Application of business intelligence for analyzing vulnerabilities to increase the security level in an academic CSIRT

Aplicación de Inteligencia de Negocios para el análisis de vulnerabilidades en pro de incrementar el nivel de seguridad en un CSIRT académico

Aplicação de Inteligência de Negócios para a análise de vulnerabilidades em prol de incrementar o nível de segurança em um CSIRT acadêmico

Francisco Xavier Reyes-Mena*
Walter Marcelo Fuertes-Díaz**
Carlos Enrique Guzmán-Jaramillo***
Ernesto Pérez-Estévez****
Paúl Fernando Bernal-Barzallo*****
César Javier Villacís-Silva******

Fecha de recepción: 5 de julio de 2017
Fecha de aprobación: 18 de noviembre de 2017

Abstract

This study aimed at designing a potential solution through Business Intelligence for acquiring data and information from a wide variety of sources and utilizing them in the decision-making of the vulnerability analysis of an Academic CSIRT (Computer Security Incident Response Team). This study was developed in a CSIRT that gathers a variety of Ecuadorian universities. We applied the Action-Research methodology with a qualitative approach, divided into three phases: First, we qualitatively evaluated two intrusion detection analysis tools (Passive Scanner and Snort) to verify their advantages and their ability to be exclusive or complementary; simultaneously, these tools recorded the real-time logs of the incidents in a MySQL related database. Second, we applied the Ralph Kimball's methodology to develop several routines that allowed applying the "Extract, Transform, and Load" process of the non-normalized logs that were subsequently processed by a graphical user interface. Third, we built a software application using Scrum to connect the obtained logs to the Pentaho BI tool, and thus, generate early alerts as a strategic factor. The results demonstrate the functionality of the designed solution, which generates early alerts, and consequently, increases the security level of the CSIRT members.

Keywords: business intelligence; cybersecurity; decision making; early alerts; electronic data processing; ETL; vulnerability analysis.

* Universidad de las Fuerzas Armadas ESPE (Sangolquí, Ecuador). fxreyes@espe.edu.ec.
** Ph. D. Universidad de las Fuerzas Armadas ESPE (Sangolquí, Ecuador). wmfuertes@espe.edu.ec. ORCID: 0000-0001-9427-5766.
*** M. Sc. Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia (Cuenca, Ecuador). carlos.guzman@cedia.org.ec.
**** M. Sc. Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia (Cuenca, Ecuador). enesto.perez@cedia.org.ec.
***** M. Sc. Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia (Cuenca, Ecuador). paul.bernal@cedia.org.ec.
****** M. Sc. Universidad de las Fuerzas Armadas ESPE (Sangolquí, Ecuador). cjvillacis@espe.edu.ec.

Resumen

Esta investigación tuvo como objetivo diseñar una solución para la toma de decisiones mediante Inteligencia de Negocios, que permite adquirir datos e información de una amplia variedad de fuentes y utilizarlos en la toma de decisiones en el análisis de vulnerabilidades de un equipo de respuesta ante incidentes informáticos (CSIRT). Este estudio se ha desarrollado en un CSIRT Académico que agrupa varias universidades miembros del Ecuador. Para llevarlo a cabo se aplicó la metodología de Investigación-Acción con un enfoque cualitativo, dividido en tres fases: Primera, se realizó una evaluación comparativa de dos herramientas de análisis de intrusos: Passive Vulnerability Scanner y Snort, que son utilizadas por el CSIRT, para verificar sus bondades y verificar si son excluyentes o complementarias; enseguida se han guardado los logs en tiempo real de los incidentes registrados por dichas herramientas en una base de datos relacional MySQL. Segunda, se aplicó la metodología de Ralph Kimball para el desarrollo de varias rutinas que permitan aplicar el proceso "Extraer, Transformar y Cargar" de los logs no normalizados, que luego serían procesados por una interfaz gráfica. Tercera, se construyó una aplicación de software mediante la metodología Ágil Scrum, que realice un análisis inteligente con los logs obtenidos mediante la herramienta Pentaho BI, con el propósito de generar alertas tempranas como un factor estratégico. Los resultados muestran la funcionalidad de esta solución que ha generado alertas tempranas y que, en consecuencia, ha incrementado el nivel de seguridad de las universidades miembros del CSIRT académico.

Palabras clave: alertas tempranas; análisis de vulnerabilidades; ETL; inteligencia de negocios; procesamiento electrónico de datos; seguridad cibernética; toma de decisiones.

Resumo

Esta pesquisa teve como objetivo desenhar uma solução para a tomada de decisões mediante Inteligência de Negócios, que permite adquirir dados e informação de uma ampla variedade de fontes e utilizá-los na tomada de decisões na análise de vulnerabilidades de um equipamento de resposta ante incidentes informáticos (CSIRT). Este estudo tem se desenvolvido em um CSIRT Acadêmico que agrupa várias universidades membros do Equador. Para realizá-lo, aplicou-se a metodologia de Pesquisa-Ação com um enfoque qualitativo, dividido em três fases: Primeira, realizou-se uma avaliação comparativa de duas ferramentas de análise de intrusos: Passive Vulnerability Scanner e Snort, que são utilizadas pelo CSIRT, para verificar seus benefícios e se são excludentes ou complementários; imediatamente são guardados os logs em tempo real dos incidentes registrados por ditas ferramentas em uma base de dados relacional MySQL. Segunda, aplicou-se a metodologia de Ralph Kimball para o desenvolvimento de várias rotinas que permitam aplicar o processo "Extrair, Transformar e Carregar" dos logs não normalizados, que logo seriam processados por uma interface gráfica. Terceira, construiu-se uma aplicação de software mediante a metodologia Ágil Scrum, que realize uma análise inteligente com os logs obtidos mediante a ferramenta Pentaho BI, com o propósito de gerar alertas precoces como um fator estratégico. Os resultados mostram a funcionalidade desta solução que tem gerado alertas precoces e que, em consequência, tem incrementado o nível de segurança das universidades membros do CSIRT acadêmico.

Palavras chave: alertas precoces; análise de vulnerabilidades; ETL; inteligência de negócios; processamento eletrônico de dados; segurança cibernética; tomada de decisões.

Para citar este artículo:
F. X. Reyes-Mena, W. M. Fuertes-Díaz, C. E. Guzmán-Jaramillo, E. Pérez-Estévez, P. F. Bernal-Barzallo, and C. J. Villacís-Silva, "Application of business intelligence for analyzing vulnerabilities to increase the security level in an academic CSIRT," *Revista Facultad de Ingeniería,* vol. 27 (47), pp. 21-29, Jan. 2018.

I. Introduction

Currently, universities and education centers are targets of cyber-attacks that focus on the alteration, extortion, and theft of sensitive information [1]. Due to such hazards, some questions arise: Do universities guarantee the confidentiality, integrity, and availability of information towards cyber threats? Does their technical staff maintain adequate security procedures to minimize vulnerabilities? Does the University have the ability to detect and respond to any cyber-attack?

For an adequate control of security incidents, organizational structures known as Computer Security Incident Response Team (CSIRT) or Computer Emergency Readiness Team (CERT) have been steadily implemented [2]. A CSIRT offers services such as analysis, coordination, support, and response to computer security incidents, based on an adequate vulnerability analysis [3]. However, when it comes to an academic CSIRT (A-CSIRT), the volume of the collected information might cause partial or total non-compliance of such services.

Based on the mentioned scenario, this study aimed at generating a novel technique that optimizes the extraction of malicious traffic, collected by intrusion detection and prevention systems in university networks in Ecuador. In order to comply with this purpose, we applied the Research-Action methodology [4]: (1) we compared Snort and Passive Vulnerability Scanner (PVS), two passive analysis tools that have been used in the CSIRT, to establish their benefits and check if they are exclusive or complementary; (2) we collected the data through extraction, transport and loading techniques (ETL); (3) we stored this information in a MySQL database; and (4) we designed an application based on Business Intelligence techniques to detect malicious events that may appear in the network, and thus act immediately.

Among the main contributions of this study is the generation of unpublished algorithms for ETL processes that allow transporting and filtering information, generating data of interest. In addition, we implemented an application using Pentaho BI [5] to generate a secure coupling.

II. Research design

This research is based on the conceptual framework illustrated in Fig. 1, and on the Ralph Kimball Methodology [6]. In an orderly way, first, we defined the requirements (upper part) and established the project planning; these two processes created the basis to determine the fulfillment of the proposed objectives. Subsequently, we developed the application, obtaining a Web product. These processes are further explained below.

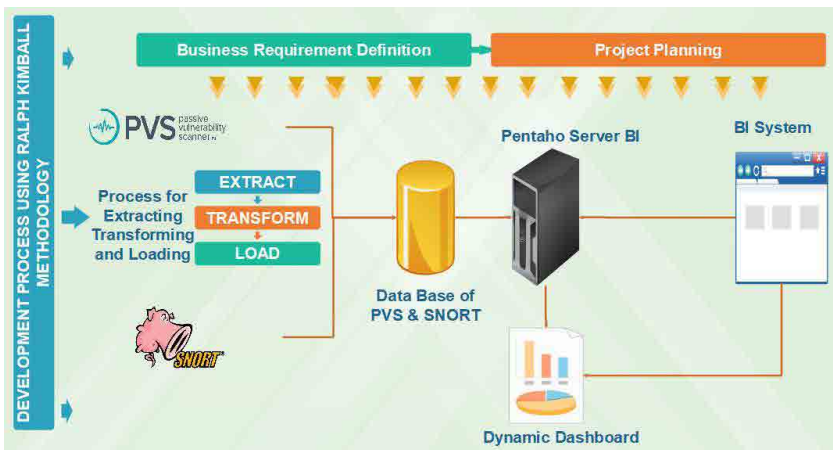

Fig. 1. Conceptual framework of this research.

A. Requirements definition

The Academic CSIRT in Ecuador is responsible for reporting computer incidents that the universities' network record and detect, and for counting with a vulnerability review service. Within this service, several points have been established such as frequent reviews of the institutions' public networks,

revisions of configuration files of routers and firewalls, configuration review, updates and security implemented in Linux and Windows servers, and policy compliance in database systems, servers, and routers, among others [7]. However, an effective automatic solution to detect and classify malicious events is still missing, due to the high volume of such events. For the CSIRT, it has been fundamental that its members have a tool that allows analyzing effectively the traffic that their network tracks, accomplishing quick decisions towards malicious events.

B. Evaluation of the traffic analysis tools

In this section, we will evaluate the Snort and PVS tools. Snort is an open source intruder detection and prevention system capable of real-time traffic analysis [8]. PVS is a product of the Tenable Network Security Company, which stands out for analyzing all the traffic transmitted within a network. PVS generates two files: the first intended to monitor real-time vulnerabilities, and the second focused on monitoring the Web and FTP activities [9]. We determined that PVS does not act as an IPS, meaning that it simply verifies the traffic that flows through the network, with no activity on the circulating packages.

To conduct an unbiased comparison, within similar conditions, we used data sets from the DARPA Intrusion, detection, and evaluation group at MIT Lincoln laboratory [10]. We analyzed the data in the alert file and the logs generated by Snort, as well as the real-time file and reports generated by the PVS. Based on this analysis we compared the results shown in Table 1.

Table 1
Comparative Analysis between Snort and PVS

Parameters	Tools	
	Snort	PVS
Virtual memory used (VIRT)	1.2 G	1.6 G
Used RAM (RES)	0.2 G	0.8 G
CPU percentage	0.3	6.0
Memory percentage	12.2	43.4
Analysis	Rules	Plugin
Detection	Attacks	Event - Vulnerabilities
Assessment	1-3 High 1 Medium 2 Low 3	1-10 Information 0 Low 0-3.9 Medium 4-6.9 High 7-9.9 Critical 10

Table 1 documents the specified consumption that the tools registered on a Centos7 distribution with a 64-bit architecture. The data demonstrated that Snort consumes less resources than PVS, mainly because Snort lacks the graphical user interface Web deployed by default in the system. In addition, the analysis in Snort was structured in "rules", in which the parameters were established subject to the packet to be captured. Furthermore, PVS analyzes the "plugin's", which provided a specification to be displayed when they do not capture the packets to be analyzed. Within the assessment, we established that the PVS lists more categories to classify the risk of the raised events in the network than Snort.

With the capture of FTP and HTTP traffic, we identified a clear difference in the volume of analyzed data between the two tools. This occurred because Snort acts with default rules. Therefore, the optimal performance of Snort was just the configuration of the rules with "Ruleset" that appeared from a variety of sources. This evaluation allowed inferring that the tools have not been exclusive, as they complement each other, providing a more forceful protection to the network.

C. Design and implementation of algorithms and the application of business intelligence

Here, we used the phases of the methodology proposed by Ralph Kimball [11], diving it into three sections: (1) specific software to analyze the benefits; (2) database design and structure of ETL; (3) specifications of the BI application and development of the product. Below, we further describe the developed process.

The ETL processes for data capture and filtering are fundamental for this study, because they represent the technique that allows organizations to move, reformat, and clean data from multiple sources, using SQL algorithms. ETL allow extracting, analyzing, and interpreting upcoming information. Nonetheless, the data formats may vary among organizations due to their source of origin. Therefore, in an unprecedented way, to homogenize the data from their sources, ETL algorithms have been designed, implemented, and subsequently loaded into a MySQL, DataMart, or Data Warehouse database to submit them to a business process [11]. In this study, the information from the Snort and PVS registers was relatively condensed and had no standardized format. However, the created ETL algorithms solved this issue, preventing to overload the database with potential irrelevant information.

We developed three solutions related to the PVS processes, based on vulnerability analysis, real-time activities, and risk filtering. Each solution processes its information from files, since the vulnerabilities have been encountered in an enriched ".XML" (.nessus) [12], and both the real-time files and the filtering work in a ".txt" file. We generated a transformation flow for the vulnerabilities, which originated from the PVS system (Fig. 2). Also, we generated flows that support the optimization of the processed packages, avoiding generating redundant or delayed data. Figure 3 illustrates the flow that which handles a "Work" generated by the Pentaho BI data integration system; in this flow, the transformation was executed establishing a parameter of repetition, and highlighting the error notification via e-mail, allowing for a subsequent call to an execution process, which manages to filter the risks according to the established parameters.

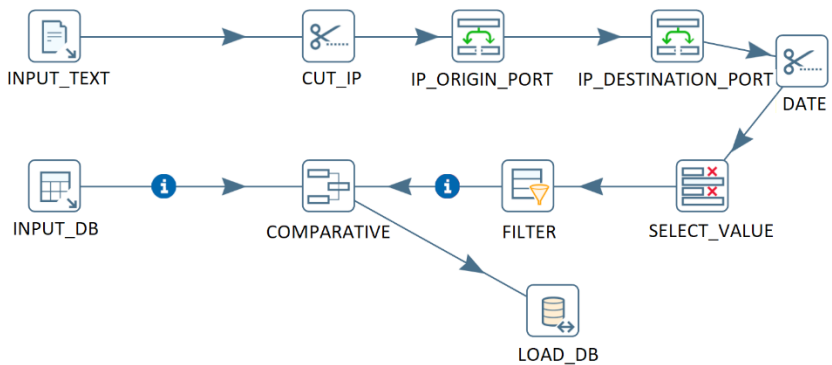

Fig. 2. ETL algorithm transformation from a flat text file.

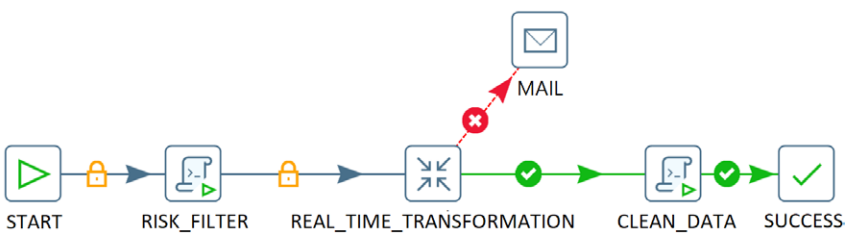

Fig. 3. A generation algorithm of an information filter with control of possible errors in the transformation.

Within the vulnerability report, the events transformation flow occurs in real time (Fig. 4). In such programming, a control of the events was established, specifying a classification that was stored in MySQL database. Like the vulnerabilities, the control of the flow transformation was highlighted by a warning via e-mail (Fig. 5).

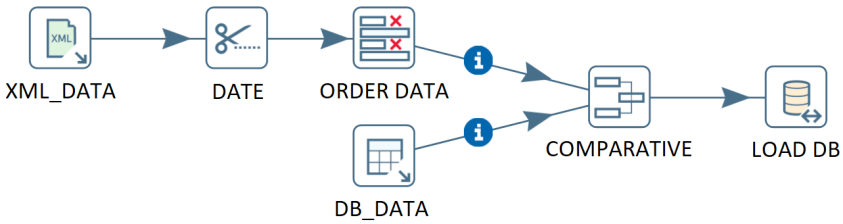

Fig. 4. A Transformation algorithm due to information extraction from an enriched XML file.

Snort is a tool with the complementary system Barnyard2, which allows migrating the alerts detected by the IDS to a MySQL database (Fig. 6 and 7). Hence, we adapted the fields to generate a relational model. For filtering, the transformation was based on programming in Pentaho BI (Spoon Data Integration) [13], which is based on the content registered in a file, allowed to catalog the obtained packages with Snort as relevant; this prevented a potential database overload with records that lack interest.

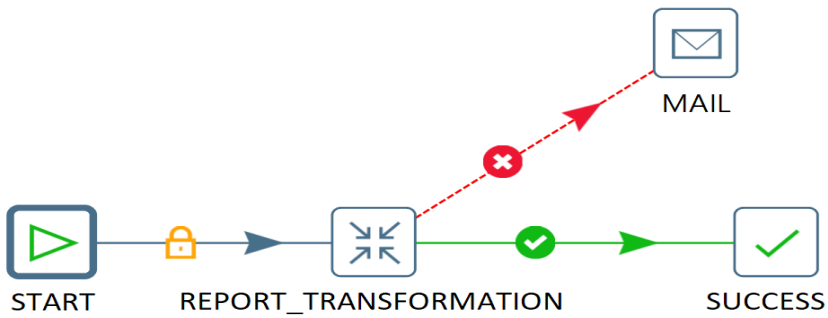

Fig. 5. A generation algorithm of the data load of any given time, controlling possible errors that may arise during the transformation.

In agreement with Stephen Few [14], who defines a Dashboard as a scorecard in which a dense array of information must be demonstrated in a reduced manner, we used the Pentaho BI with the Community Dashboard Framework (CDF) guidelines [15]. In this way, to visualize the data collected from the PVS, we developed two control panels: a first panel destined to the events in real time, and a second panel used for the vulnerability analysis. For Snort, we developed a panel that allowed visualizing a matrix with the latest detected events. Such information complements the real-time events registered in the PVS.

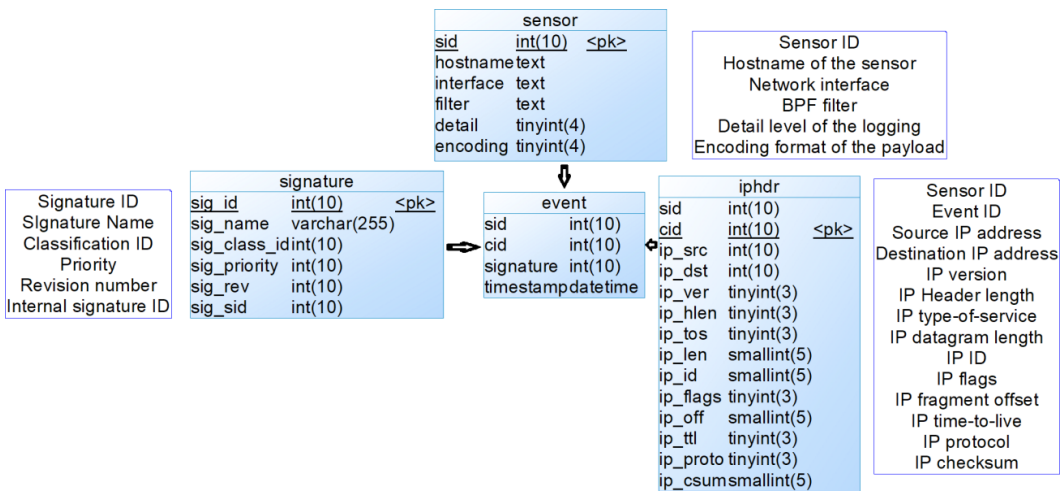

Fig. 6. Snort database diagram with the most significant tables.

Regarding the implementation of the application, we applied Scrum as a methodology to develop agile projects because it is independent of technologies and allows coupling to different programming models, focuses on people, as users and developers have defined roles, provides quick results, considers time management, is iterative, responds to changes, and the client is active [16].

We used Node.js, a platform created on Chrome's JavaScript Runtime [17] as programming language, since it allows implementing fast and scalable network applications. Node.js uses an *event-driven, non-blocking I/O* model that performs lightly and efficiently, being suitable for *data-intensive* applications in real-time.

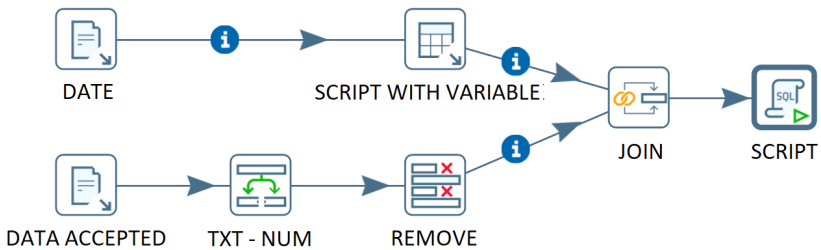

Fig. 7. Transformation algorithm that generates a filter from parameters in a flat file.

III. Results

A. Proof of concept

The environment and the flow where the concept was proven (Fig. 8) are represented by 1) the generation of a request for Dashboard or Access to the Pentaho BI system; 2) the system generating the queries to the MySQL database; 3) the system generating the control panel that will be displayed in case of a Dashboard request, while an access has been requested, depending on the user's profile and taking into account that the Dashboard may be edited; 4) the requested Dashboard being displayed in the BI System; and 5) the BI System allowing impression, as well as submission of the data to Excel, generating a proper record file.

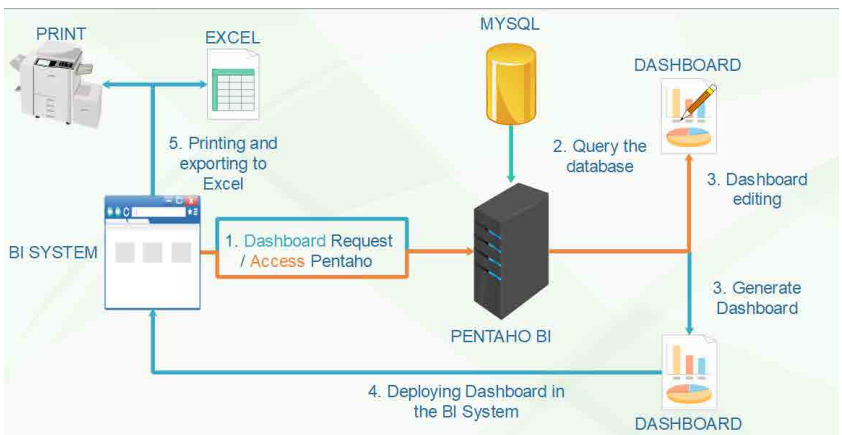

Fig. 8. Experimental diagram of the proof of concept.

B. Results assessment

The application was verified in a real-time trial period, in which the users were able to browse the entire network. Thus, for example, in the records of the last week of June, we could determine that the day with most events was the 27th (Fig. 9). Fig. 10 illustrates a variety of events recorded in the trial period. The most common event was the "ET SCAN Potential SSH Scan", which generated multiple connections to the server where the product was previously installed.

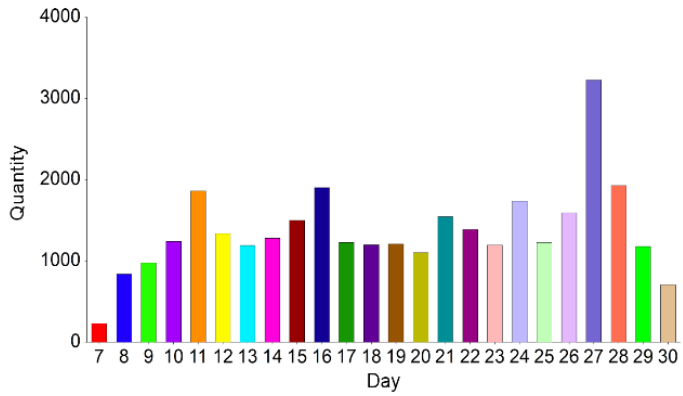

Fig. 9. Number of events per day.

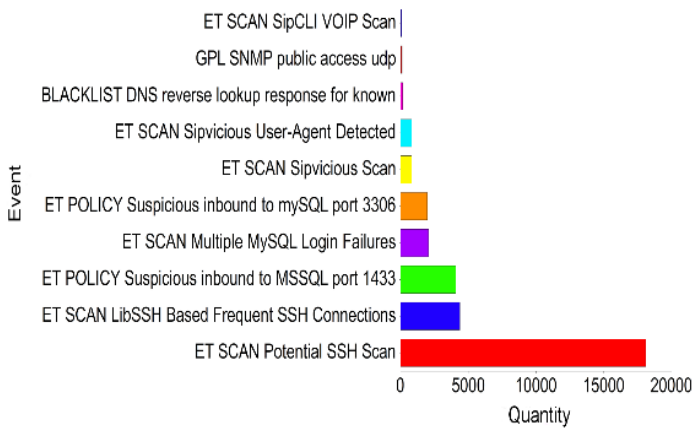

Fig. 10. Type of events and frequency.

The maximum vulnerability tracking point in the protected network was obtained on day 14 (Fig. 11). Afterwards, the flow became normal, with few days with fewer and others with more vulnerabilities; consequently, we can infer that control actions were taken on the observed vulnerabilities.

During this analysis, the false positives decreased, which implies an improvement in the fulfilment of the Academic CSIRT services. Therefore, the system focused on supporting the Academic CSIRT to improve its services for detecting vulnerabilities and tracking malicious activity on the network; this aims at uprising awareness among members to lead to a more efficient deal with incidents occurring in their network, thus achieving adequate and timely attention to their weaknesses and threats.

IV. Conclusions and Future Work

In the current study, we designed a solution implemented through Business Intelligence, which acts as a strategic factor in the vulnerability analysis of an Academic CSIRT. This was possible by applying the Action-Research methodology and the phases of Ralph Kimball. The evaluation of Passive Scanner and Snort offered security management based on network traffic and customization of its configurations, to reduce false positives and thus enhance the response to security incidents. We developed several algorithms to apply the ETL process of the non-standardized logs that the graphical interface processed. Finally, we built a software application using Scrum, which allowed linking the obtained logs in Pentaho BI to generate early alerts of vulnerabilities and malicious codes. The results demonstrate that such application has managed to help those responsible for the CSIRT to establish

immediate priorities and to allocate resources to key areas, which may be potential victims of digital attacks. Big Data techniques provide scalability in high data volume scenarios in the CSIRT, therefore, we suggest applying them in future studies.

Author's Contributions

Francisco Xavier Reyes designed and implemented the algorithms, and conducted the experiments. Walter Marcelo Fuertes analyzed the state of the art and edited the paper. Carlos Enrique Guzmán supervised the project and obtained funding. Ernesto Pérez Estévez and Paúl Fernando Bernal provided the information and technological infrastructure for proof of concept. César Javier Villacís Silva developed the BI system.

References

[1] M. Letho, "Cyber Security Education and Research in the Finland's Universities and Universities of Applied Sciences," *International Journal of Cyber Warfare and Terrorism* (IJCWT), vol. 6(2), pp. 15-31, Apr. 2016. DOI: http://doi.org/10.4018/IJCWT.2016040102.

[2] P. Cichonski, T. Millar, T. Grance, and K. Scarfone, "Computer security incident handling guide," *NIST Special Publication 800-61*, 2012

[3] M. West-Brown, *et al*. "Handbook for computer security incident response teams (CSIRTS)," No. CMU/SEI-2003-HB-002. Carnegie-Mellon Univ Pittsburgh PA software engineering institute, 2003.

[4] P. Coughlan, and D. Coghlan, "Action research for operations management," *International journal of operations & production management, vol.* 22(2), pp. 220-240, 2002. DOI: http://doi.org/10.1108/01443570210417515.

[5] R. Bouman, and J. V. Dongen. *Pentaho solutions: Business Intelligence and Data warehousing with Pentaho and MySQL*. Wiley Publishing, 2009.

[6] R. Kimball, M. Ross, J. Mundy, and W. Thornthwaite. *The kimball group reader: Relentlessly practical tools for data warehousing and BI remastered collection*. John Wiley & Sons, 2015. DOI: http://doi.org/10.1002/9781119228912.

[7] P. Valladares, W. Fuertes, F. Tapia, T. Toulkeridis, and E. Pérez, "Dimensional data model for early alerts of malicious activities in a CSIRT," in *International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)*, Seattle, 2017. DOI: http://doi.org/10.23919/SPECTS.2017.8046771.

[8] R. Gaddam, and M. Nandhini, "An analysis of various snort based techniques to detect and prevent intrusions in networks proposal with code refactoring snort tool in Kali Linux environment," in *International Conference on Inventive Communication and Computational Technologies (ICICCT)*, Coimbatore, 2017. DOI: http://doi.org/10.1109/ICICCT.2017.7975177.

[9] S. Dongkyun, and K. Lee, "Comparing security vulnerability by operating system environment," *International Journal of Services Technology and Management*, vol. 23 (1-2), pp. 154-164, 2017.

[10] H. Elshoush, and I. Osman, "An improved framework for intrusion alert correlation," *Proceedings of the World Congress on Engineering*, vol. 1, 2012.

[11] R. Kimball, and R. Margy, *The data warehouse toolkit: The definitive guide to dimensional modelling*. John Wiley & Sons, 2013.

[12] I. Sharafaldin, *et al.*, "Towards a Reliable Intrusion Detection Benchmark Dataset," *Software Networking*, vol. 1 (1), pp. 177-200, 2017. DOI: http://doi.org/10.13052/jsn2445-9739.2017.009.

[13] J.L Pereira, and M. Costa, "Decision Support in Big Data Contexts: A Business Intelligence Solution," *New Advances in Information Systems and Technologies*, vol. 444, pp. 983-992, 2016. DOI: http://doi.org/10.1007/978-3-319-31232-3_93.

[14] S. Few, "Information Dashboard Design. The Effective Visual Communication of Data," NY: O'Reilly, 2006.

[15] M. S. Gounder, V. V. Iyer, and A. A. Mazyad, "A survey on business intelligence tools for university dashboard development," in *3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, Muscat, 2016. DOI: http://doi.org/10.1109/ICBDSC.2016.7460347.

[16] J. Pajares, *et al.*, "Project Management Methodologies in the Fourth Technological Revolution," *Advances in Management Engineering*. Springer International Publishing, pp. 121-144, 2017. DOI: http://doi.org/10.1007/978-3-319-55889-9_7.

[17] R. O'Connor, V. Elger, and P. Clarke. "Continuous software engineering—A micro services architecture perspective," *Journal of Software: Evolution and Process*, vol. 29 (11), pp. e1866, Nov. 2017. DOI: http://doi.org/10.1002/smr.1866.