

# Cognitive Fingerprints

**Myriam Abramson**

Naval Research Laboratory, Code 5584  
Washington, DC 20375  
myriam.abramson@nrl.navy.mil

## Abstract

Ever since “True Names” by Vernor Vinge, identity has been recognized as our most valued possession in cyberspace. Attribution is a key concept in enabling trusted identities and deterring malicious activities. The rash of recent cyber-attacks targeting consumers, coupled with the massive amount of available digital data, has forced us to rethink our very notion of authentication as a one-shot process of verifying a claim of identity. In this context, this paper surveys the status of “cognitive fingerprints” where how you think or what you are thinking can reveal who you are. The computational approaches vary but most involve machine learning and data mining techniques and specific computational methods will be highlighted. Digital traces from social media activities are presented as an example of a cognitive fingerprint.

## 1 Introduction

The science of autonomy requires new methods for the verification and validation of highly complex and adaptive systems from two different perspectives. First, the control of autonomous systems, as virtual or physical agents, requires new methods for authenticating the human-in-the-loop in order to propagate trust. In other words, the certification of autonomous systems implies the continuous authentication of the entities on behalf of which the systems operate for attribution purposes. Second, the prediction of autonomous systems states, including the possibility of deception, requires the determination of a cognitive fingerprint to identify problem solving strategies.

Attribution is broadly defined as the assignment of an effect to a cause. We differentiate between authentication and identification as two techniques for the attribution of identity. Authentication is defined as the verification of claimed identification (Jain, Bolle, and Pankanti 1999). Identification involves recognition as a one-to-many matching problem while authentication is a one-to-one matching problem. Authentication is further distinguished from de-anonymization (also called re-identification) which seeks to identify people by name (or other personally identifiable information) from the cross reference of data sources. However, de-anonymization is only valid for a certain dataset and, unlike authentication, has no predictive power. While

traditional authentication methodologies strive to provide instant results based on something you know (e.g., a pin) or something you have (e.g., a fingerprint), behavioral biometrics try to provide continuous authentication based on something you are which can be more difficult to spoof and also to acquire. Behavioral biometrics do not claim uniqueness but may corroborate other evidence or be combined to produce uniqueness. Recently, due to the proliferation of sensor devices capturing unconscious patterns of behavior and the ubiquity of the Web, it has become possible to construct cognitive profiles defining who we are based on the digital traces of the myriad of decisions we make leading to a new type of biometrics involving cognition.

“Cognitive fingerprints” is an expression that has been coined, to the best of our knowledge, in the context of the DARPA Active Authentication (AA) program (Guidorizzi 2012). It is informally defined as the unique pattern arising from our interaction with existing technology without the need for specific sensors (Guidorizzi 2013) and thereby bypassing the need for cooperation from the user. Underlying the idea of cognitive fingerprint is “implicit learning” such as the unconscious learning occurring in the formation of skills or habits. The knowledge acquired through implicit learning does not have an explicit representation. Consequently, what we know cannot be easily stolen or divulged making implicit learning methods very attractive as an alternative to key-based encryption (Bojinov et al. 2014).

In Section 2, we describe selected modalities of cognitive fingerprints developed within the AA program and their respective methodologies. In Section 3, we introduce Web-enabled cognitive fingerprints based on Web browsing and social media user profiles. Issues and challenges can be found in Section 4.

## 2 Related Work

Previous work on operating system commands has been extended in (Salem and Stolfo 2011) to model user information search behavior on the desktop ignoring other typical activities such as networking or printing. This work is based on user studies showing different search behavior depending on familiarity with the environment. A user model is constructed based on search behavior features (e.g., accesses to desktop search tools, number of file touches, and frequency of file system navigation) aggregated in 2-minute increments

using one-class classification methods to detect deviations from normal user behavior. Other aspects of this work include the use of decoys in the file system to capture intent and thereby amplifying differences in search behavior.

Stylometry is another cognitive fingerprint that has been used extensively for authorship. This work has been extended to authentication by relating keyboard dynamics (which are a behavioral biometric) with linguistic features (Juola et al. 2013). Linguistic features include lexical statistics (e.g., word length, frequency of upper/lower case characters, frequencies of corrections, etc.) and syntactic features such as function words, part-of-speech tags, and common word n-grams. One-class classification methods and nearest-neighbor techniques were used in this work. One interesting aspect of this work is the inference of high-level features such as personality, gender, and dominant hand but those features have not been integrated to date into the construction of a unique cognitive fingerprint.

Covert games capture the cognitive fingerprint of a user by engaging the user into divulging a computational thinking strategy. Covert games, modeled as inverse Prisoner’s Dilemma games, have been developed as computer problems alerts with a choice of possible moves (Wheeler et al. 2013). A sequence of moves constitutes a user strategy. User cognitive fingerprints are distinguished with a similarity score based on move sequences and response preferences. During the enrollment phase of the authentication process, the user learns a way to solve computer problems by responding to generated alerts. An imposter, missing the enrollment phase, will not be able to respond in the same way.

Just as the way we keep our home/office can reveal something about ourselves, the way windows are arranged on our computer can reveal some fundamental feature about our cognitive fingerprint. Using pixel analysis and image processing techniques, screen fingerprints (Patel et al. 2013) have been developed in this context by extracting personal features such as how well a user sees based on the font of the text, the age of a user based on the speed by which windows are moved, the work pattern inferred from the dominant colors of the windows and background, etc.

### 3 Web-enabled Cognitive Fingerprints

Our digital footprint on the Web can be viewed from two different perspectives: Web browsing and social media. While Web browsing is considered private, social media posts and comments are often meant to be shared with others. Web browsing includes both a semantic aspect in the type of web-pages visited and a syntactic aspect as a navigation tool. Research in Web intelligence has sought to understand and predict Web behavior in order to improve a specific exogenous outcome such as the checkout of a shopping cart or improvement in meaningful search results. More interest is given now to understand and even influence the cognitive behaviors that can be inferred from our Web activities. Discovering duplicate users and fake identities are other applications of this research.

This paper extends prior work on Web browsing (Abramson and Aha 2013; Abramson 2014; Abramson and Gore

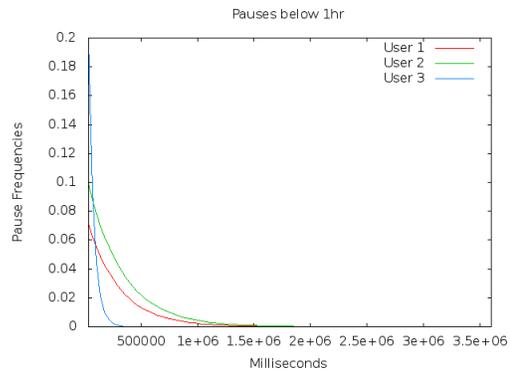


Figure 1: Pause profiles below 1hr for 3 users obtained with an exponential data fit

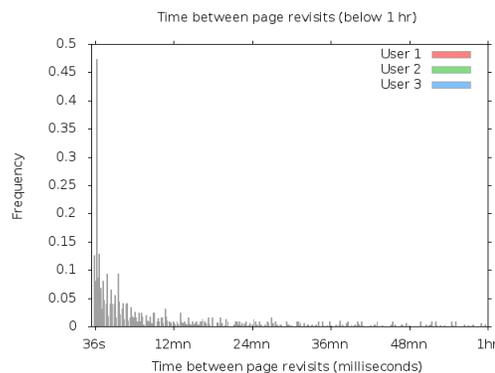


Figure 2: Time between page revisits (below 1hr) for 3 users

2013) to the social media activities of Reddit users. The data analysis shows similar aspects of Web browsing and social media behavior concerning aggregated features suggesting the possibility of similar computational approaches in capturing a cognitive fingerprint. Prior work in Web browsing identified pauses (elapsed time between page visited), time-between-revisits (page revisits within a session) and burstiness (time difference between two consecutive pauses) as typical characteristics of human activity on the Web. Figures 1, 2, and 3 illustrates those same time-variant features for 3 Reddit users. As found in Web browsing, time-variant features obey the power law of human activity. Consequently, deviations from the power law can discriminate between humans and bots in the “Internet of things”<sup>1</sup> but deviations between humans, or bots simulating humans, are harder to detect. We also found this power law manifested by the order of the activities. Figure 4 illustrates the frequency of occurrences of the most common subreddit visited and their order in the activity sequence per user. This power law distribution can be compressed using Benford’s law (Matthews 1999). Figure 5 illustrates this regularity applied to pauses. We have compressed all time-variant distributions using Benford’s law in the results below.

<sup>1</sup>[http://en.wikipedia.org/wiki/Internet\\_of\\_Things](http://en.wikipedia.org/wiki/Internet_of_Things)

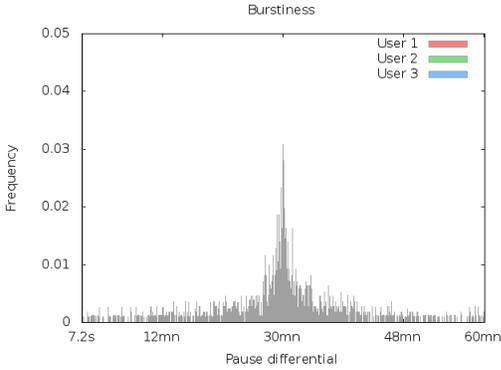


Figure 3: Burstiness (pause differential) below 1hr

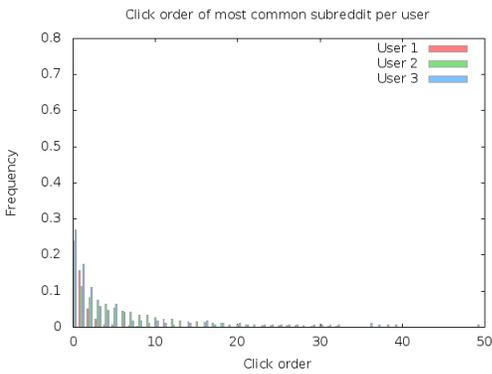


Figure 4: Frequencies of the most common subreddit by order visited for 3 users

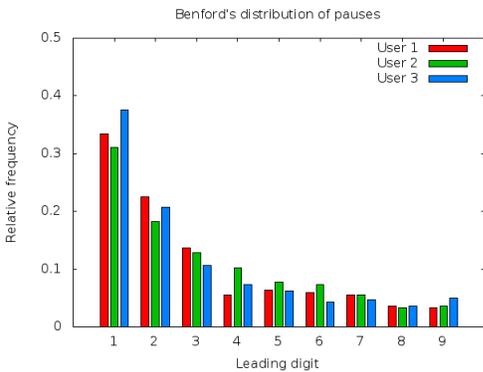


Figure 5: Benford's distribution of pauses in a session for 3 users

In addition to time-variant features, Web browsing features include the type of the webpage visited and global syntactic features of a session (including time-of-day, day-of-week, average duration, length of the session, number of unique subreddits, and post/comment rate) where a session is defined as a series of consecutive clicks delimited by pauses of 30 minutes or longer. We retrieved posts and comments using the Python Reddit API Wrapper (PRAW) from users associated to a seed user through at least one comment and from this pool of users selected at random 50 active users. There is a hard limit from Reddit to retrieve only the last 1000 post and last 1000 comments. The session pause delimiter was extended to one hour due to the sparsity of posts and/or comments and singleton sessions were pruned from the dataset. The 20 **least** frequent subreddits were used in our results below. In contrast, the 20 most frequent subreddits produced on average worse FAR results and slightly better FRR (FRR:  $50.55 \pm 2.32$ ; FAR:  $70.69 \pm 32.19$ ) Future work should calibrate the session pause delimiter to the user's behavior who might be avoiding detection by spacing out their digital traces. The subreddit of the post/comment made up its type. Table 1 describes the dataset characteristics used in our empirical evaluation.

Based on the approach outlined in (Abramson and Aha 2013), cognitive fingerprints are evaluated using the false rejection rate (FRR) or false negatives and the false acceptance rate (FAR) or false positives found in biometrics. The FRR is obtained with a 10-fold cross-validation on the user dataset while the FAR is obtained by applying the model trained with the entire user dataset against the dataset of all the other 49 users combined. The data points are the various feature distributions for one session. We evaluated each feature separately with the one-class support vector machines (OCSVMs) classification method found in LibSVM (Schölkopf et al. 2000) for comparison with the random subspace approach in an ensemble of OCSVMs over 5 iterations. The details of this approach can be found in (Abramson and Aha 2013). Here, we've improved upon this approach by randomizing the selection of a subset of learners from a pool of learners for each prediction call, in effect smoothing out large differences in prediction results. In addition, we also compare with principal components dimensionality reduction method prior to classification by OCSVMs. SVMs are not probabilistic models and therefore an authentication decision threshold is not needed. Figure 6 illustrates the comparative results. Table 2 summarizes the results for the entire dataset. There is a significant performance difference between FRR results and FAR results ( $p\text{-value} = 2.E-314$  and  $p\text{-value} = 8.E-4$  respectively) from our random subspace ensemble learning method and the OCSVMs results from the global features. In addition, 65% of the users have a combined FRR and FAR lower or equal to 20%. Fig. 7 compares cross-validation results with temporal results which are more realistic from an operational perspective. Cross-validation results are however a good approximation of short-term temporal predictions (10%). As noted in (Abramson and Aha 2013), there is a tug-of-war between the FRR representing recall and the FAR representing precision. Future work will calibrate the parameters of the

User	#Sessions	#Posts/ Comments	#Subreddits
1	266	771	43
2	137	881	23
3	164	1183	227
ALL	27933	63833	2124

Table 1: Reddit pruned dataset (no singletons) characteristics for the first 3 users and aggregated for the 50 users (ALL) with different pause interval delimiters.

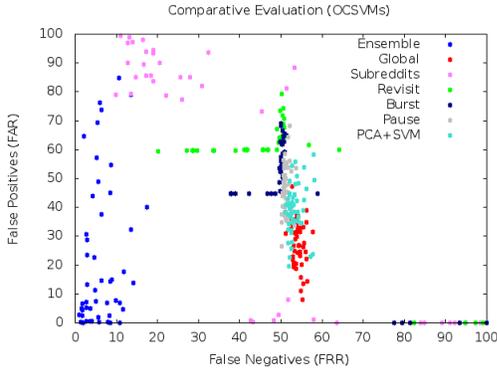


Figure 6: Comparative evaluation of OCSVMs with session intervals of 60 mins or longer. Each point represent the average of 5 user trials.

ensemble learning method (number of features, learner pool size, number of selected learners, evaluation of the learners) to adjust the results to the degree of authentication desired.

In conclusion, as found with Web browsing, some users are more recognizable than others, resulting in a high variance in the results without calibration of parameters to the user. Among our 3 users, User 3 was more recognizable (98% accuracy) with the random subspace ensemble approach. 54% of the users have a combined accuracy of 90% with the random subspace ensemble method. The compression of time-variant features with Benford’s law and the ensuing reduction in the sparsity of the data improved the performance of the random subspace ensemble. Also, as

Methods	Avg. FRR(%)	Avg. FAR(%)	Avg. Accuracy
Random Subspace	<b>6.23±4.05</b>	<b>21.86±27.25</b>	<b>0.86±0.13</b>
PCA+SVM	53.36±1.85	39.37±8.65	0.53±0.05
Pause	51.26±2.02	49.43±9.59	0.49±0.05
Revisit	60.26±28.47	40.54±30.69	0.49±0.08
Burst	41.51±25.73	63.45±20.82	0.47±0.07
Global	54.18±1.45	27.95±7.75	0.59±0.04
Subreddits	51.64±35.10	45.98±44.03	0.51±0.10

Table 2: Average comparative results for 50 users with sessions delimited by pauses of 60 mins or longer

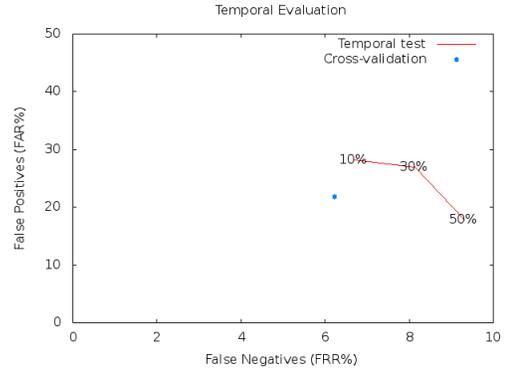


Figure 7: Temporal testing evaluation with ensemble of OCSVMs

found in Web browsing, none of the individual features by themselves are strong enough to identify someone but the global features (global characteristics of a session) rate the best overall. Moreover, the post/comment rate made a statistically significant difference in the FAR results for the global features. Interestingly, we note that subreddits are not the best identifiers overall maybe because of the commonality of certain subreddits (e.g., AskReddit, films, videos, etc). Rather, each user has personally distinguishable traits. Other approaches that have been proposed for Web browsing (Abramson 2014; Abramson and Gore 2013) will be adapted to social media behavior.

## 4 Conclusion

While the Turing test proposed to discriminate between humans and machines in general, there is a need for a personalized Turing test discriminating between thinking entities, whether humans or machines, with a cognitive fingerprint. The next challenge is to move from the evaluation of uniqueness to the inference of high-level features such as personality and intent which could further refine a cognitive fingerprint model. The similarity of results between Web browsing and social media behaviors suggests the possibility to cross-reference multiple domains in order to merge different aspects of identity into a coherent cognitive fingerprint and infer patterns of interest (e.g., fraud).

Other challenges in the determination of a cognitive fingerprint include challenges related to keyhole plan recognition – how to infer the goal from a series of observed actions – and the possibility of deception to defeat pattern matching algorithms.

## Acknowledgment

We want to acknowledge the work of Shantanu Gore in the acquisition of the Reddit dataset and the feedback of Michael P. Stein.

## References

Abramson, M., and Aha, D. W. 2013. User authentication from web browsing behavior. In *Florida Artificial Intelli-*

gence Society FLAIRS-26.

Abramson, M., and Gore, S. 2013. Associative patterns of web browsing behavior. In *2013 AAAI Fall Symposium Series*.

Abramson, M. 2014. Learning temporal user profiles of Web browsing behavior. In *6th ASE International Conference on Social Computing (SocialCom '14)*.

Bojinov, H.; Sanchez, D.; Reber, P.; Boneh, D.; and Lincoln, P. 2014. Neuroscience meets cryptography: crypto primitives secure against rubber hose attacks. *Communications of the ACM* 57(5):110–118.

Guidorizzi, R. 2012. Active authentication DARPA program. [http://www.darpa.mil/Our\\_Work/I2O/Programs/](http://www.darpa.mil/Our_Work/I2O/Programs/).

Guidorizzi, R. P. 2013. Security: Active authentication. *IT Professional* 15(4):4–7.

Jain, A.; Bolle, R.; and Pankanti, S. 1999. *Biometrics: personal identification in networked society*. kluwer academic publishers.

Juola, P.; Necker, J. I.; Stolerman, A.; Ryan, M. V.; Brennan, P.; and Greenstadt, R. 2013. Keyboard-behavior-based authentication. *IT Professional* 15(4):4–7.

Matthews, R. 1999. The power of one. *New Scientist* 163(2194):26–30.

Patel, V. M.; Yeh, T.; Fathy, M. E.; Zhang, Y.; Chen, Y.; Chellappa, R.; and Davis, L. 2013. Screen fingerprints: a novel modality for active authentication. *IT Professional* 15(4):4–7.

Salem, M. B., and Stolfo, S. J. 2011. Modeling user search behavior for masquerade detection. In *Recent Advances in Intrusion Detection*, 181–200. Springer.

Schölkopf, B.; Williamson, R.; Smola, A.; Shawe-Taylor, J.; and Platt, J. 2000. Support vector method for novelty detection. *Advances in neural information processing systems* 12(3):582–588.

Wheeler, J.; Varner, D.; Carrola, J.; Dahlberg, C.; Thornton, T.; Bohil, C.; and Terry, K. 2013. Covert cognitive games and user response patterns. *IT Professional* 15(4):4–7.