Sheffield Hallam University

An empirical biometric-based study for user identification from different roles in the online game League of Legends

DA SILVA, VR and DA COSTA ABREU, Marjory <http://orcid.org/0000-0001-7461-7570>

Available from Sheffield Hallam University Research Archive (SHURA) at:

http://shura.shu.ac.uk/25396/

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version

Copyright and re-use policy

An empirical biometric-based study for user identification from different roles in the online game League of Legends

Valmiro Ribeiro da Silva, Márjory Da Costa-Abreu[1]

[1]Departamento de Informática e Matemática Aplicada
Universidade Federal do Rio Grande do Norte (UFRN)
Natal – RN – Brazil

`marjory@dimap.ufrn.br`

Abstract. *The popularity of computer games has grown exponentially in the last few years. In some games, players can choose to play with different characters from a pre-defined list, exercising distinct roles in each match. Although such games were created to promote competition and promote self-improvement, there are several recurrent issues. One that has received the least amount of attention is the problem of "account sharing" so far is when a player pays more experienced players to progressing in the game. The companies running those games tend to punish this behaviour, but this specific case is hard to identify. The aim of this study is to use a database of mouse and keystroke dynamics biometric data of League of Legends players as a case study to understand the specific characteristics a player will keep (or not) when playing different roles and distinct characters.*

1. Introduction

Online games have become very popular and diverse since their beginning in the 80's, and each device gives us several biometric modalities to be exploited, such as gait, keystroke dynamics, mouse dynamics, touch-screen dynamics, etc. The diversity of input data is endless and it makes the game an unique experience for the player.

Even though all the previously listed biometrics are used for the same purpose in the gaming universe, they are fundamentally very different when analysed in the traditional security and authentication applications. Thus, if we intend to investigate the identity predictability of game users using these modalities, it is important to understand its differences from the traditional approaches[da Silva Beserra 2017, Camara 2017].

As a very simple example of how different the security application is from a traditional authentication task to a game authentication using biometric data, take it the keystroke dynamics modality in a continuous verification scenario:

- In a traditional verification problem, the user's behaviour is expected to suffer very little variation while typing an e-mail.
- On the other hand, in a game verification problem, the user's behaviour is expected to change and that change will be based on the configurations of the game he/she is using, e.g. the role in the game, the abilities it chose to use, the character he/she is playing with and so on.

The second case can not be considered the same as the first, because, despite the fact they are both verification problems and are using the same base data, the user's

behaviour is different which makes the security system to model it in a different way. To the best of our knowledge, no other work has tried to investigate this specific problem.

Thus, this paper aims to investigate what are the real differences (or if there really are any) in biometric behaviour when the security problem changes from the traditional systems to the gaming universe. We have chosen to investigate some users in the online game League of Legends with the idea of analysing keystroke dynamics and mouse dynamics data from the same users playing with different characters and roles in the game.

2. Biometric modalities used in desktop based game playing

The market of egames is huge and, with the recent advancement of virtual reality, the range of consoles (the hardware you need to play games) has increased greatly. The kinds of devices used to play go from simple keyboard and mouse to very expensive virtual reality glasses. However, the most popular is still the computer-based one for it is indisputable the cheapest [da Silva Beserra et al. 2016].

In a security point of view, each different kind of console will have different vulnerabilities, but the "black-box" type, the ones you buy and does not need to install any software are, until certain extent, more secure. When we are talking about private computer-based games, we have a limitation in devices that we can use to play, but the possibility of a user to play with another user's account is more evident.

Considering that we are using League of Legends as our case study, the modalities chosen to investigate are mouse and keystroke dynamics, because both peripherals are mandatory used together during the matches.

Keystroke dynamics is the unique timing patterns embedded in an individual's typing and is most often developed in a personal way, hence the use of keyboard dynamics as a biometrics-based identification modality. Processing of such data includes extracting keystroke timing features such as the duration of a key press and the time elapsed between successive key presses [Bergadano et al. 2002, Banerjee and Woodard 2012].

Mouse dynamics is the unique speed movements and frequency of clicks generated by a user using the mouse. The move speed is how fast the user moves the mouse in the 8 possible mapped directions and frequency of clicks is the amount of clicks the users performs in a time interval [Bours and Fullu 2009].

Since we are using League of Legends as our case study, it is important to understand how keystroke and mouse dynamics are used in the context of the game. The next subsection will introduce the basics of the game, as well as how our modalities can be used in context, followed by subsections enumerating the related work to mouse and keystroke dynamics.

2.1. League of Legends

League of Legends is a Multiplayer Online Battle Arena (MOBA) game. The game is based around matches with two teams of (normally) five players each, where each team tries to destroy the main base of the other. Before each match starts, each player chooses a champion to play, which is an avatar that already exists in the game, with predetermined statistics (stats) and skills. Two players of the same team cannot choose the same champion.

2

Each champion has four unique skills, where three are common skills and the last is a ultimate skill. Skills can grant passive or active abilities, where each skill is activated by the keys 'Q', 'W', 'E' and 'R', the last one used to activate the ultimate ability.

Each team member follows one of the defined roles during a match:

- Top Laner: know as "top", this player starts at the top of the map, and usually is a melee attacker;
- Jungler: This player spends most of his time defeating jungle's monsters in order to gain bonus statistics to the team. Champions with high mobility usually take this role;
- Mid Laner: Also known as mid, this player starts in the middle of the map, and uses his skill set to create combos to deal damage. Champions with synergic skill set usually take the role;
- Carry: also known as ADC (attack damage carry) is responsible take down buildings and clear minions waves. Ranged champions with high damage usually take the role.
- Support: Starts in the bottom lane and is responsible to support the ADC ans later the whole team. Champions with good supporting abilities or tanks usually take the role;

The main features used in League of Legends selected for the analysis of this paper from both keyboard and mouse dynamics can be described as follows:

- Keystroke dynamics: 'Q', 'W', E', 'R' (for the unique skills) and 'SPACE' (used to make the camera follow the player's champion);
- Mouse dynamics: Move the character (point and click in a empty space using the right mouse button), basic attacks (clicking in a enemy using the right mouse button) and target skills (using the left mouse button);

As already said previously, to the best of our knowledge, there is no other work which has investigates the individual variations of keystroke dynamics and mouse dynamics (biometrics) in the context of online games. Section 2.2 will present the main works that can be found using mouse and keystroke dynamics.

2.2. Keystroke dynamics, mouse dynamics and game-related work

Keystroke dynamics is a much older biometrics modality than the mouse dynamics, thus the number of databases available is larger. This is expected because the use of "mouse" is very much associated with the personal computer whereas the keystroke exists since the use of Morse code.

In [Idrus et al. 2013] a database containing soft-biometrics and keystroke dynamics from 110 volunteers from France and Norway is presented, where users were classified using Support Vector Machine (SVM) with an EER (Equal Error Rate) of 21% to 4% when the soft-biometrics were added. The work was expanded in [Idrus et al. 2015] using a fusion approach, where the SVM algorithm was used to classify the fused data, with an EER of 10%.

In [Lv et al. 2008] a new approach to emotion recognition using pressure sensor keyboards was described. Fear, happiness, anger, sadness, surprise and neutral emotions

were tested, with a EER of 12.02% when using only traditional keystroke methods to classify the subjects with the KNN algorithm.

In [Thanganayagam and Thangadurai 2015], a database using various fusion approach on keystroke dynamics was collected. Each user was allowed to choose their preferable username and password during the enrolment process and they were asked to type one fixed text for fifteen consecutive times, with an EER of 9% using SVM and combining features.

A login method for accessing computer systems using mouse dynamics was described by [Bours and Fullu 2009]. 28 users performed a fixed task of moving the mouse between two lines. They were classified using Levenshtein distance to calculate similarities, also know as edit distance, with an EER of 26.8%.

Pattern-growth-based mining was used to extract frequent behavior segments in obtaining stable mouse characteristics in [Shen et al. 2012], using classification algorithms to perform continuous user authentication. 22 users performed Internet surfing, word processing, online chatting and programming for 30 minutes. The best result was an EER of 1.49% using a One-Class SVM detector.

The literature does not have a large amount of multimodal systems using mouse and keystroke biometric data. Additionally, work related to online games are very limited.

An approach using game-play activities was proposed in [Chen and Hong 2007] with the purpose to attack the account sharing problem, where the idle time distribution of a player in-game was proved to be a representative feature, and the RET scheme was proposed for user identification, which is based on the Kullback-Leibler divergence between idle time distributions. The results showed that the RET scheme achieves higher than 90% accuracy with a 20-minute detection time given a 200-minute history size.

According to [Yampolskiy and Govindaraju 2006], the behavior of a player in a match can be used as a metric for identification in some cases. The authors used poker as case study, calculating the percentage of folds, calls, checks, raises, re-raises and all-ins, using euclidean distance to calculate similarity to verify 30 players identities, with an EER of 22.67%.

For this work we have used the biometrics database collected in [da Silva Beserra et al. 2016] using League of Legends as case study. Data from 56 different users were collected, using the same type of keyboard and mouse to all volunteers, where 18 users played more than one time, sometimes using different characters and/or positions. Our analysis will focus on this group.

The goal in [da Silva Beserra et al. 2016], and later in [da Silva Beserra 2017] and [Camara 2017], was to use the database for identification. For this purpose, the software WEKA was utilised in order to run machine learning algorithms trying to identify correctly each user. The best result combining keystroke and mouse dynamics in [da Silva Beserra 2017] and [Camara 2017] was 90.77% using the Random Forest algorithm, as shown in both works.

Each sample collected in these works have data of 33 different features:

- 13 keystroke features:
    - Three combination of keys, using the distance between keys, C1(Q W, W E, E R), C2(Q E, W R) and C3(Q R) , also called combos;
    - Frequency (per minute of match) for each key pressed (FQ, FW, FE, FR and FSPACE);
    - Latency for each key pressed (Q, W, E, R and SPACE).
- 20 mouse features:
    - Move speed of the 8 directions - 'Down', 'Down + Left', 'Left', 'Up + Left', 'Up', 'Up + Right', 'Right' and 'Down + Right - represented by D1, D2, D3, D4, D5, D6, D7 and D8, respectively;
    - The acceleration for each direction, represented by AD1, AD2, AD3, AD4, AD5, AD6, AD7 and AD8, respectively;
    - Frequency and Latency for right and left clicks, represented as CFR, CFL (for frequency) and CTR and CTL (for latency).

3. Experimental and statistical analysis

Numerous online games are marketed around the idea of "different characters for different people", and these games lead the players to do one of these things:

- Always pick the same character for every match, or;
- Pick a different character for each match.

The first point may imply that using the same character for every match leads the players to stay in the same role, but that not always true. A similar assumption could be made about the second point, inferring that changing a character always changes the players' role is not correct. For both cases, this can lead us to the idea that characters defines how the player behave, but this idea may not be true when we are dealing with biometrics. As suggested in [Leavitt et al. 2016] we can not assume that a player can be represented by the characters he/she uses.

Table 1 shows the 18 users who played more than one match in our database, the champions they used and the roles they played. **Users 20 and 48 are the most representative**, because they played with at least four of the five roles and played every match with a different character. The other users in this group also have a good value because some of then changed roles between matches, while the others remained in the same role, even when they changed characters. User 55, for example, played both matches in the top lane, first using a melee tank champion (Gnar) and then using a ranged mask-man (Kennen), while user 16 also played his first match using a melee tank (Darius) and then a ranged mask-man (Lucian), but played in different roles.

In order to examine whether a sample x is similar to another sample y statistically, we used the Mann-Whitney test, which tests the null hypothesis that data in x and y are samples from continuous distributions with equal medians, against the alternative that they are not. The test assumes that the two samples are independent, and x and y can have different lengths [Hart 2001]. This test can be particularly useful when behavioural effects are being studied [Tallarida and Murray 1987]. Mann-Whitney test is equivalent to the Wilcoxon signed-rank test. Other statistical tests were discarded because the conditions to perform them were not always satisfied.

For our experiments, we will use the Mann-Whitney test to observe if the data from attribute *i* collected from a user *A* is similar to *i* from user *B*. Ideally, if *A* = *B* the

Table 1. Users who played multiple matches of League of Legends

User	Character	Role
user3	Braum, Leona	sup, sup
user6	Karma, Zyra	sup, sup
user10	Fizz, Zilean	mid, mid
user14	Braum, LeeSin	sup, jng
user16	Darius, Lucian	top, adc
user20	AurelionSol, Caitlyn, Illaoi, Jinx, LeeSin, Leona, Malphite, Thresh, XinZhao	jng, adc, top, adc, jng, sup, jng, jng, sup, jng
user23	Sejuani, Sejuani	jng,jng
user24	Hecarim, Kindred	jng,jng
user36	ChoGath, Tristana	top, adc
user42	Rammus, Shyvana	jng, jng
user43	Azir, Orianna	mid, mid
user45	Irelia, Kennen	top, top
user48	Fizz, Jinx, Lucian, Morgana Taric, Thresh, Tryndamere, Twitch, Vi	mid, adc ,adc, sup sup, sup, top, adc, jng
user49	Ashe, Jinx	adc, adc
user51	Sivir, Vayne	adc, adc
user52	Caitlyn, Sivir	adc, adc
user53	Corki, Yasou	adc, mid
user55	Gnar, Kennen	top, top

test will always accept the null hypothesis, however, this will not be possible for every case, because different characters can play differently, even when both characters have the same role.

All the possible pairs [α, β] of the samples will be tested with selected users, where α is a sample from user *A* and β is a sample from user *B*, for all 33 attributes in the database. As each sample can have different sizes and are independent (each sample is from a different match), Mann-Whitney test fits perfectly to conduct our analysis. All tests were conducted with a 5% significance level, using the two most representative users (*user20* and *user48*).

3.1. Results when the samples are from the same user

For this experiment, each sample α from *user20* was compared to his 9 other samples, and then the amount of times the null hypothesis *H* was denied (*H* = 1) when comparing α to the others was counted, for every feature. After comparing all the samples, the median and average mean of the amount of times where *H* = 1 for *user 20* was measured.

Figure 1 A) shows us the results of comparing *user 20* with himself. **The vertical bars represent the number of times the null hypothesis was denied for a certain feature**, in other words, the smaller the bar, the better the results.
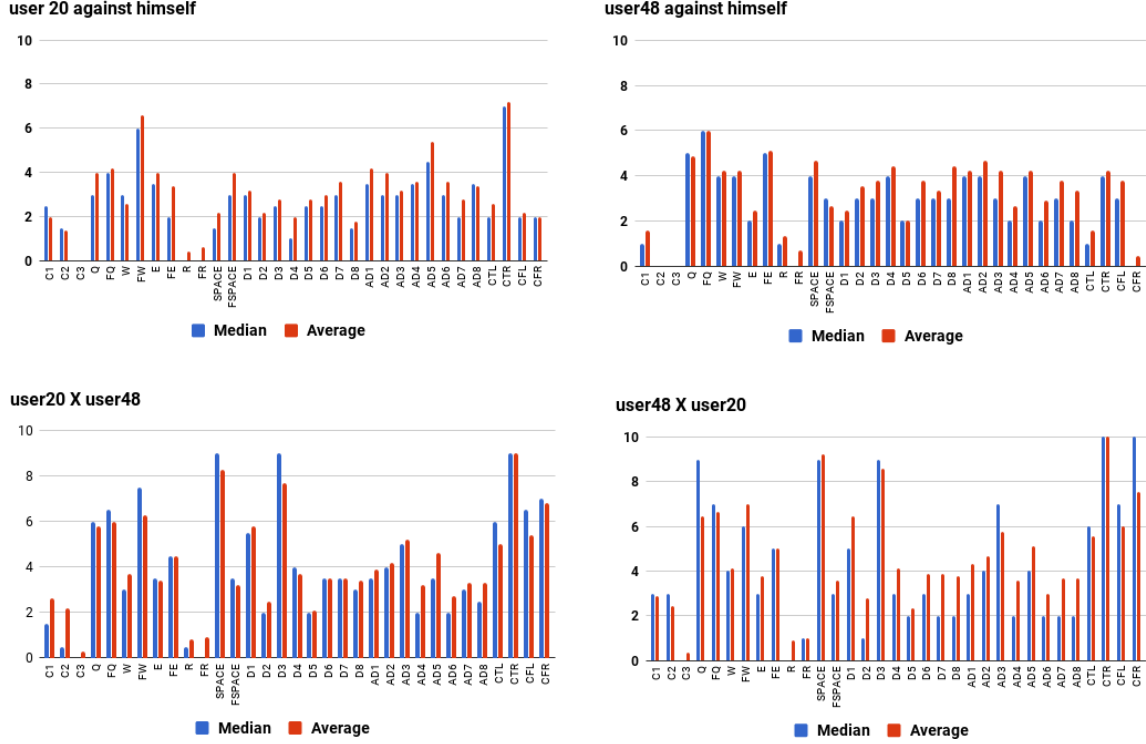

Figure 1. Median and average of A) samples from user 20 against himself, B) samples from user 48 against himself, C) samples from user 20 agains user 48 and D) samples from user 48 against user 20.

We can see that the null hypothesis was accepted on average at least half the times for almost all the characteristics, with a "high" standard deviation caused when two very different characters were compared (for example, a melee jungler with high mobility and a ranged ADC with low mobility). The median shows us a more reliable result, because half of the time the Mann-Whitney test accepted the null hypothesis for all but two features. These two features - CTR and FW - can be explained by the player's role in a match. Players taking roles like support and jungler need to be constantly moving all over the map, while the other roles usually stay in their positions for longer periods, with junglers and supports always changing their paths to fit the match, whether by attacking enemies in different lanes or conquering neutral objectives, like map visibility, thus, affecting the mouse usage. 'FW' high disparity is explained by the difference between characters. The 'W' key is often associated with passive skills that does not require the constant pressing of a button to be activated, or with an ability that does not fit every situation, explaining why the null hypothesis was not accepted.

Figure 1 B) shows the results of comparing *user 48* with himself. Much like the previous experiment, each sample β from *user 48* was compared with his other 8 samples at a level of significance of 5%.

Resembling the previous comparisons, the medians in Figure 1 shows that the null hypothesis was accepted at least half of the times for almost every feature tested, with some features being more accurate within user 48 samples, like C2 and CFR, and other being more divergent, like Q, FQ and SPACE.

From Figure 1 A) and B), we can see that *user 20* has a better accuracy related to SPACE when compared to *user 48*, implying that *user 20* uses the 'SPACE' key more consistently. It would be hard to infer some information about the other features without proper tests because they are strongly related to roles and characters. The next session presents tests between users, where more information can be gathered.

3.2. Results when the samples are from distinct users

In order to make these tests, each sample α from *user 20* was tested against all samples β from *user 48* using the Mann-Whitney test, with a level of significance of 5%, then, the opposite tests were made, putting *user 48* against *user 20*.

Figure 1 C) and D) shows samples from an user against the other. We can see that the median here is greater than half of the maximum H for more than two features, unlike the the previous experiment. This indicates that these features with a bigger H sum have a greater impact comparing *user 20* and *user 48*. We can also see the 'SPACE' features with a high value, reinforcing the affirmative that both players use the key distinctly, no matter the match.

The opposite comparison, putting *user 48* opposed to *user 20* gives us similar results. The set of characteristics with a median H greater than half of the the maximum H is almost the same as the previous comparison, reinforcing the value of these features when comparing these two users. One could argument that other features have H results similar to those shown previously, however, the set of characteristics with a high discrepancy have more value, due to exposing the differences between the behaviour of the users, increasing their biometric value.

This first analysis putting users against each others exposed what may be a pattern for every two distinct players, pointing that some characteristics have a high variability between them, exposing what are the most significant features of an individual. It is imperative to note that comparing every sample from a user α with a user β to calculate the median and average mean is not commutative, but a simple comparison between two given samples is. For example, both medians of 'Q' are high in Figure 1 C) and D), but the results of comparing a sample of *user 48* with all the other samples of *user 20* tend to negate the null hypothesis of the Mann-Whitney test more times than the opposite, indicating that *user 20* 'Q' feature is closer to *user 48* than *user 48* is closer to *user 20*. If the general comparison were commutative this would not be the case.

4. Final Remarks

The results of this work showed us that distinct users can be statistically compared by using the Mann-Whitney test to verify if their characteristics resemble one another, or even if the same player using different characters in different roles resemble himself. The comparisons made here demonstrate how to identify distinctions between users, revealing the value of their behaviour through comparisons. The case study using *user20* and *user 48* can be expanded to other users, identifying what features have more biometric value to a given user. With a bigger number of samples from other users, this study could be expanded to have a better understanding of how these important features can be identified.

This work point us in a direction where players can be identified no matter what characters or roles they played, dismissing the idea that a player is only defined by the characters they play, thus, reinforcing the idea that biometrics can be used to combat the problem of "account sharing".

References

Banerjee, S. P. and Woodard, D. L. (2012). Biometric authentication and identification using keystroke dynamics: A survey. *Journal of Pattern Recognition Research*, 7(1):116–139.

Bergadano, F., Gunetti, D., and Picardi, C. (2002). User authentication through keystroke dynamics. *ACM Transactions on Information and System Security (TIS-SEC)*, 5(4):367–397.

Bours, P. and Fullu, C. J. (2009). A login system using mouse dynamics. In *Intelligent Information Hiding and Multimedia Signal Processing, 2009. IIH-MSP'09. Fifth International Conference on*, pages 1072–1077. IEEE.

Camara, L. (2017). Acquisition and analysis of the first mouse dynamics biometrics database for user identification in the online collaborative game League of Legends. Master's Thesis (Systems and Computing), UFRN (Universidade Federal do Rio Grande do Norte), Natal, Brazil.

Chen, K.-T. and Hong, L.-W. (2007). User identification based on game-play activity patterns. In *Proceedings of the 6th ACM SIGCOMM workshop on Network and system support for games*, pages 7–12. ACM.

da Silva Beserra, I. (2017). Using keystroke dynamics for user identification in the online collaborative game League of Legends. Master's Thesis (Systems and Computing), UFRN (Universidade Federal do Rio Grande do Norte), Natal, Brazil.

da Silva Beserra, I., Camara, L., and Da Costa-Abreu, M. (2016). Using keystroke and mouse dynamics for user identification in the online collaborative game league of legends.

Hart, A. (2001). Mann-whitney test is not just a test of medians: differences in spread can be important. *BMJ: British Medical Journal*, 323(7309):391.

Idrus, S. Z. S., Cherrier, E., Rosenberger, C., and Bours, P. (2013). Soft biometrics database: a benchmark for keystroke dynamics biometric systems. In *Biometrics Special Interest Group (BIOSIG), 2013 international conference of the*, pages 1–8. IEEE.

Idrus, S. Z. S., Cherrier, E., Rosenberger, C., Mondal, S., and Bours, P. (2015). Keystroke dynamics performance enhancement with soft biometrics. In *Identity, Security and Behavior Analysis (ISBA), 2015 IEEE International Conference on*, pages 1–7. IEEE.

Leavitt, A., Clark, J., and Wixon, D. (2016). Uses of multiple characters in online games and their implications for social network methods. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 648–663. ACM.

Lv, H.-R., Lin, Z.-L., Yin, W.-J., and Dong, J. (2008). Emotion recognition based on pressure sensor keyboards. In *Multimedia and Expo, 2008 IEEE International Conference on*, pages 1089–1092. IEEE.

Shen, C., Cai, Z., and Guan, X. (2012). Continuous authentication for mouse dynamics:

A pattern-growth approach. In *Dependable Systems and Networks (DSN), 2012 42nd Annual IEEE/IFIP International Conference on*, pages 1–12. IEEE.

Tallarida, R. J. and Murray, R. B. (1987). Mann-whitney test. In *Manual of Pharmacologic Calculations*, pages 149–153. Springer.

Thanganayagam, R. and Thangadurai, A. (2015). Fusion approach on keystroke dynamics to enhance the performance of password authentication. In *Electrical, Computer and Communication Technologies (ICECCT), 2015 IEEE International Conference on*, pages 1–6. IEEE.

Yampolskiy, R. V. and Govindaraju, V. (2006). Use of behavioral biometrics in intrusion detection and online gaming. In *Proc. of SPIE Vol*, volume 6202, pages 62020U–1.