COMPUTERS & SECURITY 53 (2015) 215–233

Available online at www.sciencedirect.com

ScienceDirect

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

Computers & Security

Reconciling user privacy and implicit authentication for mobile devices☆

Siamak F. Shahandashti [a,*], Reihaneh Safavi-Naini [b,*], Nashad Ahmed Safa [b]

[a] School of Computing Science, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom
[b] Department of Computer Science, University of Calgary, 2500 University Drive NW, Calgary T2N 1N4, Canada

ARTICLE INFO

Article history:
Received 20 December 2014
Received in revised form
28 April 2015
Accepted 22 May 2015
Available online 5 June 2015

2010 MSC:
94A60
94A62
68P25

Keywords:
Implicit authentication
User privacy
Homomorphic encryption
Provable security
Behavioural features

ABSTRACT

In an implicit authentication system, a user profile is used as an additional factor to strengthen the authentication of mobile users. The profile consists of features that are constructed using the history of user actions on her mobile device over time. The profile is stored on the server and is used to authenticate an access request originated from the device at a later time. An access request will include a vector of recent measurements of the features on the device, that will be subsequently matched against the features stored at the server, to accept or reject the request. The features however include private information such as user location or web sites that have been visited. We propose a *privacy-preserving implicit authentication* system that achieves implicit authentication without revealing information about the usage profiles of the users to the server. We propose an architecture, give a formal security model and a construction with provable security in two settings where: (i) the device follows the protocol, and (ii) the device is captured and behaves maliciously.

1. Introduction

In applications such as mobile commerce, users often provide authentication information using Mobile Internet Devices (MIDs) such as cell phones, tablets, and notebooks. In most cases, password is the primary method of authentication. The weaknesses of password-based authentication systems, including widespread use of weak passwords, have been widely studied (see e.g. (Tsai et al., 2006) and references within). In addition to these weaknesses, limitations of user interface on MIDs results in an error-prone process for inputting passwords, encouraging even poorer choices of password by users.

Two-factor authentication systems can potentially provide higher security. Second factors that use special hardware such

☆ This is an extended version of a paper that appeared in the proceedings of the 29th International Information Security and Privacy Conference IFIP SEC 2014 (Safa et al., 2014).

* *Corresponding authors.*
E-mail addresses: siamak.shahandashti@ncl.ac.uk (S.F. Shahandashti), rei@ucalgary.ca (R. Safavi-Naini).

http://dx.doi.org/10.1016/j.cose.2015.05.009
0167-4048/

as RSA SecurID[1] tokens or biometrics, incur additional cost which limit their wide usage. An attractive method of strengthening password systems is to use *implicit authentication* (Jakobsson et al., 2009) as an additional factor for authentication. The idea is to use the history of a user's actions on the device, to construct a profile for the user consisting of a set of features, and employ it to verify a future authentication request. In the authentication phase, the device reports recent user behaviour, and authentication succeeds if the reported recent user behaviour "matches" her stored profile. Experiments by Jakobsson et al. (2009) showed that the features collected from the device history can be effectively used to distinguish users. Although the approach is general, it is primarily used to enhance security of mobile users carrying MIDs because of the richness of sensor and other data that can be collected on these devices. In such a scenario, a network service provider (the carrier) wishes to authenticate a user in possession of the MID.

An important distinction one needs to make is that the goal in implicit authentication is authenticating the user in possession of the device rather than the device itself. Consequently, the user profile needs to be stored at the carrier side to ensure that a compromised device cannot be used to impersonate the legitimate user.

The collected data about the user's actions can be divided into the following categories: (i) device data, such as GPS location data, WiFi/Bluetooth connections, and other sensor data, (ii) carrier data, such as information on cell towers seen by the device, or Internet access points, and (iii) third party data, such as cloud data, app usage data, and calendar entries. As discussed, the user profile including data from a mixture of these categories is stored at the carrier side. This profile however includes private and potentially sensitive user data, including device and third party data, that must be protected. One might be lead to think that there is an inherent trade-off between user privacy on one hand and the effectiveness of implicit authentication on the other. In this paper we show that this is a false trade-off; i.e., user privacy and effective implicit authentication can coexist. In particular, we propose an efficient privacy-preserving implicit authentication systems with verifiable security.

We consider a network-based implicit authentication system where user authentication is performed collaboratively by the *device* (the MID) and the *carrier* (network service provider), and will be used by *application servers* to authenticate users. The implicit authentication protocol generates a score for each feature representing the confidence level of the authentication based on that individual feature. Individual scores are subsequently combined based on the carrier authentication policy to accept or reject the user. Individual scores are obtained through a secure two party computation between the device and the carrier. Secure two party protocols can be constructed using generic constructions based on secure circuit evaluation, e.g. (Yao, 1986; Goldreich et al., 1987), or fully homomorphic encryption (Gentry, 2009). We however opt to design a special-purpose protocol fit for the type of computations needed in implicit authentication. This allows us to achieve a level of efficiency which is practical and higher than those provided by generic constructions.

1.1. Our contributions

We propose an implicit authentication system in which user data is encrypted and stored at the carrier, and an interactive protocol between the MID and the carrier is used to compute the authentication score. Data privacy is guaranteed since user data is stored in encrypted form. Because no data is stored on the MID, user data stays protected even if the device is lost or stolen. The main contributions of this paper are proposing a profile matching function that uses the statistics of features to accept or reject a new sample presented by a user, and designing a privacy-preserving protocol for computing a score function for newly presented data.

We assume the user profile is a vector of multiple features $(V_1, \ldots, V_n)$, each corresponding to a random variable with an associated probability distribution. Samples from the distribution of V_i is stored as the set of values of the variables in the last ℓ_i successful logins. A new log in attempt generates a vector of values, one for each feature. The verification function must decide if this vector indeed has been generated by the claimed user. Our proposed verification algorithm takes considers feature separately and computes a score for each feature indicating the confidence level in the presented value being from the claimed user. The final verdict is reached by combining the scores from all features.

To determine if a new value presented for a feature v_i matches the model (stored distribution of the feature), we will use a statistical decision making approach that uses the *Average Absolute Deviation (AAD)* of the distribution. We use AAD to define an interval around the reported value v_i given by $[v_i - \mathrm{AAD}(V_i), v_i + \mathrm{AAD}(V_i)]$ and then determine the score representing the concentration of past user behaviour observations close to the reported value v_i by counting the number of the stored values in the user profile that fall within the interval: the higher the number is the higher the score for that feature will be. Eventually the scores from all features are considered and the outcome of the authentication according to a certain policy is decided. AAD and standard deviation are commonly used statistical measures of dispersion, estimating the "spread" of a distribution. Our verification algorithm effectively measures similarity of the presented value with the "most common" readings of the variable. Using AAD allows more efficient private computation.

Constructing User Profiles. A user profile is a feature vector $(V_1, \ldots, V_n)$, where feature V_i is modelled by a vector of ℓ_i past samples. The vector can be seen as a sliding window that considers the latest ℓ_i successful authentication data. Using different ℓ_i is allowed for better estimation of the feature distribution. Possible features are the frequency of phone calls made or received by the user, user's typical locations at a particular time, commonly used WiFi access-points, websites that the user frequently visits, and the like. We survey the literature and find several features that are appropriate for our protocols. These are listed in Section 3.2. Some features might be dependent on other ones. For example, given that the user is in his office and it is lunch time, then there is a higher chance that he receives a call from home. We do not consider

[1] www.emc.com/security/rsa-securid.htm.