Anomaly Detection Under a Cognitive Security Model

Detección de Anomalías Bajo un Modelo de Seguridad Cognitiva

ARTICLE HISTORY

Received 01 October 2020
Accepted 16 October 2020

Jonathan Herrera
Departamento de Ciencias de la Computación
Escuela Politécnica Nacional
Quito, Ecuador
jonathan.herrera01@epn.edu.ec

Roberto Omar Andrade
Departamento de Ciencias de la Computación
Escuela Politécnica Nacional
Quito, Ecuador
roberto.andrade@epn.edu.ec

Miguel Flores
Departamento de Matemáticas
Escuela Politécnica Nacional
Quito, Ecuador
miguel.flores@epn.edu.ec

Susana Cadena
Facultad de Ciencias Administrativas
Universidad Central del Ecuador
Quito, Ecuador
scadena@uce.edu.ec

J. Herrera, R. Andrade, M. Flores and S. Cadena, "Anomaly Detection Under a Cognitive Security Model", Latin-American Journal of Computing (LAJC), vol. 7, no. 2, 2020.

Anomaly Detection Under a Cognitive Security Model

Detección de Anomalías Bajo un Modelo de Seguridad Cognitiva

Jonathan Herrera
Departamento de Ciencias de la Computación
Escuela Politécnica Nacional
Quito, Ecuador
jonathan.herrera01@epn.edu.ec

Roberto Omar Andrade
Departamento de Ciencias de la Computación
Escuela Politécnica Nacional
Quito, Ecuador
roberto.andrade@epn.edu.ec

Miguel Flores
Departamento de Matemáticas
Escuela Politécnica Nacional
Quito, Ecuador
miguel.flores@epn.edu.ec

Susana Cadena
Facultad de Ciencias Administrativas
Universidad Central del Ecuador
Quito, Ecuador
scadena@epn.edu.ec

Abstract — Cybersecurity attacks are considered among the top five of risks worldwide, according to the World Economic Forum in the year 2019. This context has generated the need to improve the tasks of cybersecurity defense in organizations. Improving the effectiveness in executing a cybersecurity task requires three pillars: people, processes and technologies. The proposal in this work is to analyze the integration of these three components as a strategy to improve the effectiveness of the execution of operational tasks in cyber defense, specifically the detection of anomalies. Based on the foundation that: cybersecurity operational tasks carried out daily by analysts require the use of cognitive processes, and that the use of techniques based on technologies such as machine learning, data mining and data science have generally been used to automate cybersecurity tasks, we have considered the use of cognitive security, as a strategy to improve the anomaly detection process, taking into account the cognitive processes and skills that are executed by the security analyst.

Keywords — cyber-defense, cognitive security, cybersecurity.

Resumen — Los ataques de ciberseguridad están considerados entre los 5 principales riesgos alrededor del mundo, de acuerdo al World Economic Forum en el año 2019. Este contexto ha generado la necesidad de mejorar las tareas de defensa de ciberseguridad en las organizaciones. Mejorar la efectividad en ejecutar una tarea de ciberseguridad requiere tres pilares: personas, procesos y tecnologías. El propósito de este trabajo, es analizar la integración de estos tres componentes como una estrategia para mejorar la efectividad de la ejecución de tareas operacionales de ciberdefensa, específicamente la detección de anomalías. Basados en el fundamento de: que las tareas operacionales de ciberseguridad llevadas a cabo diariamente por analistas, requieren de procesos cognitivos, y que el uso de técnicas basadas en tecnologías como el aprendizaje de máquina, la minería de datos y la ciencia de datos han sido usadas generalmente para automatizar las tareas de ciberseguridad como la detección de anomalías, hemos considerado el uso de la seguridad cognitiva, como una estrategia para mejorar el proceso de detección de anomalías, teniendo en cuenta los procesos y habilidades cognitivas que son ejecutadas por el analista de seguridad.

Palabras clave — ciber-defensa, seguridad cognitiva, ciberseguridad

I. INTRODUCTION

The development of computer applications and emerging technologies, such as smart cities or industry 4.0, have optimized the technology services provided by organizations around the world. Additionally, the improvement of decision-making processes in real time has become a challenge for security managers. Furthermore, cyber security problems have been increasing constantly; attacks on system availability, information theft and loss of privacy have risen in recent years, according to the World Economic Forum [1].

In this context, the need to improve the cognitive skills of security analysts, and the use of technological tools to support cyber operations has increased. From our perspective, a direct relationship between the development of technological solutions and the skills of security analysts is important. Having an effective tool, but lacking a person capable of handling it, would limit the effectiveness of said tool, and on the other hand, an analyst with extensive experience who does not have effective tools would be limited in his or her ability to perform cybernetic operational tasks.

In this study, it is our interest to analyze the importance of cognitive security regarding anomaly detection. Based on the literature review carried out in a previous work [2], "Self-Awareness as Enabler of cognitive Security", cognitive security is considered as: the ability to generate cognition in order to take decisions effectively in real time, by either a human or a computational system. This ability is based on the perception generated by the system from its environment and the knowledge about itself, which is obtained from the analysis of any type of information, structured or unstructured, by using artificial intelligence techniques and data analysis. Therefore, the system can emulate the human thought process for continuous learning, decision making and, ultimately, security analysis.

Given the fact that we intend to work with enormous amounts of data, we propose the use of the Elastic Stack (ELK). ELK is a set of tools that aim to help improve the indexation, searching and manipulation of the data [3]. In this study, we implement the ELK stack in order to index and analyze the data obtained from the IoT-23 dataset, which were downloaded directly from [4]. This data is analyzed in order to asses the connection and future development of a cognitive model, either supervised or unsupervised, which could be used to classify the data in a more reliable way. The results obtained showed a clear distinction between the behavior of the malicious traffic and the benign traffic, which facilitates the development of a future cognitive model based on the data.

II. THEORETICAL FOUNDATION

There are a set of security operational tasks that are developed daily by security analysts and that requires of the use of cognitive skills for their execution. These tasks were indicated by IBM at the RSA conference in 2017 [5]. On previous related work, we had grouped these tasks on base of cognitive processes and the Observe–Orient–Decide–Act (OODA) cognitive model, proposed by Breton (See Table I, Table II and Table III) [6-7].

TABLE I. OODA COGNITIVE PHASE AND COGNITIVE PROCESS

Phase	Process
Observe	Feature matching /Perception
Orient	Mental models / Comprehension
Decide	Evaluating / Projection

TABLE II. COGNITIVE PROCESS AND ITS ATTRIBUTES.

Process	Attribute
Perception	Identification of relevant data
Comprehension	Interpretation of data
Projection	Prediction of future events

The first two cognitive tasks "Review incident data" and "Review the events by aspect of interest" require the cognitive skill of perception by security analyst. The third task "Pivot in the data to find atypical values, or outliers" requires the cognitive process called comprehension. One of the daily operations carried out by security analysts, and related with these three tasks, is anomaly detection, i.e., the analyst's ability to identify outliers that could be possible security attacks, based on spikes in data traffic or by connections from uncommon services. This detection could allow the security analyst to take the appropriate actions, and reduce the impact of the attack. Within cybersecurity, one of the concerns of security managers are anomalies, which are defined as the detection of surprising or unusual events [5]. There are several techniques to detect anomalies, which use a range of different methods, however, it should be considered that, when the malicious actions are caused by perpetrators, the abnormal observations tend to be adapted in order to appear normal. Another major problem is the availability of labeled data, which is needed for training and validating the computational models. To determine an anomaly, several factors should be considered such as: the nature of the input data (object, record, point, vector, pattern, event, case, sample, observation or entity), the availability or not of labels, as well as the limitations and requirements induced by the domain of the application [6].

TABLE III. COGNITIVE PROCESS RELATED TO CYBER-ATTACKS

Cybersecurity operations tasks	Cognitive process
Review incident data	Perception
Review the events by aspects of interest	Perception
Pivot in the data to find atypical values or outliers	Comprehension
Expand the search to find more data	Projection
Investigate the treath to develop experience	Comprehension
Discover new treaths	Projection
Determine indicators of commitment in other sources	Comprehension
Apply intelligence to investigate the incident	Projection
Discover potentially infected IPs	Comprehension
Qualify the incident based on the knowledge generated while investiganting the threat	Comprehension
Prescriptive analysis based on the profile of the attack	Comprehension
Analysis of the lessons learned, based on the dispersion map of the attack	Comprehension

Riveiro mentions a five-phase process related to the anomaly detection process [8]:

1. Overview: continuous traffic control in real-time.
2. Filter: define if something is abnormal (make a judgment based on their experiences).
3. Waiting Time: in which the event is observed in order to evaluate its behavior.
4. More detail: the situation is analyzed in more detail.
5. Taking-action: define the best responsive action.

The first

The first step according to [8], is establishing the potentially dangerous situation, where the analyst should compare it, and update it, according to what would be the normal situation in the supervised zone. The second step correspond to narrowing the set of items, where the analyst zooms in and out the data to detect anomalous behavior, comparing it with real-time information. In this step, [8] suggests the use of data-mining techniques to decrease the time needed to identify anomalous situations. The third step is critical, because the analyst leaves the current traffic, in order to evaluate the status of the situation. This step is stressful for the analyst, due to the potential increase of traffic, or any impact in the services of the supervised zone. These three steps, defined by [8], require the analysts' cognitive processes, which we have grouped together in this work within the cognitive process called "perception". Increasing the analyst's cognitive skills can be carried out by several lines of action:

- Training,
- Experience gained from work,
- Technological solutions, or
- Cognitive security

III. ANOMALY DETECTION BASED ON COGNITIVE SECURITY

The objective of this work is to define the anomaly detection process by the security analyst, in order to translate it to a cognitive security model. The first cognitive process, used by an analyst for the detection, is the perception. The analyst must have the ability to detect outliers, by reviewing the data generated in the ecosystem. Based on the three steps proposed by [8], we propose in Figure 1, an anomaly detection process based in a cognitive model.

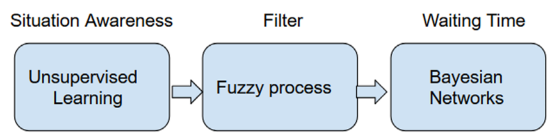


Fig. 1. Anomaly detection process based on a cognitive model

The security analyst first establishes an awareness of the situation. According to [9], unsupervised models are more

efficient and appropriate for the detection of abnormalities, compared to human analysts. The use of supervised algorithms requires the establishment of labels, and continuous training by a security analyst. However, new security attacks may not be identified by the analyst, which would limit the effectiveness of the algorithm. In this context, unsupervised algorithms can have a greater advantage.

A security analyst requires three cognitive functions in order to execute anomaly detection:

1. The ability to detect patterns and outliers from a dataset.
2. The ability to identify the time when an event occurred based on timestamps.
3. The ability to establish a geographical or spatial locality from where the attack was originated.

So, spatial and temporal reasoning are essential in the field of cybersecurity. According to [10], temporal and spatial representations are not independent between them, and different modalities are sensitive to the processing of time and space. [10] analyzes the hypotheses on the symmetric and asymmetric relationships of temporal and spatial information, and how they affect the decision-making process. In his study, he mentions that the relationship between these attributes can affect the time taken to make a judgement. Additionally, he mentions that visual analysis is more sensitive to space information that to temporal information.

In Figure 2, we show some of the attributes that could be considered as part of the temporal reasoning done by the security analyst, in order to perform an anomaly detection process. Based on the arguments of the cognitive processes proposed by [10], the use of algorithms must consider the space-time attributes, in order to improve the duration and effectiveness in the decision-making process. This is particularly important for a security analyst, when trying to identify anomalies. Additionally, the scheme using unsupervised algorithms could be hybrid, given the fact that there exists a sufficiently wide range of algorithms that can be used for this context. In Figure 3, we present an overview of some unsupervised algorithms, which could be used for space-temporal analysis in anomaly detection, among cybersecurity domains.

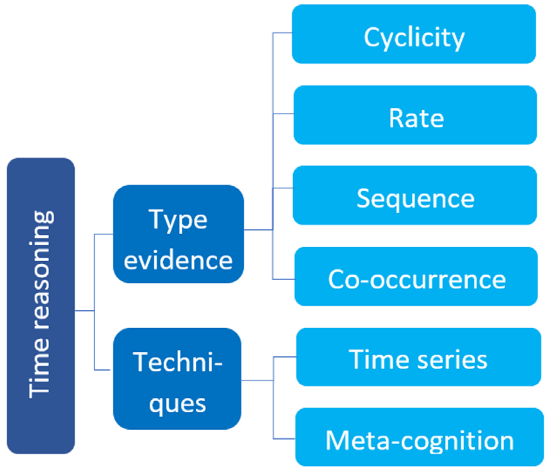


Fig. 2. Temporal Reasoning in cybersecurity anomaly detection

Some unsupervised algorithms have been used in anomaly detection before. For example, hierarchical k-means algorithms improve the clustering performance in large datasets [11]. Additionally, partition algorithms allow to reduce the complexity of time and space produced by hierarchical density [12]. There are also grid-based algorithms, which allow the detection of frequent spatial patterns [13]. These grid structures, that use space partitioning that converts space into smaller intervals, help in reducing computational overhead. Finally, in density-base clustering data, objects are categorized according to their regions of density, connectivity and boundaries [14].

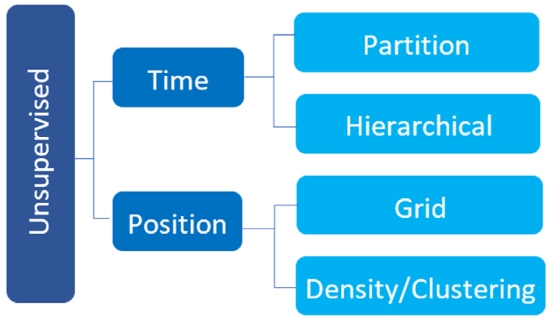


Fig. 3. Unsupervised algorithms for anomaly detection

The second step, according to [8] for anomaly detection is related to the filtering process. Anomaly detection in this step is based on the analyst's experience, so there is a high degree of subjectivity. Furthermore, the filtering process requires the processing of collected data, including their cleaning and profiling; therefore, is common in this type of process to select a methodology such as CRISP-DM, which has been widely used in Data-Mining projects. In this study, we have established the premise that, it is not feasible for the security analyst to have knowledge of all possible attacks, and that there are new attack vectors every day; the security analyst, therefore, starts from a state of uncertainty and lack of relevant information. From this point, the security analyst must build the necessary knowledge for decision making.

Regarding uncertainty, [15] mentions that uncertainty is a property of human in his/her mental processes. In this context, cognitive uncertainty is caused by the incompleteness of the important knowledge required (vague events), where fuzziness is also included as a type of non-statistical uncertainty. [15] also mentions that, categorization, is an essential function of the cognitive system when performing important processes, such as reasoning and problem solving. She proposes a fuzzy membership function for categorization, where the function measures the degree of similarity of different objects. We can use fuzzy techniques to reduce the subjectivity of the security analysis. The fuzzy techniques could measure the relationship between two vague events: when an attack occurs (time), and the source of the attack (space). So, a fuzzy technique must be aligned with the temporal-spatial principle.

In the third step, there is a need to reduce the waiting time and select the best option, considering most of the possible scenarios. The use of Bayesian networks, with a cause-effect scheme, could support a security analyst in order to help him/her consider better alternatives. Additionally, Bayesian networks are effective while modelling large datasets, where spatial and temporal information might be included. Furthermore, the Integrated Nested Laplace Approximations (INLA) approach, is a computationally effective and an extremely powerful alternative to Monte Carlo (MCMC) methods.

IV. EXPERIMENTATION

The architecture, implemented for this study, is made up of three nodes. These 3 nodes have specific functions and will be known from now on as "collector", "index" and "reporter". The "collector" node has an Intel (R) Xeon (R) E5-2640 v4 2.40GHz processor, 4GB of DDR4 ECC memory and a 200GiB NUTANIX VDISK disk. The "index" node has the same processor and memory characteristics, with the same disk model, but with 500GB capacity. Similarly, in the "reporter" node, there are the same characteristics in processor and memory, with storage of 100 GB. All three nodes are running the Ubuntu 18.04.4 operating system. Additionally, all the nodes are connected through an Ethernet network, as shown in Figure 4, with a firewall that only allows the passage for port 22, to be able to connect using the SSH protocol. The functions of each node are differentiated for each of the components of the ELK stack. In this way, the "index" node is in charge of keeping Elasticsearch services running, indexing the data and searching for them. The "collector" node is in charge of separating, organizing and coding the information for its correct indexing within the "index" node. Finally, the "reporter" node is in charge of keeping the Kibana service running, to allow access and corresponding visualization of the data that are indexed in Elasticsearch. This node is in charge of maintaining a cache with the data that need to be shown in each query, in order to facilitate viewing and searching in real time.

To allow the visualization of the Kibana service from a computer external to the network, the "reporter" node has been provided with a public IP address, where the ports 80 and 443 are open to receive requests. The aforementioned ports are routed through a proxy, implemented using the Nginx service, which allows us to redirect incoming traffic to the corresponding Kibana port This port (5600) has been arbitrarily chosen for communication, the default port (5601) would also have the same functionalities.

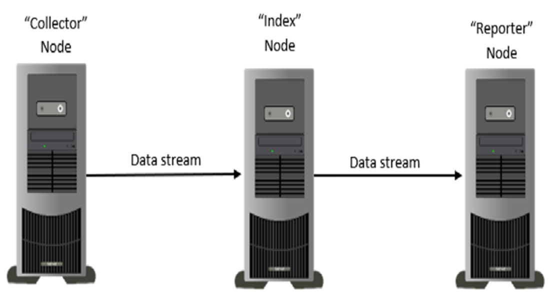


Fig. 4. Big Data Architecture

ISSN: 1390 -9266 – 2020 LAJC

A. Configuration

Los algoritmos de OCR permiten convertir el texto en una imagen a caracteres (códigos de 8 bits, como los representados en tablas ASCII). El motor o algoritmo OCR adoptado es el motor Tesseract OCR.

Index node

In the "index" node, Elasticsearch 7.9.0 was installed as indicated in [16], through a Debian package, which was downloaded directly from the Elastic servers. Once the node was installed, we proceeded to configure the elasticsearch.yml file, which is located in the /etc/elasticsearch directory. In this file, the name of the cluster, the name of the node, the address for the data, the address for the records and the port where it will receive HTTP requests were configured. Additionally, x-pack was configured, following the steps indicated in [17]. For this purpose, the variable xpack.security.enabled was set to true in the elasticsearch.yml file, in order to enable the x-pack configuration for within the Elasticsearch instance. The x-pack service contains essential users included; whose respective passwords were configured by running the elasticsearch-setup-passwords service. This service was located in the elastic installation folder, and it was run using the interactive option.

Reporter Node

In the "reporter" node, the Kibana 7.9.0 service was installed, following the steps indicated in [18]. Similarly, a Debian package was used, which was downloaded from the Elastic servers. Once Kibana was installed, the kibana.yml file was configured, which is located in the /etc/kibana folder. The options configured in this file were: the server port, the host address, the elasticsearch host address, the elasticsearch listening port, the existing username of "kibana" for elasticsearch and the password that was configured in a previous step. The Kibana service was started, and kept running in background mode.

Collector Node

The Logstash 7.9.0 service was installed in the "collector" node, following the steps specified in [19]; likewise, the installation was carried out with the use of a Debian package, downloaded from the Elastic servers. Once the installation was ready, we proceeded to configure only the location of the data in the logstash.yml file, since the rest is configured in the configuration file of the corresponding pipeline. Additionally, in order to connect Logstash to Elasticsearch using x-pack, a specific user was first configured for Logstash, as specified in [19]. This user was configured using Kibana's interface, where an Index Pattern called logstash- and a user role called "logstash_writer" were created, with the following privileges:

- Cluster: "monitor" and "manage_index_templates"
- Index: "write", "create", "delete" and "create_index" in the indexes that meet the pattern logstash- *

Finally, a user named logstash_internal was created, with the role of logstash writer, and a password was configured for its access.

B. Data Indexing

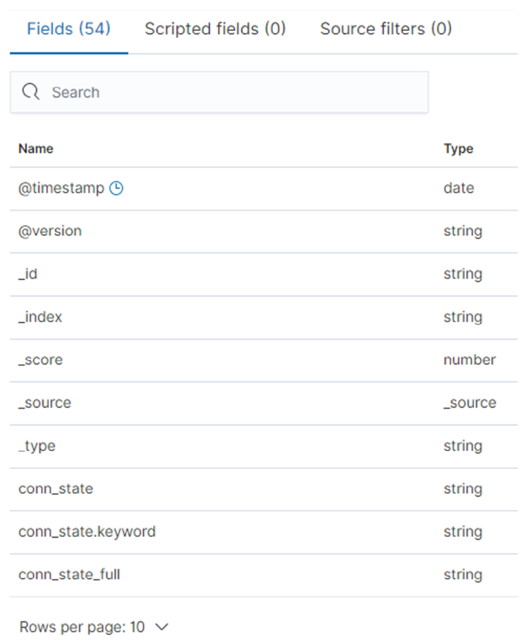
The data that were used belongs to the IoT-23 dataset, which is obtained from PCAP files and transformed into a connection log by using Zeek, which gives them a high-level format. These files were downloaded directly from the repository mentioned in [4]. In order to index them, a .conf file was used, located in the /etc/logstash/conf.d/ folder; the configuration of this file was done according to what is indicated in the documentation for logstash 7.9.0 [20]. In the input section, the location of the x-pack files within the file plugin was configured, which is part of the logstash plugins and is specified in [20]. Additionally, the start_position variable was set to "beginning" and the "sincedb_path" variable to "/dev/null". This was done so that logstash reads the entire file instead of using the default configuration, which only reads the last lines and monitors changes.

The filter section of the configuration file was started by checking the received rows with "if [message] = ~ / ^ # /", which then executes the "drop" plugin, specified in [21]; this is done to filter those lines that are comments in the obtained record. Next, the Grok plugin was used to filter the rows and separate the data into the corresponding attributes, assigning each attribute the corresponding value as indicated in the header of the log file. This Grok filter was tested using the filter test function that is available among the Kibana plugins. Additionally, the "translate" plugin was used to obtain an additional attribute, which better explains the connection status from the codes that are given in the file. Later, making use of several conditionals, the values of '-' were replaced in the following attributes: "id.orig_p", "id.resp_p", "orig_bytes", "resp_bytes", "missed_bytes", "orig_pkts", "orig_ip_bytes", "resp_pkts", "resp_ip_bytes " and "duration". Additionally, the aforementioned values were converted to "integer" using the "mutate" plugin, specified in [21], with the exception of the "duration" attribute, which was transformed to "float". Finally, the last plugin used in this section is the date plugin, which is detailed in [21], where it was indicated that the time value is in the "ts" field, and that it is formatted in UNIX time.

In the output section, the "Elasticsearch" plugin was used, which is detailed in [22]. which was configured with the address of the "index" node and the Elasticsearch port. In this section, the user created previously for the use of Logstash was entered, along with its respective password. Additionally, in this section the location of a "template" that should be used for the creation of the index was indicated, specifying its name and that it should not be overwritten. In this template, which is structured as indicated in [23], we proceeded to specify the "mapping" of the index fields. This mapping was used to specify the data type for all the values that were previously encoded in numeric format. The Logstash service was then started using the "sudo systemctl start logstash" command, to run in background mode. Finally, the index was modified through the Elasticsearch API to reduce the number of replicas to zero, because there is only one node working with Elasticsearch and there is no possibility of implementing replicas of it.

V. RESULTS AND DISCUSSION

Once the data had finished loading into the index, an Index Pattern was created in Elasticsearch from Kibana. This pattern was used to verify the loaded data. As can be seen in Figure 5, the loaded data has been correctly indexed, with the format established in the template that was previously defined. With this Index Pattern, a dashboard was created, allowing a clearer view of the data that was indexed. To create the dashboard, the Kibana tool was used, and six visualizations were obtained.



Fields (54) Scripted fields (0) Source filters (0)

Name	Type
@timestamp	date
@version	string
_id	string
_index	string
_score	number
_source	_source
_type	string
conn_state	string
conn_state.keyword	string
conn_state_full	string

Rows per page: 10

Fig. 5. Data according to Index pattern.

In Figure 6, the first visualization can be seen, which consists of the average size of a request and a response according to the type of traffic. It can be clearly seen that there are marked differences between benign and malicious traffic. In the case of the request, this tends to be smaller for benign traffic; while the answer is larger for the malicious traffic. This can be easily interpreted because the attackers tend to request information from the device that they are targeting in order to understand its vulnerabilities.

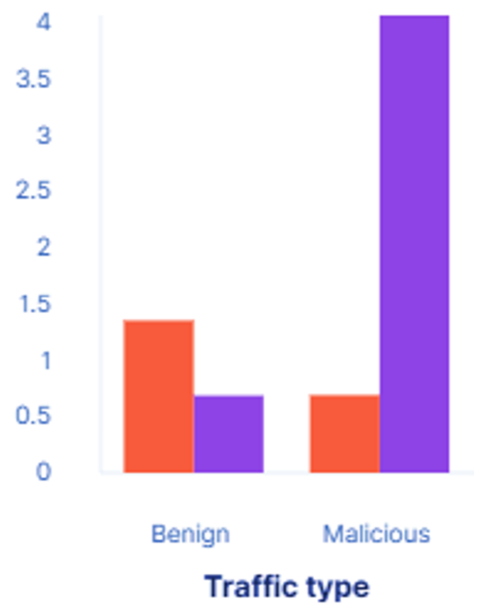


Fig. 6. Average data size according to traffic

Figure 7, allows us to appreciate the ports at which the malicious traffic is being directed. As it can be seen, the traffic tends to attack specific ports of some well-known protocols, such as telnet and HTTP. This is obviously done, in order to try to obtain access via these ports since they tend to be open in most cases. In contrast, in Figure 8, a huge difference can be seen in the ports that are being targeted by the benign traffic. Since this traffic has legitimate petitions, that are used to perform communications between the IoT devices, this clear difference was expected to be seen, according to the type of traffic.

ISSN: 1390 -9266 – 2020 LAJC

The data obtained from the IoT dataset have allowed us to appreciate the differences between the behavior of benign and malicious traffic more easily. As it was previously seen, the differences in the traffic is clear enough to perform a detection by a security analyst only by visualizing the behavior of the data. Based on this principle, a Bayesian Network could be developed in order to detect the anomalies in the traffic. One of the obvious rules that could be used, would be a detection based on the size of the petition, relative to the size of the answer, in order to detect a possible attack that tries to gather information of the system.

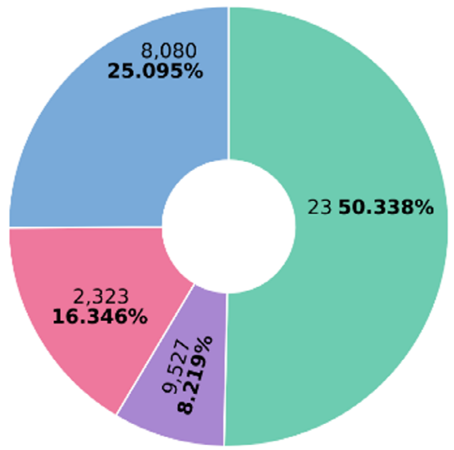


Fig. 7. Ports targeted by malicious traffic.

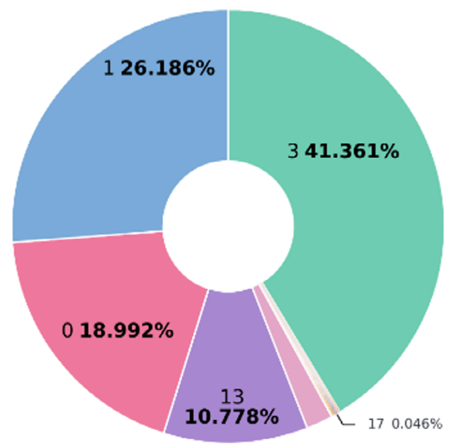


Fig. 8. Ports targeted by benign traffic.

Figure 9 shows the capabilities of ELK to display an anomaly timeline, created in base of multiple metrics. For instance, count of events versus IP address. This figure is also allowing us to identify the time of a day at which an anomaly is more likely to appear. Therefore, a system could be deployed to alert analysts of incoming traffic during those specific times of the day. The analyst could, consequently, prepare itself for a more focused and intensive analysis of the incoming traffic, since it has a higher probability of being a possible threat, due to the time of the day at which it is currently being performed. Additionally, the data clearly shows a difference in the ports that are being targeted by the malicious traffic, compared to the ports targeted by benign traffic. This difference is so prevalent that most security analysts already know which ports need to be secured, since those ports are more likely to be the target for future attacks. This data, therefore, allow analysts to protect the most vulnerable ports, and prevent them for the future, in order to focus their attention in the most vulnerable parts of the system.

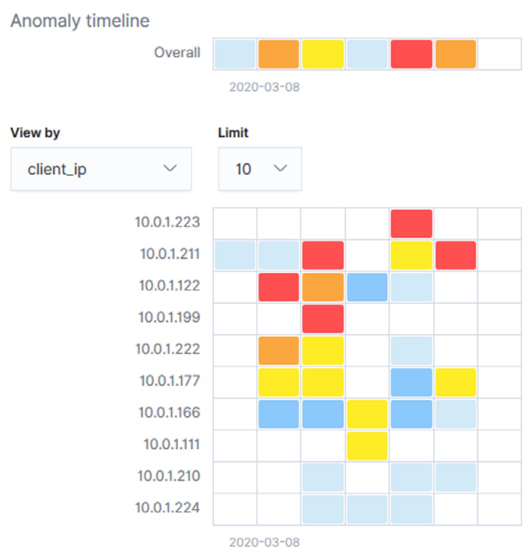


Fig. 9. Anomaly timeline based on multiple metrics.

VI. CONCLUSION

Currently, not only computers generate data, but IoT solutions also present an additional challenge by having objects connected to the internet, such as sensors and appliances (televisions, refrigerators or cars), which are capable of generating data.

The substantial difference from our point of view of cognitive security as a cybersecurity strategy, versus other strategies, is the augmentation of the analysts' cognitive skills. Coupling technological, and data science solutions, to the cognitive process can generate a bidirectional contribution to the technological tool or algorithm, because it can be better trained and configured. Additionally, the analyst also obtains a bidirectional contribution, because they will get support in the development of their own operational tasks.

The efficiency at executing tasks in cybersecurity, depends of the cognitive processes made by a security analyst. Additionally, the detection of anomalies also requires a cognitive process, and its effectiveness depends mostly on the degree of expertise that the analyst has. In this study, we have tried to highlight the importance of spatial-temporal reasoning, as one of the cognitive skills that must be developed by analysts to improve the anomaly detection processes.

Based on the spatial-temporal reasoning, we have proposed the use of data-science solutions such as unsupervised machine learning algorithms, fuzzy techniques and Bayesian networks, which can improve anomaly detection processes. In this study we wanted to highlight the importance of cognitive security, by integrating the point of view of the analyst's cognitive processes with the technological and analytical solutions available to support them.
The ELK stack has great capabilities for the analysis and indexing of data. In addition, it allows searching and storing data very easily, and has enormous possibilities for the analysis of large amounts of data. The use of this stack for the analysis of traffic data in IoT devices, has allowed us to improve the visualization of a huge number of records in a very fast time. The Kibana tool has great capabilities, which can be used to display large amounts of data in an intuitive way; with this tool, it has been possible to create visualizations with great potential, and that take very little time to update, thanks to the optimized operations performed in the indexes at Elasticsearch. Additionally, the implementation of Logstash increases the capabilities of the ELK stack, by allowing it to perform the indexing of the data in a simple way. Logstash functions to perform the reading, filtering and writing of information, allow the processing of any type of record easily and quickly. Finally, the use of the Grok plugin for structuring previously unstructured files, is the main tool that allows encoding and organizing the information in the records.

Future works on this field could be focused in the development of an automated system, which would use the data collected for this study to train an unsupervised anomaly detection algorithm. This algorithm could use various techniques, including Isolation Forest, Clustering and Histogram based algorithms to perform an automatic detection of anomalies, in order to help aid the security analyst in his work. Additionally, the relative simplicity of the previously mentioned algorithms during the deployment phase, will make it easy to perform real time intrusion detection, without causing any significant overhead in the system. Therefore, an automated system deployed to perform anomaly detection will efficiently use the resources and, hopefully, reduce the workload of a security analyst in a significant manner, without losing efficiency in detecting possible intrusions.

VII. REFERENCES

[1] World Economic Forum®, "The Global Risks Report 2019", The Global Risks Report, 2019. [online] Geneva: World Economic Forum. Available at: <http://www3.weforum.org/docs/WEF_Global_Risks_Report_2019.pdf> [Accessed 28 September 2020].

[2] R. Andrade and J. Torres, J. "Self-Awareness as an enabler of Cognitive Security", IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2018.

[3] Elasticsearch B.V., "Elasticsearch: The Official Distributed Search & Analytics Engine", 2020. [Online]. Available: https://www.elastic.co/elasticsearch/. [Accessed: 28- Sep- 2020].

[4] A. Parmisano, S. Garcia and M. Erquiaga, "Stratosphere Laboratory", A labeled dataset with malicious and benign IoT network traffic, 2020. [Online]. Available: https://www.stratosphereips.org/datasets-iot23. [Accessed: 28- Sep- 2020].

[5] IBM, "Applied cognitive security complementing the security analyst.", 2017. [Online]. Available: https://www.rsaconference.com

[6] R. Breton and R. Rousseau, "The c-ooda: A cognitive version of the ooda loop to represent c2 activities. topic: C2 process modelling," 03 2019.

[7] M. D. Mumford, E. Todd, C. Higgs, and T. Mcintosh, "Cognitive skills and leadership performance: The nine critical skills," The Leadership Quarterly, vol. 28, 11 2016.

[8] Riveiro, M., Falkman, G., Ziemke, T., & Kronhamn, T. "Reasoning about anomalies: a study of the analytical process of detecting and identifying anomalous behavior in maritime traffic data", Proceedings of SPIE - The International Society for Optical Engineering, 2009.

[9] J. Ma and S. Perkins, S., "Time-series novelty detection using one-class support vector machines", Proceedings of the International Joint Conference on Neural Networks, 2003. doi:10.1109/ijcnn.2003.1223670

[10] J. Loeffler, R. Cañal-Bruland, A. Schroeger, J. W. Tolentino-Castro, and M. Raab, "Interrelations between temporal and spatial cognition: The role of modality-specific processing", Frontiers in Psychology, 9, Article 2609, 2018. https://doi.org/10.3389/fpsyg.2018.02609 T. S.

[11] Xu, H. D. Chiang, G. Y. Liu, and C. W. Tan. Hierarchical k-means method for clustering large-scale advanced metering infrastructure data. IEEETransactions on Power Delivery, 32(2), 609–616, 2017.

[12] S. Dhandapani, G. Gupta, and J. Ghosh. Design and implementation of scalable hierarchical density based clustering (Doctoral dissertation, University ofTexas), 2010.

[13] P. Berkin. A survey of clustering data mining techniques, in: J. Kogan, C. Nicholas and M. Teboulle (Eds.), Grouping Multidimensional Data: RecentAdvances in Clustering, pp. 25–71, 2006.

[14] D. Pascual, F. Pla, and J. S. Sanchez. Nonparametric local density-based clustering for multimodal overlapping distributions, in: Proceedings of the 7IntligentDataEngineeringandAutomated Learning (IDEAL), Burgos, Spain, pp. 671–678, 2006.

[15] C. M. Varachiu and N. Varachiu, "A fuzzy paradigm approach for the cognitive process of categorization," Proceedings First IEEE International Conference on Cognitive Informatics, Calgary, Alberta, Canada, 2002, pp. 229-232, doi: 10.1109/COGINF.2002.1039302.

[16] Elasticsearch B.V., "Installing Elasticsearch | Elasticsearch Reference [7.9] | Elastic", Elastic.co, 2020. [Online]. Available: https://www.elastic.co/guide/en/elasticsearch/reference/current/install-elasticsearch.html. [Accessed: 28- Sep- 2020]

[17] Elasticsearch B.V., "Configuring security in Elasticsearch | Elasticsearch Reference [7.9] | Elastic", Elastic.co, 2020. [Online]. Available: https://www.elastic.co/guide/en/elasticsearch/reference/7.9/configuring-security.html. [Accessed: 28- Sep- 2020].

[18] Elasticsearch B.V., "Install Kibana | Kibana Guide [7.9] | Elastic", Elastic.co, 2020. [Online]. Available: https://www.elastic.co/guide/en/kibana/current/install.html. [Accessed: 28- Sep- 2020].

[19] Elasticsearch B.V., "Installing Logstash | Logstash Reference [7.9] | Elastic", Elastic.co, 2020. [Online]. Available: https://www.elastic.co/guide/en/logstash/current/installing-logstash.html. [Accessed: 28- Sep- 2020].

[20] ElasticSearch B.V., "Input plugins | Logstash Reference [7.9] | Elastic", Elastic.co, 2020. [Online]. Available: https://www.elastic.co/guide/en/logstash/current/input-plugins.html. [Accessed: 28- Sep- 2020].

[21] Elasticsearch B.V., "Filter plugins | Logstash Reference [7.9] | Elastic", Elastic.co, 2020. [Online]. Available: https://www.elastic.co/guide/en/logstash/current/filter-plugins.html. [Accessed: 28- Sep- 2020].

[22] Elasticsearch B.V., "Output plugins | Logstash Reference [7.9] | Elastic", Elastic.co, 2020. [Online]. Available: https://www.elastic.co/guide/en/logstash/current/output-plugins.html. [Accessed: 28- Sep- 2020].

[23] Elasticsearch B.V., "Index templates | Elasticsearch Reference [master] | Elastic", Elastic.co, 2020. [Online]. Available: https://www.elastic.co/guide/en/elasticsearch/reference/master/index-templates.html. [Accessed: 28- Sep- 2020].

ISSN: 1390 -9266 – 2020 LAJC

AUTHORS

Jonathan Herrera

Was born in Quito, Ecuador in 1994. He did his high school studies in "Unidad Educativa FAE" and is currently obtaining his degree in Engineering in Computer Science in "Escuela Politécnica Nacional". With a scholarship awarded by the Ecuadorian government as part of the GAR group, he studied 1 year in King's College London, completing the course of International Science Foundation Programme from September 21 of 2015 to June 10 of 2016. During his years of study, he has obtained first place in numerous contests in the field of mathematics.

Areas of interest: Artificial Intelligence, Business, Mathematics, Data Science, Machine Learning

Miguel Flores

Ph.D. en Estadística e Investigación de Operaciones, Máster en Técnicas Estadísticas (Universidad de La Coruña). Es consultor asociado de LOGIKA Inteligencia de Mercados, consultor asociado en ISVOS que ofrece servicios de investigación estadística en el ámbito económico, social, ambiental y cultural. Es consultor senior asociado en DS ANALYTICS en el área de Big Data y Ciencia de Datos.
Tiene experiencia en Educación y Formación profesional superior, universitaria y empresarial en el campo de la Statistics & Machine Learning aplicado a
los negocios e industrias. Profesor Titular de la cátedra Probabilidad y Estadística, en la Escuela Politécnica Nacional, Departamento de Matemática. Miembro del Grupo de Investigación Multidisciplinar en Sistemas de Información, Gestión de la Tecnología e Innovación (SIGTI) de la Escuela Politécnica Nacional y del Grupo de Modelización, Optimización e Inferencia Estadística (MODES) de la Universidad de La Coruña. Especialista en el desarrollo de metodologías y paquetes estadísticos en R (qcr y ILS)

ISSN: 1390 -9266 – 2020 LAJC

Roberto Omar Andrade

Estudiante de doctorado en Sistemas de Seguridad en la Facultad de Ingeniería de Sistemas en la Escuela Politécnica Nacional (EPN), su maestría es en Gestión de Redes y Telecomunicaciones en la Escuela Politécnica del Ejército en 2013 y su título de ingeniero es en Electrónica y Telecomunicaciones en la Escuela Politécnica Nacional (EPN) en 2007. Oficial de Seguridad del Ministerio de Educación de Ecuador (MINEDUC) en 2015, Coordinador de Infraestructura Tecnológica en la Secretaría Nacional de Planificación SENPLADES 2013-2014, Centro de datos, seguridad y administración de redes en SENPLADES y Tecnología Sucre 2009-2013 e Ingeniería Técnica para sistemas VoIP en SERATVoIP 2007-2011. Es instructor técnico certificado de CCNA, CCNP y CCNA Security en EPN desde 2010 hasta la fecha.

Susana Cadena

Professor at the Universidad Central del Ecuador (UCE), Doctor in Computer Science, in the line of Data Quality and Open Data. Member of the research groups: Indicators for the Management of the Ecuadorian University, State of the IT of the Ecuadorian Universities sponsored by the Ecuadorian Consortium for the Development of Research and Academy (CEDIA) and Group of Analytics and Big Data for the Cybersecurity, in addition to the groups Ecuadorian Network of Open Data and Metadata (REDAM) and Open Science Research Group.

J. Herrera, R. Andrade, M. Flores and S. Cadena, "Anomaly Detection Under a Cognitive Security Model", Latin-American Journal of Computing (LAJC), vol. 7, no. 2, 2020.