Doing User Behaviour Analytics through Interactive Visual User Profiles

Phong H. Nguyen* City, University of London, UK
Siming Chen† Fraunhofer IAIS
Natalia Andrienko‡ Fraunhofer IAIS / City, University of London
Gennady Andrienko§ Fraunhofer IAIS / City, University of London
Olivier Thonnard¶ Amadeus, France
Cagatay Turkay‖ City, University of London, UK

ABSTRACT

User behaviour analytics (UBA) systems offer sophisticated models that capture users' behaviour over time to identify fraudulent activities that do not match with their profiles. Our paper presents a visual analytics approach to build a comprehensive, multi-faceted understanding of user behaviour. We observe that with the aid of interactive visual interfaces, analysts are able to conduct exploratory and investigative analysis effectively.

1 INTRODUCTION

Fraudsters in online systems are using increasingly more sophisticated and complex approaches that are becoming challenging to identify such as by disguising themselves as legitimate users and mimicking users' activities to stay under the radar inside the system [1]. To identify such insider threats, computational UBA solutions aim to learn probabilistic models of users' behaviour through their past activities, and trigger alerts when users start behaving in unusual, unexpected ways. However, they are limited in providing in-depth views into users' behaviour, and valuable analyst time is being lost in identifying the causes of the issues and deciding on whether the alerted signals are indeed problematic cases.

In this paper, we propose a visual analytics approach to facilitate effective decision making for cyber security analysts using UBA systems through a multi-faceted understanding of user behaviour. We take a user-centred approach and characterise the problem domain thorough a study of the challenges, goals, and the analytical tasks within the context of user behaviour analytics systems. We then design a visual analytics framework that encompasses a multi-faceted interactive visual user profile to help investigate the various perspectives of user behaviour concurrently. We demonstrate the efficacy of our approach through a number of use cases conducted as a multidisciplinary team and discuss the limitations observed.

2 DOMAIN CHARACTERISATION

In this paper, we work with a UBA model that uses the logs recorded within a login and security server. The log data is split into sessions, each containing an ordered list of timestamped actions performed by the user conducting that session, and meta-information such as IP address, user ID and organisation. We also get an *anomaly score* that is provided by the underlying UBA. Actions in this context are semantically labelled such as "*CreateLoginArea*", "*SearchUsr*" and "*DisplayOrgaDetails*". The dataset comprises of *18,956 sessions* performed by *1,666 users* during a *15 day* time window, with *305 unique action types*.

*e-mail: p.nguyen@city.ac.uk
†e-mail: siming.chen@iais.fraunhofer.de
‡e-mail: natalia.andrienko@iais.fraunhofer.de
§e-mail: gennady.andrienko@iais.fraunhofer.de
¶e-mail: olivier.thonnard@amadeus.com
‖e-mail: Cagatay.Turkay@city.ac.uk

We take a user-centred approach to the design and development of our solutions. To understand the application domain, we conducted a series of workshops with analysts, who all have considerable experience with these types of investigation mentioned above. Based on these observations, we identified analytical tasks, designed and implemented initial versions of the solutions, presented them to the analysts in follow-up sessions, and gathered feedback to iteratively improve the designs.

T1 Identifying users of interest.
T2 Understanding the different facets of user behaviour.
T3 Investigating patterns and deviations in user behaviour.
T4 Putting sessions in the context of a user.

3 VASABI: VISUAL ANALYTICS FOR UNDERSTANDING USER BEHAVIOUR

3.1 Building User Profiles

We categorise features of a profile into three broad categories:

- *Inherent features* that describe the low-level characteristics of how users utilise the application. Examples: meta-information associated with each session such as IP address and browser.
- *Derived features* that are extracted through straightforward statistical computation. Examples: the number of sessions and the number of unique actions.
- *Latent features* that provide an in-depth understanding into the behaviour of users through the application of a sophisticated computation. Examples: UBA modelling score.

Users' behaviour in digital systems are largely determined by different *roles* that they have in their organisations. Each role could be responsible for a few particular tasks. We consider *task* as a latent feature of user profile and apply a text-based clustering analysis to extract the tasks automatically. Each action is mapped to a word and each session is mapped to a document. A k-means clustering ($k = 10$) is used to extract the clusters, i.e., tasks.

3.2 Visualising User Profiles

We design two views, one for exploring multiple facets of user profile and one for investigating performed tasks in more detail.

3.2.1 Profile View

It is essential for analysts to observe many facets of user profile concurrently (**T2**). For each facet, we use a standard chart to show a summary of user sessions and composite the charts to make a compact representation of a user profile. The visual profiles are stacked together to enable comparison. Analysing anomalous sessions within the context of the users performing them helps identify the deviation of user behaviour from the norm (**T3, T4**). We support this analytical task by superimposing the sessions of interest, as small orange circles, on top of the visual summary. A random noise is added to the vertical position of the dots to avoid overplotting. Mouse hovering a session highlights the same session in other facets, enabling to observe different facets concurrently.

3.2.2 Task Overview

Each task, or cluster, can be represented by five dominant actions within a cluster. A task is shown as a set of five squares, each for

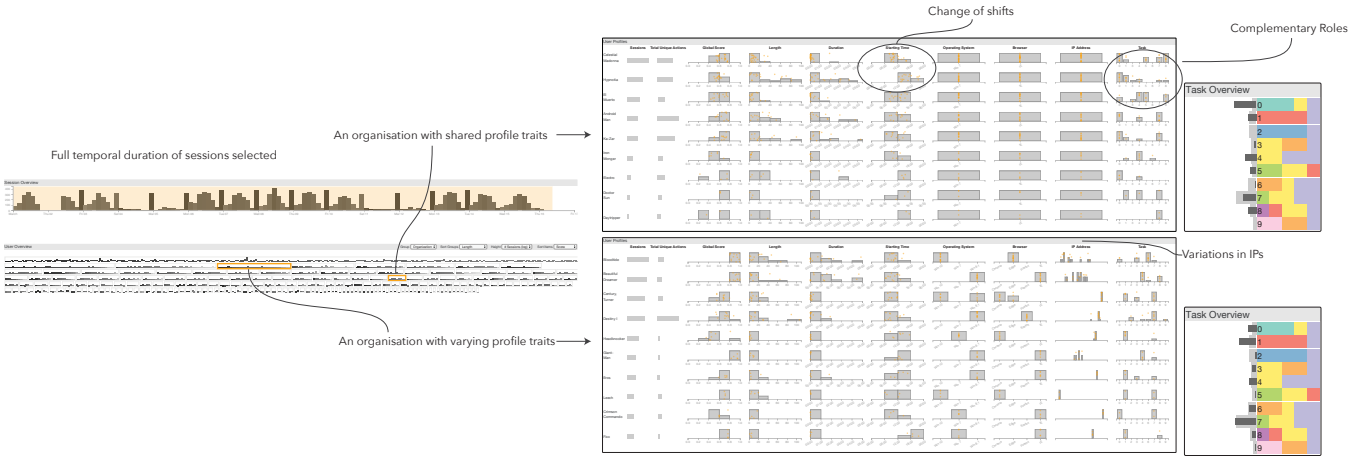

Figure 1: Our visual analytics approach VASABI as used for behaviour analysis on users from two organisations. The organisations with different characteristics are investigated together with the users working within them. Using the full temporal duration of the sessions (top-left corner: **Session Overview**), one organisation is identified as having common behaviour traits across its members, whereas the other with varying characteristics. In the first organisation (selected through the **User Overview** at the bottom-left corner), all users seem to use the same systems and share IPs (observed in the **User Profile view** at the middle). In the second, users have unique IPs and are more varied in terms of the systems they use. Looking closely, interesting patterns emerge: what could potentially be a shift change between two employees is observed in the top organisation with two users (*Celestial Madonna* and *El Muerto*) having complementary roles. **Task Overviews** (two views on the right) indicate the variations within the computationally extracted tasks as carried out in the organisations.

an action, coloured based on the action's group. This grouping is done semi-automatically based on a t-sne projection of a word2vec transformation. The left of the view shows two statistics: the distribution of tasks in the entire dataset (lighter bars) and the distribution of tasks within a selection of interest, such as sessions performed by a particular user in a given time range (darker bars). This enables comparison across tasks within a user, or two sets of sessions (**T3**).

3.3 Visual Analytics Environment

The two User Profile views are integrated into a visual analytics environment for facilitating exploration of user behaviour. The environment is described in our previous papers [2, 3], including the following three views:

- *Session Overview* provides a temporal distribution of sessions in the entire dataset.
- *User Overview* explores sessions of interest (selected in the Session view) with user and organisation perspectives.
- *Session Timeline* shows actions along a time axis according to when they occur.

4 Evaluation

We conducted two one-hour long sessions with domain analysts to understand how they would use VASABI to perform their analyses. Here, we describe one of the use cases that were identified during the evaluation session: *Organisational working patterns*. The goal of this use case is to observe the variations in the behaviour profiles across the users within organisations.

The analysis starts with the analyst selecting the whole time range of the data in the Session Overview due to goal of identifying overarching behaviour (Fig. 1). Two different organisations are selected in the User Overview (figure edited here to concurrently display two different organisations). The first organisation is observed to have highly regular behavioural traits. The users in this organisation often work with similar operating systems, browsers and use a shared IP across the organisation (**T1**). This is contrasted with another organisation where these traits are less regular and users using a wider range of IPs and browsers. This implies that the models built for these organisations need to consider these characteristics, e.g., a deviation from the regular IP should be a more significant source of unexpectedness for the first organisation compared to the other (**T1**). Another interesting observation is what might be referred to as a *shift change* between the two most active users (Fig. 1, top middle) in the first organisation by observing their working routines – aspects that can be factored in for a more sophisticated user model (**T2**, **T3**). Observing the per-user and per-organisation task profiles also reveals a further insight. Some users in the first organisation – *Celestial Madonna* and *El Muerto* seem to have complementary roles and perform tasks that are visibly different (see the difference in the two Task Overviews).

5 Conclusion

In this paper, we investigate how a visual analytics approach can provide a comprehensive, multi-faceted understanding of user behaviour to facilitate effective decision making in UBA-enabled cyber security systems. Through a preliminary evaluation, the analysts found the VASABI tool help them perform real-world tasks using their analytical strategies, explain cases that would not otherwise be possible, and reduce their analysis time. We also identify limitations and rooms for improvement including restricted analysis workflow and several visualisation design issues.

Acknowledgments

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

References

[1] J. Care and T. Phillips. Market guide for online fraud detection. `https://www.gartner.com/doc/3849295/market-guide-online-fraud-detection`, January 2018.

[2] 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.

[3] 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.