Lessons learned while supporting Cyber Situational Awareness

Graziano Blasilli\textsuperscript{1} © Emiliano De Paoli\textsuperscript{2} © Simone Lenti\textsuperscript{1} © Sergio Picca\textsuperscript{1} ©

\textsuperscript{1}Sapienza University of Rome, Italy – \textsuperscript{2}MBDA Italia

Abstract

The increasing number of cyberattacks against critical infrastructures has pushed researchers to develop many Visual Analytics solutions to provide valid defensive approaches and improve the situational awareness of the security operators. Applying such solutions to complex infrastructures is often challenging, and existing tools can present limitations and exhibit various issues. In this paper, supported by cybersecurity experts of a world leader company in the military domain, we apply an existing Visual Analytics solution, MAD, to a complex network of a critical infrastructure, highlighting its limitations in this scenario and proposing further solutions to improve the cyber situational awareness in both proactive and reactive risk analyses. The results of this research contribute to characterize the activities performed by domain experts in this domain and their implications for the design of Visual Analytics solutions that aim at supporting them.

CCS Concepts

• Human-centered computing → Visual analytics; • Security and privacy → Network security;

1. Introduction

Critical infrastructures are the systems and services of a country whose operational continuity is essential to guarantee its economic and social well-being. These infrastructures strongly rely on ICT technologies and Industrial Control Systems; the adoption of standard embedded systems platforms and commercial off-the-shelf software have contributed to lowering costs and improving ease of use, at the cost of increasing their exposure to computer network-based attacks. Defending against these threats requires a deep knowledge of the infrastructure, including its connectivity, weaknesses, and business dependencies on supporting systems, and the application of complex and structured strategies to reduce the attack surface. Furthermore, mission constraints restrict the applicability of the countermeasures, limiting the applicability of fully automated solutions and requiring security operators to identify the best trade-off between the operational impact on the organization’s business and the reduction of the attack surface.

Supported by cybersecurity experts of a world leader company in the military domain, we started to apply an existing Visual Analytics solution, named MAD\textsuperscript{[ABL*19, ABB*18]}, to a critical infrastructure of that company. While this activity has highlighted some limitations of MAD in meeting their requirements, it pushed us to design new analytical and visual solutions to satisfy them. In this paper, we present two years lasting research activities, describing the requirements elicited during the process and the solutions proposed to satisfy them. Furthermore, we discuss their broader implications on the design of Visual Analytics techniques to improve cyber situational awareness in both proactive and reactive risk analyses of critical infrastructures.

The MAD System MAD\textsuperscript{[ABL*19, ABB*18]} is a Visual Analytics solution for risk management developed in the context of the EU-FP7 Panoptesec project\textsuperscript{†} along a three years user-centered design activity involving two security managers and four operators of ACEA\textsuperscript{‡}, a large Italian public organization that supplies energy and water to millions of people. It uses a topological attack graph model in which the nodes of the graph are the hosts of the network and the edges are the possible vulnerabilities exploits between them. According to the terminology introduced by the NIST in the Cybersecurity Framework\textsuperscript{[Nat]}, it supports the Identify and Protect functions providing a visual environment for analyzing the multi-step attacks and evaluating the related mitigation actions, and the Detect and Respond functions using the underlying model to predict the evolution of ongoing attacks and to identify suitable countermeasures. The tool proved its effectiveness by supporting the identification of critical vulnerabilities during the project and, later on, it was commercially used by RHEA Group\textsuperscript{§}, one of the partner involved in its development, for training and risk assessment. Although the MAD system was born targeting the ACEA critical infrastructure, it can be used on different domains, such as the military context, as shown in the rest of this paper.

\textsuperscript{†} http://www.panoptesec.eu
\textsuperscript{‡} https://www.acea.it
\textsuperscript{§} https://www.rheagroup.com

© 2021 The Author(s)

DOI: 10.2312/eurova.20211093
2. The MBDA scenario

MBDA, world leader in the military aviation sector, is a multinational company with almost 10,000 employees working in France, the UK, Italy, Germany, Spain, and United States. The continuous application of cutting-edge technologies is an advantage in the development and production of new products, and a means of guaranteeing to customers that innovation can always be present in existing products during their lifespan in order to satisfy the continuous evolution of military scenarios. The interest of MBDA towards MAD is directly tied to the notion of dual-use technology, i.e. technologies that can be used in more than one context, for example serving both civilian and military purposes. The successful adoption of MAD into the civil context critical infrastructures of ACEA, pushed MBDA to explore the system usage on the typical company critical infrastructures such as missile systems.

The company, after having anonymized real data for security reasons, provided us with a pseudo-real network, composed of 242 devices and 62 subnetworks, which represents a common MBDA military scenario. Supported by two MBDA cybersecurity experts, we started to analyze the case study network with an already developed Visual Analytics system [ABLv1] (MADv1, for the rest of this paper). With respect to the ACEA networks, the MBDA network shows an average number of vulnerabilities on devices and more dense topological connections. While the built-in attack graph generator of MADv1 generated hundreds of possible attack paths in the ACEA network, it generated up to millions of attack paths in the latter one. This cardinality causes in MADv1 both computational and visual issues, see Figure 1, that highly impact the situational awareness for the security operators. The goal of our research was to analyze MADv1 issues and to define possible solution to the challenges posed by this new network.

![Figure 1: The MBDA scenario represented in the Proactive Environment of MADv1. The network view, on the left, shows the attack graph superimposed over the network topology. The high cardinality of devices and attack paths generates computational problems and visual cluttering issues.](https://www.mbda-systems.com)

3. Related Work

Attack graphs have been extensively studied as an efficient solution to model the paths that a potential attacker can use to intrude into a target network. Kaynar [Kay16] proposed a taxonomy of the methods applied to generate them, and their usage in network security identifying the growing size of the target network and attack graphs as one of the main problem for their comprehension and practical usage. More in general, many of the proposed solutions incorporate Visual Analytics techniques for enhancing the cyber situational awareness [End88, JLSW09] of security operators in the process of monitoring and maintaining the network health [GFS*16]. Many of these solutions evolved over a long period of time. Noel and Jajodia [NJ04] proposed aggregation and interaction techniques to support attack graph usability, while Noel et al. [NJP05] extended this work integrating adjacency matrices to evaluate the impact of changes in the network configuration. Further extensions were presented to evaluate the consequences of hardening on the attack graph [ONP08] and to simplify it by trimming redundant paths [HVOM98]. GARNET [WL08] and NAVIGATOR [CIL*10] rely on NetSPA adding, respectively, a treemap-based visualization reflecting physical or logical topology for evaluating the node reachability and a finer granularity of the analysis up to the node level. The requirements presented in the next sections demand to enrich the graph by grouping its elements and encoding additional attributes. Many contributions have been proposed for the visualization of multivariate graphs [NM19b] both for node-link representations, e.g., pie-chart-like icons encoding attributes value [CLLT15], and for matrix and hybrid approaches. At the same time, different contributions propose aggregation techniques for node-link representations, for example, the replacement of common topological patterns with glyphs [DS13] and the application of clustering algorithms [vBR16]. Despite the generality of these solutions, we adapted them to the peculiarities of our scenario as presented in the following.

4. System Evolution

The application of MAD for MBDA had the objective of supporting the cyber situational awareness in the context presented in Section 2. The features required in such context are many, starting from a wider information representation, up to attack simulation and evaluating their effects or anticipating countermeasures through the use of what-if analysis techniques. In the following we report requirements deemed useful by MBDA to achieve its objectives, discussing how and why they have been thought, analyzed and then developed into the enhanced system MADv2.

R1 – Providing Threats Impact Information – The proactive and reactive environments of MAD represent the network as a node-link diagram, by superimposing the attack paths and their attributes over the network topology (see Figure 1). A specific color coding is used to represent the roles that a graph element plays in every path: red color identifies the final step of an attack, green is used to identify attack paths source nodes, and blue represents every intermediate step. Although red has negative connotations in the cybersecurity context [RHH*11], the experts noticed that blue and green are not a good choice since they are often used for communicating safety [Dav99]. Therefore they suggested to use red, yellow, and orange for encoding attack path roles because they perform better in expressing risk information [JC20].

To deal with the cardinality of attack paths MADv1 encodes information through nodes and edge thickness representing only the
number and types of paths that cross them. Although this solution is still a good approach, the domain experts highlighted that additional information about vulnerabilities exploit consequences, can be useful to improve risk analysis. The CIA Triad (Confidentiality, Integrity, and Availability) is a well-known model for the development of security policies in the cybersecurity domain and, since vulnerability exploits can make a breach in any of the three aspects, their understanding can be crucial for improving the situational awareness [KHLT18]. Vulnerability exploits can have different types of impact on devices in which they lie [KSL18], and can be grouped according to the types of privileges gained on the machine: none privileges, user privileges, and root privileges. An attack path that involves two or more vulnerabilities on the same device allowing to increase the level of the gained privileges is called a privilege escalation attack. These types of exploits are critical being typically not revealed by network intrusion detection systems (IDS) because they do not often generate network traffic. To better support the cybersecurity operators needs when dealing with such a type of system, we improved the device encodings, see Figure 2. The proposed solution have been designed and evaluated by collecting the experts’ feedback during various meetings that we had.

Concerning the proactive analysis, background color represents the higher privilege reached by all the attack paths involving that device: gray, blue and purple stand for none, user and root privileges. The attack path proportions on nodes is shown with the internal donut chart, using the new color coding. The external donut chart represents in gray the proportion of vulnerabilities of the node used by the current attack paths, while in blue the subset of them which can be used for performing privilege escalation. Concerning the reactive analysis, while MADv1 provides only information about which devices were compromised by an attack, MADv2 shows also the impact in term of degradation of CIA levels for each compromised device, see Figure 2. It is worth noting that domain experts have generally shown more interest on the possible vulnerability exploits in the proactive scenario, and more interest on the operational impacts in the reactive scenario during the entire design process. Such a type of approach aims at providing the operators with all the necessary information and to avoiding overwhelming them.

R2 – Semantic Topology Aggregation – After having introduced the new encodings satisfying R1, cluttering issues and high computational problem, already noticed by the experts in MADv1 (see Figure 1), become more evident since the additional information we added, see Figure 3-C2. The complexity in terms of cardinality (number of paths to represent) and the increased dimensionality (attributes of paths and devices) make the attack graph comprehension difficult. Issues come mainly from the network topology since it is extremely connected. According to our experts, in their military context, it is common to have firewall policies that limiting communication among subnetworks, while no rules exists inside a subnetwork since devices have to communicate with each other. In this scenario, distinct attack paths over the network can be millions and the security operator could be overwhelmed by this huge mass of information to analyze.

The aforementioned issues cannot be mitigated by standard anti-clutter solutions like random sampling [ED07, BS04] because they can exclude critical information. The security operator must be aware of each single attack path on the network since each of them can pose high risks. We decided to aggregate devices and attack paths for both improving the attack graph rendering load and mitigating the visualization issues. The aggregation of network nodes considers both the network topology and geography, plus the business functionalities provided by the devices.

These needs were transformed into requirements for a clustering process: I) geographically close devices, or II) devices connected to the same subnetwork, or III) devices supporting the same business services should be in the same cluster. Once having modeled this information on the network graph, we applied the Clauset-Newman-Moore [CNM04] clustering algorithm to have a first raw aggregation. After that, the company experts made manual refinement using their domain knowledge of the network. As a result, starting from 242 devices, we obtained 12 clusters. As visible in Figure 3, this aggregation mitigates the cluttering issues and in the meanwhile provides a full overview of the network. To support a deep exploration of the network, each cluster can be expanded to

Figure 2: Evolution of proactive and reactive visual encodings of a device. In proactive analysis it is possible to see that the device is traversed by 75% of intermediate step and 25% of final step attack paths. The blue-red piechart on MADv1 and the orange-red donut chart in MADv2 represent path proportions. The latter shows that attack paths allow reaching user privileges on the node (blue background), involving 20% of device CVE, and 5% allow internal escalation (outer blue ring). Concerning the reactive encodings, while MADv1 shows only information about exploited vulnerabilities, MADv2 shows also an overall level of how it has been compromised and details on the CIA impacts.

Figure 3: The MBDA scenario represented in the Proactive Environment of MADv2, showing the node encodings developed in requirements R1 and R2. Clusters of devices (C1, C2, C3, …) are represented as nodes of the network. Cluster C2 is expanded and its devices are visible. When clusters are shown aggregated (C1, C3, C4), attack paths are visible at cluster level, while when a cluster is expanded attack paths are visible at device level (C2).

© 2021 The Author(s)
show its devices. The nodes and attack paths encoding remains the same. When a cluster is closed, a cluster node shows the aggregate information of all the devices of the cluster. Instead, when the cluster is opened, the information is visualized at the device level. It is worth noting that this aggregation technique helps to maintain devices that form a clique into the same cluster, thus minimizing the number of attack paths that involve more than one cluster and improving the visualization when a cluster is shown closed. Attack graphs are highly influenced by aspects such as devices cardinality, topology settings, and vulnerability profiles. To avoid cognitive overload, visual or computational issues, scalability is an important aspect to consider during the design. The experts found useful such a type of aggregation. It helped them to identify the network areas having major risks. Once having identified those clusters, further drill-down explorations allow a device-level analysis, improving the comprehension of the vulnerable surface.

R3 – Mitigation Strategies – The identification and the simulation of mitigation strategies is an important activity of a security operator of critical infrastructures. In this scenario, what-if analyses have been proven to be effective in reducing vulnerability exposure and related risks [ABC*19]. By using their knowledge about resources, costs, and business constraints, security operators can identify and test mitigation strategies without working with the real machines and risking to interrupt the business continuity. To support this requirement we included and improved the attack graph-based vulnerability fixing strategy proposed in VULNUS [ABC*19], which allows reducing the attack surface by pruning attack paths from the whole attack graph.

The VULNUS strategy, based on the unweighted set cover algorithm, generates an ordered list of vulnerabilities to fix, by prioritizing them according to the cardinality of attack paths involved. We improved the original strategy after that the experts pointed out the need of prioritizing the fix also by considering the role that devices play in the business continuity. We switched to the weighted set cover algorithm, while the experts defined ad-hoc devices weighting parameters based on their domain knowledge. As a result, the revised strategy gives priority to mitigations that reduce risks on critical nodes. While VULNUS supports this what-if analysis by suggesting mitigation strategies and computing their expected result only in terms of cardinality reduction of attack paths, in MADv2 we included detailed information about attack paths and relative mitigations. The system allows, in both proactive and reactive analysis, to inspect attack path mitigations, to simulate effects of their application, and to further analyze attack paths attributes by using visual filtering techniques proposed in [ABL*20], see Figure 4.

Threats analysis and mitigation planning are two complementary aspects of cyber defense both in proactive and reactive scenarios. Very often in existing solutions these two phases are supported one at a time in proactive scenarios and at the same time in reactive ones; we noticed that in our use case the analysis procedures of the domain experts required support for these two phases concurrently also in the proactive scenario.

R4 – Combining Topology and Geography – The company network is spread on a wide geographical area, and the domain experts expressed the need of showing also geographical data on the prototype. In fact, a comprehensive view of the whole network, showing topology information along with the geographical ones, is able to increase the situational awareness of the security operators (see e.g., [ABK*19, AS17, ABG*18]). While the geographical view is mainly focused on displaying information about the network, the topological view is the main tool for taking actions such as attack countermeasures or what-if analysis. Since distances among nodes can considerably differ when considering topology or geography, addressing such a type of requirement is still a challenging problem in visualization research [JMO*16].

To minimize differences between topology and geography we decided to represent geography information only for clusters of devices. Geographically close devices appear in the same cluster (R2), so we locate the cluster using its centroid position, while a dark gray convex-hull behind of them cover the portion of the map with the meaning that geography is not represented anymore, see Figure 3-C2. According to our company experts, cybersecurity experts of military domains use to see geographical information of the network since the position of the devices strongly influences their operational role. This is in general not only true for military context, but also for wide geographically spread critical infrastructures, such as energy supply companies. Thus geographic information is critical and should be easily accessible even during analyses mostly driven by ICT functionalities.

Figure 4: Mitigation strategies simulation. The original attack graph (left) is composed of around 3000 attack paths. After having applied a mitigation strategy involving six CVEs (right), the attack paths have been reduced to almost 3000.

5. Conclusions

The increasing number of cyberattacks against critical infrastructures has pushed researchers to develop many Visual Analytics solutions aiming at providing valid defensive approaches while improving the situational awareness of the security operators. Supported by cybersecurity experts of a world leader company in the military domain, we analyzed an existing solution [ABL*19] highlighting and discussing its limitations when applied to a complex network of a military critical infrastructure. After having discussed limitations, we proposed and implemented solutions aiming at mitigating them. From a broader perspective, these activities have contributed to characterize the activities performed by domain experts in operative scenarios, defining their implications on the design of Visual Analytics solutions that aim at supporting them. This research has also highlighted that the deployment of existing solutions in a real operative scenario can be a very long process that leads to a radical rethinking of some of the design choices that have proved to be effective in the past.
References


