User Behavior Map: Visual Exploration for Cyber Security Session Data

Siming Chen* Fraunhofer IAIS
Shuai Chen Peking University
Natalia Andrienko Fraunhofer IAIS / City, University of London
Gennady Andrienko Fraunhofer IAIS / City, University of London
Phong H. Nguyen City, University of London
Cagatay Turkay City, University of London
Olivier Thonnard Amadeus
Xiaoru Yuan Peking University

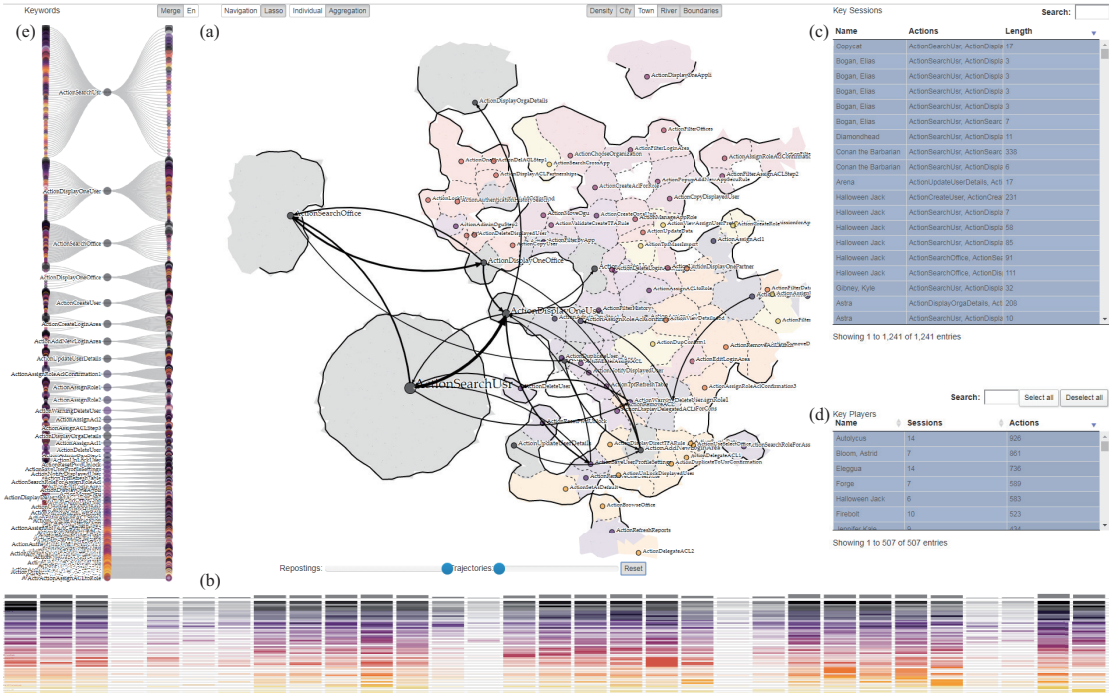

Figure 1: Visual analytics system for cyber security session data. (a) Behavior map. Actions are cities and people's action sequences (sessions) are mapped to trajectories. (b) Temporal visualization, (c) Session table view, (d) User view and (e) Sequential detail view.

Abstract

User behavior analysis is complex and especially crucial in the cyber security domain. Understanding dynamic and multi-variate user behavior are challenging. Traditional sequential and timeline based method cannot easily address the complexity of temporal and relational features of user behaviors. We propose a map-based visual metaphor and create an interactive map for encoding user behaviors. It enables analysts to explore and identify user behavior patterns and helps them to understand why some behaviors are regarded as anomalous. We experiment with a real dataset containing multiple user sessions, consisting of sequences of diverse types of actions. In the behavior map, we encode an action as a city and user sessions as trajectories going through the cities. The position of the cities is determined by the sequential and temporal relationship of actions. Spatial and temporal patterns on the map reflect behavior patterns in the action space. In the case study, we illustrate how we explore relationships between actions, identify patterns of the typical session and detect anomaly behaviors.

*e-mail:siming.chen@iais.fraunhofer.de. S. Chen is also with University of Bonn. X.Yuan is with Key Laboratory of Machine Perception (Ministry of Education), and School of EECS, Peking University.

Keywords: Behavior Analysis, Map Metaphor, Cyber Security

1 Introduction

We target on user behavior analysis in the cyber security domain. It is crucial to gain better situation awareness during the protecting, alerting, managing and fixing stages of the cyber security. Cyber security analysts need to identify typical user behaviors from a large amount of user behavior data, e.g. sequential session data. Such data contains multi-variate and dynamic behaviors. Specifically, there are two application research questions:

1. How to help analysts gain an overview of user behavior patterns, understanding the distribution and relationship among user actions?
2. How to interpret the results of anomaly detection and explain why the detected behaviors are suspicious?

Visual analytics involving domain experts' knowledge can help address on the above problems. Classical visualization approaches to understand user behaviors are based on the information about co-occurrence and sequential order of the performed actions. The advantage of these approaches is that they produce patterns of actions that are easy to interpret. However, it is difficult to identify the relationship among users and their behaviors. Recently, we proposed a map-like visual analytics approach to understand the event evolution in social media [4]. We observe the opportunities and desire for understanding the multi-variate and dynamic features of

978-1-5386-8194-7/18/$31.00 ©2018 IEEE

user behaviors in cyber security domains. In this paper, we propose to adapt ideas of the map-like visual analytics approach for enabling better situation awareness in security analysis.

Requirements from our domain experts are summarized into three tasks: understanding action distribution and relationship, identifying typical user behaviors and interpreting anomalous user behaviors. To support these tasks, we propose the User-Behavior Map, visually encoding the user behavior data into a map-like representation, which supports interactive exploration of action patterns and user behaviors. We map actions into cities and place nearby actions according to their sequential relationship and co-occurrence features. If two actions are often performed in a sequence and/or during the same period of time, they will tend to be in near places. Sessions are mapped as trajectories on the behavior map. With such a map, analysts can understand the multi-variate action distribution and interpret sequential relationships. Analysts can understand users' spatial-temporal behavior patterns.

2 Related Work

Cyber security is an important domain for applying visual analytics techniques, see a survey by Shiravi et al. [9]. Applications in network security include anomaly behavior analysis in netflow data [5], malware behaviors [11], etc. Techniques such as parallel coordinates [5], time series visualization [11] and graph visualizations [8] are commonly used in this domain. In our case study, we focus on user behavior analysis using sequential session data, for identifying user behavior patterns using a map-like visual metaphor.

User behavior visual analytics was developed in multiple domains such as trajectory analysis [2], social media [10] and security [7]. Previous works conducted temporal pattern analysis using sequential and time-series visualization [6]. We argue that map-like visualization can address the relational features in behavior analysis. Andrienko et al. proposed to construct a semantic space by projecting locations in behavior analysis [3]. Previously, we also proposed a new map-like visualization for event evolution analysis in social media [4]. With similar concepts, we design a behavior map and analyze user behaviors for security analysts. To the best of our knowledge, it is a first attempt to build a behavior map based on real data in the cyber security domain.

3 Method Overview

In this section, we introduce the motivating tasks, visual encoding, and our visual analytics system.

3.1 Data and Analysis Tasks

We define a *behavior* as a combination of *actions* performed by some agent during a period of time. Our case study uses a real data, Logon and Security Server (LSS) data, which is based on logs from a digital application where the user base of an organization is managed through user authentication, access control and sophisticated user rights for individuals and offices. The data is gathered in the monitoring logs and is stored in sessions. Our LSS dataset consists of 14,360 sessions performed by system operators in one month. Each session is considered as one behavior instance. A session consists of multiple actions, for example, "Search User", "Display One User", "Create Login Area", etc. There are 296 distinct actions. More details about the data can be found in [7]. Our domain experts collect, monitor and analyze the data. A probabilistic model was previously built to compute anomaly scores of sessions by our domain experts. We summarize the following analysis tasks.

- **Understanding action distribution and relationships.** Analysts are interested in what are typical actions and their relationship in the operation process.
- **Identifying typical session patterns.** Users with different functional responsibility will perform different behaviors. Analysts want to identify typical patterns of users.
- **Interpreting anomaly patterns.** Applying anomaly detection models, analysts detect anomaly behaviors but not sure why they are suspicious. They want to understand/interpret them.

3.2 Visual design

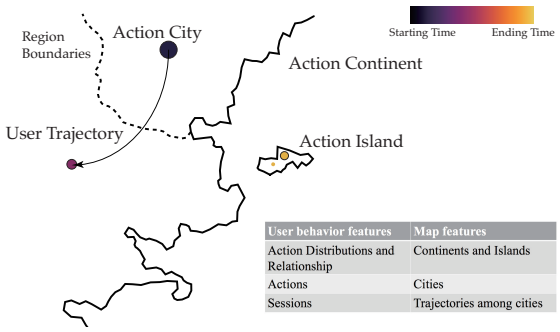

Figure 2: Visual mapping of the behavior map.

The core element of our visual design is the map metaphor. People are familiar with maps. Interactive maps provide intuitive interactions such as panning and zooming. The map metaphor helps analysts understand the abstracted behavior space. With the appropriate design and transformation from the abstracted space to the 2D map space, we help analysts understand and explore user behaviors.

We map each distinct action as a city (Figure 2). The city is surrounded by a region, the size of which indicates the frequency of the mapped action. The distance between cities encodes sequential and temporal relationships. If two actions are performed frequently in a consecutive manner or usually happen at relatively close times, they will be located closely on the map. The color of town symbols and city regions encodes relative times of the actions. Trajectories that represent user behaviors are encoded on the map as directed curved links connecting different cities. The thickness of a trajectory link encodes the number of the consecutive action pair. Details of algorithms for constructing the map are described in [4].

3.3 Interactive visual analytics system

We provide additional views to enable analysts to explore the behavior map (Figure 1). Analysts can gain an overview of the distribution of actions. Natural geographic map interactions such as panning and zooming are supported. Semantic zooming is also supported to show multiple levels of details. Analysts can brush a region of interest for exploring users who performed actions. Additional filters by sessions, users and times are applicable. Selected trajectories are visualized on the map (Figure 3). Thus, our system supports the exploration process from overview to details. The timeline shows the temporal distribution of actions (Figure 1b). In addition, we provide a detailed sequential view for showing predecessors and successors of the actions ranked by frequencies (Figure 1e). By iterating with the system, analysts can explore user behaviors.

4 Case Study

We conducted a case study with real-world data and analysis tasks, as described in Sec 3.1. Before applying our approach, the domain experts constructed a model for labeling sessions according to their anomaly. However, interpretation and explanation of model results, understanding user behavior patterns, and identification of false alarms were desired.

4.1 Action space analysis

In the overview on the map, we can immediately identify the top frequent actions, such as searching user, searching office, displaying

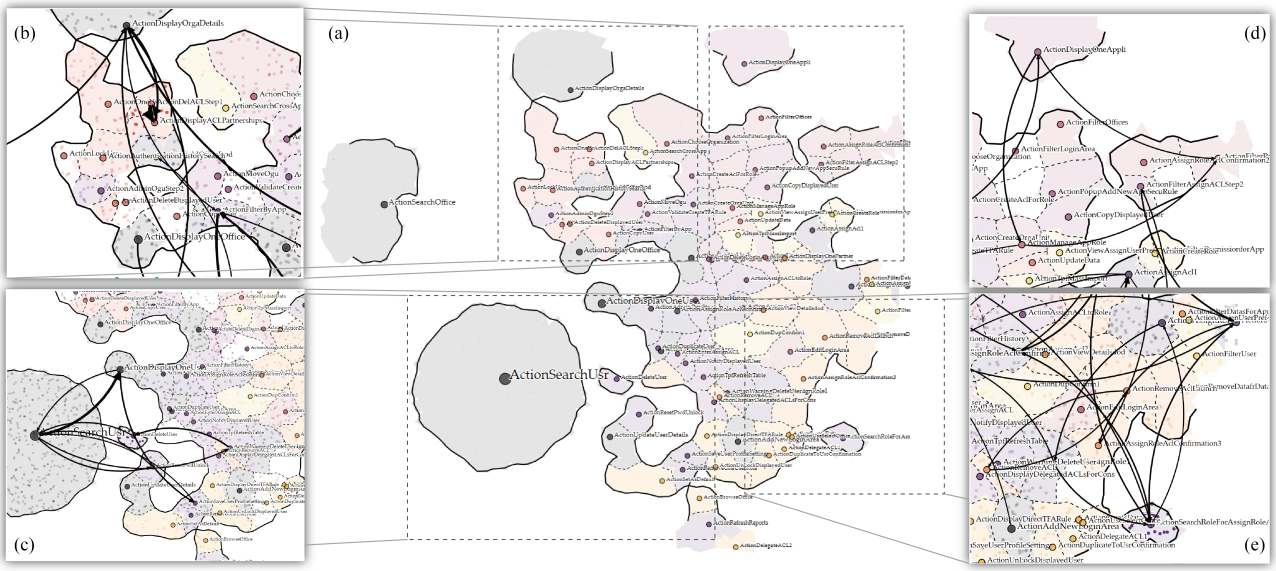

Figure 3: Overview (a) and detailed exploration (b-e) of action distributions and users' trajectories of behavior map.

organization details, etc. Moreover, we can identify the correlations among the actions (Figure 1). Frequent sequential patterns include sequences like search user - displaying user, search office - display office information, etc. One step further, we find an important role of the action *ActionDisplayOneUser*, which serves a bridge between search and subsequent actions such as deleting, editing and updating. Visually, it is positioned in an important peninsula connecting to the big islands and the mainland (Figure 3b). However, if we directly observe the statistics about the amount of *ActionDisplayOneUser*, it is relatively small because displaying only happens when search successfully hits the results. With the map-like metaphor, we can help analysts identify not only frequent actions but also "hub actions".

For progressing from the overview to details, the analyst can interactively zoom and pan to explore different regions of the map, thus focusing on behaviors of interest. In the north-west part of the map (Figure 3b), users starts with *ActionDisplayOrgDetails* and check the following actions. Interestingly, we find that *ActionDelACL* and *ActionDisplayACLPartnerships* usually happen sequentially in a loop. In the south-west part (Figure 3c), besides the important role of the *Display* action, we also identified a frequent behavior pattern consisting of "*Search*, *WarningDelete*, *Delete* and *Search*". In the north-east part (Figure 3d), frequent pattern is "*Display one application*, *Manage App role*, *Filter App role*". In the south-east part (Figure 3e), *ActionSearchRoleForAssignRoleACL* is a hub action.

4.2 Typical session analysis

With the overview, we can use the sequential view for examining details of the selected sessions. For example, we can identify the loop of "display office, search office and display office" session patterns (Figure 1). Different distributions of trajectories and focus regions on maps of different users indicate different types of their behaviors. The map acts as a profiling function for typical user behaviors. For example, in the east part of the map, there are two typical sessions. In Figure 4a, we identify a loop of *ActionFilterDataForApp* and *ActionCreateData* as a frequent pair happened 15 times. We can also trace more complicated session, such as Figure 4b. The session started from "search user" and "display user". Once the agent found a user of interest, he started to assign roles and ACL confirmation to him and repeated such operations several times. Thus, we can see a shift patterns on the map, and the arrows on the east part of the map emphasize the frequency. Furthermore, we can explore and identify the focusing regions of specific users using aggregated sessions. For example, the selected user is responsible for user editing, including role setting, login data editing, etc. All of his 8 sessions are focusing on the highlighted regions (Figure 4c).

4.3 Anomaly behavior interpretation

Analysts need to examine, interpret and explain detected anomalies, check whether they are false alarms and understand why they are suspicious. Based on the understanding of anomaly patterns, analysts can further improve their model and conduct protective actions. Analysts can identify the consistency of their behaviors among all sessions. If a user starts doing something unusual in respect to his prior session, it might be suspicious. With the trajectories on the map, the analyst can see the distribution patterns for users and thus examine and explain the resulting models. Using such procedure, the analyst studied two users focusing on office-related operations. For such users normal operations are mainly on searching office, displaying organization details and editing organization information, etc. Respectively, normal trajectories of all sessions are distributed in local areas of the map (Figure 5a). A potentially anomalous user behavior (as detected by the experts' model) not only conducts office-related actions but also performs actions that are located in many areas of the map (Figure 5b). Such pattern attracts the attention of the analyst. After exploring the results, he confirms that such behavior is suspicious because the user performs activities that are characteristic for different roles: regular user, office, organization stuff. Normally, users are not supposed to mix roles and conduct so diverse actions. Thus, the suspicious user should be monitored.

The case study demonstrates that our approach enables the analysts (One of the co-authors) to gain situation awareness of the action distributions and their relationships. We present our results to them and are verified by them. Progressing from the overview to details, the analysts can gain further insights of frequent patterns and different roles of the users and interpret their behaviors. Such analysis helps analysts to improve their original models.

5 Discussion and Conclusion

In this work, we propose and apply the map visual metaphor for user behavior analysis in cyber security. We address the required tasks and confirm the capability of exploring user behaviors. However, there are several issues to be solved in the future. First, the current implementation is not scalable in respect to the number of cities (actions). We consider adding hierarchical organization to resolve

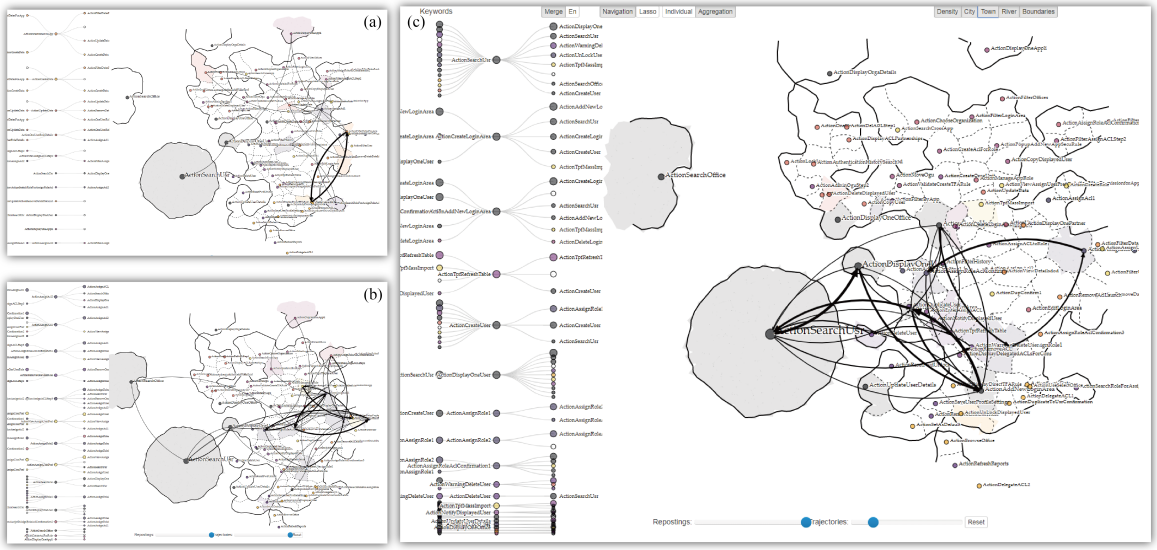

Figure 4: Selected typical individual sessions for exploring user behaviors (a, b). Aggregated multiple sessions of typical user behaviors (c).

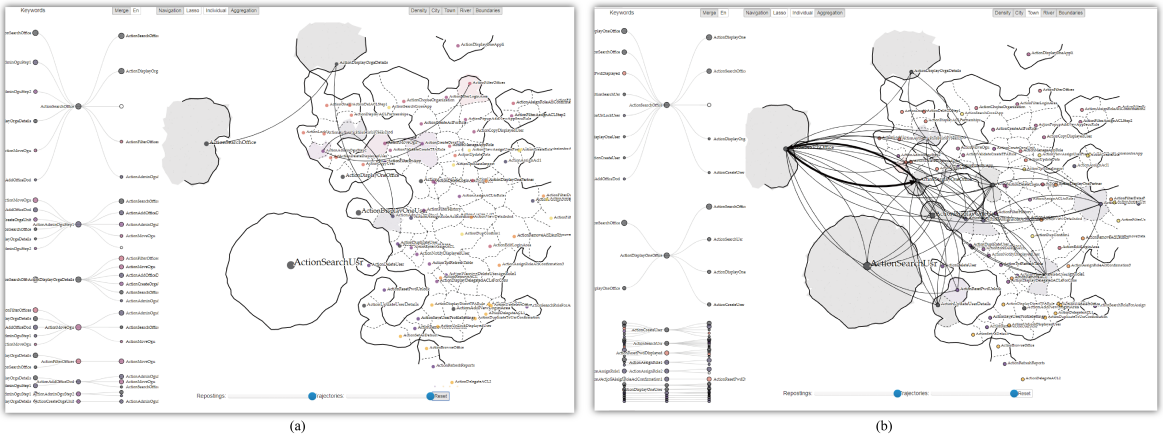

Figure 5: (a) Typical behaviors for office-related management operations. (b) Atypical behaviors visualization.

the scalability issues. Second, we aim at supporting the exploration of large amounts of data in streaming settings. We also consider adopting further methods of visual analytics of movement [1] for enhancing the capabilities and thus supporting further tasks.

Acknowledgments

This research was supported by National Key Research and Development Program of China (2016QY02D0304) and National Nature Science Foundation of China (Grant No. 61672055). This research was also supported by Fraunhofer Cluster of Excellence on "Cognitive Internet Technologies" and by EU in project DiSIEM (700692). We thank Michael Kamp from Fraunhofer IAIS. We also thank PKU-Qihoo Joint Data Visual Analytics Research Center.

References

[1] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel. *Visual Analytics of Movement*. Springer, 2013.

[2] N. Andrienko, G. Andrienko, L. Barrett, M. Dostie, and P. Henzi. Space transformation for understanding group movement. *IEEE TVCG*, 19(12):2169–2178, 2013.

[3] N. Andrienko, G. Andrienko, G. Fuchs, and P. Jankowski. Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces. *Information Visualization*, 15(2):117–153, 2016.

[4] S. Chen, S. Chen, L. Lin, X. Yuan, J. Liang, and X. L. Zhang. E-map: A visual analytics approach for exploring significant event evolutions in social media. In *Proc. of IEEE VAST*, 2017.

[5] S. Chen, C. Guo, X. Yuan, F. Merkle, H. Schaefer, and T. Ertl. Oceans: Online collaborative explorative analysis on network security. In *Proc. of VizSec*, pages 1–8, 2014.

[6] F. Du, B. Shneiderman, C. Plaisant, S. Malik, and A. Perer. Coping with volume and variety in temporal event sequences: Strategies for sharpening analytic focus. *IEEE TVCG*, 23(6):1636–1649, 2017.

[7] 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 TVCG*, 2018.

[8] S. Peryt, J. A. Morales, W. Casey, A. Volkmann, B. Mishra, and Y. Cai. Visualizing a malware distribution network. In *Proc. of IEEE VizSec*, pages 1–4, Oct 2016.

[9] H. Shiravi, A. Shiravi, and A. Ghorbani. A survey of visualization systems for network security. *IEEE TVCG*, 18(8):1313–1329, 2012.

[10] G. Sun, Y. Wu, S. Liu, T.-Q. Peng, J. J. Zhu, and R. Liang. Evoriver: Visual analysis of topic coopetition on social media. *IEEE TVCG*, 20(12):1753–1762, 2014.

[11] M. Wagner, W. Aigner, A. Rind, H. Dornhackl, K. Kadletz, R. Luh, and P. Tavolato. Problem characterization and abstraction for visual analytics in behavior-based malware pattern analysis. In *Proc. of VizSec*, pages 9–16. ACM, 2014.