Using Honeypots in Network Intelligence Architectures – *The University of Trás-os-Montes e Alto Douro case study*

José Bessa[1], Hugo Coelho[1], Pedro Monteiro[1], José Brito[1], António Costa[1]
jmiguelbessa16@gmail.com, *coelho.hu@gmail.com*, *monteiro.p@outlook.pt*, *jbrito@utad.pt*, *acosta@utad.pt*

[1]University of Trás-os-Montes e Alto Douro, Quinta de Prados, 5000-801 Vila Real, Portugal

Keywords
Network Intelligence Architectures, Business Intelligence, Big Data, ETL, Honeypots, Security, Intrusion Detection Systems, Monitoring.

1. ABSTRACT

Organizations have increasingly data that needs to be processed to produce information and knowledge in a timely manner. To assist this, Information Systems in the decision-making processes have been developed, which are the case of Decision Support Systems and the Business Intelligence architectures. Many of the organizations' services are supported by communication infrastructures and it is necessary to guarantee that these are prepared to ensure the proper functioning and availability of services and applications. According to this, there is a lot of data being produced indicating their operational status in order to monitor and track their performance.

Considering the Information's importance to organizations, security and confidentiality of the data needs to be assured, therefore Intrusion Detection Systems are an integral part in systems management, with particular importance on safety and infrastructure's reliability and the business data supported by them. This case study was conducted to assess the contribution that Honeypots can have in the creation of security intelligence. The use of Honeypots, Decision Support Systems and Business Intelligence architectures in the organizations' network infrastructures allows to build Network Intelligence Architectures for management and more efficient protection.

With the NIA built for the case study of the University of Tras-os-Montes e Alto Douro, IT and Communications Services have greater insight into their network infrastructure, allowing them to take informed and consented decisions based on the service and security indicators provided by the NIA and create policies to protect the infrastructure.

2. Introduction and Scope

Nowadays, Information, Knowledge and Intelligence represents a key to the success of an organization, and allows it to gain competitive advantage and gives the best possible way to manage the business. In order to achieve these levels of abstraction, first it is necessary to consider the data from the organizations daily activity. Given the exponential growth of data, there is a need to streamline its treatment, which has encouraged the use of Information Technologies (IT) to support the organization's Information Systems (IS). These systems do not work isolated and they need to be integrated, which involves the use of communication infrastructures and support systems, particularly when organic units are geographically dispersed. More and more organizations use communication infrastructures to support their business, and as such it is important that they are monitored and managed so they can sustain in the best way the various activities of organizations, not only for their normal operation, but also for strategic activities (Taylor, Gresty & Askwith, 2001).

Monitoring is crucial for the management operations, as it allows to early on identify trends and patterns in data traffic, network devices' behavior (anomalies, performance issues), in order to ensure infrastructure's reliability, robustness, high productivity and quality of service to those who use it (Avizienis et al., 2004). After handling the data generated by IT, which is part of the network

infrastructure and its systems, it is possible to get knowledge about it. In order to know their behavior, a tailored SI such as a Decision Support Systems (DSS), can be used to enable informed and consensual decision making.

The use of DSS and Business Intelligence (BI) architectures to obtain Knowledge on management and monitoring network infrastructures, can be associated to the concept of Network Intelligence. On the other hand, the Computer Security issues are closely linked to the performance of these infrastructures. Therefore, it is important to use mechanisms to assess whether it is safe and vulnerability free, including them in a Network Intelligence Architectures (NIA) (Russell & Cohn, 2012).

In order to reduce the risk of compromising information, security policies should be created to ensure business continuity in its normal flow. These policies should be regularly reviewed so that there is a constant evaluation. They must ensure that the information is accessed only by authorized entities and that is always available, while the data is real and consistent. Additionally it is necessary that vulnerability and threat detection methods are present in the infrastructure. This will help to improve the security policies already established. Thus, it becomes possible to calculate the risk that the compromised Information may cause to the organization (Ahmed, Mahmood & Hu, 2016).

One of the mechanisms used to increase IT Security is the use of Honeypots. A Honeypot is another security tool, which along with other tools of this kind, helps in network intrusion detection, giving network and system administrators the necessary knowledge to treat these network security incidents, either by detection and prevention of attacks, or to its reaction. If then, this Knowledge of the intrusion is integrated into a Knowledge Management platform, it is possible to create a DSS, able to assess on creating standards and security policies to be applied to the infrastructure, so they remain in a stable state of security (Spitzner, 2002).

This work aims to demonstrate the applicability of probes based on Honeypot models for intrusion detection in an attempt to improve the Security Information strategy carried out by the University of Trás-os-Montes e Alto Douro's (UTAD) IT and Communications Services (SIC-UTAD), given the considerable number of detected intrusion attempts this institution receives. These, in successful cases, could compromise their services and consequently their reliability.

This paper is divided into four sections: in section 2 it is provided the scope of the paper and in section 3 a conceptual aproach around some of the key concepts that can be associated with Network Intelligence architectures. In the 4th section the case study is explained, the identified needs that led to the execution of this work, the proposed Network Intelligence architecture to UTAD and the used technologies. The paper ends with a section of tests and achieved results, as well as some final considerations and perspectives for future work.

3. Conceptual Aproach

The development of a NIA that make feasible the intentions of SIC-UTAD for the management of the network infrastructure, required the necessity to understand the concepts/technologies that can be applied over the data generated by the network, to obtain Knowledge and Intelligence, namely (Russell & Cohn, 2012): the DSS and BI architectures (focusing on Extract-Transforme-Load processes and Data Visualization). On the other hand, assuming data security as essential, there was a need to understand the Intrusion Detection Systems (IDS) and, in this article, the application of Honeypots to ensure this assumption, allowing its incorporation in this type of architectures.

3.1. From Big Data to Managing and Monitoring Network Communication Infrastructures

Nowadays, a great number of organizations have network communication infrastructures to support their several activities. In this way, many organizations can only support their business and activities, if they have an infrastructure of this kind (Sterbenz et al., 2010).

Given the importance of the network infrastructures mentioned above, it is important to manage and monitor them. Monitoring is crucial for the management operations, since it allows to early identify trends and patterns in data traffic, behavior of network devices (anomalies, performance issues), in order to ensure the reliability, robustness, high productivity of infrastructure and quality of service for those who use it. Monitoring represents a key element in network management, being this

composed by activities, methods, processes and tools that support the operation and maintenance of the network infrastructure (Lee, Levanti & Kim, 2014).

Similar to what happens with remaining organizational data, also data from the network infrastructure have exponentially grown, being this phenomenon associated with the concept of Big Data. This concept have appeared from the need to improve the management of large volumes of data, expediting the decision-making process.

Big Data is applied to the organization's ability to manage, store and analyze data. The data storage process is one of the most important aspects contemplated in this concept, hence it is necessary to create new ways to do that, given that the traditional ways fall outside this new paradigm. On the other hand, it is essential that not only the data is stored but also used in the organization routines, in order to speed up and sustain their activities (Chen & Zhang, 2014).

With Big Data it is possible to identify the most important customers and their needs to strengthen relationships, maximize profits and sales, mining data associated to clients in order to generate new marketing campaigns, analyze data from media and banking to detect new socio-economic market trends, analyze data from social networks, among others. (Bolón-Canedo, Sánchez-Maroño & Alonso-Betanzos, 2015).

When it comes to Big Data it is critical to consider the 5V's that allow to realize their essence and relevance that there is a concept explaining this phenomenon, which are: Volume, Variety, Velocity, Value and Veracity. Explaining each of these "V's" (Hashem et al., 2015.):

- **Volume** is associated with the large amounts of data in multiple formats, from different sources, internal (e.g. internal IS) and external (e.g. newspapers, Internet, partners IS, social networks), and it is still constantly growing. The greater the data volume is, the greater the ability to detect hidden patterns and optimize analytical models;
- **Variety** closely associated with the data format irregularity, since these may be structured (e.g. relational databases), semi-structured (e.g. XML, RDFS) or without structure (e.g. SMS, tweets, audio, video, images), which it is expected given the variety of data sources (e.g. sensors, mobile data, operational IS);
- **Velocity** refers to the transmission of data, its handling, delivery and availability in the data storage structures. This factor is important so that the availability of data is made in a timely manner, which is becoming real-time, since the dynamics of organizations requires that the time for decision-making becomes shorter;
- **Value** is the factor that explains the importance of data for relevant business information to be identified and extracted, as well as standards for defining action strategies and to influence the creation of new products/services;
- **Veracity** of data is required so that when analytical models are applied, decisions can be made based on real and consistent data, not creating ambiguities, fraud, questions, inconsistencies or incoherencies in these management activities.

Despite the current adoption and exploration of the concept of Big Data in the organizational environment, simple data collection by itself is no guarantee of competitive advantages. The data are mere representations of real facts, and only have value when contextualized and delivered to whom can interpret, for generate Information. For process management it is of interest to achieve a higher level of abstraction, namely Knowledge, passing through the analysis and extraction of conclusions from the Information (Kebede, 2010).

In order to manage the network infrastructure is important to have Knowledge and Intelligence (crossing Knowledge with the personal experience of those who use it) to take consensual and timely decisions. Thus, we can say that the NIA are architectures that in its genesis include mechanisms to obtain Knowledge and later Intelligence from network infrastructures.

A NIA contemplates the use of DSS to facilitate the processes of obtainment, retention and management of Knowledge and Intelligence enhancing and maximizing organizational resources (Scheibe et al., 2006). DSS belong to the class of information systems that are designed to facilitate decision-making processes, and have advanced analytical capabilities. They are often tailored to the person or group that uses them, so they should be flexible, interactive and dynamic (Vizecky & El-Gayar, 2011). Its main tasks consists in the production and dissemination of information, hence they can be considered an information system.

This type of systems have other features such as the creation of Intelligence and organizational Knowledge, through the use of analytical data models with possibility of simulation scenarios and friendly interface with users over the creation of dashboards that represent graphically the data complexity using Data Visualization (DV) techniques (Grierson, Corney & Hatcher, 2015). Dashboards created in these kind of systems, in this case for the manage and monitor network infrastructures, use the DV techniques, which allow the application of analytical models and statistical functions, whose results are presented visually through interactive dashboards composed by tables, diagrams, graphs, histograms, maps and pilot panels (Janvrin, Raschke & Dilla, 2014).

The decision-making implies that there is organizational intelligence, that involves gathering information, analyze it, disseminate it (to who knows the business), get new opportunities, react in time and adapting. Everyday, this capacity should be an integral part of any organization, for business decision-making to be appropriated and have quality. However, with organizations storing increasing amounts of data, there is a need for the use of IT to its storage, delivery and processing to extract organizational intelligence. In order to fullfil that need, the concept of BI emerged (Ramakrishnan, Jones & Sidorova, 2012).

The concept of BI can be understood as a set of techniques and tools used to extract intelligence from a data set. Related to BI is a set of concepts, methods and processes to improve business decisions, which uses data from multiple sources and applies the experience and assumptions to develop, in a accurate way, knowledge of business dynamics (Petrini & Pozzebon, 2009).

Given the importance of the concept of BI to organizations, this has been exhaustively studied and defined by several authors. For this work, we assumed the definition of BI developed by the authors Sharda, Delen & Turban (2013), in which it is said that BI is an aggregator concept that combines architectures, tools, data sources, analytical tools, applications and methodologies. Integrates data analysis with decision support systems to provide information for all persons in the organization, who need to make tactical and strategic decisions. With the development of an appropriate BI architecture, an organization may be able to develop intelligent decision support systems for competitive advantage in the industry in which it operates (Wu, Chen & Olson, 2014).

The ability to combine business activities, IT and analytics, should be on the focus of the organization's work, with the creation of an adequate infrastructure that will achieve the organization's goals and allowing it to take competitive advantages; agility in decision-making processes; ability to handle sensitive information; identify patterns, trends or hidden behavior in the organization's data that may represent propitious situations related to customers and/or new business opportunities (Chen, Chiang & Storey, 2012).

A BI architecture must be aligned in the global infrastructure of the IS of an organization. Typically a BI architecture consists of (Olszak & Ziemba, 2007): internal data sources (data from operational systems) or external, Extract-Transform-Load (ETL), data storage source (e.g. database, Data Warehouse, Data Marts, ad-hoc queries), analytical tools for data processing (e.g. OLAP servers, Data Mining) and front-end applications where the data analysis results will be presented (e.g. dashboards, reports, graphics).

The ETL tools are a fundamental part of BI systems, since they are used to simplify and standardize the data migration process. In this work, this process has a particular focus, given the complexity and variety of data formats of the systems used in this case study. This module, aims to gather several data from heterogeneous platforms of the organization in a standard format that can be used by DSS, and at the same time should be able to integrate and standardize data. Also, it must be flexible because the requirements can be changed and the dynamics of databases should be considered as well (Guo et al., 2015).

The ETL concept, as the name suggests, involves three steps (Prema & Pethalakshmi, 2013): **Extraction**, where a copy of the new data is obtained since the last application of ETL tools, which are present in the various data sources (which may have different shapes and structures); in **Transform**, data is processed into information, a task consisting in the translation of encoded values; application of this process to certain categories of rows and/or columns; merge (merging) or aggregation of data in order to create a uniform structure which can be stored in the database. Finally the **Load** is done, which consists in the activity of populate the target tables present in the database, which can be a simple step, namely, rewrite new data over the oldest, or a more complex process in which data is held for historical purposes, keeping all records of all changes made.

Scalability is one of the most important factors to consider when this type of modules are constructed, since in some situations, the ETL modules have to process millions of data and update databases with millions of records. The exponential growth of data (Big Data) and the increasing data sources require that the ETL technologies become more advanced and flexible (Karagiannis, Vassiliadis & Simitsis, 2013).

3.2. Security

When we are talking about IS in organizations and Big Data, it is also necessary to take into account the policies and computer security mechanisms that prevent systems and organizational data of being accessed by unauthorized entities, ensuring their integrity. In this sense IDS emerged, which as the name suggests, are designed to detect and prevent intrusions (Liao et. al, 2013).

To understand the applicability of such systems it is necessary to understand what an intrusion is. An intrusion consists of a set of activities with the goal of misrepresent and/or compromise the integrity, confidentiality and availability of a resource. This type of illegal action comes from an attack or set of attacks and may or may not distort the data in a permanent way (Zuquete, 2013).

IDS try to assess this type of illegal activity and can detect intrusions from an early stage of trial or suspicious activity, to the consummation of this act. They act independently of the mechanisms that support the correct use of the systems in which they are applied, and to build an intrusion profile, they collect data on the various activities associated with a System allowing to conclude on likely past, present or future intrusions, this being a dynamic and constantly updated process (Jaiganesh, Mangayarkarasi & Sumathi, 2013).

Regarding intrusion reaction, this systems may be of two types: passive or active, whereas the former only react with alerts and reports to a specialist in this area to apply defensive measures, the latter have automatic response mechanisms to intrusions (Sekhar et. al, 2015). Despite the benefits of these systems, IDS like any other system, have limitations such as: false positives by false alarms and false negatives when an intrusion is not detected, which can lead to loss of trust in these systems; overhead on the network and its systems; a lack of standardizationon on the specification of intrusions and attacks; outdated attack databases, since all days new attacks are created; High technology requirements to support their action mechanisms(Spathoulas & Katsikas, 2010; Corona, Giacinto & Roli, 2013).

With the use of IDS for information security, illusion systems have been created, which is the case of Honeypots. Although there are several definitions of what a Honeypot is (Pouget, Dacier & Debar, 2003), the most consensual explains that this is a security resource without production value and whose true value lies in being probed, attacked or compromised, thus obtaining various information about the attackers and will allow to create preventive measures against attacks (Spitzner, 2002). In other words, a honeypot is a resource on the network, that although it is not a production system, has the ability to seem like one and to look vulnerable.

As a direct consequence of the definition, any traffic directed to a Honeypot is considered abnormal. This assumption means that, unlike the Network Intrusion Detection Systems (NIDS), the data collected is of great value and private of noise (Briffaut, Lalande & Toinard, 2009). On the other hand, the monitoring is isolated, making a Honeypot unaware of other network behavior, proving the need for other security tools, as wel as the creation of Honeynets (Curran et al., 2005). This simplicity in how a honeypot works, allows its support systems to not require large computational resources.

By logging and analysing adverse events, a Honeypot be used for the treatment of security incidents it captures, allowing preventive actions (resource deterrence), detection (event alerts), reaction (attacker data) and research (increase Knowledge). The fact that there are worldwide research Honeypots allows the information obtained by them to be correlated, identifying attack patterns (West-Brown et al., 2003). These patterns can then be crossed with the information obtained by prevention and reaction Honeypots existing in the organization networks, leading to security policies being created and applied in other network resources, such as traffic filtering policies applied to Firewalls, based on previously detected events (Spitzner, 2002; Pouget, Dacier & Debar, 2003).

Honeypots can be classified in three ways that are directly related to how an attacker interacts with it:

- **Low Interaction:** These Honeypots are easier to deploy and configure, since they only simulate services with which the attacker can interact and their importance is based on the detention of unauthorized scans and connection attempts (Briffaut, Lalande & Toinard, 2009). The risk of this type of honeypots is very low, since an operating system (OS) is not given to an attacker to interact with, but only a set of services configured to operate in a predetermined way (Spitzner, 2002);
- **Medium Interaction:** These consist of isolated subsystems monitored in real systems, giving the attacker the ability to interact with an OS that is pre-configured to not expose all the features expected from a real system, reducing the risk of being compromised (Honeynet Project, 2004). The information given by these honeypots are more extensive, having the ability to record the behavior of attackers,viruses and worms after they take over the System;
- **High Interaction:** Although these honeypots give us the most information about the attacker, they are very complex in their implementation and maintenance, and become a high risk (Briffaut, Lalande & Toinard, 2009), because the more complex a system is the greater is its probability of failure. This type of honeypots provides a real system to the attacker without any restrictions, being important the controlled environment in which they are placed (Honeynet Project, 2004).

In summary, the greater the level of interaction, the greater the freedom of the attacker and the produced information, however the risk is also bigger.

For a honeypot to be efficient producing relevant data to improve security, it is important that it is implemented in a way it is not identified (signatures problem), and is placed pertinently in the network infrastructure, since the location will determine its value. A common location for placement of honeypots is the Demilitarized Zone (DMZ), since that is where public production systems of a organization are usually located. In the DMZ, a honeypot will be able to monitor and capture external attacks targeted at the public infrastructure of the organization (Salimi & Kabiri, 2012).

Honeypots for research purposes are placed directly in contact with the public so that they can record all kinds of information regardless the organization. Another possible location is the organization's internal network. When placed in the same network as the user devices it will allow the detection of internal attacks, i.e. it will know what devices may be compromised and are possibly being used as a source of other attacks (Pham & Dacier, 2011).

4. Proposed Architecture

4.1. Instituition and Needs

UTAD is a public Portuguese university, established in 1973 in the city of Vila Real, and is one of the leading institutions promoting the development of this region, both by its scientific and technological contribution, and by encouraging the entrepreneurship. This institution consists in several organic units, and the one directly addressed in this work, is SIC-UTAD. This unit is responsible for all technological support of the information and infrastructure systems as well as the Internet access of UTAD having the responsibility to monitor and manage both the equipment that make up the infrastructure and the traffic passing through UTAD's network.

Within the decision-making processes of these services, a set of requirements related to the IT infrastructure has been identified, which led to the composition of an NIA, which allow acquire Knowledge and Intelligence of the network infrastructure. As with any system, it is also necessary to ensure its safety and integrity and so, for the intrusion detection, came the need to install a Honeypot on the network infrastructure and assess its contribution in the NIA. Therefore, the considered NIA data sources will be support services such as DNS, RADIUS, DHCP, SNMP, Firewall. Given the extensiveness of this work and given the importance of implementing a Honeypot in NIA, only the results from the implemented Honeypot will be presented.

4.2. Deploying the Technologies on Network Infrastructure

A prerequisite for the implementation of this NIA was the need to use technologies that follow the Free Open Source Software (FOSS) philosophy, to ensure that not only the monetary implications that could come from the implementation are reduced, but also to make use of the documentation and support communities that are often associated with these technologies. The fact that these

technologies are open source allows a better understanding of its operations. Thus, technologies used in this case study and their respective support systems follow this philosophy.

The architecture proposed in this section for the UTAD's network infrastructure management (**Figure 1**), took into account all the aspects mentioned in the third section of this article, respecting the NI architectures principles. This architecture is divided into four levels / layers: data, ETL, DV and Knowledge / Intelligence, the latter being the highest level.

In order to build a robust NIA and manage the communication infrastructures and its systems, there was a need to create and implement a tool that allows to obtain operating logs from honeypots and other systems that constitute the communication infrastructure and other UTAD systems (e.g. RADIUS, DHCP and DNS) as well as their processing, storage in a uniform data structure with the ability to produce visual metrics (via dashboards). Given the complexity of the situation in question, it was used the Elastic stack composed by ElasticSearch (for data storage), Logstash (for acquisition and processing of data) and Kibana (for building dashboards and obtaining Knowledge).

The ELK® stack (ElasticSearch, Logstash and Kibana) was installed on several virtual machines in a data center server. At the ETL level, each data type has a virtual machine (keeping them isolated), using the Linux operating system Ubuntu 14.03 Server and each have installed: ElasticSearch (database engine), Logstash (used to filter the data and populate the database) and shield (used for database security). Before using this stack of technologies, the logs go through a cleaning script that is designed to create a uniform structure, which can be used in Logstash filters. In the DV level, the technologies used (and operating system) are the same that the ETL level uses, with the addition of Kibana used for data analysis and consequent construction of views (dashboards) and also the installation of plugins for managing visual form of databases (and ElasticSearch), Kopf and HQ.

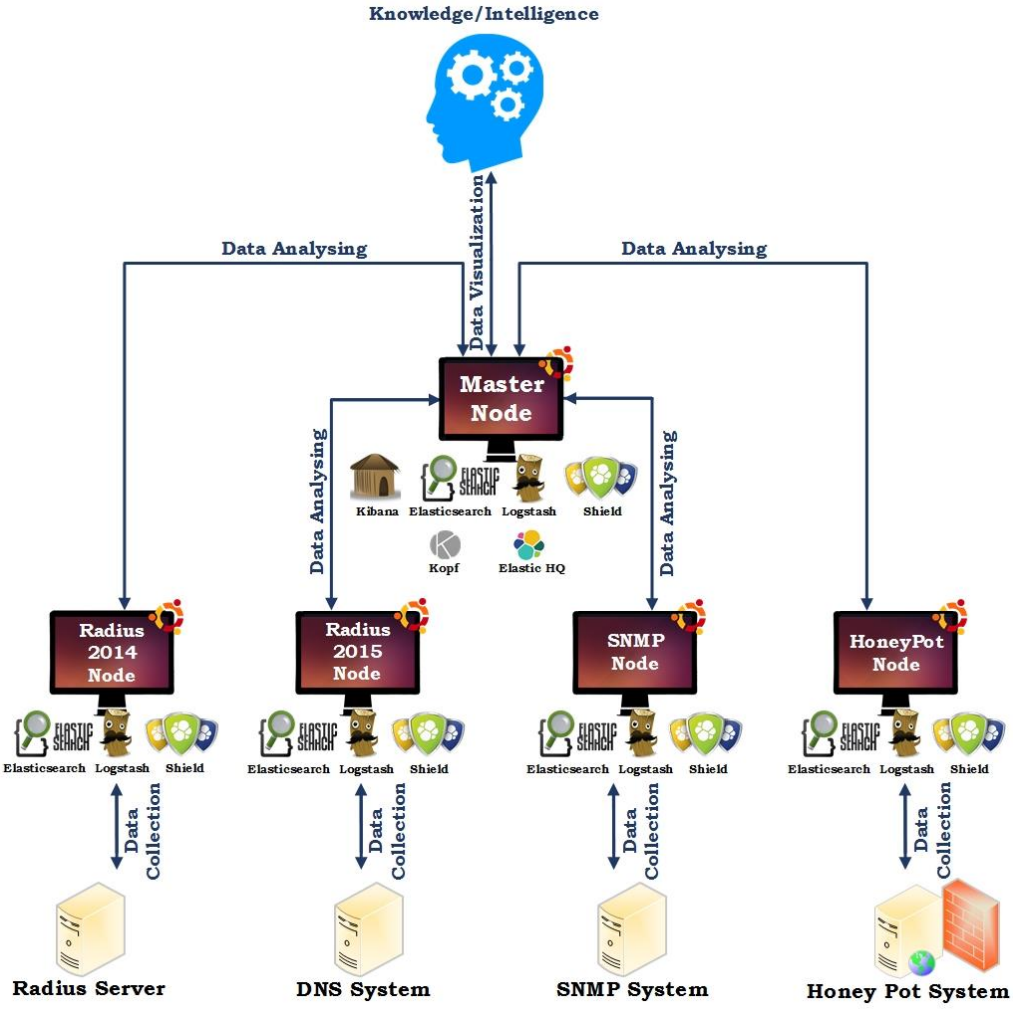

Figure 1 – NI architecture proposed to UTAD's network infrastructure

Focusing on the Honeypot node, the solution used for detection and prevention of the attacks was the HoneyD framework as a Honeypot. This product has shown to be the right one for the needs of the institution and in this case study, one of the most stable honeypot solutions recognized nationally and internationally. In addition, it is a scalable tool, highly customizable with the ability to simulate a complete network architecture and able to act as low and medium interaction Honeypot.
In this case, the HoneyD was installed to behave as a low interaction honeypot by opening some common TCP ports in order to register connections, as is the case of ports 22 (SSH) and 21 (FTP). Other ports such as 80 (HTTP) and 23 (TELNET) were configured in medium interaction mode, by associating scripts that simulate these services' behavior. While in the first case only connection data is obtained, in the second you can record some of the attackers behavior. HoneyD is also configured to respond for all Internet Protocol (IP) that are not used in the network in which it is deployed. The **Figure 2** briefly shows the place where the HoneyD server is in the network.

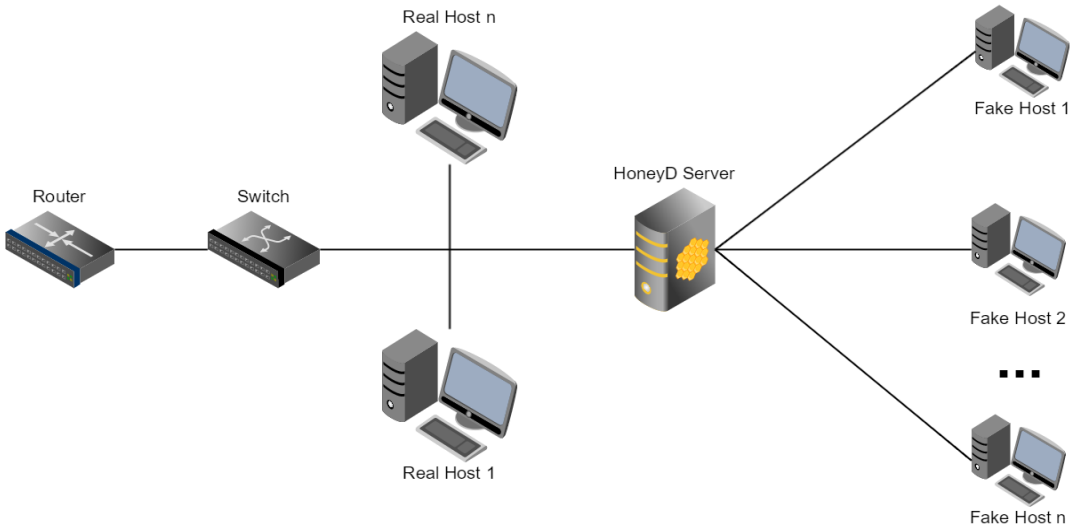

Figure 2 - HoneyD server on the network

5. Results, Final Considerations and Future Work

The proposed architecture, already serves the staff responsible for managing the IT infrastructure in their operational management activities, and also serves to support the strategic planning of the entity object of the case study. For the visual presentation of the results obtained, a set of dashboards were created that reflect more quickly and intuitively the necessary information to streamline decision-making processes, considering the systems mentioned in the proposed architecture.

Regarding the security systems (Honeypot and Firewall) discussed in the practical component of this work, they allowed to identify which infrastructure machines/IS have more attack attempts, the zones of the world where most attacks come from (China, Indonesia, Hong Kong, United States, South Korea, Vietnam, Turkey and Russia), and which are the operating systems used by the attackers (Linux for Asian countries, Russia and Ukraine, and Windows to the North American continent, Africa and Central Europe), allowing to adapt the firewall that filters the traffic, aiming for maximum protection of the infrastructure and also proposing new security measures.

In order to test the security solution implemented in the proposal NIA, the logs obtained from Honeyd in the period of **12/09/2015** to **16/12/2015** were considered. In **Figure 3**, we can see the number of connections per protocol and as a result we can see that the majority of attempted attacks are from the ICMP and TCP protocols having some from the IGMP and UDP protocols. Given the amount of ICMP and TCP connections, these will be seen in greater detail through the dashboards observed in **Figure 4** and **Figure 5**, respectively.

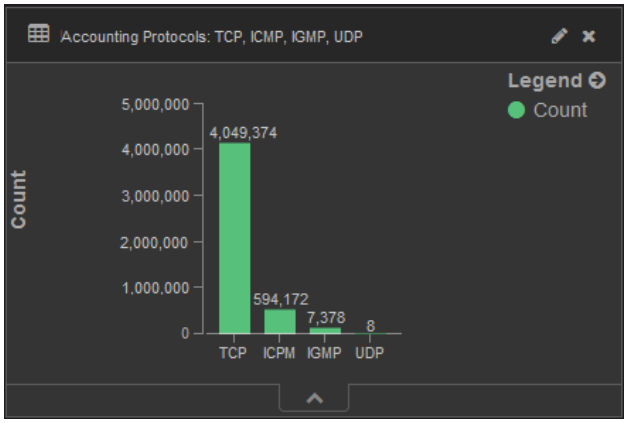

Figure 3 – Attack frequency by protocol

The analytics produced, allows assessing some information about the intrusions in the UTAD's network. The **Figure 4** shows a dashboard with some visualizations of the ICMP protocol, where it can be seen that, in the said time period, the Honeypot received 594.172 ICMP connections, and considering the timeline, the daily average was of 3,500 connections, maintaining a constant behavior, without high peaks. The biggest source of this kind of scans was the Southeast Asian region, the main countries were China, Hong Kong, Singapore and Japan. At the Americas quadrant, we found the United States and Canada as major sources of attacks, and at Europe the largest representatives of this type of illicit activity are the United Kingdom and Russian Federation. It is also possible to acquire some knowledge about the operating systems used by the attackers, it has been observed a great use of Windows XP, especially for the attackers from France, Russia, the Netherlands, China and the United States.

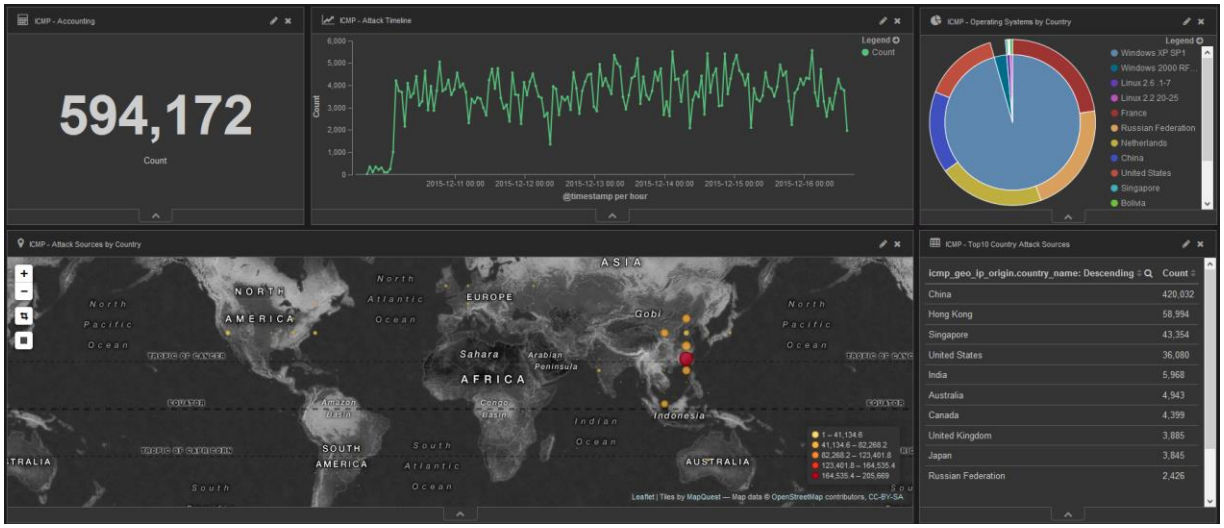

Figure 4 – ICMP - Total of attacks, number of attacks by day, operating systems used, geolocation of source attacks, top10 countries by number of attacks

In the case of TCP (**Figure 5**), the honeypot registered about 4,050,000M anomalous connections, a number greater than any other protocol. The port with more received intrusion attempts was the Telnet service port (23), with over than 75% of connections. In this protocol, there is a greater geographical focus of the attacks from Asian countries such as China, Hong Kong, Singapore and India, included in this Top10 are the United Kingdom in Europe, the United States in North America and Australia in Oceania. Unlike the ICMP protocol, with the TCP protocol, the main operating system used for the attacks execution was Linux. Analyzing the timeline, in the first two days of the period of evaluation, the number of attacks increased, maintaining a constant behavior in the remaining period of time.

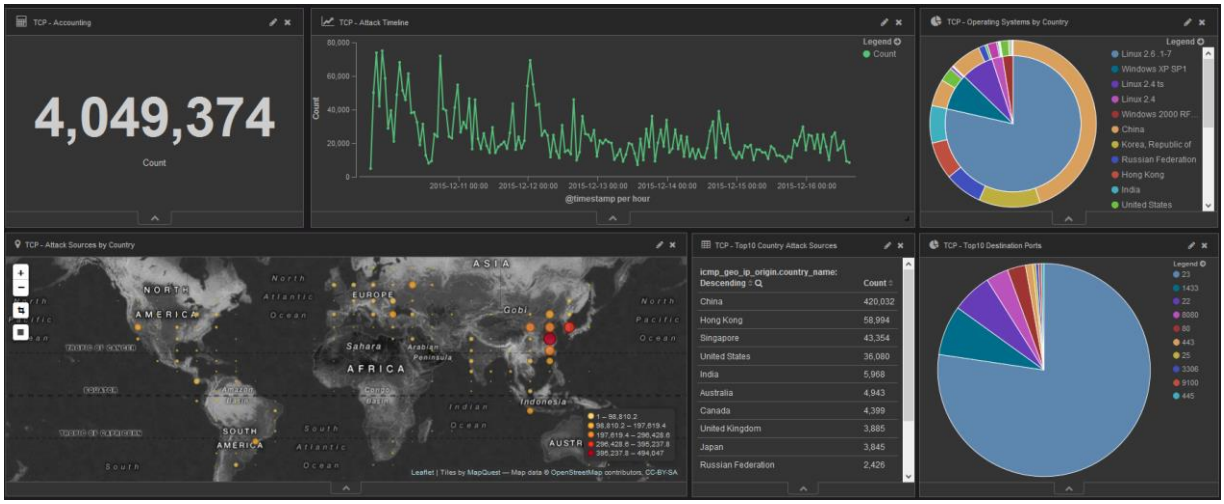

Figure 5 – TCP - Total of attacks, number of attacks by day, operating systems used, geolocation of source attacks, top10 countries by number of attacks, top10 destination ports

Nowadays, the network communication infrastructures are essential to support the activity of organizations, and thus, it is important that they are monitored and managed in the best possible way in order to guarantee quality of service to those who use them. This context is applied to UTAD and to the management areas of SIC-UTAD, hence the need to create an architecture that would allow to get Knowledge and Intelligence from the network infrastructure, that makes use of concepts associated with NIA above referred in the conceptual approach of this article. On the other hand, the organizations data is of great value to them, since it is part of their private property and so it is of most importance to ensure their protection and assure damage prevention in risk situations.

In the network infrastructure was introduced a security mechanism that allows SIC-UTAD to have a better perception of anomalous traffic flowing on the network infrastructure of the university. Beyond this detection of attack attempts, the honeypot installed allowed dissuasion of production resources, decreasing the probability of them to be attacked, and also allowed adapt Firewall rules which filters the traffic of university, based on the generated security indicators. Until now, the results are very positive, since it is possible to obtain knowledge of the generated data by the infrastructure services, previously described, allowing SIC-UTAD to have a proactive approach in the monitoring and management of UTAD's network infrastructure, having been made some adjustments in it.

Regarding future work, it is intended to continue research in this area of Network Intelligence, given the development that have been made in this. It is also intended to create new dashboards for decision support and adapt the proposed architecture to all data sources used in UTAD's network infrastructure. Associated to the Honeypot it is intended to optimize the data treatment process, in order to make it more robust, efficient and fault-tolerant. In addition to this optimization, it is expected to expand the applicability of this mechanism through the creation of several honeypots in the various university's subnets allowing a holistic view of the intrusions and attacks in the entire infrastructure.

6. REFERENCES

Ahmed, M., Mahmood, A. & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 60, 19-31.

Avizienis, A., Laprie, J., Randell, B. & Landwehr, C. (2004). Basic concepts and taxonomy of dependable and secure computing. *IEEE Transactions on Dependable and Secure Computing*, 1(1), 11-33.

Bolón-Canedo, V., Sánchez-Maroño, N. & Alonso-Betanzos, A. (2015). Recent advantages and emerging challenges of features selection in the context of Big Data. *Knowledge-Based Systems*, 86, 33-45.

Briffaut, J., Lalande, J. & Toinard, C. (2009). Security and results of a large-scale high-interaction Honeypot. *Journal of Computers*, 4(5), 395-404.

Chen, C. & Zhang, C. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314-347.

Chen, H., Chiang, R. & Storey, V. (2012). Business Intelligence and analytics: From Big Data to big impact. *MIS Quarterly*, 36(4), 1165-1188.

Corona, I., Giacinto, G. & Roli, F. (2013). Adversarial attacks against intrusion detection systems: Taxonomy, solutions and open issues. *Information Sciences*, 239, 201-225.

Curran, K., Morrisey, C., Fagan, C., Murphy, C., O'Donnell, B., Fitzpatrick, G. & Condit, S. (2005). Monitoring hacker activity with a Honeynet. *International Journal of Network Management*, 15(2), 123-134.

Grierson, H., Corney, J. & Hatcher, G. (2015). Using visual representations for the searching and browsing of large, complex, multimedia data sets. *International Journal of Information Management*, 35(2), 244-252.

Guo, S., Yuan, Z., Sun, A. & Yue, Q. (2015). A new ETL approach based on data virtualization. *Journal of Computer Science and Technology*, 30(2), 311-323.

Hashem, I., Yaqoob, I., Anuar, N., Mokhtar, S., Gani, A. & Khan, S. (2015). The rise of "Big Data" on Cloud Computing: Review and open research issues. *Information Systems*, 47, 98-115.

Honeynet Project, The. (2004). *Know your enemy: Learning about security threats* (2ª ed.). Boston: Addison-Wesley.

Jaiganesh, V., Mangayarkarasi, S. & Sumathi, P. (2013). Intrusion detection systems: A survey and analysis of classification techniques. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(4), 1629-1635.

Janvrin, D., Raschke, R. & Dilla, W. (2014). Making sense of complex data using interactive data visualization. *Journal of Accounting Education*, 32(4), 31-48.

Karagiannis, A., Vassiliadis, P. & Simitsis, A. (2013). Scheduling strategies for efficient ETL execution. *Information Systems*, 38(6), 927-945.

Kebede, G. (2010). Knowledge management: An information science perspective. *International Journal of Information Management*, 30(5), 416-424.

Lee, S., Levanti, K., Kim, H. (2014). Network monitoring: Present and future. *Computer Networks*, 65, 84-98.

Liao, H., Lin, C., Lin, Y. & Tung, K. (2013). Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications*, 36(1), 16-24.

Olszak, C. & Ziemba, E. (2007). Approach to building and implementing Business Intelligence systems. *Interdisciplinary Journal of Information, Knowledge and Management*, 2, 134-148.

Petrini, M. & Pozzebon, M. (2009). Managing sustainability with the support of Business Intelligence: Integrating socio-environmental indicators and organisational context. *The Journal of Strategic Information Systems*, 18(4), 178-191.

Pham, V. & Dacier, M. (2011). Honeypot trace forensics: The observation viewpoint matters. *Future Generation Computer Systems*, 27(5), 539-546.

Pouget, F., Dacier, M. & Debar, H. (2003). *Honeypot, Honeynet, Honeytoken: Terminological issues*. Boston: Addison-Wesley.

Prema, A. & Pethalakshmi, A. (2013). Novel approach in ETL. In *Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013. International Conference on, Salem: India* (pp. 429-434). IEEE.

Ramakrishnan, T., Jones, M., Sidorova, A. (2012). Factors influencing Business Intelligence (BI) data collection strategies: An empirical investigation. *Decision Support Systems*, 52(2), 486-496.

Russell, J., Cohn, R. (2012). *Network Intelligence*. Wisconsin: Book on Demand.

Salimi, A. & Kabiri, P. (2012). Avoiding Cyber-attacks to DMZ and capturing forensics from intruders using Honeypots. *Journal of Advances in Computer Research*, 3(1), 65-79.

Scheibe, K., Carstensen Jr., L., Rakes, T. & Rees, L. (2006). Going the last mile: A spatial decision support system for wireless broadband communications. *Decision Support Systems*, 42(2), 557-570.

Sekhar, M., Tulasi, K., Amulya, V., Teja, D. & Kumar, M. (2015). Implementation of IDS using Snort on bayesian network. *International Journal of Computer Science and Mobile Computing*, 4(4), 790-795.

Sharda, R., Delen, D. & Turban, E. (2013). *Business Intelligence: A managerial perspective on analytics* (3ª ed.). New Jersey: Prentice Hall.

Spathoulas, G. & Katsikas, S. (2010). Reducing false positives in intrusion detection systems. *Computers & Security*, 29(1), 35-44.

Spitzner, L. (2002). *Honeypots: Tracking Hackers*. London: Addison-Wesley.

Sterbenz, J., Hutchison, D., Çetinkaya, E., Jabbar, A., Rohrer, J., Scholler, M. & Smith, P. (2010). Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines. *Computer Networks*, 54(8), 1245-1265.

Taylor, M., Gresty, D. & Askwith, R. (2001). Knowledge for network support. *Information and Software Technology*, 43(8), 469-475.

Vizecky, K. & El-Gayar, O. (2011). Increasing research relevance in DSS: Looking forward by reflecting on 40 years of progress. In System Sciences, 2011. *HICSS'44th Hawaii International Conference on*, *Hawaii: United States*, (pp. 1-9). IEEE.

West-Brown, M., Stikvoort, D., Kossakowski, K., Killcrece, G., Ruefle, R. & Zajicek, M. (2003). *Handbook for Computer Security Incident Response Teams (CSIRTs)* (2ª ed.). Pittsburgh: Software Engineering Institute.

Wu, D., Chen, S. & Olson, D. (2014). Business Intelligence in risk management: Some recent progresses. *Information Sciences*, 256, 1-7.

Zuquete, A. (2013). *Segurança em Redes Informáticas* (4ª ed.). Lisbon: FCA.

7. AUTHORS' BIOGRAPHIES

José Bessa is an Information Systems student at University of Trás-os-Montes e Alto Douro (UTAD), Vila Real, Portugal, where he got his bachelor's degree in 2013 and is now completing his master's degree in Computer Science. He is currently researching Information Systems Architectures and Network Communications. Besides this, implements Business Intelligence (BI) and Self-Service BI solutions at UTAD. Further information is available at www.linkedin.com/in/jmiguelbessa.

Hugo Coelho is currently studying Computer Science at the University of Trás-os-Montes e Alto Douro (UTAD), Vila Real, Portugal, where he acquired his bachelor's degree in 2015 and he is now completing his master's degree. He is also researching the applications that a security information and event management have in the context of the university that he attends. Further information is available at www.linkedin.com/in/coelhohu.

Pedro Monteiro is a Computer Science student at University of Trás-os-Montes and Alto Douro (UTAD), Vila Real, Portugal, where he got his bachelor's degree in 2015 and is now completing his master's degree. Besides studies he is currently researching security information and event management systems and their applications in the context of the university he attends. Further information is available at www.linkedin.com/in/monteirop.

José Brito is a Computer Science student at University of Trás-os-Montes and Alto Douro (UTAD), Vila Real, Portugal, he is currently researching security information, IoT, Big Data and Data Visualization Services and their applications in the context of the university he attends. Further information is available at www.linkedin.com/in/josepedrobrito.

António Costa is a ICT specialist at the University of Trás-os-Montes e Alto Douro (UTAD), Vila Real, Portugal, and is responsible for the coordination the areas of core infrastructure and communications, computer security areas, data center, VoIP and communications networks. He collaborates in teaching on different degrees of Computer Courses, as well as in research, extension and development projects. Holds a degree in Electrical Engineering (specialization in Electronics, Instrumentation and Computation) and a post-graduate degree in engineering area. Currently, he is in the final research stage to complete the PhD in Computer Sciences. He made several made courses or specializations which includes the Upper Course Director for Public Administration; Diploma of specialization of the Information Society for Public Administration, SIP Masterclasses and a OpenStack specialization. Further information is available at www.linkedin.com/in/ariocosta.