Security Analytics: Dispelling the Fog

Ilze Birzniece [0000-0002-5775-6138]

Riga Technical University, Kalku 1, Riga, Latvia
ilze.birzniece@rtu.lv

Abstract. This article explores the term "Security analytics" and its different perspectives in up-to-date literature. The research start point is the literature analysis on security analytics topic. Situation awareness leads to the first conclusion that there is no unified definition of what security analytics is and what are its boundaries. Chosen keywords "Security analytics" allowed to select the papers and books from four notable digital libraries to study the perception of the concept. As a result, a classification of security analytics viewpoints is proposed which organizes current research and practical application directions touching security analytics area. The model includes three views: (1*) Analytics of the security*, containing the most popular and the narrower view to security analytics as cyber threat analysis, (2) *Analytics for the security*, addressing security issues beyond the cyberspace, and (3) *Security of the analytics*, considering analytical threats in traditional or machine learning environment. This review paper helps to build a common understanding of the multifaceted area called "Security analytics".

Keywords: Security Analytics, Machine Learning, Cybersecurity, Analytics of the Security, Analytics for the Security, Security of the Analytics.

1 Introduction

Nowadays security has gained full attention due to different kinds of global threats. These threats exist in the physical world, cyberspace or are cultivated in people minds. Security may refer to personal, information or infrastructure assets. It is emphasized that the current generations are more related to the cyberspace than any other. The so-called 'Generation Z' which involves cohorts of people who were born between 1995-2010 are the first native in the Internet world and are suspected to be the most individualistic and technology-dependent generation. Their social network and information usage habits are so much different from the generations before them. Therefore, the threats they are encountering come from a cyber world much more often than physical ones.

Getting the data is not the main challenge anymore. Coordination between data flows and human capabilities to comprehend them and act responsively is much more demanding. Lance James in the foreword of book "Information Security Analytics: Finding Security Insights, Patterns, and Anomalies in Big Data" [1] confirms that the information security field is a challenging with unsolved questions and unclear ground which, unlike physics, astronomy and similar sciences has not had a chance to elaborate theoretical background before these problems came into our lives.

This review paper aims to clarify the current understanding of the term “Security analytics” in both scientific and professional communities. Section 2 seeks for definitions of “Security analytics”. Section 3 gives the methodology and describes outcomes of the literature review regarding the context in which security analytics is mentioned. Section 4 amalgamates the findings and presents a model for classifying different viewpoints of security analytics, followed by conclusions in Section 5.

2 What Is Security Analytics?

The term “Security analytics” is not well grounded yet, therefore, it is rather vaguely understood based on one’s personal experience and background. Neither Cambridge Dictionary [2] nor Oxford Dictionary [3] do not contain the phrase “Security analytics”. Even Wikipedia holds only two sentences describing security analytics as a subsection under Analytics Article, namely, “Security analytics refers to information technology (IT) to gather and analyze security events to understand and analyze events that pose the greatest risk. Products in this area include security information and event management and user behaviour analytics.”

One of a few books calling security analytics its name [1] points at the strength when data science is applied to security issues. In [4] security analytics is defined similar way but narrower as “the application of Big Data Analytics techniques to cybersecurity”. Big data is also a part of [5] definition, which can be generalized for describing security analytics as “area of research and practice aimed at protecting networks, computers, and data from unauthorized access by analyzing security event data”. We get “Big data cybersecurity analytics” when big data tools and technologies are applied for analyzing the event data.

Tech news are up to date and valuable source in describing the actual situation with security analytics. In this way [6] frankly speaks out that there is a misunderstanding between security analytics, security information and event management (SIEM) and user and entity behaviour analytics (UEBA). He refers to SIEM and UEBA tools as a part of security analytics; a security analytics platform should analyze and create the full picture from the company’s existing security tools, which includes SIEM and UEBA.

Another definition comes from [7]: “The ultimate goal of security analytics is to deliver technology solutions that assist human security analysts in detecting, responding to and mitigating cyber threats.”

If we cut down the concept of “Security analytics”, we come to an obvious inference that it involves at least two expertise subjects, *security* and *analytics*, namely. Security is defined as „the state of being free from danger or threat” [3]. Analytics is defined as „the systematic computational analysis of data or statistics” [3]. The typical extensions of “security” and “analytics” are “information security” and “data analytics”, consequently leading to an understanding that data and information are the connection between these two [8].

Analytics within the information security domain is not limited to cyber threat analysis as it is often perceived. Analytic techniques can be applied to mine data and identify patterns and connections in many forms of security-related data.

Different aspects of security analytics are evaluated in a systematic literature review which is described in the next Section.

3 Literature Review

In order to investigate the nature of security analytics and clarify the aspects that this term includes, primary, secondary and tertiary sources are examined. Scientific literature comes from IEEE Xplore, ACM, Elsevier and SpringerLink digital libraries.

3.1 Methodology

The approach includes three search cycles with increasing depth of exploration with the following outcomes of each cycle:

- The appearance of the term "Security analytics" in the scientific literature, using the databases which are available under Riga Technical University e-library, i.e. metadata analysis;
- Analysis of the context in which the term "security analytics" is used to form a more precise division;
- Synthesis of security analytics viewpoints based on findings in the literature and formation of classification.

3.2 Metadata Analysis

There are numerous journals and conference or workshop proceedings in security area which at some extent touches analytical part of security, e.g. ACM Transactions on Information and System Security (TISSEC), Symposium on Bio-inspired Learning and Intelligent Systems for Security, International Conference on Information Security and Cyber Forensics (InfoSec), Journal of Information Security and Applications, IEEE Symposium on Computational Intelligence in Cyber Security (CICS), and many others. However, "security" and "analytics" together in the title appears much fewer.

The search on digital libraries was run on the phrase "Security analytics" delimiting search in (1) Title or (2) Keywords (or author keywords, where such division was possible). The goal of retrieval was to identify the number of individual research papers and publication years to understand the presence of security analytics as a common term. If term appeared in workshop proceeding or symposium metadata, or a book title, they were omitted from results represented in Table 1. Books are analyzed separately afterwards.

For analyzing the appearance of the term "Security analytics", four scientific databases were explored. Table 1 gives a summary of results, showing the number of times when the term appears in keywords or publication title and the corresponding years.

Table 1. The appearance of the term "Security analytics" in the scientific literature

Source	Appearance in key words	Appearance in title
IEEE Xplore	19 (y. 2009-2018)	20 (y. 2010-2018)
ACM	6 (y. 2015-2016)	9 (y. 2011-2018)
Elsevier	4 (y. 2016-2018)	2 (y. 2016-2017)
SpringerLink	No keyword search	7 (y. 2012-2018)

Analysis of search results reveals some interesting facts. First, the intersection of retrieved results, which holds appearance of the term "Security analytics" in both keywords and title, is very small. Second, while the term is present in so-called author keywords, it is not included in IEEE, INSPEC Controlled Indexing or The ACM Computing Classification System (CCS rev.2012) keyword notation. Most typical IEEE keywords associated with these papers are "information management", "data handling", "computer security", "monitoring" etc. This may tell us how the term "security analytics" is perceived in other words. Third, the first appearance of the term "security analytics" was in 2009 in keywords and in 2010 in the publication title. Therefore, we can see that "Security analytics" in scientific publications is used for less than ten years and still lacks maturity.

Results retrieved from SpringerLink include conference proceedings and book chapters since they could be recognized similarly as a journal or conference publications. The keyword search, to the author's knowledge, in SpringerLink, is not provided.

Several books (workshop proceedings, paper collections) have been devoted to Security Analytics and are summarized in Table 2.

Table 2. Security analytics in books

Author/Editor	Title	Year
Editors: Martti Lehto, Pekka Neittaanmäki	Cyber Security: Analytics, Technology and Automation *Intelligent Systems, Control and Automation: Science and Engineering book series*	2015 [9]
Editors: Izzat M Alsmadi, George Karabatis, Ahmed Aleroud	Information Fusion for Cyber-Security Analytics *Studies in Computational Intelligence book series*	2017 [10]
Ehab Al-Shaer, Mohammad Ashiqur Rahman	Security and Resiliency Analytics for Smart Grids *Advances in Information Security book series*	2016 [11]
Program Chairs: Ehab Al-Shaer, Krishna Kant	Proceedings of the 2014 ACM Workshop on Cyber Security Analytics, Intelligence and Automation	2014 [12]
Editors: Elisa Shahbazian, Galina Rogova	Meeting Security Challenges Through Data Analytics and Decision Support *NATO Science for Peace and Security Series*	2016 [13]

Program Chair: Rakesh Verma	Proceedings of the 2015 ACM International Workshop on Security and Privacy Analytics	2015 [14]
Program Chair: Rakesh Verma	Proceedings of the 2016 ACM International Workshop on Security and Privacy Analytics	2016 [15]
Editors: Onur Savas, Julia Deng	Big Data Analytics in Cybersecurity *Data Analytics Applications book series*	2017 [16]
Mark Ryan M. Talabis, Robert McPherson, I. Miyamoto, and Jason L. Martin	Information Security Analytics: Finding Security Insights, Patterns, and Anomalies in Big Data	2014 [1]

The first proceedings or book-wide publications regarding security analytics are dated back to 2014 which allows us to assume that the critical mass for this topic has been received about five years ago.

3.3 Analysis of Security Analytics Context

This section remarks the main findings from IEEE Xplore, SpringerLink, ACM and Elsevier digital libraries as well as other popular internet sources before organizing them in meaningful groups.

It can be noted that analytics is often used in a form “information security analytics” and is usually connected to cybersecurity issues being explained as extending analytical capabilities in information security. The typical sources from which security event data is collected are network traffic data and logs (firewall, web, system, router or database access) [5]. Despite a majority of works devoted to security analytics perception in this way, here the other perspectives will be mentioned as well.

Wider look to security analytics gives a book from *NATO Science for Peace and Security Series* [13] where cybersecurity is just one of the parts, whereas other two are devoted to counter-terrorism and maritime and border security.

Big data technologies are often involved and are mentioned both in cause [17] and the solution [4, 5] side for security topics. In conjunction with big data, security of the smart grid is discussed in several papers [11, 17].

Managerial perspective is also a part of security maintenance. The so-called data-driven security includes analysis, visualization, and dashboards [19].

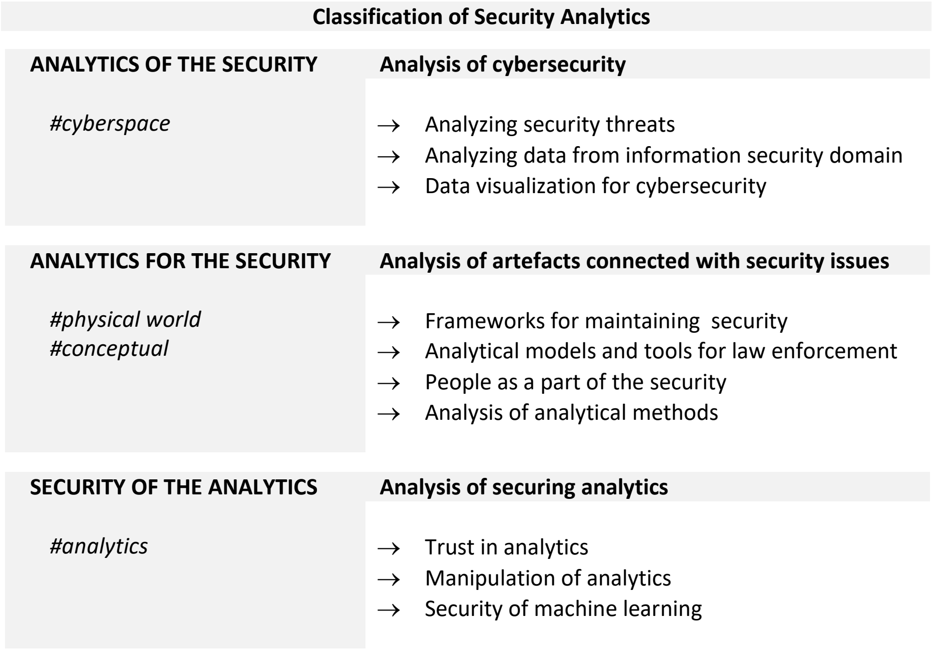
Trust is one more topic associated with both security and analytics. If we define a trust as a willingness to rely on another [20], increasing trust in data analytics include several aspects: trust in (1) initial data, (2) predictive models and (3) outcomes [21].

Security and privacy are terms often used nearby, at the same time reminding that a security does not mean the same as privacy. Both security and privacy are the issues of intensive data transmission and processing tasks, e.g. networking, cloud computing, Internet of Things. Edge and fog computing is the part of the answers to those issues and are described in [22] and other sources.

To draw the full picture we should not forget about the human aspect of security analytics. Cybersecurity is not just about technology; it is also about people, their behaviour, beliefs, and faults. As a result of human multi-faceted nature, people are the weakest point in cybersecurity [23]. If we look from social engineering perspective, we have a recent headline of the *Cambridge Analytica* scandal over its interference in the 2016 US presidential election and social network data sharing practice leading to the question of data usage for analytical purposes. Our privacy is threatened and the security might be as well. These matters are discussed in [24, 25].

4 Security Analytics Viewpoints

From the analysis of related work, several main perception tracks of security analytics can be derived which are represented in a model with three main views to security analytics concept. To specify the nature of security analytics perception they are called (1) *Analytics OF the security*, containing real-life applications for analyzing security data, (2) *Analytics FOR the security*, addressing security issues beyond the cyberspace, and (3) *Security of the analytics* considering analytical threats in traditional or machine learning environment. The following subsections define the organization and contents of each the viewpoint in detail. Figure 1 depicts a proposed classification of security analytics topics.



Classification of Security Analytics	
ANALYTICS OF THE SECURITY	**Analysis of cybersecurity**
#cyberspace	→ Analyzing security threats → Analyzing data from information security domain → Data visualization for cybersecurity
ANALYTICS FOR THE SECURITY	**Analysis of artefacts connected with security issues**
#physical world *#conceptual*	→ Frameworks for maintaining security → Analytical models and tools for law enforcement → People as a part of the security → Analysis of analytical methods
SECURITY OF THE ANALYTICS	**Analysis of securing analytics**
#analytics	→ Trust in analytics → Manipulation of analytics → Security of machine learning

Fig. 1. Security analytics viewpoints

4.1 Analytics of the Security

The narrower view (in Fig.1 gathered under name *Analyzing security threats*) to security analytics in this section includes analyzing security threats from log files with emphasis on data and well-known applications mostly in cyberspace, e.g. security of the network, fraud detection etc. These problems can be broken down into two separate considerations: anomaly detection and threat determination [7].

The broader view (in Fig.1 called *Analyzing data from information security domain*) analyze varied kind of data related to information security and puts the emphasis on the selection of analytical methods as well. At this point, machine learning[1] techniques are mentioned as a help for different analytical tasks. Machine learning is capable of automating several human routine activities or at least assist human expert in a semi-automatic manner [26]. Currently, a large majority of machine learning approaches in security is used as a type of warning system and include a human as a final decision maker [27]. Deep learning is currently hype in machine learning and inevitably involved in security applications as well [28]. Visual inspection for analysis of security issues, data visualization, etc. fits the broader sense of analytics of the security.

This view of security analytics tends to be reactive, analyzing data of actual or past events related to maintaining systems (typically cyber systems) secure. However, applications can also be proactive.

4.2 Analytics for the Security

This view is an umbrella for analyzing artefacts connected with security issues. Analytics for the security purposes covers creation and evaluation of different frameworks for maintaining security and/or privacy, analytical models and tools for law enforcement and people as a part of the security chain.

There fits technically oriented research papers regarding the implementation of infrastructure of security, development of security analytics business strategy or implementation of analysis platform [4, 11]. Architectures and techniques for protecting critical infrastructures like smart grids are the scope of various sections in the book [11], a paradigm of security-by-design is discussed in [29].

Between the frameworks, we should mention purposefully developed analytical frameworks which contain comprehensive tools for supervising law enforcement, helping police and other governmental and private institutions analyze and monitor real-life artefacts, people and their connections in order to maintain our physical and cyber security, e.g. IBM i2 [30]. There are also general-purpose analytical platforms, which could be tailored to process security-related data and analyze security threats.

[1] In this paper the term “machine learning” is used whereas in different sources “data mining” is also present. We assume that data mining makes use of machine learning methods for business purposes.

A systematic literature review regarding architectural tactics for big data cybersecurity analytic systems has been carried out in [5]. Big data cybersecurity system is defined as a combination of traditional cybersecurity solution with big data tools.

On the borderline with *Security of the analytics* is the analysis of analytical methods for particular issues, e.g. framework for automatically investigating security alerts with the goal of understanding whether and which anomaly detection approaches can be adopted for identifying relevant security events [31].

This view of security analytics tends to be proactive and more conceptual than *Analytics of the security*.

4.3 Security of the Analytics

Apart from analyzing security, security of analytics shows an issue itself. We can talk about the means by which we are intended to guarantee the security of other artefacts. This aspect includes security of machine learning, the legitimacy of data harvesting and usage of analytical results.

If we are using machine learning methods for security purposes, one has to consider vulnerabilities they have. Attacks against machine learning systems have been analyzed in [32]. A unifying threat model that allows structuring reasoning about the security and privacy machine learning systems is given in [33]. Towards securing data also works the concept of Machine Unlearning, which wipes out unwanted data [16].

Showing the emergence of the topic there is a *Special Issue on Security and Privacy in Machine Learning* from Elsevier to be published by the end of 2018.

With the security of analytics comes together ties from our own information literacy in a digital age, trust, ensuring privacy, exploiting machine learning capabilities and ethics of their application.

5 Conclusions

The aim of this paper was to dispel the fog from the term “Security analytics” and illuminate different aspects of it. This review paper was intended to summarize existing viewpoints to security analytics and contribute to common understanding towards a comprehension of the multifaceted area called security analytics. Both academic and technical review materials have been examined to catch the various aspects of security analytics. The first conclusion is that there is no unified definition of “Security analytics” term and the boundaries of it. As identified from literature review, perception of security analytics include varied aspects which are organized and classified under three views, namely, (1) *Analytics of the security*, (2) *Analytics for the security* and (3) *Security of analytics*. Each of them gathers several topics.

The question for further refinement of proposed viewpoints of security analytics is to determine whether this model is linear or cyclic and we can join together *Security of machine learning* with the initial beginning – *Analyzing security threats* as a special case of security data. Additional topics could be considered for inclusion in the presented model with extended the literature review.

References

1. Talabis, M.R.M., McPherson, R., Miyamoto, I., Martin, J.L.: Information Security Analytics: Finding Security Insights, Patterns, and Anomalies in Big Data. Syngress Publishing (2014)
2. Cambridge Dictionary | English Dictionary, Translations & Thesaurus, https://dictionary.cambridge.org/
3. English Dictionary, Thesaurus, & grammar help | Oxford Dictionaries, https://en.oxforddictionaries.com/
4. Mahmood, T., Afzal, U.: Security Analytics: Big Data Analytics for Cybersecurity. In: 2013 2nd National Conference on Information Assurance (NCIA) (2013)
5. Ullah, F., Ali, M.: Architectural Tactics for Big Data Cybersecurity Analytic Systems : A Review. Arxiv.Org. 1–48 (2018)
6. Grossman, S.: Security Analytics: What it is and what it is not, https://betanews.com/2016/05/16/security-analytics-what-it-is-and-what-it-is-not/
7. Villella, P.: A Practical Approach to Effective Security Analytics | LogRhythm, https://logrhythm.com/blog/a-practical-approach-to-effective-security-analytics/
8. Birzniece, I.: Artificial Intelligence in Knowledge Management: Overview and Trends. Applied Computer Systems 43, 5–11 (2011). doi: 10.2478/v10143-011-0001-x
9. Lehto, M., Neittaanmäki, P. eds: Cyber Security: Analytics, Technology and Automation. Springer International Publishing, Cham (2015)
10. Alsmadi, I.M., Karabatis, G., Aleroud, A. eds: Information Fusion for Cyber-Security Analytics. Springer International Publishing, Cham (2017)
11. Al-Shaer, E., Rahman, M.A.: Security and Resiliency Analytics for Smart Grids: Static and Dynamic Approaches. Springer International Publishing Switzerland (2016)
12. Al-Shaer, E., Kant, K. eds: SafeConfig'14 : : November 3, 2014, Scottsdale, Arizona, USA. In: Proceedings of the Cyber Security Analytics, Intelligence and Automation Workshop. p. 40. ACM (2014)
13. Shahbazian, E., Rogova, G. eds: Meeting Security Challenges Through Data Analytics and Decision Support: EBSCOhost. In: Meeting Security Challenges Through Data Analytics and Decision Support. IOS Press (2016)
14. Verma, R. ed: IWSPA '15: In: Proceedings of the 2015 ACM International Workshop on International Workshop on Security and Privacy Analytics. ACM, New York, NY, USA (2015)
15. Verma, R. ed: IWSPA '16: Proceedings of the 2016 ACM on International Workshop on Security And Privacy Analytics. In: Proceedings of the 2016 ACM on International Workshop on Security And Privacy Analytics. ACM, New York, NY, USA (2016)
16. Savas, O., Deng, J. eds: Big Data Analytics in Cybersecurity. Auerbach Publications, New York (2018)
17. Benjelloun, F.-Z., Ait Lahcen, A.: Big Data Security: Challenges, Recommendations and Solutions. In: Handbook of Research on Security Considerations in Cloud Computing (2015)
18. Al-Shaer, E., Rahman, M.A.: Smart Grids and Security Challenges. Presented at the (2016)
19. Jacobs, J., Rudis, B.: Data-Driven Security: Analysis, Visualization, Dashboards. Wiley

(2014)

20. Doney, P.M., Cannon, J.P.: An Examination of the Nature of Trust in Buyer-Seller Relationships. J. Mark. 61, 35 (1997)
21. Schneider, J., Handali, J.P., vom Brocke, J.: Increasing Trust in (Big) Data Analytics. In: Matulevičius, R. and Dijkman, R. (eds.) CAiSE 2018 Workshops. pp. 70–84. Springer International Publishing (2018)
22. Osia, S.A., Shamsabadi, A.S., Taheri, A., Rabiee, H.R., Haddadi, H.: Private and Scalable Personal Data Analytics Using Hybrid Edge-to-Cloud Deep Learning. Computer (Long. Beach. Calif). 51, 42–49 (2018)
23. The Australian Computer Society: Cybersecurity - Threats Challenges Opportunities. (2016)
24. O'Neil, C.: Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group (2016)
25. Berghel, H.: Malice Domestic: The Cambridge Analytica Dystopia. Computer (Long. Beach. Calif). 51, 84–89 (2018)
26. Birzniece, I., Rudzajs, P., Kalibatiene, D.: Evaluating the application of interactive classification system in university study course comparison. In: 13th International Conference on Perspectives in Business Informatics Research, pp. 335–346. Springer (2014). doi: 10.1007/978-3-319-11370-8_24
27. Wolff, M.: Applying Machine Learning to Advance Cyber Security Analytics – Cyber Security Review, http://www.cybersecurity-review.com/industry-perspective/applying-machine-learning-to-advance-cyber-security-analytics/
28. Drinkwater, D.: 5 top machine learning use cases for security | CSO Online, https://www.csoonline.com/article/3240925/machine-learning/5-top-machine-learning-use-cases-for-security.html
29. Salnitri, M., Alizadeh, M., Giovanella, D., Zannone, N.: From Security-by-Design to the Identification of Security-Critical Deviations in Process Executions. In: CAiSE Forum. pp. 218–234. Springer International Publishing (2018)
30. IBM i2 Analyze, https://www.ibm.com/us-en/marketplace/enterprise-intelligence-analysis/details#product-header-top
31. Pierazzi, F., Casolari, S., Colajanni, M., Marchetti, M.: Exploratory security analytics for anomaly detection. Comput. Secur. 56, 28–49 (2016)
32. Barreno, M., Nelson, B., Joseph, A.D., Tygar, J.D.: The security of machine learning. Mach. Learn. 81, 121–148 (2010)
33. Papernot, N., McDaniel, P., Sinha, A., Wellman, M.: SoK: Towards the Science of Security and Privacy in Machine Learning. ArXiv e-prints. (2016)