

# Twitter fingerprints as active authenticators

Alex Brown  
College of Information Sciences and Technology  
The Pennsylvania State University  
University Park, PA 16802  
atb5184@ist.psu.edu

Myriam Abramson  
Building 34  
Naval Research Lab  
Washington, D.C. 20375  
myriam.abramson@nrl.navy.mil

**Abstract**—Leveraging data drawn from the Web, or rather web analytics, has been used to gain business intelligence, increase sales, and optimize websites. Yet beyond the domain of ecommerce that web analytics is typically associated with, authentication based upon user interactions with the Web is also obtainable. Authentication is able to be achieved because just as individuals display unique mannerisms in everyday life, users interact with technology in unique manners. Leveraging these unique patterns, or “cognitive fingerprints”, for security purposes can be referred to as active authentication. Active authentication stands to add extra security without added burden, as users are allowed the capability to simply interact with technology in their natural manner. Past research on active authentication has looked at areas such as mouse pattern movements, screen touch patterns on smartphones, and web browsing behavior. Our focus here is web browsing behavior. Specifically, we seek to extend past active authentication research done on Reddit. In this research, we examine the ability of Twitter-specific features to serve as authenticators, by examining the behavior of 50 random Twitter users. Through leveraging data mining and machine learning techniques, we conduct three levels of analysis: (1) we survey the ability of Twitter-specific behavioral features from a broad perspective to determine the feasibility Twitter fingerprints as a form of active authentication; (2) we compare aggregated and non-aggregated datasets to determine whether it is better to aggregate user behavior or look at posts individually; and (3) we examine whether certain features are more important for discrimination than others. The first level of analysis suggests that the posting behavior on Twitter follows the power law of human activity and that users can be uniquely identified with a fairly decent level of accuracy. Second, we find that aggregating the data significantly improves F-scores. Lastly, our examination suggests that there is not any specific feature that serves as more discriminative than others. Rather, what is discriminative for one user may not be for another user.

**Keywords**—*active authentication; cognitive fingerprints; web analytics*

## I. INTRODUCTION

As we interact with technology, we leave data known as “trace data” [1]. Trace data about an individual browsing the Web, for example, may tell us that the user spent 5 minutes on a website, visited 3 webpages on the site, and clicked 2 videos. Simply put, every interaction with technology leaves some sort of trail. These trails can then be leveraged for various purposes.

<sup>1</sup> For the purposes of this paper, we will use the term “active” authentication.

Web masters of ecommerce sites, for instance, may utilize this information in attempt to increase sales.

Outside the realm of ecommerce, another way trace data can be used is as a soft biometric tool. Rather, just as an individual displays unique personality traits, an individual also interacts with technology in a unique manner and leaves “cognitive fingerprints” [2]. As pointed out by [2], cognitive fingerprints provide a way to engage in authentication of users, as they interact with technology in a natural manner.

Using cognitive fingerprints as a way to authenticate users has been referred to as active, continuous, implicit, and passive authentication<sup>1</sup>, depending upon the perspective taken [3-6]. Nonetheless, these terms are hinged upon the same underpinning: humans are unique in their behavior, consequently behavioral interactions with technology can act as a security mechanism. Hence, the capability exists to authenticate users without burdensome control procedures. Even beyond the benefit of no extra burden, using a myriad of sensors that monitor natural human behavior stands to serve as a greater form of security than passwords or even fingerprints [7].

One area of focus in regards to leveraging cognitive fingerprints for active authentication has been Web browsing [8-12]. This research seeks to extend research done by [8], which focused on cognitive fingerprints of Reddit. Here we seek to leverage data mining and machine learning techniques to examine the ability of Twitter-specific characteristics to serve as authentication features. By gaining a better understanding of the ability to actively authenticate individuals based upon their social media behavior, we can improve social media security, such as the ability to prevent wrongful posts displayed through hacked accounts. The hacking of high profile accounts has potential to cause chaos, disruption, and terror, as can be seen through Figures 1 and 2. Figure 1 shows an example from 2013 whenever the Twitter account of the Associated Press was hacked and reported explosions at the White House; DOW consequently received a negative impact [13]. Figure 2 is an example from earlier this year (2015) whenever Islamic State supporters hacked the Twitter account of U.S. Central Command [14].



Figure 1. A Tweet from the Associated Press’ Twitter, which caused stocks to drop [13].



Figure 2. A Tweet from U.S. Central Command’s Twitter, upon being hacked by Islamic State supporters [14].

For this research we will conduct three levels of analysis. First, from a broad perspective, we will survey the capability to authenticate users based upon the patterns they leave on Twitter and offer comparisons based upon [8]’s research with Reddit. The second level of analysis will explore whether aggregating data into sessions has benefit over examining posts individually. Lastly, we will examine whether certain Twitter features are the more important for discriminating users.

## II. RELATED WORK

Behavioral interactions with technology have been documented as being able to act as authenticators. For example, [3] achieved a false acceptance rate (FAR) and false rejection rate (FRR) of 5.9% by tracking user mouse movements. Xu, Zhou, and Lyu [6] and De Luca, Hang, Brudy, Lindner, and Hussman [15] show the potential of being able to authenticate users based upon the way they touch a smartphone. The early exhibited success, however, is not met without challenges. Particularly, questions remain about practical implications. Overcoming the challenges and proving the viability of this method is currently of focus for DARPA’s Active Authentication program [2].

Outside the realm of behavioral interactions with physical pieces of technology, our virtual interactions also leave a trail. Soon after the creation of the World Wide Web, it was discovered that transaction logs, which captured data left by users when interacting with the Web, could be leveraged to learn about user interaction with the Web. This practice became known as transaction log analysis (TLA), which then brought about the emergence of several different areas of research within the Internet research domain [16-18]. Companies were able to gain business intelligence, optimize their websites, and even predict what users were going to buy [18, 19]. The analysis of transaction logs was largely driven by ecommerce. Yet, similarly to mouse movement patterns and touches on smartphones, the trails left behind by users can also be used for other purposes, such as security.

Padmanabhan and Yang [12] explored the idea of using clicktrail patterns to authenticate users. They found that with the proper level of aggregation, users do display unique characteristics that make them reasonably identifiable. Banse,

Herrmann, and Federrath [11] have shown that users can be tied to specific Web sessions with close to 90% accuracy.

Given that Web-based digital footprints can come from two sources, web browsing or social media, much of the research to date has only begun to scrape half the potential in leveraging cognitive fingerprints for security purposes. An exception being [8] who showed that similar to web browsing behavior, users on a social media platform (Reddit) can be authenticated with a certain accuracy through their social media behavioral characteristics.

Data drawn from Twitter has been widely documented as being able to serve a variety of tasks. From predicting the stock market [20] to identifying latent attributes (i.e. regional orientation or business affinity) [21, 22], researchers are continually finding new ways to utilize and manipulate Twitter data. It is therefore the overarching purpose of this research to leverage Twitter data and extend past research done on active authentication. Motivating questions driving this research are: *How do different social media platforms perform in regards to active authentication? How does the aggregation of posting behavior affect the authentication accuracy? What Twitter features serve as the best identifiers?*

## III. METHODOLOGY

As originally stated, this research will focus on three levels of analysis. The first level of analysis will survey the ability of Twitter-specific features to serve as authenticators. Second, the results between aggregated data and non-aggregated data will be compared. Lastly, whether certain features are more discriminative than others will be explored. The pursuit of this research is carried out in three phases: data collection, data preparation, and data analysis.

### A. Data Collection

Randomization was obtained using Twitter’s search function to obtain the last 5,000 Tweets containing the letter ‘a’. From those 5,000 Tweets, 100 users were randomly selected. Posts (i.e. Tweets and Retweets) on the selected users’ timelines from the month of May were retrieved. The 50 most active users were then chosen for examination. Overall characteristics for the dataset can be seen in Table 1.

TABLE I.

Dataset Characteristics	
Time	May 1 - 31, 2015
Users	50
Posts	53,108
Retweets	24,311
Tweets	28,797
Sessions	6,008
Singletons	1,661

### B. Data Preparation

Within Web analytics, a commonly accepted definition of a session is the time from whenever a user starts interacting with a website to the time they stop interacting with the website, or after 30 minutes of inactivity (whichever comes first). Since time of login and logoff is not accessible information via Twitter’s API, sessions for the aggregated dataset must be delineated in a different manner. The method of choice here is to log the timestamps of each post, calculate the time between each post, and delineate sessions by pauses that last longer than one hour. Moreover, singeltons (i.e. single post sessions) were removed to reduce noise. Features of the aggregated dataset are: day of the week, hour of day the session started, length of session, posts per session, sentiment (positive, negative, and neutral counts), retweets, tweets, retweet to tweet ratio, average tweet length, and hashtags per tweet. Features of the non-aggregated (i.e. taking each post individually, as opposed to grouping them into sessions) dataset are: day of the week, time of the day, pause between posts, the number of retweets and favorites received by the post, Tweet length, whether the post was a Retweet or not, and sentiment. All values are then standardized using Weka [23] before the start of analysis.

### C. Data Analysis

Analysis will start with time-variant features to examine whether, as with the Reddit research [8], users obey the power law of human activity. In particular we will examine the pauses and burstiness (i.e. difference between the pauses) of the five most active users. We will then use support vector machine classification (SVMC), within Weka’s workbench [23, 24], to obtain the average FRR and FAR. Specifically, these metrics are obtained by running 5 iterations for each user (for both the aggregated and non-aggregated datasets) against a random subset of other users. Based upon the resulting averages, F-scores for each user will then be calculated as well.

FRR and FAR are commonly used in biometrics for user authentication. FRR refers to the probability that a user is incorrectly denied access. FAR refers to the probability that a user is incorrectly allowed access [25]. An F-score can be thought of as the harmonic mean of precision and recall, with a score of 1 being the best and 0 being the worst. It is calculated as follows in equation 1:

$$F = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (1)$$

The last level of analysis will be conducted using J48, part of Weka’s workbench [23]. J48 is selected as opposed to SVMC, since J48 is a tree and we can therefore see what the top discriminating feature is. Every user, for both the aggregated and non-aggregated datasets, will be run through J48.

## IV. RESULTS AND ANALYSIS

Abramson [8] examined Reddit posts and found that time-variant features of user behavior obeyed the power law of human activity. Results displayed in Figure 3 show the same observation for Twitter. Figure 4 also reiterates this through the display of burstiness, or the difference among pauses. This information can be used to discern humans from bots [8]. It is also interesting that this holds true with Twitter, given the perceived spontaneity Twitter users exhibit. Hence, even though one may think that the timing of posts on social media is likely to be random, Figures 3 and 4 suggest otherwise.

Figure 5 shows an overall depiction of the FAR and FRR results obtained from the testing. As expected, certain users are

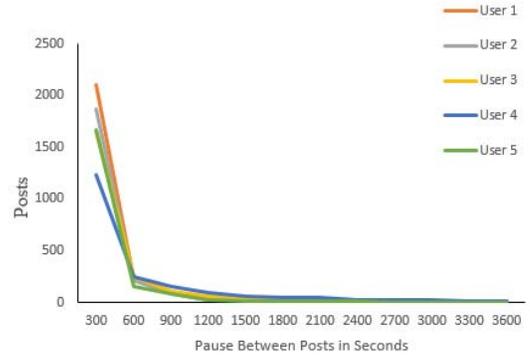


Figure 3. Posting behavior (i.e. pause profiles) within session for the 5 most active users.

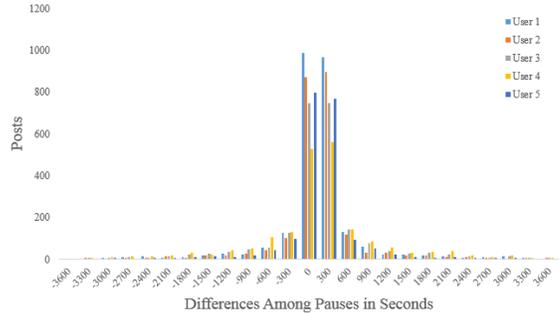


Figure 4. Burstiness (differences among the pauses) for the 5 most active users.

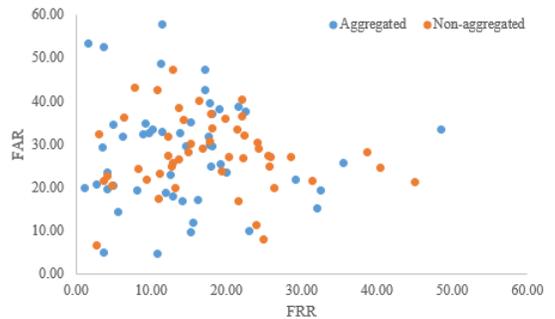


Figure 5. A comparison of the aggregated and non-aggregated results.

more predictable than other users. Tables 2 and 3 (aggregated and non-aggregated, respectively) show the top five and bottom five users, based upon how identifiable they are (based upon F-scores). The results suggest that, similarly to [8]’s Reddit research, solely using the session features utilized in this research may not be ideal as a soft biometric tool for a social media platform like Twitter. Nonetheless, the results were decently accurate and promising potential is shown. Future research may look to add more features.

TABLE II.

Aggregated Results			
User	FRR	FAR	F-score
31	3.70 ± 0.00	4.80 ± 0.00	0.96
22	10.82 ± 0.42	4.70 ± 0.51	0.92
17	5.58 ± 0.43	14.34 ± 0.45	0.90
10	1.22 ± 0.58	19.72 ± 0.62	0.90
8	2.68 ± 0.55	20.76 ± 0.51	0.89
.	.	.	.
.	.	.	.
.	.	.	.
2	22.50 ± 0.46	37.40 ± 0.43	0.72
42	17.24 ± 0.92	47.06 ± 0.51	0.72
38	11.50 ± 0.40	57.68 ± 0.92	0.72
20	35.52 ± 1.48	25.60 ± 0.98	0.68
7	48.50 ± 2.15	33.42 ± 1.45	0.56
Averages	14.50	28.01	0.80

TABLE III.

Non-aggregated Results			
User	FRR	FAR	F-score
31	2.74 ± 0.02	6.42 ± 0.04	0.96
16	3.66 ± 0.10	21.52 ± 0.15	0.88
22	4.74 ± 0.06	20.26 ± 0.02	0.88
10	4.24 ± 0.18	22.62 ± 0.08	0.88
35	10.98 ± 0.16	17.24 ± 0.17	0.86
.	.	.	.
.	.	.	.
.	.	.	.
29	28.58 ± 0.40	26.86 ± 0.31	0.72
2	22.10 ± 0.10	40.40 ± 0.20	0.71
39	38.68 ± 0.20	27.98 ± 0.13	0.65
32	40.44 ± 0.11	24.44 ± 0.12	0.65
42	45.08 ± 0.31	21.08 ± 0.23	0.62
Averages	17.89	27.75	0.78

TABLE IV.

Aggregated J48 Top Feature Results	
Feature	% Top Feature
Day of Week	0
Start of Session	2
Session Length	0
Tweets per Session	2
Positive Count	12
Negative Count	4
Neutral Count	4
Retweets	14
Tweets	0
Retweet to Tweet Ratio	20
Average Tweet Length	32
Hashtags per Tweet	10

TABLE V.

Non-aggregated J48 Top Feature Results	
Feature	% Top Feature
Day of Week	0
Time of Day	24
Pause Between Posts	2
Retweets	26
Favorites	4
Tweet Length	12
Retweet T/F	14
Sentiment	18

To determine whether aggregating the data into sessions makes a difference in regards to predictability, a paired sample t-test was employed. It was found that the F-scores of the aggregated data were significantly higher  $t(49) = 2.29, p < .05$ . To restate, we aggregated data by grouping Tweets together that were posted less than 1 hour apart. Future research may look into what the ideal session length time should be when examining social media.

Results from Tables IV and V suggest that there may not necessarily be one feature that is ideal for discrimination among users. Rather, what may be discriminative for one user may not be discriminative for another user. Therefore, the more features that are captured, the greater the likelihood that a discriminative feature for that specific user will be captured. This reaffirms the idea that more features (e.g. political orientation or gender) will likely equate to even greater results.

## V. DISCUSSION AND CONCLUSION

As we interact with technology, the cognitive fingerprints we leave behind can be leveraged to authenticate users. Past research has largely focused on how to more effectively

authenticate users using active authentication, but not as much research exists exploring different forums of technology usage. An exception to this is [8]’s work, which surveyed Reddit usage and the ability of publicly available Reddit data to be used for authentication purposes. Subsequently, motivation existed to explore Twitter and examine whether similar findings would hold true, despite the spontaneity Twitter users are expected to display.

First, it was discovered that users follow a power law distribution in regards to the timing of their posting behavior. We then used SVMC to overview the feasibility of using publicly available Twitter data to uniquely identify users. We used two different datasets: aggregated and non-aggregated, and achieved fairly accurate results. For the aggregated dataset we achieved an average FRR of 14.50, FAR of 28.01, and F-score of 0.80. With the non-aggregated dataset we obtained an average FRR of 17.89, FAR of 27.75, and F-score of 0.78. A paired sample t-test revealed that F-scores from the aggregated dataset were significantly higher  $t(49) = 2.29, p < .05$ . Lastly, we looked at all 50 users from the aggregated dataset and the non-aggregated dataset within a J48 tree. We found that there was not necessarily one feature that served as a top discriminator. Rather, different features are likely to be discriminative for different users.

While the results here were not ideal, there is believed to be room for significant improvement. For example, this research largely focused upon features already existent from Twitter. However, more in-depth analysis of the posts themselves (e.g. political orientation, gender, etc.) could allow for greater accuracy. Moreover, since grouping the Tweets into sessions was found to make a significant difference, it may be worthwhile to explore what the optimal aggregation time is for social media. By optimizing aggregation and adding more features, it is believed that active authentication within Twitter is practical for creating a social media active authentication mechanism. Also take into account that the results were achieved completely through publicly available posting behavior. Other non-publicly accessible features are also likely to improve the results.

By improving upon the techniques used here, we could potentially prevent hacked social media accounts from having wrongful information posted. This stands to prevent chaos, disruption, and terror from those who seek any way possible to inflict it.

#### ACKNOWLEDGMENT

This research was conducted at the Naval Research Lab (NRL), as part of a summer internship through the Department of Homeland Security (DHS) HS-STEM program. I would like to thank everybody at the NRL for their support, and those in charge of the DHS HS-STEM program for selecting me as a participant.

#### REFERENCES

- [1] B. J. Jansen, "Understanding User-Web Interactions via Web Analytics," *Synthesis Lectures on Information Concepts, Retrieval, and Services*, vol. 1, pp. 1-102, 2009.
- [2] R. P. Guidorizzi, "Security: Active authentication," *IT Professional*, vol. 15, pp. 4-7, 2013.
- [3] Y. Aksari and H. Artuner, "Active authentication by mouse movements," in *Computer and Information Sciences, 2009. ISCIS 2009. 24th International Symposium on*, 2009, pp. 571-574.
- [4] M. Jakobsson, E. Shi, P. Golle, and R. Chow, "Implicit authentication for mobile devices," in *Proceedings of the 4th USENIX conference on Hot topics in security*, 2009, pp. 9-9.
- [5] S. Hashiaa, C. Pollett, M. Stampc, and M. Hall, "On using mouse movements as a biometric," 2005.
- [6] H. Xu, Y. Zhou, and M. R. Lyu, "Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones," in *Symposium On Usable Privacy and Security, SOUPS*, 2014, pp. 187-198.
- [7] Google Developers, "Google I/O 2015 - A little badass. Beautiful. Tech and human. Work and love. ATAP.," ed, 2015.
- [8] M. Abramson, "Cognitive fingerprints," presented at the 2015 AAAI Spring Symposium, 2015.
- [9] M. Abramson and D. W. Aha, "User authentication from web browsing behavior," presented at the The Twenty-Sixth International FLAIRS Conference, 2013.
- [10] M. Abramson and S. Gore, "Associative patterns of web browsing behavior," presented at the 2013 AAAI Fall Symposium Series, 2013.
- [11] C. Banse, D. Herrmann, and H. Federrath, "Tracking users on the internet with behavioral patterns: Evaluation of its practical feasibility," presented at the Information Security and Privacy Research, 2012.
- [12] B. Padmanabhan and Y. C. Yang, "Clickprints on the web: Are there signatures in web browsing data?," *Available at SSRN 931057*, 2007.
- [13] (2013, July 20). *AP Twitter account hacked, 'explosions at White House' tweet crashes DOW*. Available: <http://www.rt.com/usa/hackers-associated-press-obama-282/>
- [14] D. Lamothe. (2015, July 20). *U.S. military social media accounts apparently hacked by Islamic State sympathizers*. Available: <https://www.washingtonpost.com/news/checkpoint/wp/2015/01/12/centcom-twitter-account-apparently-hacked-by-islamic-state-sympathizers/>
- [15] A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann, "Touch me once and I know it's you!: Implicit authentication based on touch screen patterns," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 987-996.
- [16] A. Kaushik, *Web Analytics: An Hour a Day*. New York: Wiley Publishing, 2007.

- [17] B. J. Jansen, "Search log analysis: What it is, what's been done, how to do it," *Library & Information Science Research*, vol. 28, pp. 407-432, 2006.
- [18] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "Web usage mining: Discovery and applications of usage patterns from web data," *ACM SIGKDD Explorations Newsletter*, vol. 1, pp. 12-23, 2000.
- [19] A. L. Montgomery, S. Li, K. Srinivasan, and J. C. Liechty, "Modeling online browsing and path analysis using clickstream data," *Marketing Science*, vol. 23, pp. 579-595, 2004.
- [20] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, pp. 1-8, 2011.
- [21] D. Rao, D. Yarowsky, A. Shreevats, and M. Gupta, "Classifying latent user attributes in twitter," in *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, 2010, pp. 37-44.
- [22] M. Pennacchiotti and A. M. Popescu, "A machine learning approach to Twitter user classification," *ICWSM*, vol. 11, pp. 281-288, 2011.
- [23] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, 2009.
- [24] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, p. 27, 2011.
- [25] Biometric-solutions. (2013, July 16). *Glossary*. Available: <http://www.biometric-solutions.com/glossary.php>