Usability Testing of Visual Policy Evaluation for Network Security Event Detection*

Volker Ahlers and Bastian Hellmann

University of Applied Sciences and Arts Hannover, Germany
Faculty IV, Dept. of Computer Science
volker.ahlers@hs-hannover.de

Abstract. The network security framework VisITMeta allows the visual evaluation and management of security event detection policies. By means of a "what-if" simulation the sensitivity of policies to specific events can be tested and adjusted. This paper presents the results of a user study for testing the usability of the approach by measuring the correct completion of given tasks as well as the user satisfaction by means of the system usability scale.

Kurzfassung. Das Netzwerksicherheits-Framework VisITMeta ermöglicht die visuelle Evaluation und Verwaltung von Erkennungsregeln für Sicherheitsvorfälle. Mit Hilfe einer "Was-wäre-wenn"-Simulation kann die Sensitivität der Regeln für bestimmte Vorfälle getestet und angepasst werden. Diese Arbeit stellt die Ergebnisse einer Nutzerstudie zur Untersuchung der Gebrauchstauglichkeit des Ansatzes vor, in der die korrekte Bearbeitung vorgegebener Aufgaben sowie die Nutzerzufriedenheit mittels des *System Usability Scale* gemessen werden.

Keywords: Network Security · Policy Evaluation · Information Visualization · Visual Analytics · User Interfaces · Usability Testing.

1 Visual Analytics of Network Security

The monitoring and protection of network security is usually performed by a combination of different detection systems such as firewalls, network access control (NAC), and intrusion detection systems (IDS). Several approaches have been proposed for the visualization of the – in general time-varying – security information, e. g., by means of graphs showing connections between hosts and users or between security-relevant events [8,4], or by means of information dashboards displaying the information of separate security detection components within a single screen [6].

* This work was supported by German Federal Ministry of Education and Research (BMBF) within the projects VisITMeta and SIMU (grant nos. 17PNT032, 16KIS0045). The fruitful collaboration with Gabi Dreo Rodosek, Felix Heine, Carsten Kleiner, the Trust@HsH group, and our industry partners is gratefully acknowledged.

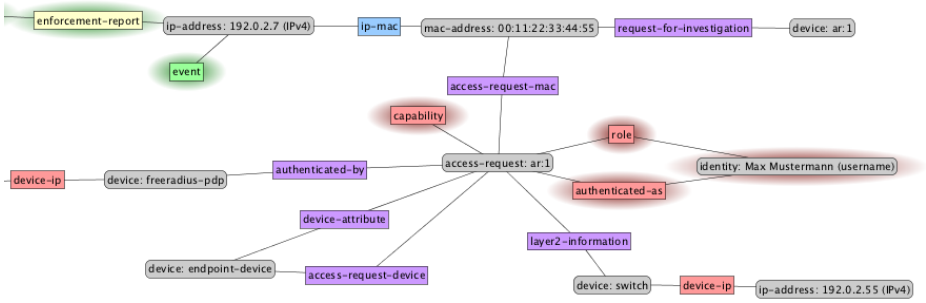

Fig. 1. Part of an example IF-MAP graph visualized with force-directed layout. Identifier nodes are drawn with rounded edges, metadata nodes with sharp edges. The VisITMeta user interface adds a green and red glow effect to nodes that have been added and removed during a selected time interval, respectively.

In order to overcome the limitations of separate information sources, the *Interface for Metadata Access Points* (IF-MAP) specification defines a system architecture for the collection of data from various physical and logical network components (MAP clients) by a central MAP server [9]. The graph-based data model consists of two types of nodes named identifiers and metadata, which are connected by edges. Identifiers represent physical and logical entities such as devices, IP and MAC addresses, or users. Metadata represent different types of connections between identifiers as well as detected network security events.

Our visual analytics framework VisITMeta employs the IF-MAP data model and architecture to visualize network security information, as shown in Fig. 1. A core feature is the storage of historic network data in a graph database, which allows the comparison of network states at different time instances as well as the visualization of network changes within a selected time period. Details on the underlying concepts and the implementation can be found in our previous works [2,3]. The open-source software VisITMeta and related projects are available via GitHub [10].

2 Visual Policy Evaluation

By an extension of the IF-MAP specification, the VisITMeta framework offers the visualization of a detected security event together with the current network state and the detection policy that triggered the event, as shown in Fig. 2. This enables the security administrator to trace the network state and check the plausibility of the detected event, thereby keeping the mental model of the graph-based visualization. In case of false positives, i. e., wrongly detected events, or false negatives, i. e., known events that remain undetected, the administrator can adjust the rules of the policy and re-evaluate it against the historical network state. The underlying “what-if” simulation uses the data model and the data exchange mechanism of the IF-MAP specification [7].

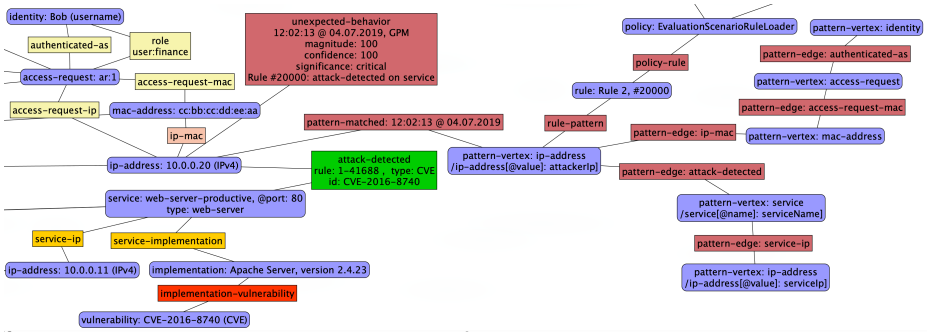

Fig. 2. Example of a detected security event (green) visualized together with the current network state (left) and the policy that triggered the event (right). The relevant identifiers of the network state and the policy are connected via a "pattern-matched" metadata node.

3 Usability Testing

Study Design. In order to test the usability of our IF-MAP-based visualization framework VisITMeta as well as the visual policy evaluation, we conducted a user study consisting of three scenarios with different tasks to be solved on a laptop computer [1]:

A: Find out information on an enterprise network from log files and security component reports (variant A1) as well as from the VisITMeta visualization (variant A2) in random order A1 $\leftrightarrow$ A2.
B: Examine the correctness of policy rules for a detected event.
C: Adapt the detection policy using the "what-if" simulation.

The details and questions of the tasks will be described below.

The two principle variables that have been measured are the task correctness, i. e., the proportion of correctly solved tasks, and the processing time, i. e., the time required to process all three tasks. Furthermore, a retrospect evaluation of scenario 1 based on the system usability scale as well as a general questionnaire for all three scenarios have been applied after the completion of the tasks.

Study Execution. The study has been conducted with $n = 12$ test subjects: 7 computer science students, 2 research assistants, 2 IT administrators, and 1 IT security consultant. The test subjects were asked to self-assess their prior aggregated knowledge (AK) and experience in the following fields on a scale from 0 (no prior knowledge) to 4 (extensive prior knowledge):

AK1: analysis of log output,
AK2: usage of general software and/or hardware for network security,
AK3: knowledge of the IF-MAP specification,
AK4: experience with the VisITMeta framework.

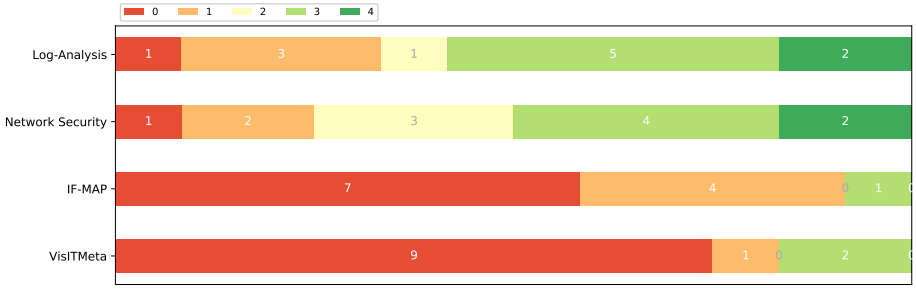

Fig. 3. Prior aggregated knowledge of test subjects ($n = 12$) in different fields by self-assessment on a scale from 0 (no prior knowledge) to 4 (maximum prior knowledge).

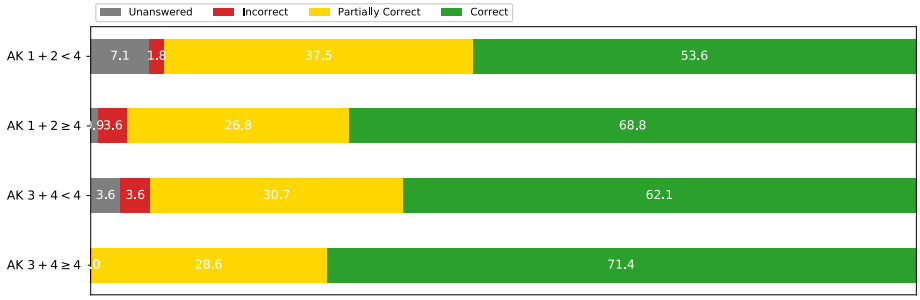

Fig. 4. Overall task correctness (percentage) for all three scenarios in dependence of prior aggregated knowledge AK1 + AK2 and AK3 + AK4, respectively (cf. text for explanation).

The distribution of prior knowledge is shown in Fig. 3. While more than half of the test subjects judge themselves to have good knowledge of log analysis and general network security, very few have prior experience with IF-MAP or even the VisITMeta framework.

Influence of Prior Aggregated Knowledge. The proportion of correctly solved tasks for all three scenarios in dependence of prior knowledge is shown in Fig. 4. While a higher degree of prior knowledge leads to better results (as expected), the difference is more pronounced for the fields AK1 + AK2 (log analysis and general network security) than for the fields AK3 + AK4 (IF-MAP and VisITMeta), indicating that the visualization does not require experience with the underlying technologies.

Processing Time. The median processing time for all tasks of the three scenarios has been 106 minutes, with a minimum of 86 minutes and a maximum of 130 minutes. The processing time includes a short introduction of the study supervisor. Furthermore the test subjects had the opportunity to ask the supervisor if they had problems understanding scenarios or the tasks.

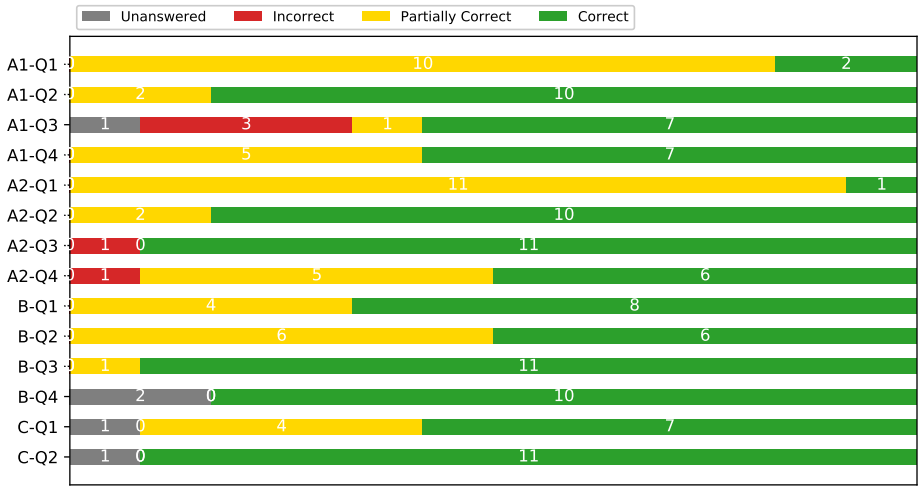

Fig. 5. Task correctness ($n = 12$) for questions Q1 to Q4 of scenarios A and B and Q1 to Q2 of scenario C. Scenario A has the variants A1: monitoring based on log files and security component reports, and A2: monitoring using the VisITMeta visualization (cf. text for further explanation).

Task Correctness. The tasks of the three scenarios A to C included two to four questions for each scenario that had to be answered by the test subjects. These questions can be summarized as follows.

A (network monitoring): List all IP addresses (Q1) and users (Q2) with associated parameters, identify a network component that authenticated a certain user (Q3), describe any detected vulnerabilities of the network (Q4).

B (security event analysis): Identify any events detected by a certain security component (Q1), state the timestamps of each event (Q2), find out which software was running on target device (Q3), decide which of two possible rules produces more false positives (Q4).

C (policy evaluation): Identify the policy rule responsible for a false positive event (Q1), modify the rule to prevent further false positives.

As explained above, scenario A has two variants: In variant A1 the network monitoring is based on the analysis of log files and security component reports, while in variant A2 the VisITMeta visualization is used. The two variants are later compared based on the system usability scale (see below).

The correctness of the answers to the individual questions is displayed in Fig. 5. A few incorrect or partially correct results were found to be the result of a misunderstanding of the task descriptions.

For scenario A (network monitoring), the task correctness for the two variants A1 (log data and security component reports) and A2 (VisITMeta visualization) is comparable. A significant difference only exists for question Q3 (identification of authentication component). One possible explanation for the close results

Fig. 6. System usability scale for scenario A (network monitoring) with variants A1: monitoring based on log files and security component reports, and A2: monitoring using the VisITMeta visualization, in dependence of the random order of presenting both variants to the test subjects.

is that most of the test subjects were used to work with log data and security reports, but had to familiarize themselves with the VisITMeta application in a very short time.

The results of scenarios B (security event analysis) and C (policy evaluation) can be considered good, given the fact that most test subjects (9 out of 12) had no prior experience with the VisITMeta framework.

System Usability Scale. The two variants of scenario A (network monitoring) have further be compared by means of the system usability scale (SUS) [5], which consists of 10 standard questions that have been slightly adopted to the prototypical nature of the software. The results are shown in Fig. 6. Summarizing over all results independent of the order the of presenting both variants to the test subjects, the median SUS is 26.3 for variant A1 (log data and security component reports) compared to a median SUS of 73.8 for variant A2 (VisITMeta visualization). Given the common interpretation that a system is usable if the SUS is higher than 68, this is a clear support of the visualization approach. The difference is more pronounced for the order A2 $\rightarrow$ A1, i. e., first using VisITMeta and afterwards the log data and security reports, but it is also significant for the order A1 $\rightarrow$ A2, i. e., the other way round.

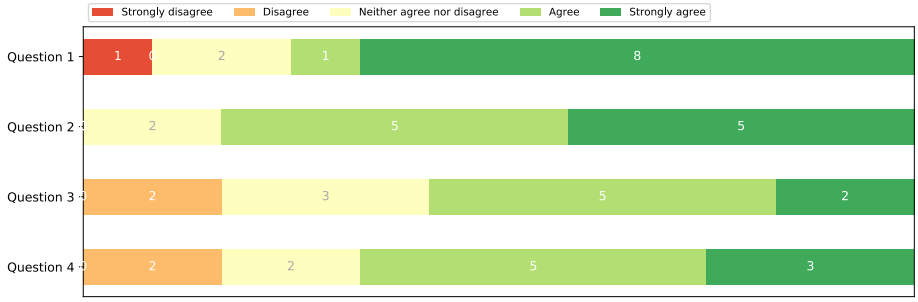

Fig. 7. Answers to the questionnaire with the following four questions ($n = 12$):
1. Can the first scenario be easier solved with visualization rather than with log output?
2. Is the consistent visualization between different scenarios helpful?
3. Were you aware of the fact that data was gathered from different components?
4. Is the integration of detection rules as well as their interconnections helpful?

Questionnaire. Finally, the test subjects were asked to fill in a questionnaire with four questions as stated in the caption of Fig. 7. The results also give a clear indication that the VisITMeta visualization as well as the visual policy evaluation approach are considered helpful.

4 Conclusion

We have presented results of a user study for the usability testing of the visual analytics framework VisITMeta for monitoring the network security. The framework is based on the IF-MAP specification und employs a graph-based visualization of physical and logical network components. It furthermore offers a visual evaluation and management of security event detection policies by means of a “what-if” simulation using historical network data.

While the correctness of typical network monitoring tasks was found to be comparable when using either classical log files and security component reports or the VisITMeta visualization, the usability in measures of the system usability scale was assessed significantly higher for the visualization solution. This as well as the helpfulness of the visual policy evaluation approach were further supported by the answers to a questionnaire.

The scalability to large and highly dynamic networks remains a major challenge of our approach. Finding solutions for these issues will be the subject of future work.

References

1. Ahlers, V., Hellmann, B., Dreo Rodosek, G.: A user study of the visualization-assisted evaluation and management of network security detection events and policies. In: 2019 10th IEEE International Conference on Intelligent Data Acquisition and

Advanced Computing Systems: Technology and Applications (IDAACS). vol. 2, pp. 668–673. IEEE, Piscataway, NJ, USA (2019). https://doi.org/10.1109/IDAACS.2019.8924439
2. Ahlers, V., Heine, F., Hellmann, B., Kleiner, C., Renners, L., Rossow, T., Steuerwald, R.: Integrated visualization of network security metadata from heterogeneous data sources. In: Proc. GraMSec 2015. pp. 18–34. Springer International Publishing, Cham, Switzerland (2016). https://doi.org/10.1007/978-3-319-29968-6_2
3. Ahlers, V., Hellmann, B.: Visual analytics of network security metadata (invited paper). In: Proceedings of IWEIC 2019 Hiroshima: International Workshop on Electronics, Information and Communication. pp. 1–4. Hiroshima City University, Hiroshima, Japan (2019)
4. Angelini, M., Prigent, N., Santucci, G.: Percival: proactive and reactive attack and response assessment for cyber incidents using visual analytics. In: Proc. VizSec 2015. pp. 1–8. IEEE, Piscataway, NJ, USA (2015)
5. Brooke, J.: SUS – a quick and dirty usability scale. Usability Evaluation in Industry **189**(194), 4–7 (1996)
6. Goodall, J.R., Ragan, E.D., Steed, C.A., Reed, J.W., Richardson, G.D., Huffer, K.M.T., Bridges, R.A., Laska, J.A.: Situ: Identifying and explaining suspicious behavior in networks. IEEE Transactions on Visualization and Computer Graphics **25**(1), 204–214 (2019). https://doi.org/10.1109/TVCG.2018.2865029
7. Hellmann, B., Ahlers, V., Dreo Rodosek, G.: Integrating visual analysis of network security and management of detection system configurations. In: 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). vol. 2, pp. 1020–1025. IEEE, Piscataway, NJ, USA (2017). https://doi.org/10.1109/IDAACS.2017.8095240
8. Liao, Q., Striegel, A., Chawla, N.: Visualizing graph dynamics and similarity for enterprise network security and management. In: Proc. VizSec 2010. pp. 34–45. ACM, New York, NY, USA (2010). https://doi.org/10.1145/1850795.1850799
9. Trusted Network Communications Working Group: TNC IF-MAP binding for SOAP, version 2.2, revision 10 (March 2014), `https://trustedcomputinggroup.org/wp-content/uploads/TNC_IFMAP_v2_2r10.pdf`
10. Trust@HsH Group: Iron/VisITMeta project suite on GitHub, `https://github.com/trustathsh/`