part I. Fundamentals”, *The Journal of Industrial Engineering.*
10. Pritsker, A.A.B.(1979), *Modeling and analysis using Q-GERT networks*, New York: Wiley : Distributed by Halsted Press.
11. Semenov, S.G., Bos'ko, V.V. and Berezyuk, Í.A. (2012), “Issledovaniya veroyatnostno-vremennykh kharakteristik mul'tiservisnogo kanala svyazi s ispol'zovaniyem matematicheskogo apparata GERT-seti”, *Sistemi obrobki ínformatsíí*, Kharkiv : KHU PS, Vol. 1. Ic. 3 (101). – pp. 139–142.
12. Semenov, S.G. (2012), “Metodika matematicheskogo modelirovaniya zashchishchennoy ITS na osnove mnogosloynoy GERT-seti”, *Vísnik Natsíonal'nogo tekhníchnogo uníversitetu «KPIt»,* KH.:NTU «Kharkívs'kiy polítekhníchniy ínstitut»,. № 62 (968), pp. 173–181.
13. Sung Gyeong Bae, Hyunghun Cho, Inho Lim and Sukyoung Ryu (2014) “SAFEWAPI: Web API Misuse Detector for Web Applications”, *Proceedings of the 22Nd ACM SIGSOFT International Symposium on Foundations of Software Engineering*, pp. 507–517, available at : https://pdfs.semanticscholar.org/ (last accessed December 26, 2016).

Надійшла (received) 22.01.2017
Прийнята до друку (accepted for publication) 16.05.2017

GERT-модель алгоритму аналізу безпеки web-додатку

А. С. Семенова, М. С. Дубровський, В. В. Савицький

Предметом дослідження в статті є математичний мережевий алгоритм GERT для аналізу безпеки веб-додатків, який дозволяє знайти довільну функцію розподілу і щільність ймовірностей часу виконання алгоритму аналізу безпеки веб-додатків. **Мета:** аналіз проблеми та формулювання завдання, вирішення завдання, розробка блок-схеми аналізу безпеки веб-додатків, розробка GERT моделі алгоритму аналізу безпеки веб-додатків, знаходження розподілу і щільності ймовірностей часу виконання алгоритму аналізу безпеки веб-додатків. **Використовувані методи:** методи теорії графів, алгоритми тестування безпеки, методи теорії ймовірностей і математичної статистики. **Отримані наступні результати.** На основі експоненційної GERT-мережі розроблено математичну модель алгоритму аналізу DOM XSS уразливості, яка відрізняється від відомих, урахуванням виконання або аналізу DOM структури. Модель може бути використана для дослідження процесів в комп'ютеризованих системах, при розробці нових засобів і протоколів захисту даних. Застосування експоненційних стохастичних моделей GERT дасть можливість використання результатів, отриманих в аналітичному вигляді (функції, щільності розподілу) для проведення порівняльного аналізу і досліджень, більш складних комп'ютерних систем математичними методами.

Ключові слова: безпека веб-додатку, математична модель, модель GERT.

GERT-модель алгоритма анализа безопасности web-приложения

А. С. Семенова, М. С. Дубровский, В. В. Савицкий

Предметом исследования в статье является математический сетевой алгоритм GERT для анализа безопасности веб-приложений, который позволяет найти произвольную функцию распределения и плотность вероятностей времени выполнения алгоритма анализа безопасности веб-приложений. **Цель** – анализ проблемы и формулировка задачи, решение задачи, разработка блок-схемы анализа безопасности веб-приложений, разработка GERT модели алгоритма анализа безопасности веб-приложений, нахождение распределения и плотность вероятностей времени выполнения алгоритма анализа безопасности веб-приложений**. Используемые методы:** методы теории графов, алгоритмы тестирования безопасности, методы теории вероятностей и математической статистики. **Получены следующие результаты.** На основе экспоненциальной GERT-сети разработана математическая модель алгоритма анализа DOM XSS уязвимости, которая отличается от известных, учетом выполнения или анализа DOM структуры. Модель может быть использована для исследования процессов в компьютеризированных системах, при разработке новых средств и протоколов защиты данных. Применение экспоненциальных стохастических моделей GERT даст возможность использования результатов, полученных в аналитическом виде (функции, плотности распределения) для проведения сравнительного анализа и исследований, более сложных компьютерных систем математическими методами.

Ключевые слова: безопасность веб-приложения, математическая модель, модель GERT.