Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Robustness of keystroke-dynamics based biometrics against synthetic forgeries☆

Deian Stefan[a], Xiaokui Shu[b], Danfeng (Daphne) Yao[b,*]

[a] Department of Electrical Engineering, The Cooper Union, New York, NY 10003, United States
[b] Department of Computer Science, Virginia Tech, 2202 Kraft Dr, Blacksburg, VA 24060, United States

ARTICLE INFO

Article history:
Received 21 May 2011
Received in revised form
22 August 2011
Accepted 4 October 2011

Keywords:
Keystroke dynamics
Authentication
Malware detection
Forgery
Bot
Attack
Classification

ABSTRACT

Biometric systems including keystroke-dynamics based authentication have been well studied in the literature. The attack model in biometrics typically considers impersonation attempts launched by human imposters. However, this attack model is not adequate, as advanced attackers may utilize programs to forge data. In this paper, we consider the effects of *synthetic forgery attacks* in the context of biometric authentication systems. Our study is performed in a concrete keystroke-dynamic authentication system.

The main focus of our work is evaluating the security of keystroke-dynamics authentication against synthetic forgery attacks. Our analysis is performed in a remote authentication framework called TUBA that we design and implement for monitoring a user's typing patterns. We evaluate the robustness of TUBA through experimental evaluation including two series of simulated bots. The keystroke sequences forged by the two bots are modeled using first-order Markov chains. Support vector machine is used for classification. Our results, based on 20 users' keystroke data, are reported. Our work shows that keystroke dynamics is robust against the two specific types of synthetic forgery attacks studied, where attacker draws statistical samples from a pool of available keystroke dataset other than the target.

We also describe TUBA's use for detecting anomalous activities on remote hosts, and present its use in a specific cognition-based anomaly detection system. The use of TUBA provides high assurance on the information collected from the hosts and enables remote security diagnosis and monitoring.

1. Introduction

Keystroke-dynamics based authentication is a cheap biometric mechanism that has been proven accurate in distinguishing individuals (Bleha et al., 1990; Ilonen, 2003; Killourhy and Maxion, 2008; Monrose and Rubin, 2000; Song et al., 2001; Yu and Cho, 2003). Most of the attack models considered in keystroke-dynamics literature assume the attackers are humans, e.g., a colleague of Alice trying to log in as Alice. However, there has been little effort on studying the robustness of this technique against synthetic and automatic attacks and forgeries.

We evaluate the robustness of keystroke-based biometric authentication systems against a new type of forgery attacks. In the context of biometrics, a synthetic forgery attack is carried out by submitting generated or synthesized credentials to an authentication module. For example, an attacker writes a program that performs statistic manipulation and synthesis to produce keystroke sequences in

☆ This work was supported in part by Rutgers University DIMACS REU programs, National Science Foundation grants CNS-0831186 and CAREER CNS-0953638. Stefan is currently a graduate student at Stanford University.

* *Corresponding author.*
E-mail addresses: stefan@cooper.edu (D. Stefan), danfeng@cs.vt.edu, subx@cs.vt.edu (D. (Daphne) Yao).

0167-4048/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cose.2011.10.001

order to spoof others. These types of forgery attacks pose a serious threat. However, the research community has not extensively investigated on possible anti-forgery techniques. It is unclear from the current literature how robust keystroke dynamics is against forgery attacks. Synthetic forgery attacks may also be possible in other types of biometric systems as well.

The technical enabler for our investigation is a remote authentication framework that we design and implement. The framework called TUBA (Telling hUman and Bot Apart) monitors a user's typing patterns in a client-and-server architecture. We systematically study the robustness of TUBA through comprehensive experimental evaluation including two simulated bots. We perform a user study with 20 users and use the collected data to simulate and evaluate the difficulty and impact of synthetic forgeries.

Another contribution of this paper is that we describe the use of TUBA and keystroke dynamics to identify anomalous activities on a personal computer, e.g., activities that may be due to malware. We consider a model where a user's computer in an organization or enterprise may be infected with malicious software that may stealthily launches attacks. This model is motivated by the increasing number of infected hosts caused by organized malicious botnets. Our solution provides strong assurance of authentication results. We provide a practical solution that effectively allows a remote trusted server to monitor the integrity of a computer. The main application of TUBA is to detect stealthy malware residing on a user's computer such as application-level spyware.

Our study uniquely combines techniques from system and network security, biometrics, machine learning, and usability engineering. Our technical contributions are summarized as follows.

1. We design and implement a simple and easy-to-adopt protocol for authenticating a computer owner that utilizes the user's keyboard activities as an authentication metric. We present our protocol in a lightweight client-server architecture using the X Windows System (X11 for short).
2. We analyze the keystroke data from a group of users on a diverse set of inputs, including email addresses, a password, and web addresses. We find that performance results vary according to the strings used for authentication. We find that different types of strings give different classification accuracy when used for authentication.
3. We evaluate the robustness of keystroke-dynamics based authentication against automated bot attacks. We implement two bot programs, called *GaussianBot* and *NoiseBot*, respectively, which are capable of injecting statistically-generated keystroke event sequences on a (victim) machine. The bot programs aim to pass our keystroke authentication tests by mimicking a particular user's keystroke dynamics. The bots are capable of launching forgery attacks drawn upon the statistical analysis of collected keystroke data. Experiments show that our classification is robust against these specific attacks, and is able to correctly classify the attacks by GaussianBot and NoiseBot with low false positive rates. The GaussianBot and NoiseBot forge keystroke sequences following simple first-order Markov models.

TUBA is particularly suitable for detecting extrusion in enterprises and organizations, and protecting the integrity of hosts. Our work gives the indication that certain human behaviors, namely user inputs, may be suitable for malware detection purposes. We also give examples that illustrate the prevention of malware forgery in such human-behavior driven security systems. This study is the result of an on-going effort towards designing human-inspired security solutions. Our work also suggests the need for studying the robustness of other biometrics against synthetic forgery attacks beyond the studied keystroke-authentication problem. Because of the wide use biometrics in government, military, and enterprise environments, the better understanding of their security against sophisticated attacks is important.

1.1. *Organization of the paper*

We describe our design of a remote authentication framework and our security model in Section 2, where a use case of using TUBA to detect anomalous network activities is also described. Details of our implementation including data collection, keystroke logging, feature extraction, and classification can be found in Section 3. We implement two bots that are capable of injecting synthetic keystroke events, which are presented in Section 4. Our experimental evaluation results and user study are described in Section 5. A specific application of TUBA for liveliness detection as well as an open problem are presented in Section 6. Related work is described in Section 7. In Section 8, we conclude the paper and describe plans for future work.

2. Overview and security model

TUBA is a remote biometric authentication system based on keystroke-dynamics information. We use machine-learning techniques to detect intruders merely based on keystroke dynamics, i.e., timing information of keyboard events. We allow for certain types of key event injection by bots.

2.1. *Security assumptions and malware attack model*

We assume that the host operating system kernel, our client-side keystroke-collection modules, and cryptographic keys are secure and not compromised. The remote server for issuing keystroke challenge and data analysis is trusted and secure. Client-side malware may run as a user-level application, e.g., spyware implemented as Firefox extensions. Malware is active in making outside connections for command & control or attacks. We allow malware to inject arbitrary keystroke events and sequences synthesized from the data other than that of the owner. Thus, under this malware-attack model we assume that keylogging by spyware or by human intruders (Zhang and Wang, 2009) on the owner's computer does not exist. An attacker may also carry out conventional network attacks such as eavesdropping on the communication channel between client and server, or replaying network packets.

We note that with hardware chip TPM (trusted platform module) enabled, fake key events can be detected and removed