International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Matching Between SIEM Tools and Smart DLC Systems

Mohammed EL ARASS, Iman TIKITO, Nissrine SOUISSI

Abstract*: Nowadays, cybersecurity data management has become a challenging issue especially with the emergence of Big Data. This paper introduces the System of Systems (SoS) paradigm to design a new generation SIEM POC (Security Information Event Management Proof Of Concept) made up of an open source Big Data platform ELK and integrated with other open source security and load-balancing tools. To do this, we first focused on the Big Data and Smart Data requirements to model a data lifecycle from the literature named Smart DLC to the System of 7 Systems, So7S. Second, we used the proposed cycle as SoS tools design, implement and test the proposed SIEM POC by matching the cybersecurity tools to each system of the SoS modeled. The proposed open source SIEM is operational and meets all cybersecurity monitoring requirements with challenging results and may interest small and medium-sized companies dealing with cybersecurity issues.

Keywords* : Big Data, Cybersecurity, Data LifeCycle, Smart DLC, Security Information Event Management, System Of Systems.

I. INTRODUCTION

Nowadays, cyber-attacks have become more and more recurrent and can affect any type of business from small to large companies and whatever their way of securing their information system by installing equipment defense security like firewalls or anti-virus software. One of the solutions that proved its results is to set up a Security Information Event Management (SIEM) which has been defined in [1], [2] as a software solution that provides security information of an Information System. However, [3] defines the SIEM as a set of security software tools including the following systems: log management, log security, event management, security information management, and security event correlation.
The main functions of the SIEM can be summarized as follows:

- Log collection and analysis
- Log Transfer in a standard format
- Security threats notification
- Security incident detection
- Incident response workflow

However, classical SIEMs don't deal with Big Data cybersecurity issues because 80% of the global data is unstructured and large and up to 95% cannot be analyzed automatically [4], [5]. Commercial SIEMs that rely on relational databases struggle against the strangulation of their databases for companies with consistent IS [6]. The case study conducted by Zions Bancorporation in [7] revealed that it would take 20 minutes to an hour to query a month's security data from their traditional SIEMs. However, the same query takes only one minute using Big Data platforms such as Hadoop. So, SIEM based on Big Data technologies is the proper solution to deal with these issues in order to optimize the performance of Big Data management. The purpose of this paper is to design and implement a new generation SIEM POC (Proof Of Concept) that manages the huge amount of cybersecurity data. However, this work is very complex which makes us think of the System Of Systems (SoS) paradigm to reduce this complexity. Because data in a SIEM follow a lifecycle, we have also adopted a challenging Data LifeCycle (DLC) in the literature named Smart DLC [8] adapted to Big Data context and which was applied to the SoS paradigm in [9]. Jamshidi defines in [10] a SoS as a "Super System" consisting of several elements which themselves are operational, independent and complex systems and which interact with each other for a common purpose. Indeed, the SoS is a collection of dedicated systems that combine their resources and capabilities to create a new, more complex system that offers more features and performance than just the sum of the constituent systems [11]. Maier defined five characteristics of a SoS in [12]. This definition has been taken up by several contributions [9], [10], [13]–[17] which verify through these characteristics their Complex Systems as SoS.The rest of this paper is structured as follows: Section 2 identifies Smart DLC systems. Section 3 presents the Big Data and Smart Data requirements. Section 4 presents the proposed SIEM POC. Section 5 illustrates the SIEM POC implementation. Finally, Section 6 summarizes the contributions of the paper.

II. SMART DLC SYSTEMS

Smart DLC proposed in [8] is a data lifecycle modeled as a process cartography resulting from the ISO 9001: 2015 standard [18] and the CIGREF framework [19]. Smart DLC distinguishes three types of processes:

- **Operational processes** have a direct impact on the data received.

Manuscript published on November 30, 2019.
* Correspondence Author
Mohammed EL ARASS*, Mohammed V University in Rabat, EMI-SIWEB Team, Rabat Morocco
Iman TIKITO, Mohammed V University in Rabat, EMI-SIWEB Team, Rabat Morocco
Nissrine SOUISSI, Mines-Rabat School, Department of Computer Science, Rabat Morocco

- **Management processes** determine how operational phases should work.
- **Support processes** collaborate with the operational phases to achieve their assignment.

Smart DLC proposed in [8] adapted to Big Data consists of complex processes which made the same authors think in [9] that the paradigm "System of Systems" (SoS) represents an appropriate theoretical framework for dealing with these issues. Indeed, the phases or processes that make up the Smart DLC in [8] could be systems in its own right which makes the proposed lifecycle as SoS. Each constitutive system is generally a complex, operational and independent system but the collaboration of all these systems makes it possible to fulfil a well-defined goal which cannot be achieved if each system is set up alone [9]. Smart DLC proposed in [8] as a process cartography was modeled under a SoS standpoint. Smart DLC has been demonstrated as SoS in [9] by checking each Maier characteristic [12].

A. Systems identification

The data lifecycle Smart DLC proposed in [8] is composed of 7 operational, independent, autonomous and complex systems that interact with each other for a common goal that is to make smart data for better management and decision-making effective [20]: *Management System, Collect System, Storage System, Analysis System, Visualization System, Archiving System,* and *Support System*. The collect system has been defined in [20] and for the rest of the include systems, they are illustrated in Figure 1.

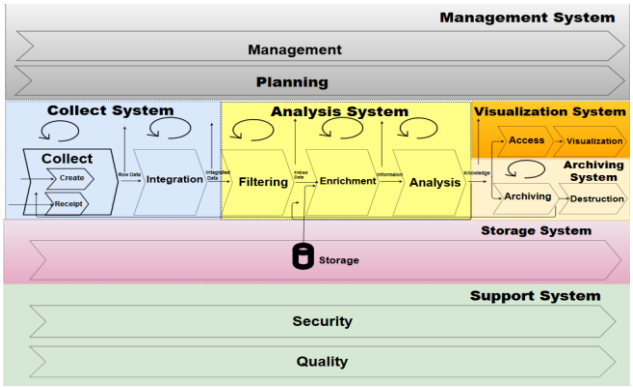

Fig. 1 Smart DLC systems [9]

- The *Management System* is the set of processes that manage the entire end-to-end lifecycle to make communication between all phases effective. It enables the identification and capitalization of good practices and the management of the internal control of the cycle. Also, it makes possible to measure the internal satisfaction of the service rendered by the lifecycle.
- The *Collect System* is a decisive system in the So7S because the DLC intelligence depends a lot on the way this system works. This system consists of receiving raw data of different natures and making the necessary conversions and modifications to organize them. Cleaning the data received in real time saves computing time and memory space. The quality of data must be carried out at this level because it optimizes the overall data processing circuit, which can be very costly in the Big Data context. It may be necessary to find a balance between speed of access to information and quality requirements [21].
- The *Analysis System* is the main element of Smart DLC, which strongly contributes to the transition from Big Data to Smart Data. It allows to drawing knowledge from raw data.
- The *Visualization System* consists of access and presenting the results of the analysis system in a comprehensible and intelligent way thus facilitating the decision-making of the strategic managers of the company. These results are either viewed or archived and then destroyed if the data becomes obsolete.
- The *Archiving System* is about long-term storage of data for possible use. In [22], effective data lifecycle management includes intelligence not only at the archive data level but also at the policy of archiving based on specific parameters or business rules, such as the age of data or the last date of their use. This system cannot detach itself from the destruction process when the use of the archived data is completed and if they will become useless and without added value.
- The *Storage System* is the set of hardware and human resources that provide data storage throughout the data lifecycle. The huge volume of data received means that this system is managed intelligently because storage is a fundamental and sensitive element of the Smart DLC, so it must be adapted to its needs.
- The *Support System* consists of the implementation of the security and quality necessary for other systems for their proper functioning.

B. Block Definition Diagram

To formalize the different systems that make up our SoS, we used SysML, which is an OMG standard used for the specification, analysis, design, verification, and validation of heterogeneous systems and SoS. According to SysML, we represent a structure of the Smart DLC asking questions of requirements. The choice of these 7 systems has been made to respect the characteristics of the systems that constitute a SoS. These 7 systems are complex, operational and independent, but their collaboration fulfils a well-defined goal of Smart DLC, which is to move from Big Data to Smart Data. The purpose of the block definition diagram is to define the hierarchy of our SoS and the system/component classifications.

Figure 2 represents a set of blocks constituting the architecture of the Smart DLC and its associations.

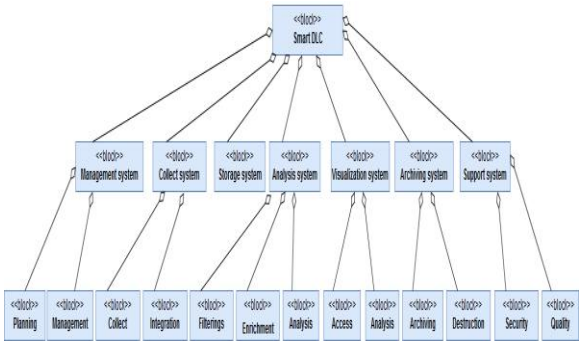

Fig.2. Smart DLC Block Definition Diagram

III. SMART DLC REQUIREMENTS

In this section, we present Big Data and Smart Data requirements that Smart DLC must meet.

Applying the SoS paradigm to the whole phases of a data lifecycle was the first attempt in the literature. [20] has applied this concept to only the first data lifecycle phase (collect) and considered it as a system, but, for us, we have applied it to the seven systems that made up Smart DLC.

For Big Data requirements presented in Table I, we used [20] which links them to its 7 Vs of Big Data: *Volume*, *Velocity*, *Variety*, *Veracity*, *Value*, *Variability*, and *Visualization*. Each requirement must meet the criteria. However, we did not take *Visualization* as a requirement for Big Data but rather for Smart Data because visualizing is results of Big Data analysis is a major challenge to have Smart Data. Another reason is that visualization is a characteristic of Smart Data rather than Big Data. Visualization is used to display knowledge extracted from raw data in order to make data Smart, so this convinced us to put visualization as a requirement for Smart Data.

We have introduced a new kind of requirements that Smart DLC must meet to be Smart. These are the requirements related to Smart Data because the main objective of data management in the Big Data context is to be able to move from Big Data to Smart Data, which means collecting raw and worthless data and transforming them into valuable insights that enable to make reliable and effective decisions. That's why we're not focused only on Big Data Vs because these Vs only deal with *Big*. However, our goal is to formalize a data lifecycle that participates in effective data management where the dominant of Smart is primordial in the Smart DLC. Thus, we based ourselves on the challenges presented in [5] to translate them into requirements and criteria of these requirements. These requirements are as follows:

• *Reliable infrastructure*: Access and data processing are fields that are affected by standards in different ways. Standards can be applied to file formats, access protocols, and system interfaces. The Smart Data infrastructure is, in most cases, provided by a Cloud environment [5]. The infrastructure that supports Smart DLC must be able to give it high service availability, even without interruption. It must offer reliable and valuable services.

• *Analysis Technologies*: Data analysis is the most important phase in a data lifecycle [23]. Algorithms, tools, or libraries can be standardized to achieve the same results across different infrastructures or data formats. This can be tested using benchmarks and other standardized tests. The results must be readable to enable the visualization system to play its role, and the analytical results must be documented, for which standardized forms can also be used [24].

• *Interoperability*: Smart DLC as SoS consists of autonomous and independent systems but they must cooperate to fulfil a well-defined objective. This is why the interoperability of the different systems is essential to effectively fulfil this objective.

• *Portability*: While interoperability focuses on technology, these aforementioned lock-in effects can also be problematic when dealing with data processed by a smart data provider. Being locked by a single provider or service can be a source of slowdowns in the growth process when a provider no longer satisfies requirements. Thus, the value of the service can be improved by interchangeability. The data itself must remain portable ie, the source, intermediate and result data must be transferable from one provider to another.

• *Visualization*: Although data analysis plays a big part in the transfer of Big Data to Smart Data, but if the results it offers are not displayed properly knowledge will lose their reliability resulting in a decrease in the intelligence of the data generated.

• *Conformity*: For successful service delivery, providers must comply with different regulations and constraints including societal, business and legal issues must be satisfied. General acceptance needs to be embraced by society, government institutions can observe if companies respect concepts for national markets, finances and systems, judicial institutions and competitors and can also be verified to comply with the law and other legal provisions [24].

Each of the above-mentioned requirements contains other requirements, which are essential criteria for the fulfilment of the requirement. Tables I and II summarize respectively the criteria for Big Data and Smart Data requirements.

Table-I: Criteria for Big Data Requirements

Requirement	Criterion	Description
Volume	• Capacity • Storage • Transactions	Bandwidth capacity Devices, Cloud, … Time dimension to describe the data
Velocity	• Speed • Real time • Frequency • Data flow	The rapidity to gather data The data should be collected once it's needed and available. The occurrence to gather data The number of data able to move from one point to another
Variety	• Structured • Unstructured	Formalized data using a clear data model Data model nor predefined organized without a predefined data

	• Different formats	Html, jpeg, PDF, video, …
	• Different sources	Connected Social objects, network, Paper, … databases, Cookies, GPS,
	• Trust	Belief in the reliability
	• Authenticity	Data received is original and without variation received
	• Root	The source of data
	• Accuracy	The results of collecting data should be exact
Veracity	• Integrity	The data records are real and were not faked or modified
	• Statistics	Characteristic or measure obtained from a snip of data
	• Events	Data actions
	• Dependencies	Define the dependencies between data
	• Issues	The value shouldn't generate any problem and be exploitable
Value	• Interpretability	The value of information should provide a meaning
	• Customer requirements	Customers request
	• Appropriate amount	Suitable quantity to have the required data
	• Quality	Good quality to provide a high level of information
Variability	• Interpretable	To provide comprehensible information

Table-II: Criteria for Smart Data Requirements

Requirement	Criteria	Description
Reliable Infrastructure	• Availability • Scalability • Performance • Security	The data lifecycle infrastructure must be reliable
Interoperability	• Services • Technologies • Management	Each system must be interoperable with the rest of the SoS Smart DLC
Portability	• Data source • Intermediate data • Result data	The data itself must remain portable, that means the source, intermediate and result data must be transferable from one provider to another
Visualization	• Facility • Access • Availability • Understandable	The lifecycle must supervise the data throughout its cycle
Conformity	• Social Regulations • Market Regulations • Legal Regulations	Data Lifecycle providers must comply with different regulations and constraints
Analytical Technologies	• Results Readability • Analysis algorithms • Tests	Analysis technologies should give the same results on different infrastructures or different data formats
Security	• Integrity • Access control • Private life	Data must be protected against unauthorized access and loss, but also against illegal use

IV. DESIGN OF A SIEM POC

While Big Data systems are able to process heterogeneous and large data, they also generate a significant amount of logs [25]. We believe that Smart DLC requirements and systems could be a basis for designing and implementing a SIEM POC. To do this, we followed a rigorous method to validate both the proposed cycle and the SIEM POC designed.

A. Method

We followed a rigorous method for choosing tools that make up our SIEM POC as well as for its design, implementation, and functional tests.

- *Step 1*: We identified several open source tools that meet the requirements of our cycle and that also meet at least one functional SIEM requirement.
- *Step 2:* We selected tools that can be integrated with each other. In the case where we found several tools for the same function, we chose the most widespread tool and offering a great availability of documentation.
- *Step 3*: We performed a mapping between the systems that make up the proposed cycle and the selected tools.

B. SIEM construction

Often, a tool is composed of several components. Each one provides a function that meets one of the Smart DLC requirements, therefore, concerns one of its systems. For example Snort, it consists of a packet decoder, a pre-processor, and a detection engine. Table III illustrates the match matrix between the selected tools and Smart DLC requirements as So7S in order to verify that all requirements are met at least by one tool.

Table-III: Functional correspondence matrix between SIEM POC tools and Smart DLC requirements

SIEM POC Tools	Definition	Smart DLC requirements												
		Volume	Velocity	Variety	Value	Veracity	Variability	Analysis technologies	Interoperability	Reliable infrastructure	Portability	Conformity	Visualization	Security
Snort	Open source intrusion detection and prevention system (NIDS & NIPS) maintained by CISCO [26].	✓			✓			✓			✓		✓	✓
Sguil	Open source desktop application that offers an intuitive interface for events, session data and raw packets visualization. It was created by Network Security Analysts [27].				✓			✓			✓		✓	✓
Bro	Network analysis system. It provides a complete platform for analyzing network traffic [28].				✓			✓			✓		✓	✓
Redis	Open source in-memory database used to manage the data queue in memory [29].	✓	✓							✓				
Beats	Open source agent platform installed on remote devices to send logs to Logstash.		✓	✓		✓	✓		✓			✓		
Logstash	Dynamic open source pipeline, collects and integrates data simultaneously from a multitude of sources to transform them and send them to another storage system generally elaticsearch [30].	✓	✓		✓			✓	✓	✓				
Elasticsearch	Open source multi-entities RESTful distributed search and analysis engine. It manages a NoSql DB.	✓	✓		✓			✓	✓	✓	✓			
Kibana	Powerful and intuitive visualization tool for data found in elasticsearch.	✓	✓		✓								✓	

After that and based on each tool functions, we identify tools that will be implemented in all Smart DLC systems. Table IV illustrates the functional correspondence matrix between the selected tools and the systems that make up Smart DLC as So7S.

Table-IV: Functional correspondence matrix between SIEM POC tools and Smart DLC as So7S

SIEM POC Tools	Smart DLC systems					
	Collect system	*Analysis system*	*Visualization system*	*Storage system*	*Support system*	*Management system*
Snort	✓	✓	✓	✓	✓	
Sguil	✓	✓	✓	✓	✓	
Bro	✓	✓	✓	✓	✓	

Redis	✓	✓		✓		✓
Beats	✓					✓
Logstash	✓	✓				✓
Elasticsearch	✓	✓				✓
Kibana			✓			

Tools that make up our SIEM POC concern only collect, analysis, visualization, storage, and support systems. For management system, it is essentially a planning system that does not process data but rather plans their collect, analysis, storage, etc. When we were in the tools implementation phase, we also rolled out this system for each integrated tool. As an illustration, when we configured the Snort tool for analysis, storage, support, and visualization systems, we defined its parameters as follows: Intrusion Detection Management Rules, Network variables, Dynamic snort library, Pre-Processor, Output plugins, and Configuration Execution Guidelines.

V. IMPLEMENTATION

We designed a distributed architecture of the SIEM POC to simplify its implementation. And then, we implemented all the selected tools in a distributed Linux environment. Finally, we tested all the SIEM POC tools.

Before starting the SIEM POC implementation, we designed a distributed architecture in which our POC will be integrated. This distributed architecture shown in Figure 3 will optimize its operation.

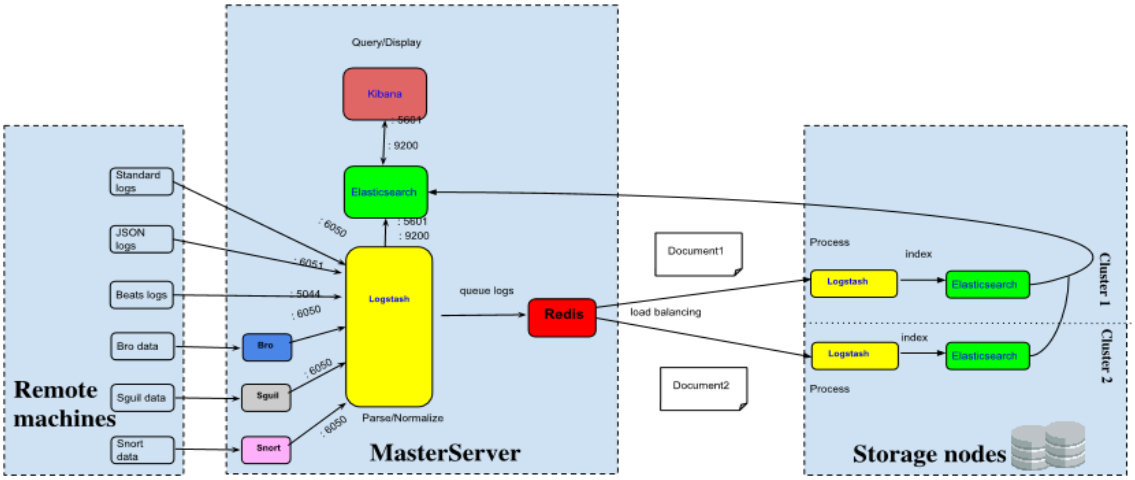

Fig. 3. SIEM POC implementation architecture

We opted for a distributed architecture composed of three types of elements:

- *Remote machines*: These devices are usually the basis for producing Big Data. Indeed, they produce a huge amount of logs and network traffic in the form of structured data (JSON objects, SQL data, etc.) and unstructured data (text, PCAP, logs, etc.). Depending on their type, these data will be received by a specific TCP port of the Master server. Beats that are compatible with ELK are received by port 5044. However, other types of logs and data that require parsing are received by port 6050, such as syslog logs, Windows logs, Bro logs, Sguil logs, Snort logs and network data like Sguil's PCAP.
- *Master server*: It receives all data (Big Data), processes them, analyzes them and visualizes them in an intelligent way (Smart Data).
- *Storage nodes*: They receive logs stored in the memory of the Master server by the Redis tool which provides a load balancing function between the two clusters. The clusters receive documents that are already standardized and parsed by the Master server and process them for presentation to the Elasticsearch engine in the Master server.

To do this, we have implemented all the selected tools in a distributed Linux environment with the technical elements **listed in Table V.**

Table-V: Technical specifications of the implementation

Hardware specifications	Software specifications
•Master server with 8 CPU cores, 16 GB RAM and 1 TB disc space •Storage Node with 4 CPU cores, 8 GB RAM and 100 TB disc space •Forward node with 2 CPU, 2 GB RAM and 500 Go disc space	•VirtualBox 6.0.4 •Linux Ubuntu 16.04 •Elasticsearch-6.6.1.tar.gz •Logstash-6.6.1.tar.gz •Kibana-6.6.1-linux-x86_64.tar.gz •Java SE 11.0.2(LTS) •Redis 5.0.3 •filebeat-6.6.1-linux-x86_64.tar.gz •packetbeat-6.6.1-linux-x86_64.tar.gz •winlogbeat-6.6.1-windows-x86_64.tar.gz •metricbeat-6.6.1-linux-x86_64.tar.gz •Sguil 0.9.0 •snort-2.9.12.tar.gz

We performed functional tests to our SIEM POC to verify that all of its components work and interact with each other without errors. For example, Figure 4 illustrates the Kibana home interface. This shows that this visualization tool has access to all the alerts that have been generated by our SIEM. Figure 5 shows some alerts that were recorded by the filebeats tool installed on a device during the period between 5 February 2019 and 5 March 2019.

Data shown in Figure 5 represent Smart Data in our context. So, the SIEM POC implemented following the proposed architecture has allowed transforming from Big Data to Smart Data.

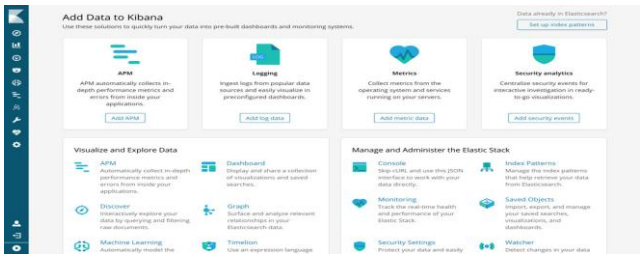

Fig. 4. Kibana home interface

Fig. 5. Visualization of alerts generated by filebeat during the period between 2/5/2019 and 3/5/2019

We have tested the proper functioning of all the tools that make up our SIEM POC. The performance of this POC will be performed later to compare them to free and commercial SIEMs that are on the market. For storage capacity reasons, we monitored a link of 50 Mbps = 6.25 MB / s = 375 MB / minute = 22,500 MB / hour = 540,000 MB / day. It remains to test its behaviour in a production environment with a much larger link. However, this POC is perfectly suited to small and medium-sized companies that typically do not exceed 20 Mpbs.

VI. CONCLUSION

In this article, we have proposed a new generation SIEM POC of made up of a Big Data platform ELK with other intrusion detection and load balancing tools. The proposed SIEM could manage the large heterogeneous networks composed of several devices with a log files centralization in order to make detection and analysis faster and efficient and so deal with the Big Data issues. Because a SIEM is a data lifecycle that consists of several systems, we have used the System of Systems (SoS) paradigm to reduce its complexity management and match intrusion detection systems with the Big Data platform ELK to Smart DLC systems proposed in [8], [9] to design an open source SIEM prototype. The proposed open source SIEM is operational and meets all cybersecurity monitoring requirements and could interest small and medium-sized companies.

Further works is needed to show SIEM POC performance and limitations in a real production environment of a government defense agency.

REFERENCES

1. E. Al-Shaer, J. Wei, K. W. Hamlen, and C. Wang, ‘Towards Intelligent Cyber Deception Systems’, in Autonomous Cyber Deception, Cham: Springer International Publishing, 2019, pp. 21–33.
2. B. Geluvaraj, P. M. Satwik, and T. A. Ashok Kumar, ‘The Future of Cybersecurity: Major Role of Artificial Intelligence, Machine Learning, and Deep Learning in Cyberspace’, in International Conference on Computer Networks and Communication Technologies, vol. 15, S. Smys, R. Bestak, J. I.-Z. Chen, and I. Kotuliak, Eds. Singapore: Springer Singapore, 2019, pp. 739–747.
3. I. Alsmadi, ‘Incident Response’, in The NICE Cyber Security Framework, Cham: Springer International Publishing, 2019, pp. 331–346.
4. M. El arass, I. Tikito, and N. Souissi, ‘An Audit Framework for Data lifecycles in Big Data context’, in Proceeding of The International conference on selected topics in Mobile and Wireless Networking, Tangier, Morocco, 2018.
5. A. Lenk, L. Bonorden, A. Hellmanns, N. Roedder, and S. Jaehnichen, ‘Towards a taxonomy of standards in smart data’, in Big Data (Big Data), 2015 IEEE International Conference on, 2015, pp. 1749–1754.
6. R. Zuech, T. M. Khoshgoftaar, and R. Wald, ‘Intrusion detection and Big Heterogeneous Data: a Survey’, Journal of Big Data, vol. 2, no. 1, Dec. 2015.
7. ‘A Case Study In Security Big Data Analysis’, 2019. [Online]. Available: https://www.darkreading.com/analytics/security-monitoring/a-case-study-in-security-big-data-analysis/d/d-id/1137299. [Accessed: 23-Mar-2019].
8. M. El arass and N. Souissi, ‘Data Lifecycle: From Big Data to SmartData’, in 2018 IEEE 5th International Congress on Information Science and Technology (CiSt), Marrakech, 2018, pp. 80–87.
9. M. El arass, K. Ouazzani Touhami, and N. Souissi, ‘The System of Systems paradigm to reduce the complexity of data lifecycle management. Case of the Security Information and Event Management’, IJSSE, in press.
10. M. Jamshidi, System of systems engineering: innovations for the twenty-first century, vol. 58. John Wiley & Sons, 2011.
11. I. E. daoui, M. Itmi, A. E. Hami, N. Hmina, and T. Mazri, ‘A study of an adaptive approach for systems-of-systems integration’, International Journal of System of Systems Engineering, vol. 9, no. 1, p. 1, 2019.
12. M. W. Maier, ‘Architecting principles for systems-of-systems’, in INCOSE International Symposium, 1996, vol. 6, pp. 565–573.
13. J. Boardman and B. Sauser, ‘System of Systems-the meaning of of’, in System of Systems Engineering, 2006 IEEE/SMC International Conference on, 2006, pp. 6–pp.
14. P. Chen and M. Unewisse, ‘SoS thinking: an approach to conceptualising and understanding military systems-of-systems’, International Journal of System of Systems Engineering, vol. 8, no. 1, p. 74, 2017.
15. F. Lahboube, S. Haidrar, O. Roudiès, N. Souissi, and A. Adil, ‘Systems of systems paradigm in a hospital environment: benefits for requirements elicitation process’, International Review on Computers and Software (IRECOS), vol. 9, no. 10, pp. 1798–1806, 2014.
16. C. B. Nielsen, P. G. Larsen, J. Fitzgerald, J. Woodcock, and J. Peleska, ‘Systems of Systems Engineering: Basic Concepts, Model-Based Techniques, and Research Directions’, ACM Comput. Surv., vol. 48, no. 2, pp. 18:1–18:41, Sep. 2015.
17. B. Nikolopoulos, A. Dimopoulos, M. Nikolaidou, G. Dimitrakopoulos, and D. Anagnostopoulos, ‘A System of Systems Architecture for the Internet of Things exploiting Autonomous Components’, IJSSE, 2019.
18. ISO, ‘ISO 9001 Quality management’, 2015. [Online]. Available: https://www.iso.org/iso-9001-quality-management.html. [Accessed: 10-Jul-2018].
19. M. B. Sophie Bouteiller, ‘Big-Data-Vision-grandes-entreprises-Opportunites-et-enjeux-CIGREF’, 2013.
20. I. Tikito and N. Souissi, ‘Data Collect Requirements Model’, in in proceeding of the 2nd International Conference on Big Data, Cloud and Applications, BDCA’2017, Morocco., 2017.
21. S. Bouteiller, Enjeux business des données. Comment gérer les données de l’entreprise pour créer de la valeur? CIGREF, 2014.
22. IBM, ‘Wrangling big data: Fundamentals of data lifecycle management’, 2013.
23. M. El arass, I. Tikito, and N. Souissi, ‘Data lifecycles analysis: towards intelligent cycle’, in Proceeding of The second International Conference on Intelligent Systems and Computer Vision, ISCV’2017, Fès 17-19 April, Fez, Morocco, 2017.

24. A. Lenk, L. Bonorden, A. Hellmanns, N. Roedder, and S. Jaehnichen, 'Towards a taxonomy of standards in smart data', in Big Data (Big Data), 2015 IEEE International Conference on, 2015, pp. 1749–1754.
25. P. Wu et al., 'Bigdata logs analysis based on seq2seq networks for cognitive Internet of Things', Future Generation Computer Systems, vol. 90, pp. 477–488, Jan. 2019.
26. CISCO, 'Snort website', 2019. [Online]. Available: https://snort.org/documents. [Accessed: 08-Mar-2019].
27. Network Security Analyst, 'Sguil - Open Source Network Security Monitoring', 2019. [Online]. Available: http://bammv.github.io/sguil/index.html. [Accessed: 08-Mar-2019].
28. Zeek, 'The Zeek Network Security Monitor', 2019. [Online]. Available: https://www.zeek.org/. [Accessed: 09-Mar-2019].
29. Redislabs, 'Redis', 2019. [Online]. Available: https://redis.io/. [Accessed: 08-Mar-2019].
30. Elasticsearch, 'Logstash', 2019. [Online]. Available: https://www.elastic.co/fr/products/logstash. [Accessed: 08-Mar-2019].

AUTHORS PROFILE

Mohammed EL ARASS is a Cyber Security Project Manager in Moroccan defense agency and member of Information System and WEB (SIWEB) in EMI School. He received his engineering degree from INSA Lyon in 2017. His research interests include Data lifecycle, Big Data management, System of Systems, and Cybersecurity.

Iman TIKITO has more than 6 years of international experience as Business Analyst and Project Lead, working for a multinational company. She holds a double Master Degree in IT Applied to Offshore Development from the University of Mohammed V at Morocco, a master degree in Offshore Development of Information Systems from University of Bretagne Occidentale at France. She's currently pursuing her Ph.D. in Science and technology for the engineer at "EMI" Mohammadia School of Engineers.

Nissrine SOUISSI is a fulltime professor at the MINES-RABAT School, Morocco. She obtained a Ph.D. in computer science from the UPEC University in 2006, France and an Engineer degree from Mohammadia School of Engineers in 2001, Morocco. Her research interests include process engineering, business process management, databases, data lifecycle, smart data, hospital information system, and information system.