Mouse Clicks Can Recognize Web Page Visitors!

Daniela,Chudá, Peter Krátky, Jozef Tvarožek
Faculty of Informatics and Information Technologies
Slovak University of Technology
Ilkovičova 5, 842 16 Bratislava, Slovakia
{daniela.chuda, peter.kratky, jozef.tvarozek}@stuba.sk

ABSTRACT

Behavioral biometrics based on mouse usage can be used to recognize one's identity, with special applications in anonymous Web browsing. Out of many features that describe browsing behavior, mouse clicks (or touches) as the most basic of navigation actions, provide a stable stream of behavioral data. The paper describes a method to recognize Web user according to three click features. The distance-based classification comparing cumulative distribution functions achieves high recognition accuracy even with hundreds of users.

Categories and Subject Descriptors

H.4.m [**Information Systems Applications**]: Miscellaneous

Keywords

Biometrics; user recognition; mouse dynamics; mouse click features; mouse movement features

1. INTRODUCTION

Differences in how users control computer have led to a new branch in biometrics - behavioral biometrics based on standard input devices. Usage of keyboard, computer mouse or touch screen has been studied from different points of view, such as latencies between pairs of pressed keys (keystroke dynamics), mouse movement speed, curvature, and other mouse dynamics. However, one thing has been mostly overlooked - a simple *mouse click*.

Similarly to other biometric characteristics, click features could contribute to methods dealing with verification or recognition of a user's identity. Especially, in the Web environment, where *point and click* is a basic navigation action. The methods could be used for effective recognition of persons accessing the Internet via shared computer, re-authentication of Web accounts, or abused for recovering deleted cookies.

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2742749.

In our work, we propose a new method for user identity recognition based on mouse clicks features using comparison of cumulative distribution functions. It shows accurate performance when identifying a user in a small user base, as well as when hundreds of users are present. The method can recognize an individual user out of 100 users with 85% success rate based on only 50 clicks. The success rate improves when more clicks are available, while it declines only slowly when more users are present.

2. RELATED WORK

Various features of mouse movements were studied in the literature papers, while clicks or double-clicks are incorporated only marginally. Pusara and Brodley [4] calculated distance, angle and velocity between two clicks. Gamboa and Fred [2] used time between click and last movement event referred as *pause to click*. However, the performance of the features is not stated there. Zheng et al. [5] found that pause to click feature is sensitive to hardware replacement significantly. Instead, three stable angular metrics (features) for mouse movement that work well for verification are presented - direction, angle of curvature, distance of curvature.

The research papers dealing with this kind of biometrics are mostly devoted to identity verification problem (proving one's identity). In the work by Monrose and Rubin [3], identity recognition (finding one's identity) is discussed. The success rate of identification based on free text typing on keyboard is 23% (91% when predefined text sample is used). To classify identity, (symmetric) distributions of feature values were compared. In our previous work [1], identification based on five mouse features (including duration of mouse click) reached 75% success rate using statistic of t-test to compare feature distributions.

3. METHOD

The proposed method assigns identity to a user according to values of three mouse click features - *pause to click* (as stated above), *click duration* (the delay between mouse button press and release), *pause after click* (the delay between click and the following movement).

In our method, user's behavior is represented as a set of measured samples. The classification algorithm is based on measuring distance between user models, meaning that the space of users is searched for the one with lowest distance to the user whose identity is unknown. This approach ensures that the retraining phase is avoided when the user provides new data or a new user comes in. The click features follow a very skewed distribution therefore the standard distance

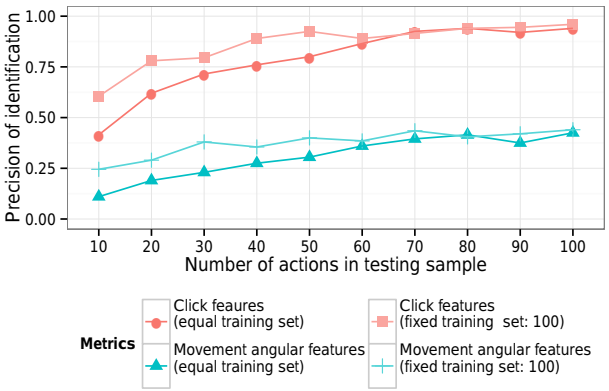

Figure 1: Performance comparison of click and movement features.

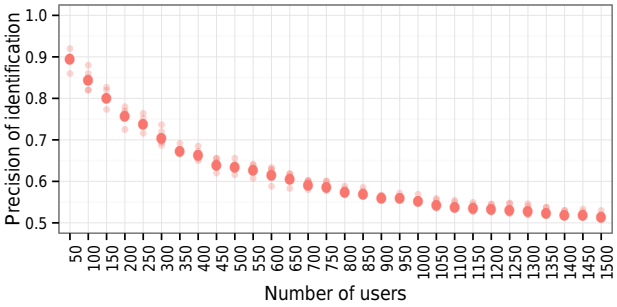

Figure 2: Influence of user pool size on performance.

measures based on mean and deviation are insufficient. We employ statistic of Kolmogorov-Smirnov test to the measure difference of two cumulative distribution functions and the distance between two user models is the sum of the differences for each feature.

4. EXPERIMENTS AND RESULTS

We evaluated the method on two datasets: a controlled experiment with 20 users, and a large scale deployment with 180700 users. To collect data, we implemented a logging (JavaScript) module that records user interface events, including mouse events (mouse down/up, mouse movement) in browser and sends them to server logging backend.

The first dataset was collected in a controlled experiment with 20 users. The users were asked to play a browser memory game, in which cards had to be flipped by mouse clicks. Each of them provided more than 200 click events in ten minutes of game play using the same hardware.

The results in Figure 1 demonstrate that mouse click features are suitable candidates for user recognition. We compared performance of our classification method on the click features and the three features adapted from [5]. The angular features were calculated within each movement stroke (movement bounded with clicks). Initially, we used our method to calculate success rate for both sets of features on exclusive training and testing sets of equal sizes. Additionally, evaluation on a fixed training set (100 samples) was used. The success rate (average of 10 iterations of 2-fold cross validation) using 100 clicks is 96%, while the success rate when using 100 movement strokes is only 44%.

For the second dataset, we collected data from a large Central European tourism portal during a period of 21 days. Total of 2286638 clicks from 180700 unique users (12.7 clicks per user) were collected. Click counts per user followed a long-tailed distribution. For evaluation, we selected users with at least 100 clicks during the collection period, resulting in a subset of 1523 users. For each user, 100 clicks selected randomly were divided into two equal sets for 2-fold cross validations: 50 clicks for training, and 50 clicks for testing (user identification).

The Figure 2 shows how identity recognition success rate (average of 5 iterations of 2-fold cross validation) depends on the size of the user pool. Based on 50 clicks, the recognition accuracy of a user in a pool of 100 users is 85%. On the far end, the accuracy declines slowly. To recognize an unknown user in a pool of 1500 users the success rate is still over 51%. Mean time between two clicks in a session was approximately 23 seconds, requiring a total of 19 minutes to achieve the stated accuracies.

5. CONCLUSIONS

In the paper we present a novel approach to user identity recognition on the web. It shows experiments results benchmarking the classification algorithm based on distance measure adapted from Kolmogorov-Smirnov non-parametric test. The method could be used for accurate recognition of users in small groups (e.g. sharing the same computer) to improve user-oriented services. Additionally, it enables relatively accurate recognition of a user in user pools containing several hundreds of users.

6. ACKNOWLEDGMENTS

This work was partially supported by the grant No. VG 1/0752/14 and grant No. VG 1/0646/15.

7. REFERENCES

[1] D. Chudá and P. Krátky. Usage of computer mouse characteristics for identification in web browsing. In *Proceedings of the 15th International Conference on Computer Systems and Technologies - CompSysTech '14*, pages 218–225, New York, New York, USA, 2014. ACM Press.

[2] H. Gamboa and A. Fred. A behavioral biometric system based on human-computer interaction. In *Proceedings of SPIE 5404, Biometric Technology for Human Identification*, pages 381–392, 2004.

[3] F. Monrose and A. Rubin. Authentication via keystroke dynamics. In *Proceedings of the 4th ACM Conference on Computer and Communications Security*, pages 48–56, 1997.

[4] M. Pusara and C. E. Brodley. User re-authentication via mouse movements. In *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security - VizSEC/DMSEC '04*, pages 1–8, New York, New York, USA, 2004. ACM Press.

[5] N. Zheng, A. Paloski, and H. Wang. An efficient user verification system via mouse movements. In *Proceedings of the 18th ACM conference on Computer and communications security - CCS '11*, pages 139–150, New York, New York, USA, 2011. ACM Press.