Poster: The Side Channel Menagerie

John V. Monaco
U.S. Army Research Laboratory
Contact: www.vmonaco.com

***Abstract*—Temporal keylogging attacks exploit between-subject similarities in typing behavior to recognize which keys were pressed based only on key-press timings. However, due to differences in behavior, such an attack may work well for some users and not for others. We examine temporal keylogging performance from the perspective of keystroke biometrics and attempt to establish some preliminary linkages between behavioral biometrics and human-based side channel attacks.**

I. Background and problem

Temporal keylogging attacks, as introduced by Song et al. [1], attempt to recognize which keys a user typed on a keyboard based only on the key press and release timings. Such an attack may target client/server applications that emit network traffic in response to keyboard events, such as SSH in interactive mode or a web search query box with autocomplete functionality. In both applications, the client responds to key-press events by immediately emitting a burst of network traffic, which, when observed remotely, reveals key-press timings. Keyboard interrupt events may also be detected from within a sandboxed environment [2] or on a multi-user system [3]. Consequently, the ability to recognize which keys were typed based only on event timings is a major threat to user privacy.

Such temporal keylogging attacks exploit the similarity with which different people type on a keyboard [4]. The touch typist is generally quicker to press keys that are far apart compared to keys that are close together, a consequence of having to reposition the hand or finger to strike neighboring keys. This inverse scaling between key-distance and key-press latency is common among touch typists [5], enabling general inferences to be made about which keys were pressed based only on the time interval between key-presses.

At the same time, between-subject differences in typing behavior enable its use in biometric applications. Keystroke biometrics exploits the differences with which different people type on a keyboard to perform user identification and authentication. It is also well-known that the performance of such a biometric authentication system will vary from user to user. This phenomenon is referred to as the biometric menagerie [6]. For many users (the "sheep"), the system works quite well; some may have difficulty authenticating (the "goats"); others may be vulnerable to impersonation (the "lambs").

Reminiscent of the biometric menagerie, we observe differences in temporal keylogging performance across users; the attack works well for some users (the side channel "lambs") and not so for others (the side channel "goats") [4]. It is not yet well understand which users are more or less vulnerable to this kind of human-based side channel attack. This abstract is a first attempt to reconcile these seemingly contradictory applications and develop user-specific criteria for side channel attack performance. We aim to determine whether the biometric "animal types" carry over to a side channel attack and establish some preliminary linkages between behavioral biometrics and side channels targeting human behavior.

II. Method

We evaluate keystroke biometric authentication performance and temporal keylogging performance for each of 1060 Amazon Mechanical Turk users in a public dataset [7]. Users were required to write truthful and deceptive short essays of at least 100 words. Each user provided between 2 and 6 freely-typed samples with 946±358 keystrokes per sample.

A. Biometric authentication

We characterize each writing sample by a 50-element feature vector. First, we determine the 50 most frequent bigrams (length-2 key sequence) across the entire dataset. For each sample, we construct a feature vector comprised of the mean key-press latency (time interval from key-press to key-press) for each bigram. If the sample contains less than 3 occurrences of a particular bigram, then the mean key-press latency of all bigrams in the sample is used. This form of feature fallback (called "backoff" in linguistics) is necessary for freely-typed samples since some bigrams may not occur within a sample.

To simulate authentication, we retain a single sample from each user as the template and the remaining samples from all users as queries. Genuine match scores are obtained by making within-subject comparisons and impostor match scores through between-subject comparisons. The match score is the negative Manhattan distance from query sample to template.

We obtain scores through a stratified 4-fold cross validation to ensure that each user appears in each fold. This procedure yields 4.5k genuine scores, 4.3M impostor scores, and overall 29.7% equal error rate (EER). Note a relatively high EER is obtained due to utilizing only key-press latencies; if key-release features are considered, such as duration and release-press latency, a 12.2% EER is obtained. The objective in this work is to examine the relationship between keystroke biometric match scores and temporal keylogging performance.

B. Temporal keylogging

A word-based temporal keylogging attack is simulated for each user. The attack scenario assumes that the key-press timings of a single word are exposed. The attacker attempts to determine what word the user typed by comparing the

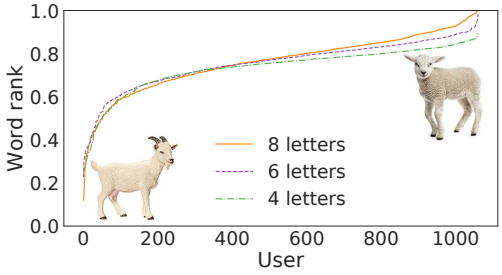

Fig. 1. Temporal keylogging performance (avg. word rank) for each user; chance is 0.50. Goats are invulnerable to attack and lambs are vulnerable.

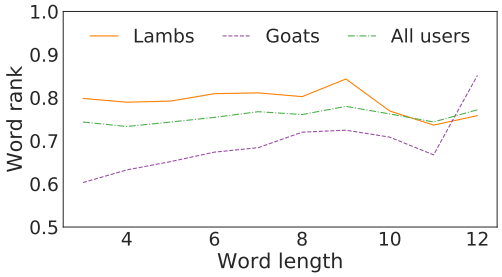

Fig. 2. Temporal keylogging performance for different biometric animal types. Biometric goats (lambs) are side channel goats (lambs).

sequence of key-press latencies to words in a dictionary without any prior knowledge of the user's typing behavior. Thus, the attack assumes word delineation is provided and that the typed word is contained in the dictionary.

The dictionary is constructed by taking the mean key-press latency vector of each word in the training dataset that occurs at least 10 times. Only words between 3 and 12 letters (inclusive) are considered. We use the *word rank* to evaluate attack performance. For each query, the dictionary words are ranked by their distance to the query word, again using negative Manhattan distance. At each word length, the average word rank for each user is determined. Note that chance is 0.5, i.e., the expected word rank from a random permutation, and perfect word recognition accuracy is achieved with rank 1.0. The average word rank for each user is determined through a group 1060-fold validation procedure to ensure that each user appears in only one fold, i.e., the user's samples do not appear in the training dataset. This procedure yields about 0.75 average word rank for all users and word lengths.

III. Results

The average word rank per user is shown in Figure 1. While some users are relatively invulnerable to attack (side channel "goats", low word rank), other users are much more vulnerable (side channel "lambs", high word rank). This scenario is reminiscent of the biometric menagerie, whereby some users have difficulty authenticating (biometric "goats"), and other users are vulnerable to impersonation (biometric "lambs").

To determine whether the biometric animal types carry over to a side channel attack, we evaluate temporal keylogging performance for the top-50 biometric goats and lambs. Goats are those users who have difficulty authenticating due to relatively low genuine match scores. Ranking the biometric templates by their average match score, we determine word rank for the 50 templates with lowest match scores. The same process is repeated for the lambs, instead taking the 50 templates with the highest impostor match scores. Shown in Figure 2, word rank for the biometric goats is lower than all users (up to 11 letters) while biometric lambs are more vulnerable than all users (up to 9 letters).

This relationship is verified by the examining the linear correlation between genuine/impostor match scores and word rank. For 4-letter words, Pearson's r between genuine match scores and word rank is 0.33 ($p = 1.7 \times 10^{-28}$); between impostor match scores and work rank, Pearson's r is 0.60 ($p = 6.9 \times 10^{-104}$). Significant correlation of genuine scores indicates that biometric "goat-like" users are less susceptible to attack. Alternatively, significant correlation of impostor scores indicates that biometric "lamb-like" users, who are easy to impersonate, are more vulnerable to attack. These relationships are significant up to 8 and 11 letters, respectively.

IV. Conclusion

Preliminary results suggest that some of the biometric "animal types" do carry over to a side channel attack. Lambs are users who are similar to other users, making them more vulnerable to impersonation in a biometric authentication system, and at the same time, more vulnerable to temporal keylogging attack. Goats are users who exhibit greater variability between samples, which degrades the usability of a biometric authentication system, and at the same time, makes them less vulnerable to temporal keylogging attack. In this regard, homogeneity in behavior and similarity to other users is an indication of vulnerability to side channel attack, i.e., it's good to be a side channel goat.

Future work will establish user-specific factors that affect both systems and examine the deeper relationship between behavioral biometrics and human-based side channel attacks: behavioral differences enable biometric identification and behavioral similarities enable human-based side channel attacks.

References

[1] D. X. Song, D. Wagner, and X. Tian, "Timing analysis of keystrokes and timing attacks on ssh.," in *Proc. Usenix Security Symp.*, vol. 2001, 2001.

[2] M. Schwarz, C. Maurice, D. Gruss, and S. Mangard, "Fantastic timers and where to find them: High-resolution microarchitectural attacks in javascript," in *Proc. 21st Intl. Conf. on Financial Cryptography and Data Security (FC)*, p. 11, IFCA, 2017.

[3] K. Zhang and X. Wang, "Peeping tom in the neighborhood: Keystroke eavesdropping on multi-user systems," *analysis*, vol. 20, p. 23, 2009.

[4] J. V. Monaco, "Sok: Keylogging side channels," in *Proc. IEEE Symp. on Security & Privacy (SP)*, IEEE, 2018.

[5] J. V. Monaco *et al.*, "Spoofing key-press latencies with a generative keystroke dynamics model," in *Proc. IEEE 7th Intl. Conf. on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–8, IEEE, 2015.

[6] N. Yager *et al.*, "The biometric menagerie," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 220–230, 2010.

[7] B. Ritwik *et al.*, "Keystroke patterns as prosody in digital writings: A case study with deceptive reviews and essays," in *Proc. Conf. on Empirical Methods in Natural Language Processing*, (Doha, Qatar), Association for Computational Linguistics, October 2014.