Journal of Engineering and Computer Sciences
Qassim University, **Vol. 14, No. 1, pp. 64-81 (January 2021 /Rabi II 1441H)**

Score level prediction of Google Chrome Vulnerability using Bayesian Network

Mafawez T. Alharbi
Department of Natural and Applied Sciences,
Buraydah Community College, Qassim University,
Saudi Arabia

Maft.alharbi@qu.edu.sa

(Received 10/3/2021, accepted for publication 10/8/2021)

Abstract. Software vulnerability is a weakness and bug that can affect security protocols of the system. The issue at hand has produced an increased software vulnerability within each lifecycle release. A web browser is used by users as the primary method to communicate with each other online via devices. People can find great content on the web. Challenges arise within security protocols when attackers access users via web browsers. Therefore, web browsers have become a target for attackers who use web browser vulnerabilities. This research focuses on Google Chrome vulnerabilities, and proposes a new approach which can provide probabilistic predictions that generate a score level of the Chrome browser vulnerability. The predication model uses artificial intelligence from the "Bayesian Network ".

Keywords: Google Chrome vulnerabilities, security protocols, provide probabilistic predictions

1- Introduction

Software systems play a very important role in our everyday life, as most users who use such systems do so in order to do their jobs. Software systems are distributed in every government and private sector around the world. These systems are used in government and private sectors in order to manage and control the operation processes and measurement performance of software, in how it manages information. Most of these sectors have the ability to develop businesses and earn profit through the usage of software systems, whereas other sectors might lose profit because of the vulnerability of the software system being used or adapted. [1]

Software vulnerability is a weakness and/or a bug in the software that can affect the security of a system. The Information Security Community (ISC) works very hard to discover and prevent such

software vulnerabilities, which are discovered every year. Software vulnerabilities that lead to data exposure is one of the major threats that face the development of software, as these vulnerabilities are exploited by attackers; this directly impacts the availability, integrity and confidentiality of users and harms the software system, which may lead to crash the system [2] .

Every half hour or so, an attack can happen on the organization's software system, and then sensitive data can be sent outside the organization. There is a need today to analysis and predicate vulnerabilities by the use of software engineering and research centers using techniques from Artificial Intelligence [3].

The Internet and web browsers, which are used to view and navigate, have become one of the most important software applications used in our daily life. Web browsers are software applications that have the ability to do most internet activities. For this reason, web browsers have become a target for attackers. The common web browsers that are used are: Internet Explorer, Safari, Mozilla Firefox and Google Chrome [4,5].

The aim of this paper is to implement a predicate score level of Google Chrome vulnerability using the Bayesian Network technique and data form NVD, while also showing the advantages of the proposed model in this research:

- The proposal in this research helps cyber security managers in organizations to predicate the score level severity which will help them to prevent such attack from happening again.
- It has the ability to identify the most likely causes of vulnerability.
- The model is presented as an easy-to-read graph.
- The model is able to deal with time-series data, which shows previous and current data of Chrome vulnerability.
- The model can handle incomplete data, in cases where it is impossible to have all the input data.

The rest of the research is organized as follows: section 2 present literature review, section 3 discusses the problem statement, section 4 shows the proposed model, and section 5 presents validation of the model and section 6 presents the conclusion.

2. Literature review

2.1 Related work

This section presents the related work on Google Chrome and Bayesian Network in security via a literature review. The research presented in [6] discusses vulnerabilities that come within the Chrome web browser, and describes the impact factors and reasons for vulnerabilities. In addition, the authors discuss the threat and the damage that comes from such vulnerabilities, from simple to complex ones. Data has been taken, assessed and collated for four years from 2010 to 2015, and the research presents different vulnerabilities found during these years.

Other research in [7] proposes a framework that addresses real change, by giving developers and users the ability to do an analysis of vulnerabilities to further develop the Chrome browser. The technique that has been used in this research uses a feature of a Chrome extension that reads data flow via the extension, looking at Java Script code, for vulnerabilities to make sure that any other Chrome browser extension is or is not malicious.

A study by [8] proves that the data in NVD is abnormal in Chrome data, where the most vulnerabilities are presented in the first version. This paper studies the reliability of data in NVD, and it considers if the vulnerability in each version of Chrome in NVD is a real vulnerability or not. The research shows that there are some errors in the Chrome vulnerability data. The study shows how these errors might affect other studies on vulnerabilities of a similar nature. The result of the research presents that there are a number of observations that can be obtained on this issue, which need to be addressed.

On [9] the uncertainty mode for cybersecurity and initial evidence that indicate it is a good methodology that has many advantages. The proposed mode takes data directory for security and analysis it. The Bayesian Network can identify uncertainty types and then build Bayesian Network modes based on the current security platform being used. The finding in this study proves the Bayesian Network can lead to new organizational development and security analysis via graphical models.

A research study in [10] proposes cyber attacker predication modes based on the Bayesian Network. An attack graph has been used to show all vulnerability and attacking path probability. The model uses impact factors on attack probability using the Bayesian Network

structure/framework. Through the experimental study, it proves that the proposed model has a high degree of accuracy.

In addition, the research in [11] focuses on their study of the limitations on the temporal factors, on measuring security networks. The paper proposes a model based on the Dynamic Bayesian Network, in order to track the change in a computer network. A research presented by [12] proposes an architecture of risk management by using the Bayesian Network. Via this proposal, the administrator has the ability to control the quantity of network flexibility on a number of levels. The proposal aims to continually analyze through developed network phases.

In [13] suggests using effective algorithms to make exact inferences in the Bayesian attack graphs, to enable the dynamics and state the risk evaluation on a network. To make sure that this proposal is valid, some experimental studies have been made by the Bayesian Network. The proposal allows for the evaluation of risk in cybersecurity on networks attacked by calculating the probability that attackers will attack such networks that lack of good cyber security.

Other research extracted feature toggles from the Google Chrome web browser, in order to use them in understanding the architecture of the system due to the lack of high-level and concrete visualization structures for Chrome, the researchers driven to derive these representations from the existing documentation and mapped into the source code [14].

[15] focused on the Chrome browser, which is the natural choice for all users because it is a widely popular open source. Web feature deprecation was analyzed via Chrome browser lens, where it is rated at a value of 2.5 years Chrome deprecation and six reasons found, clarifies reasons why developers want to deprecate the web feature.

Finally, a paper in [16] presents the possibility of using Chrome apps and extensions to acquire knowledge of mathematics, writing, reading, organizing, and planning in general. The Google Chrome browser, which is widely spread on computers today, has a large collection of premium and low-cost applications and extensions that can benefit all users.

2.2 National Vulnerabilities Database

National Vulnerabilities Database (NVD) is a public website that contains information about vulnerabilities and public details of them. It started in 1997 and is owned by the ministry of defense

in the United States of America. NVD includes very important information on a security checklist, security related software flaws, misconfiguration, production name and impact matrices. NVD is one of the products from the National Institute of Standards and Technology (NIST) and is supported by the Department of Homeland Security, National Cyber Security Division (NCSD), and it is a well-known data source for vulnerability information that can be valuable in predicating and estimating vulnerabilities in software systems [17].

2.3 Bayesian Network (BN)

Artificial intelligence has been used in many domains. One of the AI techniques is called, the Bayesian Network. This research follows the methodology of the Bayesian Network. This section will start off with a basic idea of the Bayesian Network, before discussing the methodology it uses. The Bayesian Network is one of the techniques of using artificial intelligence in cyber security. This type of technique presents a high level of probability distribution on groups of variables which are used to build problem solution platforms [18]. The BN has the capability to combine different sources of knowledge and has the ability to overcome the scarcity of historical data in cyber security modeling. The BN has been widely employed in real-world applications, especially in medical diagnosis. BN has been used in security these days, due to the capability it has to overcome data limitation [19]. In general, Bayesian Networks could be possibly used in Google chrome vulnerability, as it enables diagnostic reasoning that could help to identify the most likely cause of an attack event based on certain symptoms.

The main advantages of BN is to use these networks as powerful tools to collect different resources of knowledge with different degrees of uncertainty, in effective mathematical ways [20]. BN consists of qualitative and quantitative data gathering, where qualitative data is the part that presents directed acyclic graphs that include nodes and edges, where each node shows a variable and the edge between nodes present conditional dependences. Meanwhile, quantitative data is the part that takes conditional probability modes that decide dependency amounts between connected nodes via conditional probability between each node [21].

The Bayesian Network uses a directed acyclic graph and node framework that represents variables, and the directed edges represent dependences between variables. The relationship between variables can be described by a conditional probability table (CPT). BN is based on Bayes' theory.

3. Problem statement

Vulnerability in the Google Chrome browser has increased in recent years, and most people use this browser. Researchers and developers try their best to develop software to prevent such vulnerabilities. However, their efforts need to be increased with using new techniques such as AI, as it has the ability to predict new vulnerabilities, based on previous data. To the best of our knowledge, there is not any research being done using BN to predict Google Chrome vulnerability.

4. Proposed model

This research proposed a new model of predicating Google Chrome vulnerabilities based on BN.

Also, it will show how the author selects, collects, and analyses research data. The methodology for this research starts when the author uses NVD to selects the research data focusing on Google Chrome vulnerability. NVD has been chosen because the NVD is the U.S. government repository of standards based vulnerability management data represented, which is a platform that can collect information about discovered computer security vulnerabilities.

CVEdetails which was developed by Serkan Ozkan automatically collects vulnerability data from the NVD [23]. Soup and Scrapy are libraries in Python that were used in this research to scrap the vulnerabilities of Google Chrome from CVEdetails

To reach this goal, there are four stages needed to be followed; all these stages have to be done in order to build an effective BN model

- To identify the factor
- To draw the model
- To parametrize the model
- To apply inference algorithm

4.1 First: To identify the factor

This stage aims to identify the factor that affects the severity of a score level. The author has studied Google Chrome vulnerabilities in the National Vulnerability Database and has selected the most likely factors that can affect the score level of vulnerabilities found in Google Chrome. Six factors have been selected: Confidentiality Impact, Integrity Impact, Availability Impact, Authentication, Access Complexity and Gained Access. Confidentiality Impact means the impact

to the confidentiality of the affected system. Integrity Impact refers to the integrity of the affected system, where the attacker can modify any files or information on the target system. Availability Impact refers to the impact to the availability of the affected system. Authentication means the level of authentication needed to access the vulnerable system. Access Complexity refers to the level of difficulty to access the candidate. Each factor has three states that show the level of the factor. Table 1 below presents each factor with their possible states.

Table 1: Factors with their states

Factor	**State 1**	**State 2**	**State 3**
Confidentiality Impact	None	Partial	Complete
Integrity Impact	None	Partial	Complete
Availability Impact	None	Partial	Complete
Authentication	Not Required	Single System	Multiple
Access Complexity	Low	Medium	High
Gained Access	None	User	Admin

4.2 Second: To draw the Bayesian Network model

This research use the tool called "GeNIe," which is one application of Bayesan Network techniques that has ability to implement the BN mode that is developed by the Decision System Laboratory, at the University of Pittsburgh. Some advantages of GeNIe are that, it is free to be downloaded for academic research, it provides a graphical user interface and it has the ability to offer bar charts for each node. There are many tools that have the ability to draw the model of BN;

one of these tools is GeNIe, which is a graphical user interface. More details about this tool can be found in [21,22].

This stage finds the relationships between the selected factors that have been identified in previous sections. These factors need to be fused together to draw upon the model of the Bayesian Network. The proposal model is presented in Figure 1 below; the hypothesis node here is the score level that is needed to predicate and "score" the variability of a network. The rest of the nodes above the score level are the factor nodes. This can be described as the qualitative part of BN, which shows the model as a directed acyclic graph. In addition, the model presents the causal relationship between the factors. In addition, Figure 2 gives more details about the Bayesian Model with their states.

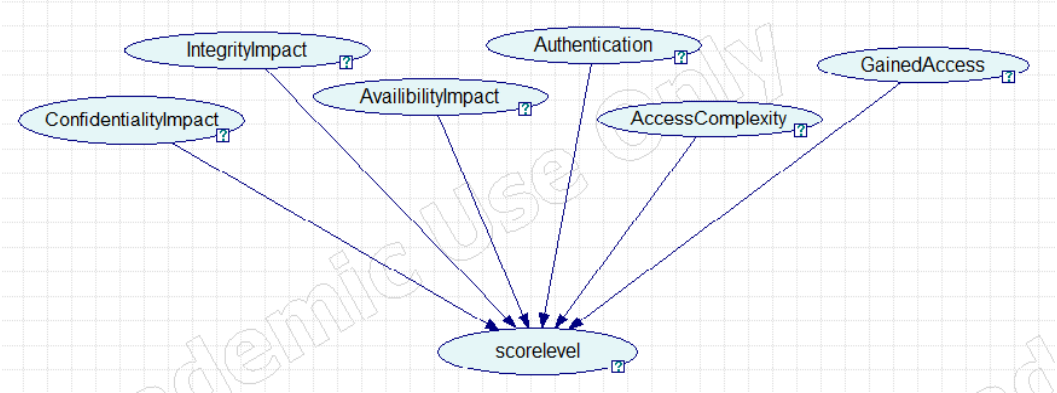

Figure 1: Bayesian Network model

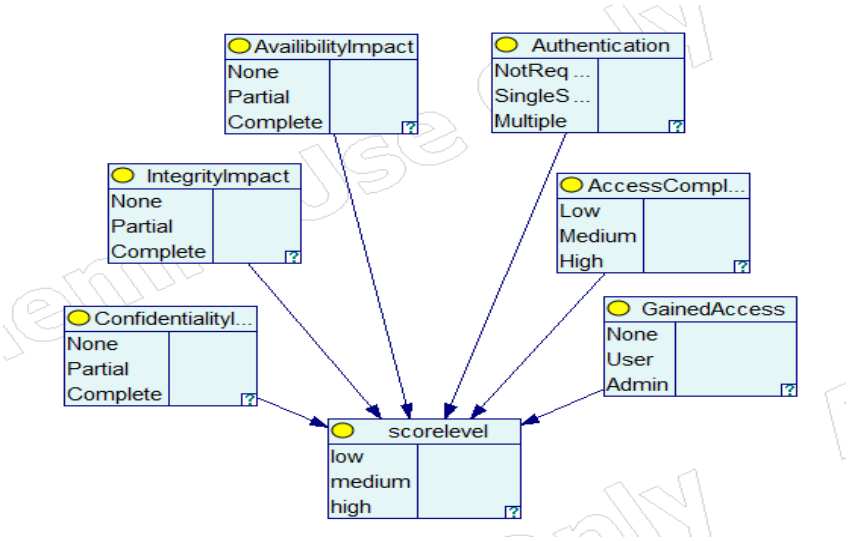

Figure 2 : Bayesian nodes with their states

4.3 Third: To parametrize the network

This stage presents parametrizing of the proposed model, which can be described as the quantitative part of the model. This stage indicates how the condition probability for each of the connected node must be calculated. Each node in a Bayesian Network must be associated with a condition probability table (CPT). This is the probability distribution of the node's possible states, conditioned on the parents' states. The model needs to fill up with the data of Google Chrome vulnerabilities that has been gathered from NVD, within the period of 2008 to 2019 and presented as an Excel file, after that the data has been cleaned and converted to CSV path, and the author has found 1,859 vulnerabilities. The data file was imported to the GeNIe tool; then it learned parameters used in GeNIe, in order to calculate the value of each state of the node. For example Figure 3 shows the parametrization of the Confidentiality Impact node.

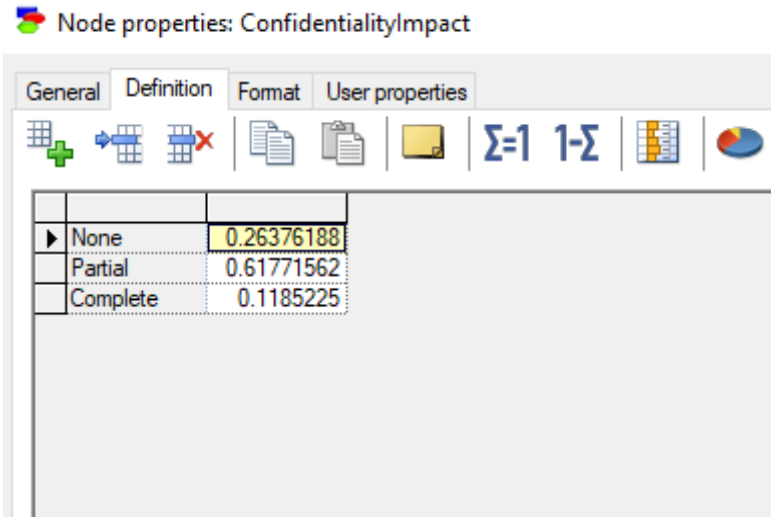

Figure 3: Parametrizing the Confidentiality Impact node

4.4 Four: To apply inference algorithm

The last stage involves applying an inference algorithm to the model where the main aim of building a BN mode is to apply inference. BN have two types of inference, exact and approximate inference. Exact inference consists of seven types: Polytree algorithms, Elimination, Clustering, Symbolic, Conditional, Different methods, and Arc reversal [22]. In this proposed model, exact inference will be used, specifically the Polytree inference. This research uses the Polytree inference because the research has one path between any two nodes in a network.

5. Validation the proposed model

This section shows the validation in using the Bayesian Network. To perform the test, there are two steps: first, selecting the evidence for all or some child nodes, and the second is to update beliefs. When more evidence becomes available for some factors in the BN, the probabilities of the score level factor in the BN could be updated (This is called inference or belief updating). Three cases have been performed to test the model as follow:

Case 1:

If we set up the state on "partial," we can access the complexity node as an evidence factor. This means that the state of evidence factor is considered as 100 %, and the rest of the selected states of each node are stated and shown in Table 2 below.

Table 2: Case 1

Nodes/factors	State
Confidentiality Impact	Partial
Integrity Impact	Partial
Availability Impact	Partial
Authentication	Not Required
Access Complexity	High
Gained Access	None
Score Level	Medium
Probability	0.94444444

As a result of this case, the probability of a proper score level as a medium is 94% as shown in Figure 4. This indicates that the access complexity factor with a state of high could lead to a medium score of severity level. This is because access to complexity measures the complexity of the attack required to exploit vulnerabilities, once an attacker has gained access to the system. In addition, there are other factors such as confidentiality, integrity, and availability with their state of partial, which have different outcomes.

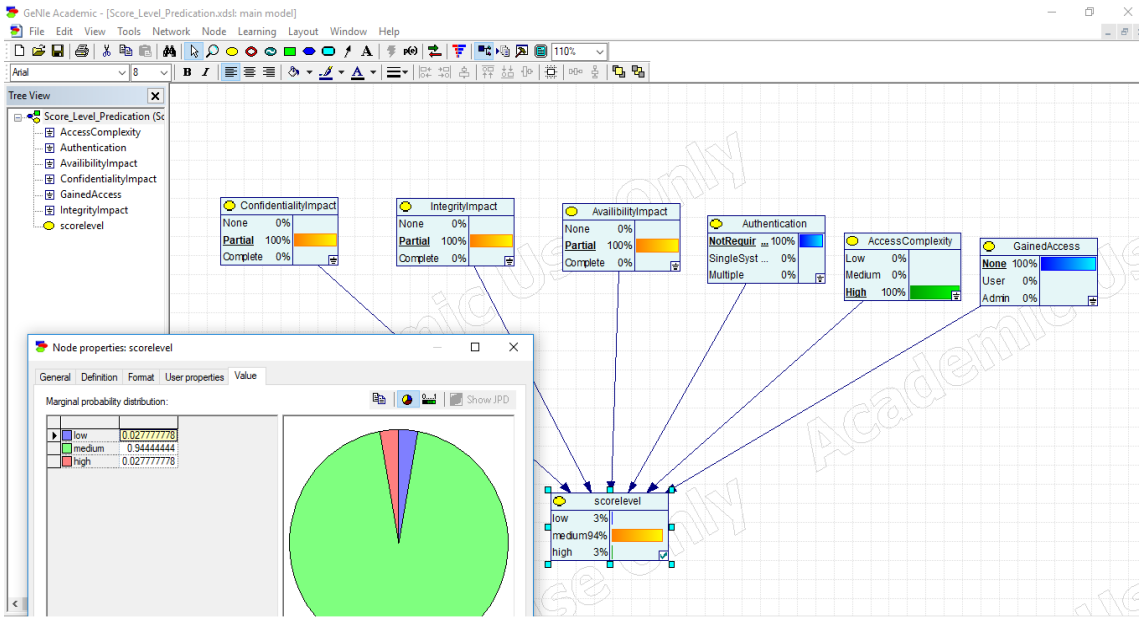

Figure 4: Validation of Case 1

Case 2:

In this case, it is assumed that the state of "confidentiality impact" and "integrity node" are on "none." These are called "evidence factors," and they are the state of evidence factor, which is considered as 100 % in this case. The rest of the selected states for each node are shown in Table 3 below. As a result of this case, the probability of a score level to be low is 67%, as shown in Figure 5; this is due to the confidentiality and integrity node with a "state" set to "none."

Table 3: Case 2

Nodes/factors	State
Confidentiality Impact	None
Integrity Impact	None
Availability Impact	Partial

Authentication	Not Required
Access Complexity	High
Gained Access	None
Score Level	Low
Probability	0.66666667

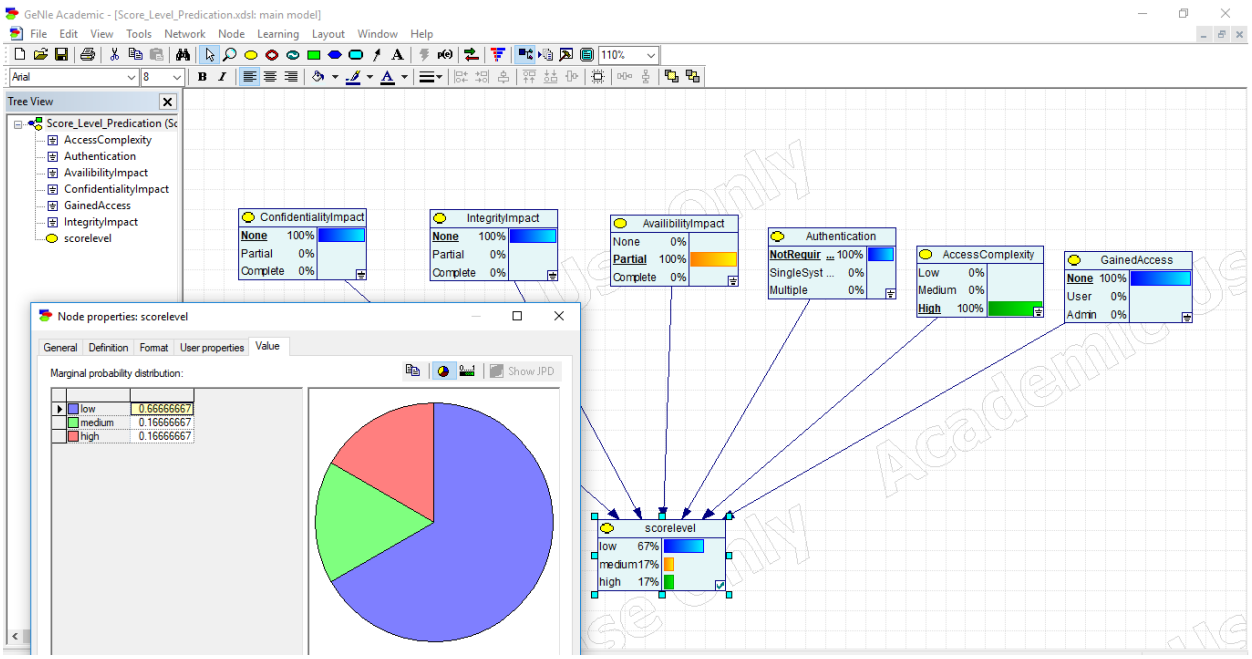

Figure 5: Validation of Case 2

Case 3:

If we set up each factor to their high state, more details can be captured in regards to confidentiality, integrity, and availability. Setting the factors to "high," means "Authentication," is multiple, "Access Complexity," is high and "Gained Access," is admin. This means that the state of evidence factor is considered to be 100 %, as shown in Table 4 below. As a result of this case, the probability

of a high score level is 90% as shown in Figure 6. In this case, all the factors are in high levels, so this could significantly affect the severity of the score level to be high.

Table 4: Case 3

Nodes/factors	State
Confidentiality Impact	Complete
Integrity Impact	Complete
Availability Impact	Complete
Authentication	Multiple
Access Complexity	High
Gained Access	Admin
Score Level	High
Probability	0.9

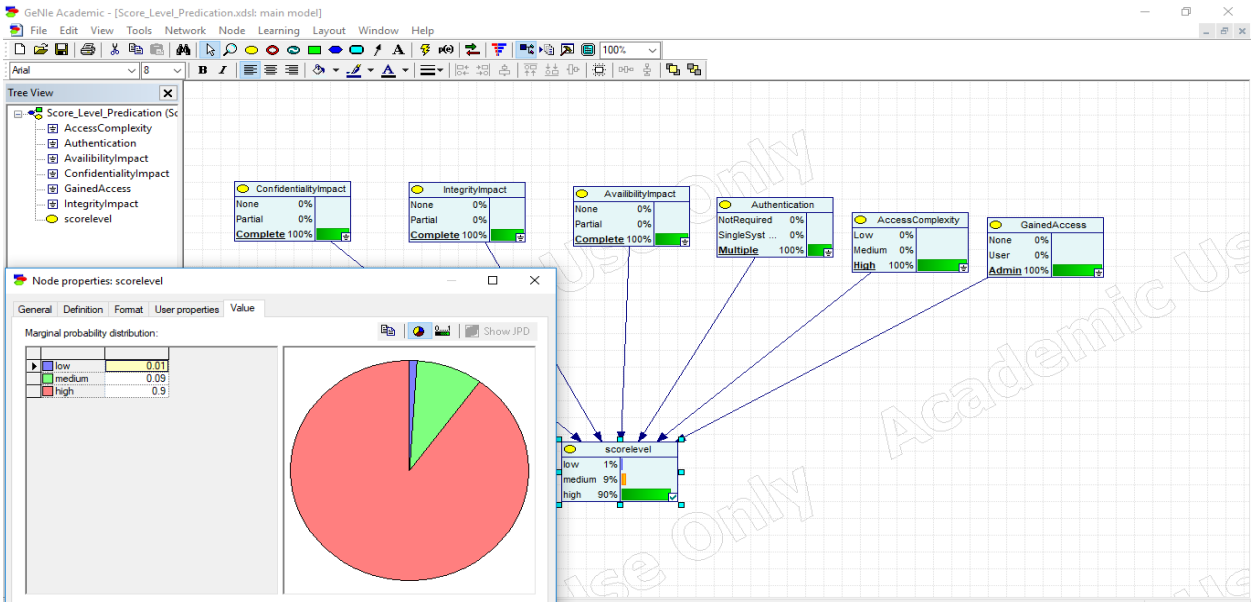

Figure 6: Validation of Case 3

6. Conclusion

This research focuses on Google Chrome vulnerabilities. The author uses the Bayesian Network to predicate a score level for the vulnerability of a cyber-attack towards a network. The research introduces vulnerabilities in general in order to look at Google Chrome vulnerabilities. Then the author defines the national vulnerabilities database. After that, related work is presented. Then the methodology of the paper is presented with an introduction of the Bayesian Network, and the proposed model is shown in four stages with validation of the model. Furthermore, the test of the Bayesian Network model is carried out. Finally, we conclude our work in the last section.

7. References

[1] Kumar, M., and Arun, S. "An integrated framework for software vulnerability detection, analysis and mitigation: an autonomic system,", *Sadhana*, Vol.42, No.9 (2017), pp.1481-1493.

[2] Spanos, G., and Lefteris A. "A multi-target approach to estimate software vulnerability characteristics and severity scores," *Journal of Systems and Software*, Vol.146, (2018), pp.152-166.

[3] Murtaza, S. S., Khreich, W., Hamou-Lhadj, A., and Bener, A. B "Mining trends and patterns of software vulnerabilities," *Journal of Systems and Software*, Vol .117, (2016), pp. 218-228.

[4] Arunagiri, J., Rakhi, S., and Jevitha, K. "A systematic review of security measures for web browser extension vulnerabilities," *Proceedings of the International Conference on Soft Computing Systems*, Springer, New Delhi, (2016), p. 99-112.

[5] Roumani, Y., Joseph, N., and Yazan, R. "Time series modeling of vulnerabilities," *Computers & Security*,Vol. 51, (2015),pp. 32-40.

[6] Manhas, S., and Swapnesh, T. "A Comparative Analysis of Various Vulnerabilities Occur in Google Chrome," *Soft Computing: Theories and Applications*, Springer, Singapore, (2018), pp. 51-59.

[7] Duebendorfer, T., and Stefan, F. "Web browser security update effectiveness," *International Workshop on Critical Information Infrastructures Security*. Springer, Berlin, Heidelberg, (2009), pp.124-137.

[8] Nguyen, H., and Fabio, M. "The (un) reliability of nvd vulnerable versions data: An empirical experiment on google chrome vulnerabilities," *Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security*, (2013), pp.493-498.

[9] Xie, P., et al. "Using Bayesian networks for cyber security analysis," *IEEE/IFIP International Conference on Dependable Systems and Networks, (2010)*, pp.211-220.

[10] Wu, J., Lihua, Y., and Yunchuan, G. "Cyber-attacks prediction model based on Bayesian network," IEEE *18th International Conference on Parallel and Distributed Systems*, (2012), pp.730-731.

[11] Frigault, M., Wang, L., Singhal, A., & Jajodia, S. "Measuring network security using dynamic bayesian network," *Proceedings of the 4th ACM workshop on Quality of protection*, (2008), pp.23-30).

[12] Poolsappasit, N., Rinku, D., and Indrajit, R. "Dynamic security risk management using bayesian attack graphs," *IEEE Transactions on Dependable and Secure Computing,Vol.* 9, No.1 (2011),pp. 61-74.

[13] Muñoz-González, L., Sgandurra, D., Barrère, M., & Lupu, E. C"Exact inference techniques for the analysis of Bayesian attack graphs," *IEEE Transactions on Dependable and Secure Computing*, Vol.16, No.2 (2017), pp. 231-244.

[14] Rahman, T., Rigby, C. and Shihab, E."The modular and feature toggle architectures of Google Chrome," *Empirical Software Engineering*, Vol.24, No.2, (2019), pp.826-853.

[15] Mirian, A., Bhagat, N., Sadowski, C., Felt, A. P., Savage, S., & Voelker, G. M. "Web feature deprecation: A case study for Chrome," In *IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice*, (2019), pp.302-311.

[16] Ok, M.W., and Rao, K. "Digital tools for the inclusive classroom: Google chrome as assistive and instructional technology," *Journal of Special Education Technology*,Vol. 34, No.3, (2019), pp.204-211.

[17] National Institute of Standards and Technology ' National Vulnerability Database ', Available: https://nvd.nist.gov [Accessed: 10- June- 2021].

[18] Mittal, A., and Kassim, A," Bayesian network technologies: applications and graphical models: applications and graphical models", New York, IGI Global, 2007.

[19] Chockalingam, S., Pieters, W., Teixeira, A., & van Gelder, P. "Bayesian network models in cyber security: a systematic review," Nordic Conference on Secure IT Systems, Springer, Cham, (2017), pp.105-122.

[20] Pourret, O., Naïm, P. and Marcot, B. eds., "Bayesian networks: a practical guide to applications", England, John Wiley & Sons, 2008.

[21]Marek, J. D., Tomek, S., 'Bayes Fusion', 2020-11-05 Available: https://www.bayesfusion.com/genie [Accessed: 02- May- 2021].

[22] Guo, Ha. and William, Hsu. "A survey of algorithms for real-time Bayesian network inference," *Join Workshop on Real Time Decision Support and Diagnosis Systems*, (2002), AAAI Press, pp.1–12.

[23] Serkan, Q.,' credentials ', Available: https://www.cvedetails.com/about-contact.php [Accessed: 19- April- 2021].

Journal of Engineering and Computer Sciences
Qassim University, **Vol. 14, No. 1, pp. 64-81 (January 2021 /Rabi II 1441H)**

توقع مستوى درجة الثغرة في متصفح جوجل كروم باستخدام شبكة بايزين

مفوز ذويبان الحربي
قسم العلوم الطبيعية والتطبيقية – كلية المجتمع ببريدة – جامعة القصيم – المملكة العربية السعودية
Maft.alharbi@qu.edu.sa

ملخص البحث. ثغرة البرامج هي نقطة ضعف وخلل تؤدي إلى التأثير على بروتوكولات أمان النظام. أدت المشكلة المطروحة إلى زيادة نقاط الضعف في البرامج في كل إصدار يتم طرحه. يستخدم متصفح الويب من قبل المستخدمين لأنه الطريقة الأساسية التي يعتمد عليها جميع الأشخاص من أجل التواصل مع كل شخص عبر الإنترنت خلال الأجهزة. حيث يمكن للأشخاص العثور على محتوى رائع على الويب. تنشأ التحديات ضمن بروتوكولات الأمان عندما يصل المهاجمون إلى المستخدمين عبر متصفحات الويب. لذلك ، أصبحت متصفحات الويب هدفًا للمهاجمين الذين يستخدمون ثغرات متصفح الويب. يركز هذا البحث على الثغرات الأمنية في جوجل كروم ، ويقترح نهجًا جديدًا يمكن أن يوفر تنبؤات احتمالية تولد مستوى درجة من نقاط الضعف في متصفح الكروم . يستخدم المودل المقترح الذكاء الاصطناعي من نوع "شبكة بايزين".