Secure Information Assets with Data: An Information Security Governance Framework Using Orchestrated Data Analytics from a Holistic Perspective

Huaying Chen
28th Institute
China Electronics Technology Group Corporation (CETC)
Nanjing, China
chenhuaying.cool@163.com

Zhijun Song
28th Institute
China Electronics Technology Group Corporation (CETC)
Nanjing, China

***Abstract*—Information Security (InfoSec) Governance Framework is applied to help in-house professionals, including top-level officers, as well as practitioners doing field work, in order to operate the organizational InfoSec business within a systematic approach. This paper proposes a framework from the perspective of data analytics, focusing on InfoSec data, which the front-line professionals are more familiar with. The framework introduces the logic from data collection to data analysis emphasizing the new generation data derived by raw data as well as corresponding analysis on it. Furthermore, it provides an application model to indicate how to apply the framework to an organizational environment.**

Keywords—Information Secutrity; goverance framwork; data analytics;

I. INTRODUCTION

Information Security (InfoSec) means the protection of information assets [1]. It becomes increasingly important as an elementary business in most organizations nowadays. As Information Technology evolves dramatically, new trends, such as Cloud computing, big data, Internet of Things has become more and more popular and been applied widely in to organizations [2]. This leads to the increase of the amount of information assets and expend the range of the protection. New challenges to information assets protection create a pressing need to establish a systematic defense for valuable information in an organization.

Common solution to counter-attack the invasion to the assets can be categorized into two main sections, which refer to technology-based and management–based methods. The technical methods focus on utilizing software and hardware detecting, diagnosing, as well as migrate risks and halt threats. The managerial approach is applying regulations with particular strategy, to influence people's behavior to prevent the illegal break of the information systems from both inside and outside the organization. InfoSec researchers realize that each approach has its own advantages and should be applied coordinately [3].

The reason of undertaking this research is twofold. On the one hand, it builds a framework in order to cut off the boundary and combine these methods from data perspective, so that the practitioners can make the best use of the advantages of both methods. Moreover, the framework may also guide them to analyze InfoSec data, then compose some solutions based on the result of the analysis to improve the existing defense. On the other hand, it provides a conceptual prototype to a software that can assist professionals with the analysis to be more comprehensively and profoundly.

This paper is structured as follows. Firstly, we present a review of previous study on InfoSec framework followed by some commons of these frameworks, which drives the research questions (RQs). We then propose our framework and application model of the framework to answer the RQs. Finally, we sum up with the outcomes, as well as the future work of meliorate the framework.

II. LITERATURE REVIEW

The focus of this literature review is on the InfoSec Governance Framework itself, as well as the method of building it. However, there is little or no previous research on InfoSec framework, particular in governance. Therefore, studies reviewed are more general in terms of framework related on InfoSec content.

Despite on detail purposes and general environment, considering the InfoSec framework, researches have constructed few frameworks from numerous perspectives. The logic of building framework varies from studies, but can be categories into two sections. One is on experience of InfoSec Engineering, while the other is to adapt business model into InfoSec content.

On a glance of engineering-based framework, scientists have proposed the framework following engineering procedure and structure. Rebollo et al. [4] have developed a framework from the process of planning/strategy definition, security analysis, then designing the system, implementing it, operating it and terminating it. Whilst Feng et al. [5] proposed another one, it structures the framework from infrastructure level (IT infrastructure), to fundamental level (OS, database etc.), to application level (workbench application, map application, entertainment application etc.). Yang et al. [6] conducted an integrated InfoSec framework based on the requirements generating by risk assessment. Also based on risk migration,

Otoom et al. [7] composed a framework with a top-level strategy and risk analysis as well. On the other hand, scientists have also contributed to establish a framework using a managerial lens. Herath et al. [8] has developed a conceptual framework applying balanced scored card. By evaluating the performance of the current InfoSec system inside organization, the framework means to find the inefficient place and providing improvement guidance to commissioners, thus enhance the InfoSec defense.

Engineering experience and business both provide profound thought of InfoSec framework, as well as standard procedure that how it operates. These kind of frameworks are proposed from a higher view in an enterprise. Furthermore, the improvement suggested by those frameworks may needs authorization from the top. Therefore, the frameworks are more suitable for top-level managers and not easily to be implemented by field-work staff.

III. Research Questions

In-house InfoSec professionals are the major operators proceeding the organizational InfoSec business. In addition, InfoSec software and hardware play a significate role in InfoSec business as well. Since the data generated by them describes the overall situation of the InfoSec environment of an organization, establishing an InfoSec governance framework from data view is necessary and reasonable. Furthermore, there is still an interesting gap of conducting an InfoSec framework from a data analysis perspective according to the literature review. Therefore, this research is trying to answer following research questions:

- RQ1: What is an InfoSec governance framework building from a data analysis perspective?
- RQ2: How to use it to meliorate the organizational InfoSec defense in field work?

IV. InfoSec Governance Framework Based on Holistic Orchestrated Data Analytics

The framework is introduced by the root process it is based on, its key elements, the idea of holistic and orchestrated data analysis, and the detail structure.

A. *Root process*

The Root process in this research is defined as the basic process that our framework is built on. It helps to explain how the framework operates. In this framework, it is a business analytics process proposed by Howes et al [9]. This process is chosen, because the analysis is as the core in our framework. The basic business analysis process presents a logic that how organizational business process is influenced by making decision through data analysis as following (Figure a).

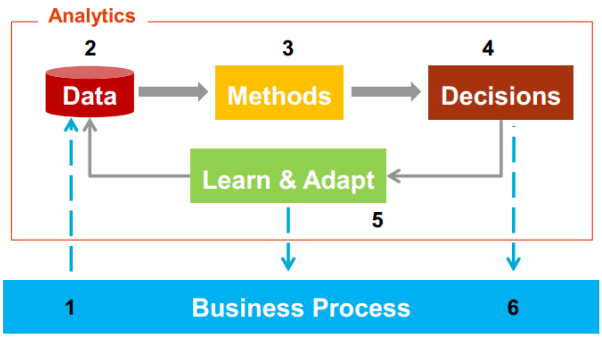

Figure a Business Analytics Process[9]

Firstly, data is generated by business process in an organization. After applying some method, the raw data become meaningful data supporting professionals in making decision by providing statistics evidence with particular subject. Then the decision makes some adjustment to meliorate business process, or to the data, which triggers a new cycle of business analytics.

Since the protection of information assets is also regarded as a common business process in an organization [10], it is rational to adapt the root process into InfoSec content. The data, in this setting, is generated from InfoSec related processes, as well as analyzed to support decision making. Our framework emphasizes data analysis, the framework, therefore, is established to demonstrate how to absorb, fusion, analyze the data, and the outcomes of data analysis. This makes the "data" and "method" become the key elements of our framework.

B. *Key Elements: InfoSec Data and Methods*

The data related to InfoSec contains a wide range and should be all collected to better support decision making. However, this requirement is far beyond the capability of today's InfoSec products in an organization, because of its cost [11]. Linking to common industrial situation, we only discuss the data that be restored by typical InfoSec products.

1) Data Absorbing

ISO 27001 provides a sets of suggestion that how an organization should consider its InfoSec business [12]. According to it, the raw data involves mostly in InfoSec business of Physical and environmental security, Communications and operations management, Access Control, Information Systems Acquisition, Development and Maintenance, Business continuity management etc., while others, such as Information Security Policy, Organization of Information Security, normally include no typical products, thus no related data. This indicates that majority of raw data obtained from two sources, one is systems' log, while the other is InfoSec libraries' data, which includes data in Firewall Rules Library, Virus Database, and IDS/IPS Rules Library etc. These libraries obtain not only from in-house professionals' experience, but also from outside experts' study. Since the data is the core of the framework, the raw InfoSec data collected

becomes the basis of the data analysis.

2) Data cleaning

While the raw data is acquired, it's still in an in-analyzable status, because of data duplication, different format etc.[13]. Plenty methods can be utilized to clean the data. Since these methods are hardly different from ETL process [14], a typical ETL process is chosen for data cleaning. In our content, ETL process is extracting systems' log and InfoSec libraries' data to acquire metadata, transforming these data into readable structure decided by specific subject of analysis and loading into a certain place preparing for analyzing. Despite the different source, we only concern the content of the data. Therefore, it is necessary to combine same log or library items which may be conducted by different products.

3) Data analyzing

Data analyzing means to consider all components of data from different parts and levels trying to answer specific question(s) [15]. In this case, questions on organizational InfoSec rise and is needed to be answered through data analyzing process. The questions and the logic between them are showed as following model (Figure b).

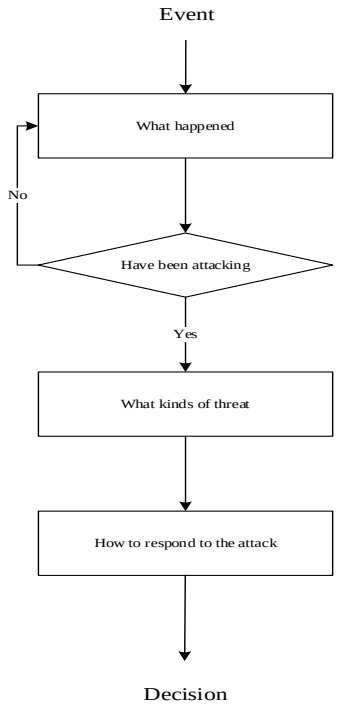

Figure b Data Analyzing Model

a) What is happened?-understanding the current situation.

Understanding the current situation is as the first and important step to data analyzing. Since the question is descriptive, it is reasonable to structure the answer with 5 "W"s -who, when, where, what and how [16]. Element of "Who", in InfoSec content, means a user, who directly or indirectly operates Information Systems. Considering from the perspective of Information Systems, a user may not correspond to a person precisely. Additionally, a user fraud case is not concerned in this level, because the main purpose is to clarify the situation. Element of "When" is to point out the start/end moment and the duration of the operation. Element "Where" focuses on the source and the destination a user comes from and goes to. Mostly, it shows the IP or an address of a user and the system it logged into. Element of "What" emphases the action a user did, e.g. accessing a system, add/delete/modify the data, download/upload files etc. Element of "How" concerns the approach that a user use to make the action done, such as using VPN to access the system.

b) Have been attacking?-assessing the situation.

Assessing the situation is to identify the illegal components among the current situation. Prevalent approaches can be categorized into two sections. One is to comparing the characteristic elements abstracting from the action among the situation with the corresponding data in the InfoSec libraries. Once there is a match, an attack is confirmed. The other is to distinguish the abnormal action from the situation. An action which does not fit in the normal rules is considered to be abnormal, e.g. login a system over a regular time. If no attack is found, the next loop starts from the top of understanding the situation. One the other hand, while an attack is identified, a further analysis need to be proceeded.

c) What kinds of attack it is? -knowing the attack.

Knowing the attack is to find out the method it utilizes, the influence it may have, the damage it may cause. This step is considered as analyzing the attack and providing more information specific to the attack.

d) How to respond to the attack? -responding the attack.

Responding the attack is to figure out what kinds of action should be taken to eliminate the attack. As soon as the attack is confirmed, the suggestion of count-attacking this attack can be obtained from InfoSec libraries. Based on this, professionals make the decision and solution to minimize the damage it may cause. After that, a new circle of data analyzing is imitated from the top of understating a new situation.

C. Holistic and Orchestrated Data Analysis

According to the model above, data analysis is considered to understanding and assessing the situation, knowing and responding to the attack. However, depending on the experience of professionals and real-time data, these processes only can identify and deal with known threat, which the organization has encountered. It is still a difficult issue to identifying and responding to the known threat, which is more likely to cause a large damage [17]. Furthermore, it is leak of suggestion to improve the overall InfoSec defense, as well as quantitative evidence to support the suggestion. It needs more data and powerful analysis on it.

Data abstracted from InfoSec business after ETL process constitutes the 1st generation organizational InfoSec data. More new data can be obtained by mining it. When more new data is gathered, new analysis can proceed. The idea of holistic and orchestrated data analytics is to generate data involving all components of InfoSec environment and emphasizing the logic between them. The Data Generation Model presents as followed (Figure c).

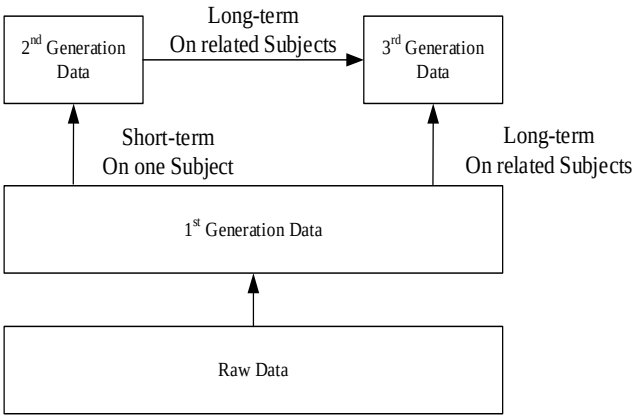

Figure c Data Generation Model

The 2nd generation data is born by combining periodical data in a short-term on a same subject, e.g., intrusion records in a week, virus records in a mouth etc. The analysis on this set of periodical data focuses on three questions:

- How many and how often the organization has encountered threats during that period?
- What characteristics those attacks have in common?
- When and how the next attack may occur?

The first two is the pre-questions of the last one. They lead a statistic analysis and provide a quantitative base for predicted analysis. The answer of the last question gives more information of the next threat that may happen, which enhance the capability of detecting and handling unknown attack.

Moreover, 3rd generation data is concreating by integrating the basic and the 2nd generation data with related subject in a long period, e.g. log in Information Systems and its related database. Linking the data with related subject provide a comprehensive understanding of InfoSec data in an organizational wide and providing more information to identify the real problem behind the threat. When the vulnerability is pointed out, the corresponding solution to improve the defense can be conducted. Thus, the whole InfoSec defense is meliorated by dispelling most vulnerability.

D. The Framework

Following the logic demonstrated above, the framework is established by investigating organizational InfoSec data, methods of absorbing data, cleaning it, the logic of analyzing it, as well as the data it derives and profound analysis against it. The Holistic Orchestrated Data Analytics InfoSec Governance Framework is shown as Figure d.

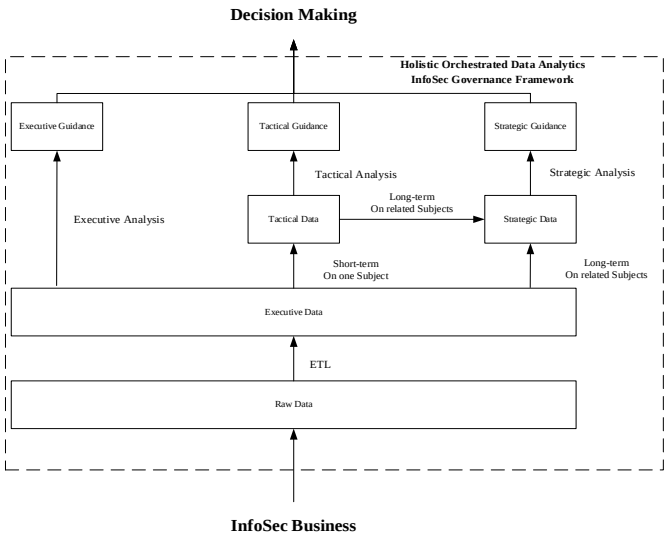

Figure d Holistic Orchestrated Data Analytics InfoSec Governance Framework

From bottom to top, the raw data is abstracted from typical products using in InfoSec business. After ETL process, the data is cleaned and prepared for analyzing. Then the data is conducted and categorized into executive, tactic, and strategic data. Correspondingly, the analysis is divided into executive, tactic and strategic level.

Three level data is composed based on the difference on period and subjects. Executive data is the first-level data, which is obtained from the products in real-time. Periodically, executive data composes tactic data, which combines the data on same subject in a short period. Then strategic data is integrating the executive data and tactical data with the logic of related subject in a long term.

Against three level data, the analysis is proceeded in three level as well. The executive analysis is conducted focusing on understanding the current situation and assessing, responding to the attack. Subsequently, the tactical analyzing is towards the tactical data emphasizing the frequency and pattern of the threat and predicting the unknowns. Then strategic analysis is to figure out the essential issue behind the threat and locating the vulnerability.

Each analysis provides corresponding guidance as result to professionals on different objectives. Executive analysis towards the current situation provides a better understanding to the situation and improves the accuracy and efficiency of evaluating the attack and count-attacking it. Tactic analysis concludes the pattern of threat outputting the multi-characteristics of the potential threat, and gives a prediction unknown threat, which can help to detecting and responding to unknown threat. Strategic analysis identifies the essential vulnerabilities of the InfoSec defense outputting the strategic suggestions of improving the overall InfoSec defense in an organizational aspect.

E. Application Model

The framework outcomes guidance as result of the data analysis. It is meaningful to apply the guidance into the

organizational environment so that the defense can be improved. The application model indicates the logic of adapting the framework into the InfoSec Business in an organization shown as followed (Figure e).

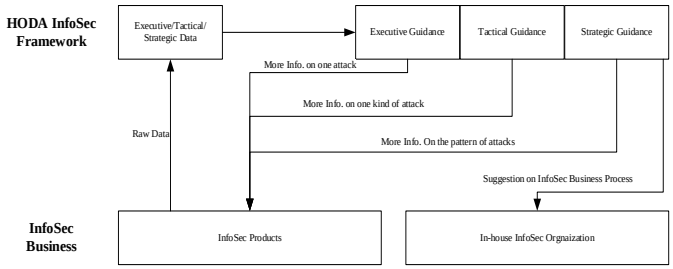

Figure e Application Model

According to the figure, different kinds of guidance drive different solution through different method. Firstly, executive guidance enriches the libs of InfoSec products with more detail characteristics of a certain attack. By acknowledging more characteristics of an attack, it improves the accuracy and efficiency of assessing and responding to a known attack. Since the analysis is in real-time, this kind of results are implied in real-time. Secondly, tactical guidance helps expand the range of the libs, and improve their capacity. The tactical results identified the pattern of a kind of attack, which add several items in libs to figure any transform of that kind of attack. The key difference of the application between executive and tactical analysis results is that executive results enrich an item with more detail information to one attack, while the tactical analysis results enrich the number of items in a lib relating to a kind of attack. Last but not least, into a top level, strategic results expend the libs within several related products. Furthermore, it can alter the deployment of the products to achieve a better capability of cooperation. Moreover, it provides evidence of deploying more product on necessary location. It influences not only products, but also provides guidance directly to InfoSec business through implementing regulations so that the processes are changed into a better way. And this may take a long-term scenario.

Generally, analysis results deliver guidance in corresponding levels improving overall InfoSec defense by enhancing the InfoSec products' performance and meliorating the InfoSec business process. Specifically, by applying the framework, the single products capability, the related products capability of collaboration, InfoSec business process, all be enhanced. Thus, the whole InfoSec defense system is improved.

V. CONCLUSION

In conclusion, the framework proposed in this paper is built on the basis of data analysis. It follows the logic of data absorbing, data cleaning, data analysis, holistic data generating, and orchestrated data analysis. Then the analysis delivers executive, tactical, strategic guidance as outcomes. Additionally, an application model is also given in this paper. It helps professionals to use the framework in the field work

More research is still needed to detail the data analysis with more comprehensive and profound analysis questions. Moreover, it is necessary to apply the framework in a real industrial content and improve it according to the feedback. In addition, based on the framework, a software platform can be design to be an intelligent tool to facilitate the data analysis.

REFERENCES

[1] ISO/IEC 27000:2009: Information technology — Security techniques — Information security management systems — Overview and vocabulary, International Organization for Standardization, Geneva, Switzerland, Available at: http://standards.iso.org/ittf/licence.html

[2] MacDonald, Neil, 2012, Information Security is Becoming aBig Data Analytic Problem, Gartner, (23 March 2012), DOI= http://www.gartner.com/id=1960615.

[3] Changxiang Shen, Huanguo Zhang, Dengguo Feng, Zhenfu Cao, Jiwu Huang, "Informayion Security Overview," Science in China, vol. 37-2, 2007, pp. 129-150.

[4] Rebollo, Oscar, et al. "Empirical evaluation of a cloud computing information security governance framework." Information & Software Technology 58(2015):44-57.

[5] Feng, Deng Guo, et al. "Study on Cloud Computing Security." Journal of Software 22.1(2011):71-83.

[6] Yang, Tsung Han. "An integrated system for information security management with the unified framework." Journal of Risk Research 19.1(2014):1-21.

[7] Otoom, Ahmed, and I. Atoum. "An Implementation Framework (IF) for the National Information Assurance and Cyber Security Strategy (NIACSS) of Jordan." International Arab Journal of Information Technology 10.4(2013):342-348.

[8] Tejaswini Herath, Hemantha Herath, and Wayne G. Bremser. "Balanced Scorecard Implementation of Security Strategies: A Framework for IT Security Performance Management. " Information Systems Management 27.1(2010):72-81.

[9] Howes, Joshua, et al. "Enabling trustworthy spaces via orchestrated analytical security." Proceedings of the Eighth Annual Cyber Security and Information Intelligence Research Workshop ACM, 2013:718-725.

[10] Yaokumah, Winfred, and S. Brown. "An empirical examination of the relationship between information security/business strategic alignment and information security governance domain areas." Journal of Business Systems Governance & Ethics 9.2(2015).

[11] Zuo, Zheng, et al. "Research on information security cost based on game-theory." Industrial Electronics and Applications 2013:1435-1436.

[12] ISO/IEC 27001:2013: Information technology- Security techniques-Information security management systems-Requirements, International Organization for Standardization, Geneva, Switzerland, Available at: http://standards.iso.org/ittf/licence.html

[13] Johnson, Theodore, and T. Dasu. "Data quality and data cleaning." ACM SIGMOD International Conference on 2003.

[14] Deshmukh, Ratnadeep R., and V. C. Wangikar. "Data Cleaning: Current Approaches and Issues." IEEE International Conference on Knowledge Engineering 2011.

[15] Bryk, Anthony S., and S. W. Raudenbush. "Hierarchical linear models:, Applications and data analysis methods. " Journal of the American Statistical Association 88.463(1992):767-768.

[16] Morse, David R., S. Armstrong, and A. K. Dey. "The what, who, where, when, why and how of context-awareness." CHI '00 Extended Abstracts on Human Factors in Computing Systems ACM, 2000:371-371.

[17] Sengupta, Sankar, M. Deneweth, and R. P. V. Til. "Cooperation of Intelligent Honeypots to Detect Unknown Malicious Codes." The Workshop on Information Security Threats Data Collection & Sharing 2008:31-39.