Designing Visualisation Enhancements for SIEM Systems

Phong H. Nguyen* City, University of London, UK
Siming Chen Fraunhofer IAIS
Natalia Andrienko Fraunhofer IAIS / City, University of London
Michael Kamp Fraunhofer IAIS
Linara Adilova Fraunhofer IAIS
Gennady Andrienko Fraunhofer IAIS / City, University of London
Olivier Thonnard Amadeus, France
Alysson Bessani LaSIGE Faculdade de Cilncias, Universidade de Lisboa, Portugal
Cagatay Turkay City, University of London, UK

ABSTRACT

DiSIEM is an ongoing EU-funded project that aims to extend existing Security Information and Event Management (SIEM) systems with a set of diversity-related components to improve their capacities. This paper focuses on the scope of visualisation research within DiSIEM and presents the objectives in relation to enhancing the visualisation capability in current SIEM systems, discusses the design approach taken, and reports the initial results from the ongoing visualisation design and development efforts.

1 INTRODUCTION

Organizations currently monitor and manage the security of their infrastructures by setting up Security Operation Centres (SOC) to make security-related decisions (e.g., which system is under attack, what has been compromised, where has an access breach occurred, how many attacks have happened in the last 12 hours). A SOC obtains an integrated view of the monitored infrastructure by employing a SIEM system. These are complex systems that incorporate the functionality to collect logs and events from multiple sources, correlate these events together and then produce summarised measurements, data trends and different types of visualisations to help system administrators and other security professionals.

Instead of proposing novel architectures for future SIEM systems or modifications to existing ones, the DiSIEM project[1] will address the aforementioned limitations by extending current systems in production, leveraging their built-in capacity for extension and customisation. As a multidisciplinary project comprising universities, research institutions, and application partners, the core idea of the project is to enhance SIEM systems with several diversity mechanisms such as diversity-enhanced monitoring and open-source intelligence data analysis. One outstanding limitation observed within existing SIEM systems is the limited capability in supporting the effective extraction of actionable insights from the huge amount of data being collected by the systems [2]. One key novelty we propose with the DiSIEM project is therefore to enhance the visualisation capability in current SIEM systems.

Within the context of DiSIEM, we take a user-centred approach to understand the limitations in existing workflows of SOC operators who work with SIEMs, identify areas and analytical tasks where visualisation can play a role, and iteratively design and develop visual anaytics solutions that we plan to integrate and validate within the current working environment of the operators in the final phase of the project. In this paper, we present the goals of the project in terms of introducing advanced visual analytics capability to current SIEMs, describe the results from our ongoing efforts to enhance the visualisation capability, and report our initial results.

*e-mail: p.nguyen@city.ac.uk

[1] http://disiem-project.eu/

2 ENHANCING SIEMS THROUGH VISUAL ANALYTICS

Although most current SIEM systems offer data visualisation capacities to their users, most often, the visual representations are generic, not designed with particular user needs in mind, or even highly rudimentary to derive any significant insight from the generated data [2]. To add to that, existing systems do not have the capacity to utilise the diverse data modalities that DiSIEM will be generating, such as statistical modelling outputs and comprehensive models of user behaviour. These novel data facets, when combined with the data that is already being gathered, offer challenges and opportunities that we build upon and investigate within DiSIEM. We identify the following objectives to enhance the visualisation capability within SIEM systems:

O1: Design and develop a rich set of specialised visualisations that can handle the diverse types of data that we are analysed within DiSIEM. The data in DiSIEM is multi-faceted having high-dimensional, temporal and relational characteristics and thus requires novel mechanisms to visualise and link these multiple facets.

O2: Develop visual analytics methods to enable analysts in gaining a deeper understanding of user behavior in digital systems. We propose to support this through the visual analysis of user behaviour models that capture the idiosynractic characteristics of user behavior and through the visual investigation of actions carried out by users in the context of their expected behaviour.

O3: Develop interactive methods to elicit experts knowledge to optimise and effectively use predictive and probabilistic models that aim to estimate the potential stability and threats to the system. We develop interaction mechanisms to enable experts to communicate their prior knowledge and preferences to the models, and incorporate methods to communicate the uncertainties in the algorithmic results.

3 DESIGNING VISUAL ANALYTICS SOLUTIONS IN DISIEM

In this section, we report one use case that is motivated by the objectives in the previous section: user behaviour analytics. We explain our approach to designing the solutions, present our early results and discuss how we evaluate the solutions.

3.1 Design Approach

We take a user-centred approach to the design and development of the solutions we present in this paper. To understand the application domain, we conducted a series of workshops with experienced analysts (5+ years). The initial sessions were primarily to understand the current working process, make observations, and informal interviews were carried out where the analysts demoed how they interpret the signals stemming from the UBA model and how they investigate sessions. The analysts used a kibana-based dashboard with off-the-shelf bar charts and pie charts showing the statistics of sessions such as frequency of sessions per IP address, and distribution of UBA modelling score. To investigate a particular session, they needed to manually go through a long list of performed actions to understand what happened in the session. The current process is ineffective, laborious and error-prone.

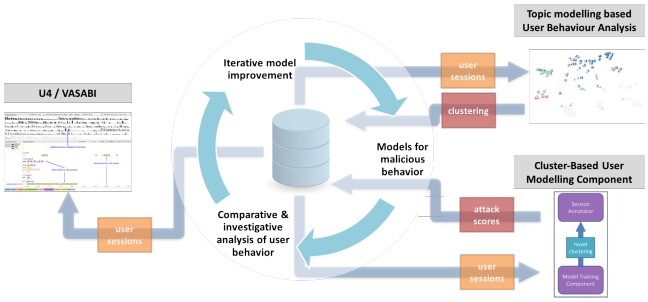

Figure 1: Illustration of how the components that implement the analytic goals work together.

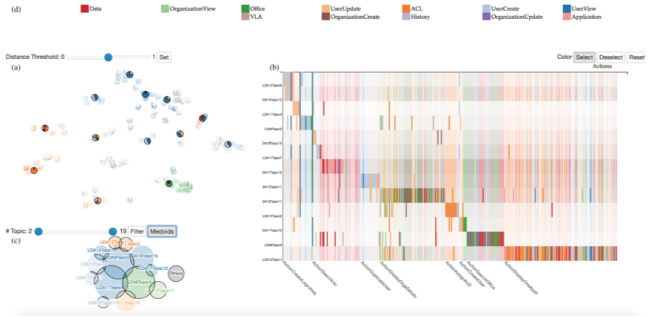

Figure 2: Visual analytics system for topic modelling based user behaviour analysis.

Based on these observations, we identify goals and tasks (as listed in the following), designed and implemented initial versions of the solutions, presented them to the analysts in follow-up sessions (in some cases, some designs were discussed in isolation), and gathered feedback to iteratively improve the designs.

G1 – Pattern analysis of user behaviour. This is to gain overall understanding of the sessions that users performed. The goal is to learn similarity and difference in those sequences of actions, and to cluster the sessions for further behaviour modelling.

G2 – Cluster-based anomalous behaviour modelling. Through the workshops, we understand that the current limitation of UBA models is to have a single global model for the entire dataset. This hinders the model performance because the ways users perform tasks are diverse, leading to noise in the data. The goal is to build a local model per cluster, identified in **G1**.

G3 – Interactive investigation of user sessions. This is to provide better visual support to the investigation of sessions that is currently laborious and error-prone.

3.2 Resulting Solutions

Fig. 2 shows how the components implementing those goals work together in an iterative manner. This section briefly discusses them.

3.2.1 Topic Modelling based User Behaviour Analysis

We develop a visual analytics system with topic modelling ensembles for user behaviour analysis. Taking the advantage of LDA process from text mining [1], we treat sessions as documents and each action in a session as a word to be able to generate the *topics*, i.e., the probability clustered results of user behaviour. We then provide a visual interface that visualises the distribution of initial behaviour clusters, the distribution of actions in each cluster, and the overlap among selected clusters. Analysts can interactively select suitable clusters by observing how representative they are and how much overlap exist between them. In such an iterative manner, the analysts can better understand the behaviour, generate the behaviour clusters to further improve the user behaviour modelling. The system has been empirically evaluated with domain experts.

3.2.2 Cluster-based Behavioural Model Building

We build probabilistic models, one for each cluster provided through the topic modelling, to assess the anomaly of user activities. A score indicating the likelihood of certain actions being harmless, or harmful, is derived from Long short-term memory models (LSTM) [3]. These models output for each action its probability given the previous actions in the session. The score is then calculated from the probabilities provided by multiple LSTM models: one general LSTM that models the general behaviour of all users in the system, and an LSTM model for each cluster of sessions. The probabilities are then weighted by the similarity of the observed user session to each cluster.

3.2.3 Interactive Investigation of User Sessions

Working with domain experts, we further derive specific "session and user investigation tasks" to guide the development of multiple linked view visualisation tools. We briefly describe here the core tasks that we support with further details in our published papers [4, 5].

Multi-perspective exploration of sessions. We provide a visualisation that displays several session attributes (e.g., score, length, user, etc) so that the analysts can examine them concurrently, avoiding to rely on the anomaly score (which can be biased) alone.

Comparative analysis of sessions. We provide a compact visual representation of a session showing actions along a timeline according to their timestamps. The view facilitates identifying similarity and difference between multiple sessions.

4 Conclusion

This paper reports the ongoing efforts in the design of visualisation solutions within the multidisciplinary research project DiSIEM. The final phase of the project will involve the tighter integration of the solutions both within themselves and with the existing SIEM environment of the application partners. The visual analytics components will be able to access the live data stores and be able to integrate any data that emerge from the interaction process, aiming for a streamlined, seamless use of the tools. Already through the ongoing evaluations of the visualisation solutions, we are getting promising results to indicate that visual analytics has the potential to be an indispensable aspect of future SIEM systems and we aim to pursue this agenda during and following the project.

Acknowledgments

This work is supported by the European Commission through the H2020 programme under grant agreement 700692 (DiSIEM).

References

[1] D. M. Blei. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.

[2] DiSIEM Consortium. D-2.1: In-depth analysis of siems extensibility, 2017. Available from `http://disiem-project.eu/wp-content/uploads/2017/03/D2.1.pdf` [accessed 29 August 2018].

[3] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[4] P. H. Nguyen, C. Turkay, G. Andrienko, N. Andrienko, and O. Thonnard. A Visual Analytics Approach for User Behaviour Understanding through Action Sequence Analysis. In *EuroVis Workshop on Visual Analytics*. The Eurographics Association, 2017.

[5] P. H. Nguyen, C. Turkay, G. Andrienko, N. Andrienko, O. Thonnard, and J. Zouaoui. Understanding user behaviour through action sequences: from the usual to the unusual. *IEEE Transactions on Visualization and Computer Graphics (in press)*, 2018.