Müller, Neumair, Reiser, Dreo Rodosek (Hrsg.): 11. DFN-Forum Kommunikationstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2017 23

Stakeholder Specific Visualization and Automated Reporting of Network Scanning Results applying Vis4Sec

Tanja Hanauer[1], Stefan Metzger [2]

Abstract: This article introduces a process framework – Visualization for Security (Vis4Sec) – that supports the generation of organizational security knowledge and awareness. Vis4Sec is used to generate stakeholder specific visualizations based on the results of regular performed network scans in a complex IT infrastructure. The process steps Ask, Prepare Data, Visualize and Interact assist to define security relevant questions, prepare a data-driven visualization, embed it into an organizational context and distribute it. A proof of concept implementation was successfully done in a network environment operated by a Higher Educational Institution data center. Scan data resulting from several e. g. Network Mapper (nmap) based scanner machines has been aggregated and analyzed automatically, was then highly-enriched with organizational, security relevant information, visualized in dashboards, adapted to stakeholder specific requirements and distributed as reports.

Keywords: Information Security; Visualization of security-related data

1 Introduction

In many higher educational institutions' (HEIs) data centers IT systems are administered by different operating teams. In addition some services, e.g. hosting virtual machines or Infrastructure as a Service (IaaS) cloud services for institutions or individual users strengthen this and require the incorporation of customers' staff. This differs significantly from the centralized service operating models common in the industrial sector. HEIs also provide open and almost unrestricted infrastructures, so users are allowed to bring and connect their own devices or install any software found somewhere on the internet. There are no data center or network infrastructure wide asset management systems or configuration management databases, that provide a complete overview of the connected devices and the installed software. To cope with the security of such an ever changing environment the authors suggest the usage of integrated network scanning techniques, that provide an overview of IT systems, services, operating systems, and application software. The results of such regular scans and their deltas are a good starting point for security reporting. Unfortunately they are hard to grasp by the human eye as the results are in plaintext or in Extensible Markup Language (XML), which makes identifying security relevant changes difficult, especially when a huge number of systems are concerned. The visualization

[1] Leibniz Supercomputing Centre, Boltzmannstraße 1, Garching n. Munich, Germany hanauer@lrz.de
[2] Leibniz Supercomputing Centre, Boltzmannstraße 1, Garching n. Munich, Germany metzger@lrz.de

process Vis4Sec collects and aggregates these scan results at one central point and uses adequate data visualization methods as an approach to meet this challenge following the proverb *A picture is worth a thousand words*. Furthermore, existing organizational and security relevant information is correlated to handle and distribute those results best. The visualization process provides a framework for data acquisition, aggregation, visualization and organizational distribution of information. It supports the tracking of changes in the environment and makes the determination of the current attack surface possible.

The rest of this article is structured as follows: Section 2 provides a short overview of relevant security controls based on ISO/IEC 27001 and the Center for Internet Security Critical Security Controls (CSC), broaches data visualization techniques and introduces requirements of relevant stakeholders, before section 3 describes the Vis4Sec process itself. Section 4 shows a proof of concept process operation with the goal to limit and control open network ports. Section 5 summarizes the benefits of our approach and gives an outlook on future work.

2 State of the Art

This approach is based on actual security controls according to ISO/IEC 27001, good practices like the Critical Security Controls, well-known guidelines for data visualization, study-based stakeholder requirements analysis and and a brief selection of existing approaches.
The international standard ISO/IEC 27001 defines minimum requirements which an information security management system (ISMS) has to fulfill. Besides the general clauses provided in sections 4 to 10 of the standard, Annex A defines in total 114 reference security controls of which two are relevant for network scans. Control A.13.1.2 requires among other things inclusion and a regular review of technical and organizational security aspects related to network services. Control A.18.2.3 requires a technical review to ensure compliance with the organization's information security policies and standards. This encompasses the usage of tools, e. g. port scanners, or conducting penetration tests. While ISO/IEC 27001 does not require a specific implementation of security measures, the Critical Security Controls (CSC) [Ce16] expand on such details. The CSC are a widely used set of actions for cyber defense recommended by the Center for Internet Security. Controls with relation to automated network scanning are CSC 9 – Limitation and Control of Network Ports (9.1, 9.3), CSC 3 – Secure Configurations for Hardware and Software (3.6), and CSC 18 – Application Software Security (18.1, 18.4). CSC 9.1 and 9.3 are introduced in detail during the exemplary process run, the remaining controls are described briefly as subsequent iterations of the process. Subclause 9.1 ensures that only ports, protocols, and services with validated business needs are running on each system, which makes it directly mappable to controls in ISO/IEC 27001. Further relevant details are found in subclause 9.3 which requires automated regular port scans against all key servers and comparison of the results to a known baseline. This allows the discovery of unlisted changes to the organization's

approved baseline.
The generated visualization takes into account general knowledge about processing of visual information like the principles of Gestalt Theory, and design basics like Tufte's Design Criteria or Shneiderman's Information Seeking Mantra. Furthermore are challenges of information visualization in large companies addressed concerning the integration of tools in daily work processes, getting the data and being in constant close cooperation next to others as described by Sedlmair et al. [Se11]. The working conditions of system administrators, security personnel and members of the higher-level management described in qualitative ethnographic field studies have been analyzed and complemented with our own observations over a timespan of five years working in a HEIs' data center. Key findings according to Anderson [An02] are that transparency, notification, automation, schedulability, simplicity and scalability are very important criteria for admin tools. Besides that, data visualization can, as highlighted by Haber et al. [HK07] and Mahendiran et al. [MHZH12], provide improved system and security monitoring, a better overview of the current status of the infrastructure and simplicity of use for admins and security personnel. Human factors like the communication of security issues, organizational culture or an open environment are just like technical factors like the complexity of systems and applications and their vulnerabilities relevant for IT security management, as stated by Werlinger et al. [WHB08]. Current approaches in visualization research exist for security tasks like (network) security alert management [CvW16, FPLB17], network security and management [LS10], or even more specific topics, like visualization of ports or firewall configurations. These approaches are very task-specific and only focused on one use case for the decision support challenge at hand. Furthermore, a lot of visualization-method-specific research for security tasks exists with a prototypical implementation for one kind of visualization like assessing cyber incidents and network security with graphs [APS15], or ensembles [HHH15], or the visualization of specific data sources like data streams, web server and other log files. These approaches are helpful for the design of single visualizations, but they are also no solution for the organizational challenge at hand. Hence, the overall generation of visualization is taken into consideration described by visualization processes and frameworks like [Fr04, Ma08] and further developed into an integrated security specific management solution.

3 Process Vis4Sec

The analysis of stakeholder requirements, the collection of relevant data as a basis for the visualization, its organizational integration and the security requirements are implemented within a process framework – Visualization for Security (Vis4Sec). This framework offers a systematic approach to improve the information security of an organization. Figure 3 shows the iterative process Vis4Sec with its Initiation and four process phases Ask, Prepare Data, Visualize and Interact.

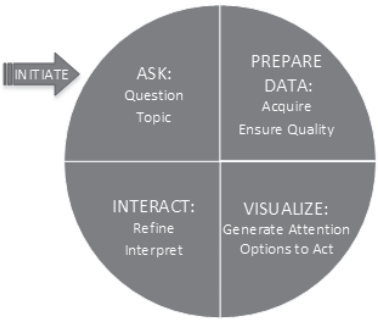

Fig. 1: Process Vis4Sec

Initiate: Vis4Sec starts by collecting and briefly describing necessary parameters for the visualization like the environment, the stakeholders, requirements and planned actions.

Ask: The initial question to be answered by the visualization has to be stated. It defines the context and it is the basis for the collection of relevant data. It is important to keep the question as simple as possible to allow for a successful process operation. In repeated iterations of the process the starting question is enhanced, refined or even completely redefined.

Prepare Data: This is by far the most extensive step as it consists of technical aspects like data collection, its pre-processing and analysis, and additional organizational aspects like the preservation, access control and disposal of data according to compliance and data protection regulations. It also includes the technical embedding into the organization like the definition of a data model, the insurance of data quality and the automation of the data collection and its quality insurance.

Visualize: The representation and presentation form for the quality ensured data is chosen with the goal to draw the attention of the recipient to the topic and offer options to act. The visualization should also generate awareness, provide a point of communication between the recipients and improve the status of the reported issue through the generated visibility.

Interact: The recipient provides feedback and adds expert knowledge that further improves the data visualization. Its utility is then evaluated according to the question asked during the question phase. New knowledge is interactively generated and the process starts over again with new or adapted questions derived from the previous result. New recipients obtain the improved or additional visualization.

Iterate: The process is iterated with modified questions stemming from the experience of the previous process runs. For example the specificity of the inquiry is enhanced, new data sources are added or existing ones are improved, and the feedback from the recipients of the visualization is built into the next process run.

4 Proof of Concept: Limitation and Control of Network Ports

In this section an exemplary process operation is shown as we deduce the question *"What are the reachable ports – externally or internally – on each system?"*. It is answered primarily by obtaining portscan data in the first run of the process. Afterwards data of the organization

is collected and its data quality is ensured using visualizations that track the data quality and initiate and accompany an organizational handling of the results. The aim here is to improve the organization's security especially towards the fulfillment of the requirements specified in CSC 9.1. In the next iteration of the process the requirements of CSC 18.1 are addressed and the false-positive rate decreased.

4.1 Initiation

The environment The Leibniz-Supercomputing Centre (LRZ) – a HEI's data center – providing more than 70 ICT and network services for institutions connected to the Munich Scientific Network (MWN), operates internally more than 130 server subnets to which more than 700 heterogeneous IT systems are connected. This environment can be specified as an overall complex and continuously changing setup. Usually, little knowledge about the concerned services and their probable vulnerabilities exists, which becomes obvious when security relevant questions are asked.

Requirements concern running services to be known, new services to be timely detected and that potentially vulnerable services to be recognized and patched as fast as possible. To fulfill those the distribution of stakeholder specific reports that provide an overview, impart knowledge and offer options to act, seems to be useful.

Stakeholder specific reports are designed for system administrators, security practitioners and IT management staff in this example scenario:
`Internal IT System Operations Teams (aka system administrators):` They configure, maintain and provide the IT systems and services. They are expert users with specific needs in terms of complexity, collaboration and risk. This group requires frequent reports with technical details about the configuration of the systems.
`IT Security Personnel (aka security practitioners):` Their focus lies on the security of the IT systems and the infrastructure. They have to deal with more complexity than system administrators, due to the high environmental rate of change, and the trade-off between usability, security and costs is also highly relevant. They occupy mostly a position in between since IT security often depends on the commitment of the management and the cooperativeness of the system administrators. This group requires frequent reports providing an overview and specific details to security related topics.
`IT Management Staff:` They are usually business-driven, and mainly responsible for making high level decisions. So they need abstraction from the technical aspects in form of information that allows them to make decisions based on correct data. This group requires infrequent reports in form of a high-level management view.

Planned Actions are the automation of network scans on a fine-granular level, followed by the collection and analysis of the scan results, their annotation with organizational and security-related information, and their stakeholder specific visualization and distribution, which lead to an enhancement of the organization's security level.

4.2 First Iteration: Open Network Ports on each System

Ask-1 According to the clauses in CSC 9 the question is: "What are the open ports, protocols and services with validated business needs running on each system?" To get a more practicable starting point the business need is left out, because it is not vital to answer in the academic environment of a HEIs' data center. Also the protocols and services are set aside for the next iterations and the question is thus simplified to: "What are the open ports on each system?" We will specify our question even more to further minimize the amount of results we get:[3]
"What are the reachable ports – externally or internally – on each system?"

Prepare Data-1 Resulting data from port scans is created, and organizational data from an already existing server configuration management database with detailed information about each system is collected and prepared. Furthermore, the quality of the data ensured. In the next iteration additional data, like the SSL configuration, is added to enrich the results.

`Data Source: Port Scanners` are tools easy to install and a simple network scan is also an easy task. But to do this in a more structured manner requires a comprehensive concept for deployment of several scanning machines, their operations and proper result processing. Answering questions about deployed scanning tools, usage of more than one scanning machine and their placement inside or outside the scanned network infrastructure, the scope of each scan performed and the repetition intervals is needed as Hommel et al. described [HSM15]. The complexity of this setup (several scanners, different locations) and various responsible contacts for the scanned subnets and machines bears the challenge of how to deal with the results, how to filter and distribute them among the stakeholders. This is done with the organizational data introduced in the following:

`Data Source: Organizational Data` is data that is almost static, e. g. basic information about the machines like the version of the operating system installed, its IP address, the system administrator's name and the department operating it. This information is extracted from a tool that functions as part of an organization wide Configuration Management Database (CMDB). It quickly became obvious that its data quality is unreliable, but this is crucial to provide useful results.

`The Data Quality` is ensured according to the six dimensions of data quality according to DAMA UK [ACea13]: completeness, uniqueness, timeliness, validity, accuracy and consistency. An overview of the data input's continuity is generated, quality checks of the organizational data are initiated, and an additional comparison with other data sources as data collected directly from the servers, in form of their installation base is also done in an iteration of the process.

Visualize-1 Visualization gives an overview or provides details if necessary. Dashboards in a WebGUI allow interactive usage of the information and PDF reports sent out by email

[3] The examples stem only from externally reachable systems.

present the information in a static form. Data is explored with use of an interactive dashboard as basis for the stakeholder specific dashboards and security enhancing visualizations. The first dashboard built functions as viewing tool. It is used to explore the data and extract areas of interest for stakeholder specific reports. The security practitioners are in most cases the user group that chooses the content of the visualizations because of their professional interest and their expert knowledge. The overview of the results from the port scans about the most exposed subnets or machines can support them to proactively enhance the overall security of the organization and in case of a newly disclosed vulnerability it provides them with the means to find the affected machines.
The interactive WebGUI makes it possible to filter the data and explore relevant subnets, groups or ports. The search capability allows to filter findings for one specific port, which is helpful to find machines providing a probably vulnerable service. So all systems providing services on that port and probably using the vulnerable product are quickly found and can be further investigated. For example in case of the Heartbleed OpenSSL flaw a search for servers providing OpenSSL services to the outside on port 443 was done on the Security Practitioners Dashboard shown in figure 2. It displays systems with an open web service

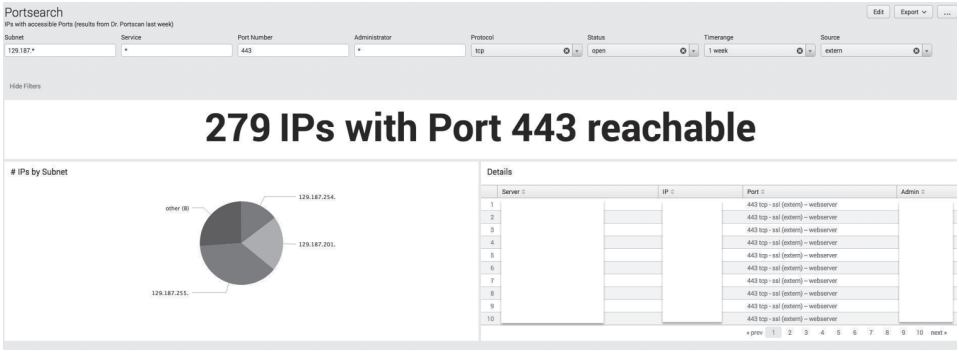

Fig. 2: Security Practitioners Dashboard: Search for web server systems on port 443 (HTTPS). The filters (top). The overall result (middle). The most exposed subnets (left). The details about each system (right).

on port 443 and the subnets where most of them are operated in. It can next to the port be filtered for subnet, service, system administrator, status (open/closed/filtered), time range or source (intern/extern). The pie chart on the left is used to highlight the most exposed subnets. It provides an easy orientation on which subnet to start with by representing the subnets with the highest number of IPs reachable on port 443.[4]

Interact-1: Before the security practitioner alerts the system administrators with a list of servers that *could be* vulnerable, further checks for the ciphers currently used and the actual exploitability of the service are initiated. The goal is to minimize the number of false positives and alert only, if there is a need for action. So the next iteration of the process operation with a redefined question starts.

[4] Additional ports like 22, 4443, 8443, . . . are handled the same way.

4.3 Second Iteration: Exploitable OpenSSL Library

Ask-2 The redefined question to answer is *"What are the externally reachable services that use an exploitable OpenSSL library?"*

Prepare Data-2 The detailed result from the search for systems with a reachable port 443 taken in the previous step is the data source at this point. But further data to identify a web server with a vulnerable OpenSSL version is necessary. This data is added once more as additional scan data – results of a SSL cipher-suite scan. The data quality of the results is ensured by comparison with the also newly added data source configuration data – package information from the servers directly, and information about the vulnerable and patched versions of OpenSSL – Common Vulnerability and Exposures (CVE) data.

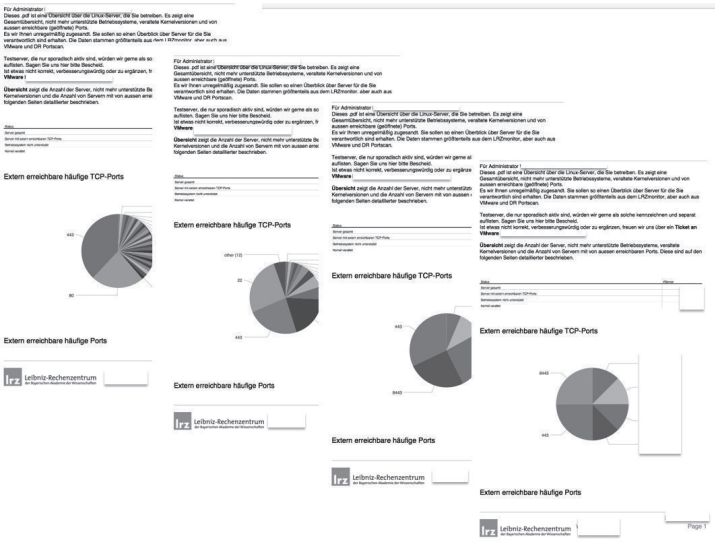

Fig. 3: Reports showing an overview of the most common services externally provided

Visualize-2 The second generated dashboard is a filtered version of the correlated data sources that functions as a request to act. It uses the data from the network scanners enhanced with the results from the cipher-suite scan, the organizational information and the configuration data and then adds a description of the vulnerability in question and how to fix it. For the system administrators it only displays the servers of a single system administrator that are externally reachable and also vulnerable to this exemplary attack. For the third stakeholder group – IT management –, that does neither need interaction with the data directly nor a lot of details, a dashboard providing an overall overview is generated and sent as a regular report. The results from the first iteration are processed to show the services in the organization. The derived report informs about the most common services and the groups providing those services. The second process iteration with OpenSSL results is used

as one of many data points to generate a quarterly report about the patch status and the reaction to actual vulnerabilities.

Interaction-2 Finally all of these dashboards and the searches behind are adapted to fulfill stakeholder specific requirements, from which reports are automatically generated and sent with ReMailS (Report Mailer for Splunk). The reports are generated on configurable intervals and they are embedded into an organization wide feedback system. The first pages of exemplary reports sent regularly to single system administrators and the IT management in an aggregated form are shown in figure 3.

4.4 Further Iterations

In a further iteration the interactivity of the dashboard is enhanced and a search is provided that enables security staff to search ad hoc for vulnerabilities. This was described for a web service on port 443, but in the same iteration it was also done for SSH (22) providing externally reachable management access, which is often an unnecessary exposure of a service.

5 Conclusion and future research

The visualization of security-related data using the Vis4Sec framework is a trigger for security enhancement in an organization. It provides a framework to track and continuously improve the security level of different areas. Based on simply obtainable data like results of network scans correlated with other data sources, the security level of the application software or the compliance to data protection regulations can be ensured. The iterative process approach ensures stepwise refinement of the questions and results meeting stakeholders' needs and the focus on feedback improves the quality of the data, generates organizational knowledge and communication points. Further iterations with refined questions like "Are the software versions used still supported by the vendor?" (CSC 18.1) stemming from the ISO/IEC 27001 and CSC are planned. A further process iteration with data on software packages, CVE data including Common Vulnerability Scoring System scores is in preparation. Also an iteration asking 'What are new listening ports, new administrative users, changes to groups and local policy objects or new services running on a system?" (CSC 3.6) sounds promising.

References

[ACea13] Askham, Nicola; Cook, Denise; et al., Martin Doyle: The Six Primary Dimensions for Data Quality Assessment. White paper, Data Management Association UK, October 2013.

[An02] Anderson, Eric Arnold: Researching system administration. Phd, University of California at Berkeley, 2002.

[APS15] Angelini, M.; Prigent, N.; Santucci, G.: Percival: proactive and reactive attack and response assessment for cyber incidents using visual analytics. In: IEEE Symposium on Visualization for Cyber Security (VizSec). pp. 1–8, Oct 2015.

[Ce16] Center for Internet Security: , The CIS Critical Security Controls for Effective Cyber Defense Version 6.1, August 2016.

[CvW16] Cappers, Bram C. M.; van Wijk, Jarke J.: Understanding the context of network traffic alerts. In: 2016 IEEE Symposium on Visualization for Cyber Security, VizSec 2016, Baltimore, MD, USA, October 24, 2016. pp. 1–8, 2016.

[FPLB17] Franklin, L.; Pirrung, M.; L. Blaha, et al.: Toward a visualization-supported workflow for cyber alert management using threat models and human-centered design. In: 2017 IEEE Symposium on Visualization for Cyber Security (VizSec). pp. 1–8, Oct 2017.

[Fr04] Fry, Benjamin Jotham: Computational Information Design. PhD thesis, Massachusetts Institute of Technology, April 2004.

[HHH15] Hao, L.; Healey, C. G.; Hutchinson, S. E.: Ensemble visualization for cyber situation awareness of network security data. In: 2015 IEEE Symposium on Visualization for Cyber Security (VizSec). pp. 1–8, Oct 2015.

[HK07] Haber, Eben M; Kandogan, Eser: Security Administration in the Wild: Security Administration in the Wild: Ethnographic Studies of Security Administrators. In: ACM SIG CHI. 2007.

[HSM15] Hommel, Wolfgang; Stefan Metzger, et al.: Improving higher education network security by automating scan result evaluation with Dr. Portscan. In: EJHEIT. volume 2 of EUNIS 2014 – 20th EUNIS Congress, Umeå, Sweden, pp. 11–20, May 2015.

[LS10] Liao, Qi; Striegel, Aaron, et al.: Visualizing Graph Dynamics and Similarity for Enterprise Network Security and Management. In: Proceedings of the Seventh International Symposium on Visualization for Cyber Security. ACM, NY, USA, pp. 34–45, 2010.

[Ma08] Marty, Raffael: Applied security visualization. Addison-Wesley, cop, http://www.conkel.net/download/Applied2008.

[MHZH12] Mahendiran, Jeevitha; Hawkey, Kirstie; Zincir Heywood, Nur: Understanding the Use of Models and Visualization Tools in System Administration Work. Dalhousie University DCSI Proceedings, 2012.

[Se11] Sedlmair, Michael; Isenberg, Petra; Baur, Dominikus; Butz, Andreas: Information visualization evaluation in large companies: Challenges, experiences and recommendations. Information Visualization, 10(3):248–266, 2011.

[WHB08] Werlinger, Rodrigo; Hawkey, Kirstie; Beznosov, Konstantin: Human, Organizational and Technological Challenges of Implementing IT Security in Organizations. In (Clarke, Nathan L.; Furnell, Steven, eds): HAISA. University of Plymouth, pp. 35–47, 2008.