Adam Zagorecki - Jozef Ristvej - Krzysztof Klupa *

ANALYTICS FOR PROTECTING CRITICAL INFRASTRUCTURE

In this paper we review the key trends related to application of information technology and, in particular, automated data analysis to the problem of protecting critical infrastructure. We focus on technologies that use automated data collection and analyses that can be exploited for improving security provision for critical infrastructure in the future. Of our particular interest are technologies that are at relatively early stage of adaptation and in our judgement have potential to significantly affect how the security is provided in the future, the technologies that support physical aspect of security. Then we discuss analytics in context of cyber-security with applications.

Keywords: *Critical infrastructure, analytics, data mining, cyber-security.*

1. Introduction

The terrorist attacks in September 2001 have changed the perception of critical infrastructure security. These attacks were characterised not only by sophistication of planning and execution but as well the selection of targets – a diverse set of critical infrastructure buildings that included government, defence and commerce [1]. Those were followed shortly by London bombings and attack on trains in Spain, proving that complex attacks on critical infrastructure are becoming a new trend in terrorism [2].

In this paper we review the key trends related to application of information technology and, in particular, automated data analysis to the challenge of protecting critical infrastructure. We focus on technologies that use automated data collection and analyses that can be exploited for improving security provision for critical infrastructure in the future. Of our particular interest are technologies that are at relatively early stage of adaptation and in our judgement have potential to significantly change how the security is provided in the future. We mostly focus our discussion on technologies that support physical aspect of security: biometric technologies, video analytics, sensors and integration/data fusion. Then we briefly discuss the big data and analytics in the context of big data. Finally, we discuss analytics in context of cyber-security. We focus on the man-made threats to critical infrastructure. These are typically divided into two categories: physical threats and cyber threats. We want to emphasise that this division does not imply that these two categories are separate; in fact, one should expect that more sophisticated attacks in the future should combine the two domains.

2. Data for analytics and security

Data analytics [3] is the process of discovery and communication of meaningful patterns in data typically with use of data-mining techniques [4]. It has become a very prominent trend especially in business context (often referred to as business intelligence). The premise behind the analytics is that the use of (objective) data gathered on multiple aspects of the problem (data fusion) should improve understanding of the problem and provide new insights. In the context of providing security to critical infrastructure, one should be able to use data coming from various sensors, cameras and other data sources (for example: authorised users database) to improve provision of security by enhanced capability of identifying security breaches, reducing costs of providing security and supporting security personnel in their tasks [5]. Below we review key classes of technologies that, in our opinion, are becoming prominent for providing security and rely upon or explicitly use analytics.

2.1 Biometrics

The biometric technologies [6] are aimed at verifying personal identity – ensuring that only authorised personnel is able to access the area or perform tasks by measuring some physical aspect of a person. The most popular biometric techniques focus on the following aspects: finger prints, face recognition, and iris scans. All of them make use of sophisticated digital data analysis. Below we will briefly discuss all three types.

* [1]Adam Zagorecki, [2]Jozef Ristvej, [3]Krzysztof Klupa
[1]Cranfield University, Defence Academy of the United Kingdom, Senior Research Fellow, Shrivenham, United Kingdom
[2]Department of Crisis Management, University of Zilina, Slovakia,
[3]Department of Security Sciences, the General Tadeusz Kosciuszko Military Academy of Land Forces, Wroclaw, Poland
E-mail: a.zagorecki@cranfield.ac.uk

Finger prints were recognised as a useful tool of identification individuals as early as 19th century. It was recognised that this method of biometrics is relatively easy to bypass and susceptible to noise while reading, often requiring multiple scans. However, it is accepted that this technique can be successfully used as an additional security measure in order to increase security by diversifying methods.

Face recognition is considered as natural and least invasive method of biometric identification. Facial recognition systems range from software based solutions to complete close circuit TV systems. The technology relies on samples of images of an individual stored in the database against which the pictures taken by cameras are compared. In practice high-quality enrolment material is essential, with quality of the enrolment material determining the performance of the system. The face recognition systems have advantage for environments with a large number of people – public transportation to mass events. However, one should remember that this technology is relatively easy to be fooled and, therefore, not suitable for situations where high reliability is required from a biometric system.

Iris recognition is believed to be the most promising of the biometric methods. The iris patterns are believed to be unique to an individual, constant over time and not subject to changes caused by medical conditions. Scanning process is performed using a camera (visible or near-infrared light) and, therefore, is non-invasive (unlike retina scanning). In terms of accuracy it has low occurrence of false positives and extremely low of false negatives. The iris recognition systems have been fielded in some environments, however at the current stage they have not been widely accepted with some systems being withdrawn (e.g. UK Iris Recognition Immigration System). Particular challenges include requirement for good quality samples and relative ease to fool the system by presenting an image.

Fig. 1 Example of biometric technologies in practice [7]

In summary, the biometric technologies (Fig. 1 is illustrative) are considered to be immature and not sufficiently reliable to provide definitive ways of authentication in large scale and general setting. However, with time they may become more mature and they are likely to provide increased levels of authentication, especially when their accuracy and speed will be increased.

2.2 Video analytics

The sophistication and costs related to digital imagery have drastically dropped due to advancements in related technologies. In the result, modern digital imagery systems may not only record high quality images and videos, but as well allow for more and more sophisticated means of image analysis [8]. The types of image analyses vary in complexity which translates into ability of automated systems to address those problems (Fig. 2 is illustrative). We can identify four key types of analytics that relate to security-based tasks:

- Motion detection – a simple task of detection of changes in an otherwise static image. This task is extremely useful for identifying situation that may potentially require security officer's attention in order to analyse the scene.
- Object detection and classification – a task of automated interpretation of images in order to identify particular types of objects of an interest (e.g. a person or a van).
- Object recognition – for example, face recognition. A task of identifying a particular instance (e.g. person) of an object.
- Object tracking – a task of following an object on an image, or even following the object in a series of images (using views from different cameras).

Fig. 2 Illustration of video analytics [7]

Automated image analysis is a relatively new domain with first more sophisticated practical applications (such as face recognition) being fielded no more than 15 years ago. More sophisticated functionalities such as object tracking are still on relatively early stage of maturity. One of interesting problems with the image processing techniques are differences between reported performance of algorithms achieved in the laboratory setting and actual fielded applications. Similarly, it has been often reported that the performance claimed by suppliers is much higher to that achieved by fielded systems. These may be not necessarily

due to intentional actions, but due to the nature of phenomena related to algorithm evaluation and it emphasises the fact that a proper scientific approach to evaluation of such systems is necessary. From practical perspective, substantial research is needed till such systems achieve desirable performance. Another challenge with video analytics is that they require relatively large computational power due to sheer amount of data encoded in an image. Therefore, further computational performance improvements can make video analytics even more prominent for providing security in the future.

2.3 Sensing technologies

During the recent decade a revolution in development of sensing technologies has taken place. Examples of different sensors include:

- Accelerometers – which are so affordable that often they are installed 'just in case' in other electronic devices.
- Digital cameras – experiencing dramatic lowering of costs and improvements in terms of resolution.
- Transmitters/receivers - which allow for communication between sensors and the information infrastructure.
- Other sensors such as temperature, pressure, light, etc.

An example of a sensor platform is a smart phone – a typical smart phone includes all the above sensors – and typically it is not strictly required by the primary function of the device (making phone calls) but the sensors are included just because they are affordable and can provide value added to the user making the model more competitive on the market. Sensors are able to produce large volumes of data which can be used by data-mining algorithms to derive automatically new knowledge of the domain. However, a common misconception should be clarified here – the data produced by sensors do not imply the useful knowledge – this should be extracted from the data through the analytical processes that are not trivial. Therefore, the sensors should be considered as an enabling technology with analytics required for making efficient use of the data generated by sensors.

2.4 Integration

The key trend in security systems is integration [9] - security systems are becoming more integrated and it can be observed at several levels:

- Monitoring of several buildings or structures from the same location by exploiting remote sensing and telecommunication infrastructure that allows for transferring video streams. The primary benefit of such integration is lowering security costs by reducing number of personnel and facilities.
- Integration with other building systems (such as HVAC, electrical systems, etc.). The purpose of it is utilising common infrastructure and useful information about the state of the monitored infrastructure that can be used to inform security.
- Integration with business processes – for example, integration with organisational data warehouse to use up to date personnel data for the security purposes. The purpose of such integration is reduction of organisational costs and enhancing security by means of data fusion.

3. Big data

'Big data' is a term that describes a set of technologies that are related to collection, storage and analysis of large volumes of data. By its definition the big data technologies exceed capabilities of a single computer, especially in terms of storage. The key characteristics of the big data that differentiate it from traditional data warehouses that store large volumes of data are the 'four Vs':

- Volume – the quantity of data should greatly exceed storage capabilities of a single computer, which means that dedicated IT infrastructure should be in place.
- Variety – this is probably the most important criterion that distinguishes the big data from traditional data storage: the data must be multidimensional, which should guarantee that it captures complexities of the domain it relates to.
- Velocity – the data should be constantly generated and keep up with the changing environment.
- Variability – the data should reflect dynamics of the environment. This aspect is particularly important from the analytics perspective, as traditional data mining algorithms assume that the data relates to the system (or at least most aspects of it) that is constant over time.

The big data is more than just a scale-up version of traditional data warehouses. The big data carries a premise of using large volumes of multidimensional data to make predictions about the world that would not be possible otherwise. This premise is based on the fact that the sheer volume and complexity of data exceeds human abilities to analyse it, and, therefore, one should expect that the algorithms that are able to handle and exploit the big data, would be able to provide knowledge and insights that are beyond human capabilities. To validate if this premise is true requires some time and maturity of the big data solutions, however, current practical applications of the big data concept show that there is at least some merit in the big data.

The big data is becoming a trend both in commercial and government sectors. The commercial applications are mostly driven by solutions that allow customisation of provided service. Typical examples are recommender systems implemented for online shops, or more from security domain tools that allow personalised risk scoring: for example tools that use data fusion of different sources of data for credit scores.

There are challenges related to the big data – in particular the cost of implementation and the fact that there is no guarantee that

the investment in massive data infrastructure will be justified by benefits that cannot be guaranteed or even estimated at the time of investment. Currently, there is a lot of optimism and hopes related to the big data, but one will need to wait till those are verified by actual implementations.

4. Cyber-security

In the recent decades the rise of cyberspace and related threats associated with this domain has been observed and widely discussed in the literature [10]. In particular, it is the networked nature of the monitoring systems and their connectivity to the Internet that creates a bridge between providing physical security and the cyberspace. This is not intention to discuss cyber-security threats and application of analytics in this paper, however automated intrusion detection systems based on constant monitoring and automated interpretation of data are an active and very promising research field [11] (Fig. 3 is illustrative).

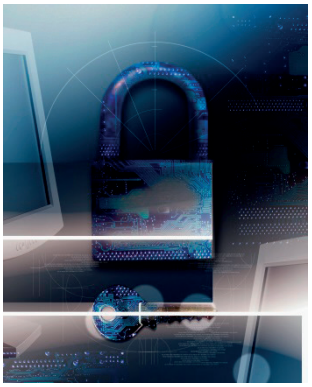

Fig. 3 Illustration of cyber-security [7]

The big data technologies are particularly relevant to cyber-security. It is because of the two aspects that characterise cyber-threats:

- The ease of collecting large volumes of data related to cyber-security – implementing data collection on IT infrastructure is relatively affordable and technically unchallenging. Large organisations are already aware of cyber-threats and a typical first response is implementation of the data collection infrastructure, often with false assumptions that the data itself will help improving security.
- The nature of cyber-threats – unlike civil unrests, bomb attacks or simple perimeter violations, cyber-threats are hidden from the eyes. The most dangerous cyber-threats may remain undetected for long periods of time and require specialised knowledge and tools to be detected. The cyber-threat detection is basically based on analyses of computer logs, process that can be naturally automated.

The most common approaches to analysis of cyber-security data can be summarised in two broad categories:

- Patterns or signature detection – this approach is based on identifying known patterns of attacks by using various kinds of pattern-matching methods. They rely on the experts identifying patterns of typical cyber-attacks and describing them in form of patterns that are later used by analytical software to match against suspected activities. The strength of this approach is simplicity and use of known facts about cyber-threats. An obvious disadvantage is their unsuitability for detecting new threats, and inability to learn.
- Anomaly detection – these methods are based on automated methods that detect unusual behaviours in the system, and flag them for interpretation by human analysts. These methods are more suitable for unknown threats, however, require input from humans.

It is likely that the development of analytical tools for cyber-security will be intensified in the future, as cyber-security is being put in the spot light by industry and governments.

5. Conclusions

In this paper we outlined the trends in analytics for the problem of providing protection for critical infrastructure, as it was highlighted in several paper before (such as: [12] and [13], [14] and [15]). The field of analytics is currently dynamically developing and in our opinion it is at a relatively immature not only for security applications but for a wide spectrum of applications in general. The trends from other domains (especially business) indicate that the data fusion and the concept of 'big data' carry a promise of revolutionary changes in analytics in general. If it is the case – it is yet to be seen. But certainly some more basic applications and concepts outlined here have a potential to substantially affect how the security for critical infrastructure will be implemented.

Cyber threats and cyber-security are emerging phenomena. Even though that the immense number of strategies, reports, white-papers and both professional and academic papers have been proposed and published, we are yet to see the development of these threats into a security daily reality.

Acknowledgements

This work was co-funded by the Slovak Research and Development Agency under the contract No. DO7RP-0025-12. And the project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 313308.

References

[1] EDWARDS, M.: *Critical Infrastructure Protection*, vol. 116 of NATO Science for Peace and Security Series - E: Human and Societal Dynamics, M. Edwards (ed.), IOS Press, 2014.

[2] LEWIS, T. G.: *Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation.* Wiley, May 2006, ISBN: 978-0-471-78628-3.

[3] KOHAVI, R., ROTHLEDER, N. J., SIMOUDIS, E.: Emerging Trends in Business Analytics. *Communications of the ACM*, 45 (8): 45–48, 2002.

[4] HAN, J., KAMBER, K., PEI, J.: Data Mining: *Concepts and Techniques*. Morgan Kaufmann, 2006.

[5] SHANKAR, M., RAO, N., BATSELL, S.: *Fusing Intrusion Data for Detection and Containment*, Proc. of the 2003 IEEE conference on Military communications (MILCOM'03), vol. II, 2003. IEEE Computer Society, Washington, DC, 741-746.

[6] JAIN, A. K., ROSS, A., PANKANTI, S.: *Biometrics: A Tool for Information Security.* IEEE Transactions on Information Forensics and Security, vol. 1, No. 2, June 2006, 125-143.

[7] Fotosearch, *Security Alert by Brand X Pictures*, 30.3.2008.

[8] KO, T.: *A Survey on behavior Analysis in Video Surveillance for Homeland Security Applications*, Applied Imagery Pattern Recognition Workshop, 2008. AIPR'08. 37th IEEE, pp. 1-8. IEEE, 2008.

[9] BASS, T.: Intrusion Detection Systems and Multi-sensor Data Fusion. *Communications ACM*, 43, 4, April 2000, 99-105.

[10] MILLER, R. A.: Cyber War Realities - What Lies Ahead, *Intern. J. of Critical Infrastructure Protection*, vol. 5, No. 2, July 2012, 84-85, ISSN 1874-5482.

[11] DI PIETRO, R., MANCINI, L. V.: *Intrusion Detection Systems*. Springer, 2008.

[12] HOLLA, K.: Dealing with Key Terms in Risk Analysis and Phenomenon of Uncertainty in this Process, *Communications - Scientific Letters of the University of Zilina*, vol. 9, No. 4, 2007, 59-61, ISSN 1335-4205.

[13] ZANICKA HOLLA, K., MORICOVA, V.: Human Factor Position in Rise and Demonstration of Accidents, *Communications - Scientific Letters of the University of Zilina*, vol. 13, No. 2, 49-52, 2011, ISSN 1335-4205.

[14] SIMAK, L.: Increasing the Security Level in the Slovak Republic, *Communications - Scientific Letters of the University of Zilina*, vol. 10, No. 2, 2008, 67-71.

[15] REHAK, D., SENOVSKY, P.: Preference Risk Assessment of Electric Power Critical Infrastructure, *Chemical Engineering Transactions*, vol. 36, 469-474, 2014, ISSN: 1974-9791.