Resolving Rubber Hose Attacks Using Implicit Learning

Shuyang Wang
shuyang@cs.ucsb.edu
Department of Computer Science
University of California Santa Barbara

June 13, 2018

Abstract

Rubber Hose Attacks are when the user is forcibly asked by the attacker to reveal the key. Many cryptographic schemes are vulnerable to this type of attack. This paper will survey a new method that uses implicit learning based on neuroscience techniques and cognitive psychology to prevent this type of coercive attacks. Participants are projected to be trained to play a game online repeatedly, and the computer will record the participant's style of playing, and use the knowledge for authentication. In this paper, the neuroscience aspect of the technique will be explained briefly, and a detailed description of the experiment will be presented.

1 Introduction

1.1 Rubber-hose Attack

Many security schemes revolve around mathematical or technical cryptanalytic attacks, making sure a secret key is impossible or extremely hard to be obtained by exhaustive calculations. However, there exists another kind of cryptoanalytic attacks, rubber-hose attacks, where the person who possesses the secret key will be forced to reveal the secret by coercion or torture, such as beating someone with a rubber hose. A new proposal for solving this problem came forward in 2012 [1]. The paper explained how it's using implicit learning concept from neuroscience to prevent the rubber-hose attack.

1.2 The Implicit Learning Game

There are multiple memory systems in human brain. Explicit memory, which is usually used in verbally expressible facts, events or episodes, depends on medial temporal lobe memory system including the hippocampus. In contrast to explicit memory, implicit memory refers to the learning of skills. Implicit learning is believed to involve te part of the brain called the basal ganglia [2]. The implicit knowledge learned is not consciously accessible to the person being trained. An example learning task would be a person learns to ride a bike by repeatedly trying to repeat the task.

In the paper [1], implicit learning is used in creating a coercion-resistant security systems. Users of this system are asked to do a specific task called Serial Interception Sequence Learning (SISL) explained in the next subsection via a computer game, and this training will plant a password in the human brain that can be detected during authentication, but cannot be explicitly described by the user. Therefore, under coercion, the user will not be able to reveal the password even if they want to.

1.3 The SISL Task

SISL was initially introduced in [3]. It utilizes human brain's implicit learning will develop sensitivity to structured information without being aware of what has been learned. The task is much like the popular game "Guitar Hero", where there are four different columns with an object falling in constant speed on each column until it reaches the bottom and disappears. The player needs to intercept the objects delivered in a predetermined sequence. A successful interception is performed by pressing the key that corresponds to the object's column when the object is in the correct position. Not pressing the key in time or pressing the wrong key will result in a failure.

In the task set up for this paper, the number of columns are increased to six. The game is designed to prevent conspicuous, easy to remember patterns. To do this, sequences are designed to contain every ordered pair of characters exactly once with no character appearing twice in a row. Therefore, the sequence length will be 6*5 = 30 when 6 characters are used to represent each column. The paper later demostrated that a participant will perform better on the trained sequence than an untrained sequence by a significant difference, and the participant does not consciously recognize the trained sequence.

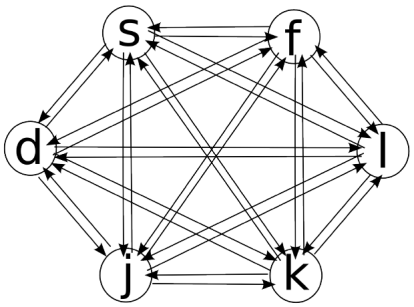

Figure 1: The secret key we generate is a random 30-character sequence from the set of Euler cycles in this directed graph. The resulting sequence contains all bigrams exactly once, excluding repeating characters.

2 Identification System

The SISL task provides a way to store a secret key within the human brain and can manifest itself during authentication, but cannot be explicitly described by the user. The identification system includes two parts, training and authentication. In the training phase, a sequence of 30 characters over the set $S = \{s, d, f, j, k, l\}$ will be learned by the user and be used as secret key. Following the rule that each sequence needs to contain every ordered pair of the characters exactly once with no character appearing twice in a row, an Euler cycle (a cycle where every edge appears exactly once) is used to describe all the possible cycles, shown in Figure 1: The number of the possible secret keys are given by the BEST theorem [4].

$$\#keys = 6^4 * 24^6 \approx 2^{37.8} \approx 2.48 * 10^{14}, \tag{1}$$

which is considered enough entropy for a safe password.

2.1 Training

Let Σ denote the set of all possible secret key. Users learn a random 30-item secret key $k \in \Sigma$ by training in the SISL game in a trusted environment. The procedure is described as follows: the 30-item secret key sequence is repeated three times and then combined with 18 items selected at random from another sequence, making a 108-item sequence. Then the sequence is repeated five time, making a 540-item sequence. At the end of this sequence there is a short pause in the SISL game and then the 540-item sequence is repeated six more times, making the entire training session 3780-item long.

2.2 Authentication

To authenticate in the system at a later time, the trained user is presented a sequence that contains elements from the trained authentication sequence and untrained sequence, where performance are being measured and contrasted. The setup is as follows: Let k_0 be the trained 30-item sequence and let k_1, k_2 be two additional 30-item sequences chosen at random from Σ. The system chooses a random permutation π of $(0, 1, 2, 0, 1, 2)$ and presents the user with the following sequence of 540 = 18*30 items:

$$k_{\pi_1}, k_{\pi_1}, k_{\pi_1}, ..., k_{\pi_6}, k_{\pi_6}, k_{\pi_6} \tag{2}$$

Therefore, each of k_0, k_1, k_2 is shown to the user exactly six times, but ordering is random. For $i = 0, 1, 2$ let p_i be the fraction for the correct keys the user entered during all plays of the sequence k_i, The system defines authentication as successful if

$$p_0 > average(p_1, p_2) + \sigma \tag{3}$$

where $\sigma > 0$ is large enough to minimize the possibility that this gap occurred by chance, but without causing authentication failures.

3 Performance

Many experiments are done to verify that 1) The trained user can successfully finish the authentication task over time; 2) attackers cannot coerce the user to reveal the secret key even if the user wants to. The following subsections will describe the result briefly.

3.1 Implicit and Explicit Knowledge

Participants have an average rate of 79.2% correct for the trained sequence and 70.6% correct for the untrained sequence. The difference 8.6% indicated reliable better performance. Figure 2 shows that as more training blocks are executed, the trained advantage also grows, showing that participants gradually begin to express knowledge of the repeating sequence.

3.2 Long-term Effect

In this experiment, the participants came back for a second authentication session after one week and two weeks respectively. Figure 3 shows that although the advantage decrease a little compared to authentication immediately after training, there is still

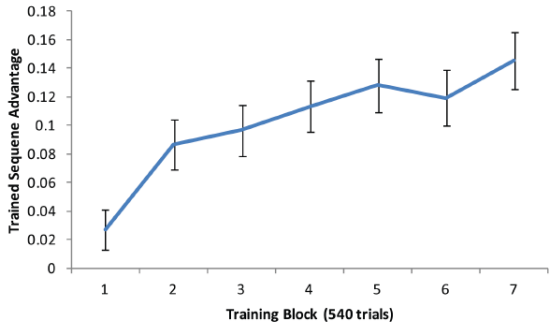

Figure 2: Trained Sequence Advantage

a clear advantage . Participants can still gradually begin to express knowledge of the repeating sequence by exhibiting a performance advantage for the trained sequence.

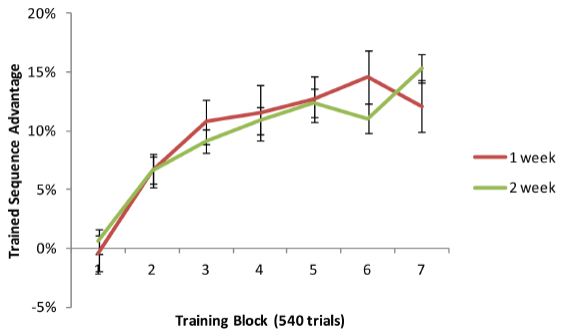

Figure 3: Recall Ability Over 1 week and 2 weeks

4 Security Analysis

4.1 Threat Model

Now that the usability of the system can be established, the security feature of this system are being examined. Implicit learning provides the following new abstract functionality: the training phase embeds a predicate

$$p : \Sigma \to \{0, 1\} \tag{4}$$

in the user's brain for some large set Σ. For $k \in \Sigma$, the predicate evaluates to 1 when k has been learned by the user and evaluates to 0 otherwise.

The adversary is assumed to have the following properties:

- The adversary will coerce the trained user to reveal the secret.
- The adversary's goal is to pass the test.
- The adversary has a single chance at the authentication test.
- The adversary need to be present to test the system.

Assume that the training procedure embedded an implicit learned predicate p in the user's brain, and the attacker intercept u trained users and subjects each one to q queries. His chance of finding a valid sequence is at most $qu/|\Sigma|$. If each test takes five minutes, the upper bound for each captured user is $q = 10^5$. If $u = 100$, the probability of the attacker succeeding is

$$100 * 10^5/|\Sigma| \approx 2^{-16} \tag{5}$$

One thing to notice is that this authentication system will still be fragile to the traditional eavesdropping attacks, so other measures must be taken to prevent eavesdropping attacks.

4.2 Coercion detection

If the attacker managed to coerce a user to authenticate while they are under the attacker's control, our current method will not work. However, through implicit learning, we can add more control to monitor the behaviors of user and thus detect if coercion happened during authentication. If the pattern of a user authenticating under stress can be recorded and recognized, then this pattern can be used in coercion detection. Also with other methods such as video monitoring, voice stress detection and skin conductance monitoring [5][6][7], the authentication system will have a more well-rounded performance.

5 Conclustion

A counter-measurement of rubber-hose attack is surveyed in this report. The idea concept of implicit learning can be a very powerful tool in modern cryptography based on the reliability of human memory system. Although this system has not been used in real-life authentication process, it shows a promising future with further work.

References

[1] H. Bojinov, D. Sanchez, P. Reber, D. Boneh, and P. Lincoln, Neuroscience meets cryptography: designing crypto primitives secure against rubber hose attacks. In Proceedings of the 21st USENIX Security Symposium (2012).

[2] A. Destrebecqz and A. Cleeremans. Can sequence learning be implicit? new evidence with the process dissociation procedure. *Psychonomic Bulletin and Review*, 8:343–350, 2001.

[3] D. Sanchez, E. Gobel, and P. Reber. Performing the unexplainable: Implicit task performance reveals individually reliable sequence learning without explicit knowledge. *Psychonomic Bulletin and Review*, 17:790–796, 2010.

[4] T. van Aardenne-Ehrenfest and N. G. de Bruijn. Circuits and trees in oriented linear graphs. *Simon Stevin*, 28:203–217, 1951.

[5] Payas Gupta and Debin Gao. Fighting coercion attacks in key generation using skin conductance. In USENIX Security Symposium, pages 469–484, 2010.

[6] Robert Ruiz, Claude Legros, and Antonio Guell. Voice analysis to predict the psychological or physical state of a speaker, 1990.

[7] J. Benaloh and D. Tuinstra. Uncoercible communication. Technical Report TR-MCS-94-1, Clarkson University, 1994.