Ubiquity Symposium

Big Data

Corporate Security is a Big Data Problem

By Louisa Saunier and Kemal A. Delic

Editor's Introduction

In modern times, we have seen a major shift toward hybrid cloud architectures, where corporations operate in a large, highly extended eco-system. Thus, the traditional enterprise security perimeter is disappearing and evolving into the concept of security intelligence where the volume, velocity/rate, and variety of data have dramatically changed. Today, to cope with the fast-changing security landscape, we need to be able to transform huge data lakes via security analytics and bid data technologies into effective security intelligence presented through a security "cockpit" to achieve a better corporate security and compliance level, support sound risk management and informed decision making. We present a high-level architecture for efficient security intelligence and the concept of a security cockpit as a point of control for the corporate security and compliance state. Therefore, we could conclude nowadays corporate security can be perceived as a big-data problem.

Ubiquity Symposium

Big Data

Corporate Security is a Big Data Problem

By Louisa Saunier and Kemal A. Delic

We live in the zetta/exa/petabyte era. If judged by the current development trends, data created, generated, stored, and analyzed in the internet space could reach an unprecedented magnitude soon. According to Cisco's global IP traffic forecast [1], the annual global IP traffic will pass the zettabyte threshold soon and will reach 2.3 ZB per year by 2020—increasing nearly threefold over the next five years. The number of devices and users connected to the internet increases exponentially with the emergence of the IoT (internet of things), oceans of mobile devices, and the omnipresence of social networks generating torrents of data each second. With the internet of everything (IoE) phenomenon, the number of devices connected to the internet is expected to more than double, from 4.9 billion in 2015 to 12.2 billion devices by 2020 according to the Cisco forecast [1].

In parallel, we see a high rate of growth in data breaches. In 2015, it was reported that 16 records are breached every second, or more than 1.3 million records are breached every single day [2]. 2016 averaged at least one health data breach per day, affecting more than 27 million patient records [3]. Therefore, we could consider all interconnected devices and users in the internet space either as resources, essentially very complex eco-systems providing valuable services, or as potential targets for security attacks. We live in an interconnected virtual world where security obeys the same law of communicating vessels as in the physical world (see Sidebar). Recently, widely reported, well-coordinated, large-scale activities turned IoT devices into zombies used for major infrastructure attacks. Using the internet or even connecting a device to internet, as well as using big data flowing out of social networks and other sources, will have to be better regulated than it is today. Some security models that are used in corporate environments (e.g. security policy and checking device compliance against policy before connection/disconnection) could be evaluated for applicability to the public internet space. Security in an interconnected world today should not be addressed with traditional means, e.g. log files analytics, SIEM (security information and event management), IDS (intrusion detection system), etc. Instead, we need to consider using big data processed using

data grids in modern/cloud data center facilities to produce effective security intelligence. Several projects or work groups aim at gaining actionable security intelligence from big data analytics addressing different aspects of the problem: Cloud Security Alliance's Big Data Working Group [4], botnet detection [5, 6], advanced persistent threats detection [7], WINE [8], and phishing attack detection [9].

In this article, we point to some critical security challenges of corporate businesses today and discuss why big data should be considered a corporate security problem and opportunity. Our approach is to use security analytics and big-data technologies to turn various data flows into effective security intelligence[1] presented through a security cockpit to support better security and compliance levels, sound risk management, and informed decision making. We present a high-level architecture for effective security intelligence and the concept of security cockpit as a point of control of the overall security and compliance state.

We first describe how security has been addressed in the past by slowly reacting and adapting IT systems behind the security parameter of the corporate network. After pointing out the current trend of the disappearing security perimeter in modern corporate IT systems, which has lead to the always post-reacting on security breaches, we outline an innovative architecture for effective security intelligence. Volumes, velocity, and variety of corporate security data is such that only large-scale facilities can cope with it in a holistic, accurate, and timely way to produce effective (trustworthy) security intelligence.

We conclude with the opinion that the security problems introduced via modern technological means should be addressed with advanced technological (big data) solutions augmented with business, legal, and social considerations. Earlier thinking about corporate IT security as an isolated system hidden behind the enterprise security perimeter protecting monolithic, enterprise applications should evolve into rethinking the corporate security as a complex eco-system immersed in social, economic, legal, and even political contexts—better described as security intelligence. The viability of the evolving and forthcoming digital economy is, and will continue to be, crucially dependent on it.

Traditional Corporate Security

Business enterprises are heavily dependent on the use of information and communication technologies to operate in uncertain and constantly changing markets. Typical IT security

[1] Effective corporate security intelligence is a concept aiming to provide better, faster and more complete execution of all corporate security operations. [3]

architecture of the past (see Figure 1, left) was centered around the corporate network with a well-delimited external security perimeter and provided a communication backbone interconnecting applications in data centers with users and devices: desktops, laptops, tablets, and phones. Operations were carefully monitored and managed from operation centers and service desks. Prevailing security logic was centered on protecting the enterprise security perimeter to try to ensure the security of users, applications, and operations.

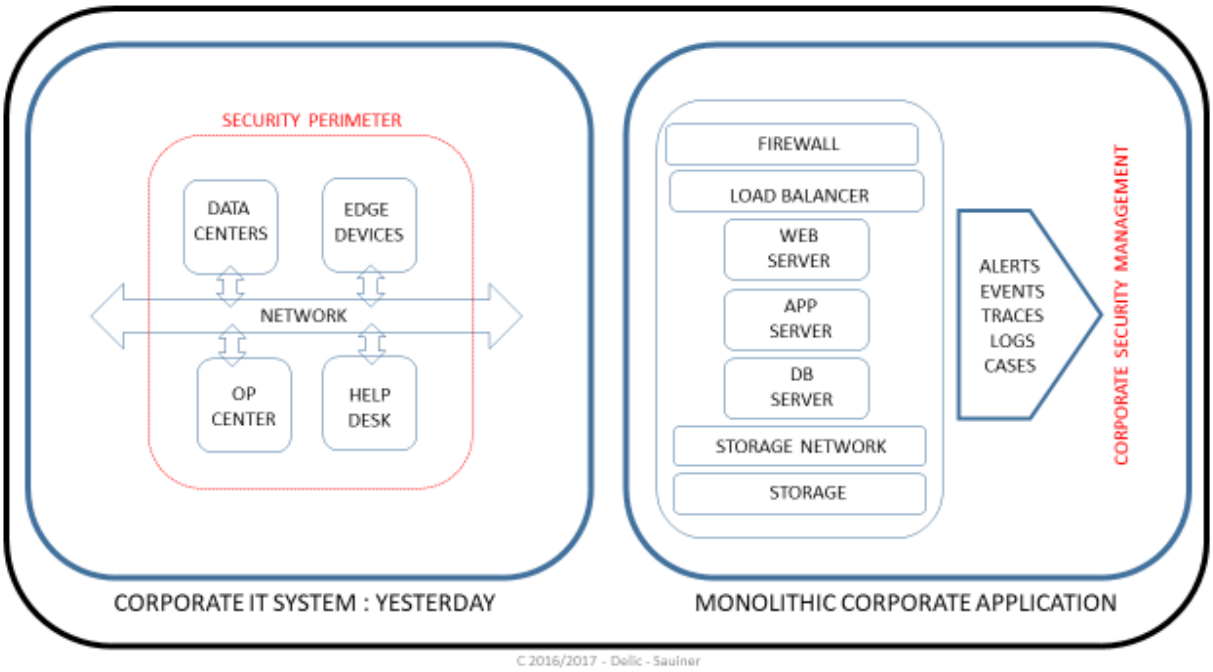

Figure 1. Traditional corporate security IT system: security perimeter.

Typical corporate applications (see Figure 1, right) are monolithic consisting of the characteristically stacked architecture, generating torrents of events, logs, traces, and logs. The key deficiency of classical corporate security approach lies in the fact that various types of security devices and applications are treated in an isolated, perimeter-centric manner slowly adapting to technological changes and new appearing security threats. For example, within the walls of big businesses detecting advanced persistent threats can take (an average of) seven months [7]. Security breaches resulting in theft of personal, financial, or confidential data are constantly reported. Furthermore, the huge variety of data formats and data of different vendors makes it that the integration of different data flows is so difficult that the creation of an accurate and complete security picture is cumbersome, if even possible, using traditional means (SIEM, intrusion detection, etc.). As we deal with a variety of security data sources, the

security approach has generally been reduced to the integration of a few data sources in a data warehouse used for execution of various analytics, canned queries, or ad-hoc inquiries.

The limits of the past IT security system can be understood easily when considering corporate businesses, e-commerce transactions, internet service providers, cloud service providers, and mobile operators are constantly creating millions of events each second, as well as producing endless streams of logs from applications, network devices, DNS, anti-malware systems, servers, and personal devices. In most cases, these data are quickly discarded and not analyzed in depth. As the volume of security logs rises and dynamics accelerates, we envision a system that makes use of big-data analytical technologies to synthesize a holistic picture of security and regulatory compliance, guide investigation/forensic analysis, and support-informed decision making. Overall, the traditional corporate IT system is usually reacting to security events "after the fact," adapting slowly to the fast changing security landscape.

We need to see the emergence of a new generation of IT security systems capable of processing large data sets containing long-term historical data where the traditional tools are no more sufficient.

In the next section, we look at how modern IT architecture creates new challenges in which old-style thinking should be radically changed.

Modern Corporate Security System

We have seen a major shift toward cloud architectures (Figure 2, left) where corporations operate in a large, highly extended ecosystem. The most frequent corporate architecture is a kind of hybrid-cloud IT (Figure 2, left), in which volumes and event rates are very high, covering large domains. Thus, enterprise security perimeter is disappearing and evolving into the concept of security intelligence where the volume, velocity, rate, and variety of data have radically changed.

Our main point is to make use of the huge data lakes, with their data flows transformed via security analytics software into effective security intelligence. As inflows of torrents of data from devices, applications, and users are large and constant, we see the development of two classes of security insights on the horizon: (1) stream processing on data in motion, i.e. real-time for immediate event detection, notification and containment, and (2) batch processing on data at rest, i.e. near real-time to support compliance, risk management, and sound decision making.

Big-data technologies (e.g. MapReduce, Hadoop DFS, Spark, and Storm for stream data processing) are enabling storage and analytical processing of large heterogeneous data sets at an unprecedented scale and rate. Such security data lakes (see Figure 2, right) are enabling, equipping, and driving big-data security analytics, allowing automatic mining of meaningful/telling security information from a wide variety of sources. Typically, data sources include firewalls, security devices, external security alerts, website traffic logs, business processes logs and day-to-day transactional logs, external alerts, both structured and unstructured data sets, and multimedia data sets flowing into a single security intelligence framework. Security analytics is usually performed by continuously scanning and examining the state of the internal IT environment as well as a wide variety of external data sources. These sources may include streams of events such as website access events, DNS (domain name server) requests, change events, firewalls events, vulnerability databases updates, CMDB (configuration management data base), virus signature updates, etc. Additionally, data sources would include system state, application/data criticality state, user privilege assignment and patching state, as well as information from operational business processes, such as access approval, change approval, incident resolution, and asset changes.

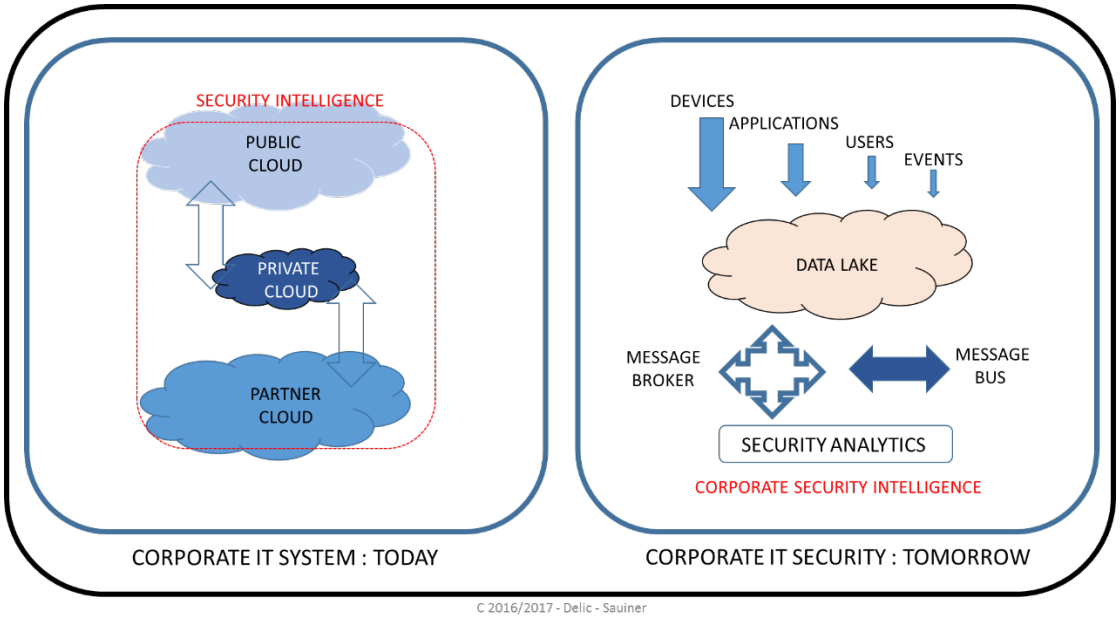

Figure 2. Modern corporate IT system: security intelligence.

Security analytics will have (at least) two dimensions: real-time insights (stream processing on data in motion) and near real-time investigations (batch processing on data at rest). Security

analytics will provide a set of interactive maps maintained and updated in real-time, thus enabling unprecedented insights into current, accurate and trustworthy state of corporate IT security.

As an example of security analytics, the concept of security maps is useful to categorize the results derived from deployment of analytical methods and techniques. We can distinguish several types of maps: control maps, state maps, animated maps, and interactive maps. Control maps depict the status of a security control. State maps reflect the overall system state or policy compliance. These maps use volume, proportions, colors, etc. to pass cognitive hints to humans, while animated maps show dynamics of security events and enable spotting of trends, patterns, and behaviors. Finally, interactive maps enable safe play of various "what if" and "what when" scenarios. The aim is to provide better decision support through fact-based decision making, making decisions timely using real-time data, and clarifying choice in advance.

All this data can be arranged into service oriented maps that can be used for risk metric reporting or for security audits, for example. The data and mappings are uploaded into a security data lake, then split and spread on the nodes of a cluster (e.g. Hadoop cluster) for parallel processing and analytics (e.g. machine learning applications) are applied on the top of it.

Security intelligence is just a part of the overall corporate IT architecture which recent technologies have turned into cloud and big data-based corporate IT.

Corporate Security Intelligence: Architecture

Established business corporations are facing a big challenge of transforming themselves into something similar to digital corporations while dealing with huge legacy IT systems and fast-changing external circumstances. Ensuring secure operations is of primordial importance for their adaptation and survival. In Figure 3, we outline a conceptual architecture of security intelligence, which is embedded into a bigger, enterprise IT architecture and contains several emerging technologies. The security cockpit is different from the typical security dashboard, as it requires timely attention and action not only passive data observation and digestion. The security cockpit is driven by the security analytics, which are being fed from the big data lakes distributed across clouds and IoT edge devices. This is conceptual security intelligence architecture within an overall enterprise architecture, which is shown for the illustrative purposes—the design and engineering of it may differ widely.

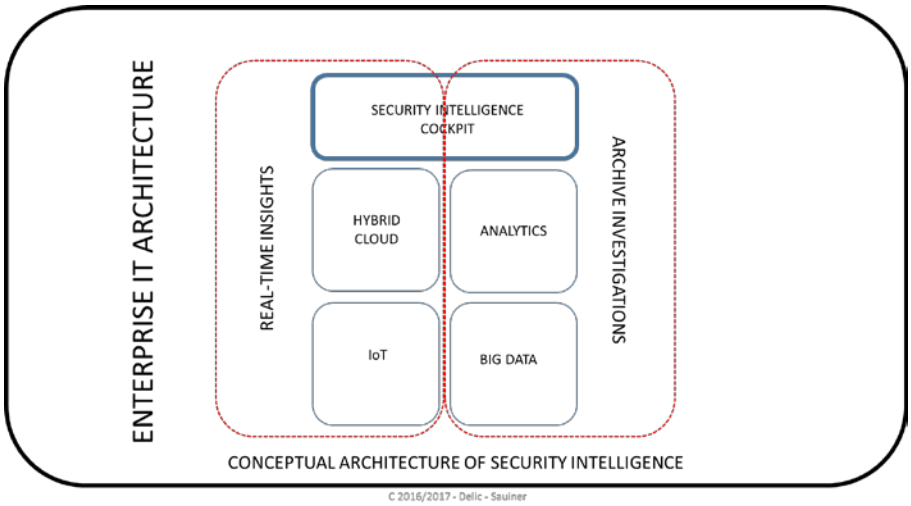

Figure 3. Corporate security intelligence: conceptual architecture.

Corporate security intelligence supports better decision making and enables sound risk management. Security cockpit maps include trending key security risk indicators and critical IT controls across multiple systems and applications over time. Knowing the importance and criticality of data and systems, as well as their interdependencies, is essential in risk management and business decision support—this is the principal role of the security cockpit.

Conclusion

Looking into the future, we believe data will continue to grow by orders of magnitude and the necessity for big data analytics for obtaining effective security intelligence will become more urgent. Attackers will also use new technological means for improving their modes of attack; the need for protecting privacy data will rise at the same pace. Aggregation of data from different sources will become easier and it will be even more important to contemplate a better-regulated internet and IoT space.

Presently, we expect a major shift toward hybrid-cloud architectures, where corporations operate in a large, highly extended eco-system. Thus, the traditional enterprise security perimeter is disappearing and evolving into the concept of security intelligence with volume, velocity/rate, and variety of data dramatically changing. Today, to cope with the fast-changing security landscape, we need to be able to transform huge data lakes via security analytics software and big-data technologies into effective security intelligence presented through a security cockpit to achieve better corporate security and compliance level, support sound risk management, and make informed decisions. We have presented a high-level architecture for

efficient security intelligence and the concept of security cockpit[2] [10], [11] as a point of control of the corporate security and compliance state. Therefore, we could conclude nowadays corporate security can be perceived as a big-data problem. Recent advances in some traditional technologies as anti-virus or vulnerability scanning of vendors, like Symantec and McAfee, now turn to cloud-based solutions harnessing the power of big-data processing and security intelligence, which is a good indicator of the current trend.

To add to the "data wish list" [12] published by the security research community in 2009, taking into account recent security breaches, we would advocate a better regulated internet space is needed, as it is not enough to protect just individual networks or organizations but consider cybersecurity in the perspective of a globally interconnected world where security obeys the law of communicated vessels [see Sidebar]. The existence of weak and insecure systems somewhere in the internet space (e.g. botnets or a set of non-protected IoT devices) may threaten the peacefulness of many people, organizations, companies, or even entire countries. We should move away from siloed thinking and perceive cybersecurity as a global, planetary topic, very much similar to the global-warming problem, with a high level of urgency.

With the recent DDoS attacks based on botnets, we need to ensure ISP, service providers, and other IT key players integrate in their systems big-data tools that allow for the discovery of botnets and produce some effective security intelligence from analyzing DNS and network traffic. We need the integration of security big-data analytics to become a mandatory requirement for service providers as well as to envision the need for a better-regulated IoT space. IoT vendors need to go through a compliance verification of the connected devices, which would depend on the device use (i.e. healthcare related would potentially require additionally data encryption or other extended requirements). All this seems as the key precondition for the rise of digital economy.

References

[1] CISCO. The Zettabyte era: trends and analysis. June 7, 2017.

[2] Gemalto. 2015 First Half Review: Findings from the Breach Level Index. 2015;

[3] Protenus, Inc. 2016 Breach Barometer Report: Year In Review. 2017.

[2] We described this concept at an academic conference some 10 years ago and obtained a European patent on the key component of the security cockpit

[4] Cloud Security Alliance (CSA) Big Data Working Group. Big Data Analytics for Security Intelligence. Report. September 2013.

[5] Francois, J. et al. BotCloud: detecting botnets using MapReduce. In *Proceedings of the 2011 IEEE International Workshop on Information Forensics and Security.* IEEE, Washington D.C., 2011.

[6] Buriya, S., Bhilare, D. S., and Singh, A. A survey of botclouds: Botnet detection using MapReduce and big data analytics. 2015.

[7] Giura, P. and W. Wang. Using large scale distributed computing to unveil advanced persistent threats *Academy of Science and Engineering (ASE) Science Journal* 1, 3 (2012), 93-105.

[8] Dumitras, T. and D. Shou. Toward a standard benchmark for computer security research: the worldwide intelligence network environment (WINE). In Proceedings *of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS '11).* ACM, New York, 2011.

[9] Las-Casas, P. H. B., Santos Dias, V., Meira, W., Jr., and Guedes, D. A big data architecture for security data and its application to phishing characterization. In *Proceedings of IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS).* IEEE, Washington D.C., 2016.

[10] Delic, K. et al. Active enterprise analytics. In *Proceedings of the 13th Annual Workshop of HP Openview University Association, Hosted by University of Nice at Côte d'Azur (HPOVUA '2006).* K. Boudaoud, N. Nobelis, and T. Nebe (eds.). Infonomics-Consulting, Stuttgart, Germany, 2006.

[11] Delic, K. and P. Leberre. Methods relating to the monitoring of computer systems. EP 1526679 A1. Patent.

[12] Camp, J., Cranor, L., Feamster, N., Feigenbaum, J., Forrest, S., Kotz, D., Lee, W., Lincoln, P., Paxson, V., Reiter, M., and Rivest, R. Data for cybersecurity research: Process and "wish list." 2009.

Sidebar: Internet Security and the Law of Communicating Vessels

Since ancient days, the concept of communicating vessels has been widely used—water will reach the same level in all parts of the system regardless of what the lowest point is of the pipes and regardless of the shape and volume of the vessels. If water is added to one vessel, it will find an equal level in all the containers.

Similar to communicating vessels, nowadays security of a single system (e.g. a single organization) is not enough to take into sole consideration to guarantee the overall risk level, because all interconnected networks and devices have become connected vessels (as billions of others ¨things¨) linked to the internet—our common liquid. The weakest points in our interconnected world would impact the security of the whole system (as recently seen in DDoS attacks using insecure IoT devices or botnets). In the 17th century Blaise Pascal stated the pressure exerted on a molecule of a liquid is transmitted in full and with the same intensity in all directions. The same could be observed in our internet world—we witnessed initially limited attacks involving a few systems spreading widely and jeopardizing many interconnected systems and organizations.

Biographies

Kemal A. Delic is an associate editor for *Ubiquity Magazine*. He is also a senior technologist with HP Enterprise Services. He serves as an adjunct professor at PMF University in Grenoble, advisor to the European Commission FET 2007–2013 Programme, and expert evaluator for Horizon 2020. He can be found on Twitter @OneDelic. Contact him at Kemal.Delic@hpe.com.

Louisa Saunier, CISSP, CISA, CISM, PMP is a security expert with HP Enterprise Services. She is also a reviewer of ISACA Journal and author of a several patents and articles in the security area. Contact her at louisa.saunier@hpe.com.

DOI: 10.1145/3158348