Am I Really Who I Claim to Be?

Olivier Coupelon[1], Diyé Dia[1,2], Yannick Loiseau[2], and Olivier Raynaud[2]

[1] Almerys, 46 rue du Ressort, 63967 Clermont-Ferrand
{olivier.coupelon,diye.dia}@almerys.com
[2] Blaise Pascal University, 24 Avenue des Landais, 63170 Aubière
{loiseau,raynaud}@isima.fr

Abstract. Faced with both identity theft and the theft of means of authentication, users of digital services are starting to look rather suspiciously at online systems. The behavior is made up of a series of observable actions of an Internet user and, taken as a whole, the most frequent of these actions amount to habit. Habit and reputation offer ways of recognizing the user. The introduction of an implicit means of authentication based upon the users' behavior allows websites and businesses to rationalize the risks they take when authorizing access to critical functionalities. This rationalization brings with it user trust as a result of increased security.

Keywords: Implicit authentication, user behavior, trust, reputation.

1 Introduction

An online survey of 772 Internet users carried out in February 2013 showed that the French are wary of digital services. To be more precise, the study concluded that 86% of Internet users use digital services, whereas trust is declining in e-commerce and in social networks, and is stabilizing in online banking [1]. Even so, trust is a powerful driving force for economic development [2], so we need to introduce ways of sustaining this trust by readjusting the relationship between users and digital services. In the field of computing, trust relates to security protocols designed to allow authentication of source of information [3]. In order to maintain users trust, we will need not only to incorporate historic security protocols into new application systems, but also to make them acceptable to users and take their use into account. The context for our study is a digital environment offering a set of services, including a digital safe[1]. This space is only accessible using strong authentication (smart card + PIN code), which is highly inconvenient for users. Indeed, this service space includes critical functionalities which must only be accessible to a user who has been authenticated and authorized to access them. Weak authentication (login + password) will soon become available. Although it is more convenient, it is less secure and lays the critical functionalities open to malicious use. The aim of this project is to make this weak authentication "stronger" by carrying out additional implicit authentication. If the means of authentication is lost or stolen, our study ensures that the system will

[1] http://www.ebeeoffice.ca/ebee-home/public

Á. Rocha et al. (eds.), *New Contributions in Information Systems and Technologies*,
Advances in Intelligent Systems and Computing 353, DOI: 10.1007/978-3-319-16486-1_60

be able to detect the intruder's presence and react. The unique feature of this project lies in studying a user's behavior on a given website for implicit authentication purposes. Our research will allow practical but secure access by means of implicit authentication, by answering the question: is the user who she claims to be? Indeed, others works explored the browsing behavior of users on the Web [4], for either identification or page recommendation purposes. The outline for the article is as follows: in section 2, we shall offer a state-of-the-art before moving on to propose a solution to the set of problems we are dealing with in section 3. Finally, in section 4, we shall discuss our future work.

2 Related Work

Trust in computer science is a huge subject and is dealt with in depth in [3]. Even so, we can identify a few major areas which are directly linked to our set of problems. By a "trust policy", we mean a whole series of authentication processes which may involve means such as an electronic signature or the exchanging of certificates. Trust based on the policy is defined as the trust obtained by adhering to the rules and constraints laid down by the trust policy. In the digital environment created by the business which provides the context for our study, the trust policy has already been set up. Trust built up based upon a reputation [5] is defined by the study of past interactions between the various components of a digital system. In our own context, a user who is logged into the digital environment and the services offered by this environment are the components of the system. The reputation system could draw upon the recommendations made by each service about a user. This area of study dealing with trust lies at the heart of our set of problems. In [6], the author takes a particularly close look at the representation of trust and the associated risks which are taken. So it will inspire us in order to rationalize our decision-making system. *Behavioral patterns* based on Web browsing sessions are studied in [4]. Two techniques for the construction of users' profiles are proposed: the *support* algorithm and the *lift* algorithm based upon the *a priori* method, in order to look for common behavioral models. These techniques are compared to decision-making trees and support vector machines (SVMs). Yang has some interesting results but recognizes that her method is limited. It can be applied only when there is a small number of a user. The [7] study resolves the problem when the system has a large number of users and proposes other methods to be used to build user profiles based upon their Web browsing histories. The authors compare the algorithms used by [4], the closest neighbor algorithm and the Naive Bayes Classification. For our problem, we will use these algorithms and compare them. *Implicit authentication*, as defined in [8], allows a user to be authenticated without asking her for a password, provided that she behaves in a way which matches her usual behavior. This definition is given in the context of mobile telephones. The applications found on the telephone only log out or freeze if the user's behavior is different from usual [9]. The behavior is studied based upon factors including browsing between the applications on the telephone (we can make an analogy with browsing on the Web), geographical location and the gait data from acceleration sensors of the telephone, amongst others [10]. In [10], factors that define the behavior are closely related to the telephone. For our study, we need to find relevant factors that are available whatever the device used. Implicit authentication is a kind of "trust policy".

Table 1. Means of authentication and service

Means of authentication	**Service levels (uncon-trolled risk-taking)**	**Service levels (uncon-trolled risk-taking)**
Strong (smart card + PIN code)	Unrestricted access	Level 3
Weak (login + password)	Restricted access	Level 1
Strong (smart card + PIN code) & implicit		Level 4
Weak (login + password) & implicit		Level 2

3 Approach and Methodology

Each service in the digital environment is made up of a set of functionalities, some of which are more critical than others (e.g. electronic signature of official documents). A business expert will sort these functionalities into various service levels, according to how sensitive they are, and then the user will have either partial or total access to the functionalities, according to their level. This classification can be carried out based upon the proposals in [9]. Table 1 summarizes the system which we plan to set up in the digital environment and includes the implicit authentication module. Level 1 offers access to basic, not critical functionalities. Low criticality functionalities are on level 2. Level 3 is made up of high criticality functionalities and finally level 4 involves very high criticality functionalities. All of the functionalities of a level are also to be found in the next one. The aim of our projects is to set up an implicit authentication module on a particular website. This module is made up of a user profile construction model and a decision-making model. The user profile is built incrementally from behavioral patterns which are discovered using sequence mining algorithms on the user's session's data. A session is defined by the logging in and out stages. The analysis begins as soon as the user logs into the digital environment. The profile is updated from session to session. Using the Weka application, we build test databases and experiment with the algorithms used in the literature [4] [7]. This work is currently underway. We analyze two case studies involving the use of the decision-making model. The module makes a decision for each level n+1 functionality call. If the behavior differs from usual, the system denies access to the requested functionality, and requires the user to identify her formally. In the second case, the decision-making model is constantly in operation. The system is constantly checking whether the user is actually who her claims to be and triggers an alert whenever the behavior is unusual. The modules reaction needs to be immediate in all cases. Depending upon the criticity and the certainty on user identity, the decision-making model will either take the risk of opening up access to a requested functionality or else triggering an alert. We will be looking at the relevant thresholds in order for the module to take the right decisions in 90% of cases. *Input data*, characterizing of each users behavior, are as follows: the session number, the dates and times that the user logged both in and out, which Web page inside the website the user visited, the date and time that the Web page visited was called and the IP address for the connection. Multiple visits to a single page and how many of them there were are also included.

In [4], the session number, the name of the site visited, the number of pages viewed, the session start time and length are used. The purpose of our additional data is to increase results' quality.

4 Conclusions and Future Work

The security of traditional means of authentication can be strengthened by setting up a new implicit authentication module. This module takes rational decisions based upon the user's behavior and reputation. We have modeled our set of problems and, in the work which still lies ahead, we are going to be using sequence mining algorithms to look for relevant behavioral models. One of our future works will be to adapt a reputation model and an inter-service recommendation system for the network layer [5] to the application layer in order to build up the user's reputation. The unique feature of our approach is the use of browsing data within a unique website. Our system needs to be able to react if attacked, so we need to optimize the systems learning period. We also need to set up a feedback system to evaluate the accuracy of the decision making module.

References

1. CDC-ACSEL, B.: Résultats de la 3ème édition du baromètre caisse des dépôts/acsel sur la confiance des français dans le numérique (2013)
2. Nissenbaum, H.: Securing trust online: wisdom or oxymoron? Boston University Law Review 81(3) 635-664 (2001)
3. Donovan, A., Yolanda, G.: A survey of trust in computer science and the semantic web. Web Semantics: Science, Services and Agents on the WWW 5, 58–71 (2007)
4. Yang, Y.C.: Web user behavioral profiling for user identification. Decision Support Systems (49), 261–271 (2010)
5. Basu, A.: A Reputation Framework for Behavioral History. PhD thesis (2010)
6. Marsh, S.P.: Formalizing Trust as a Computational Concept. PhD thesis (1994)
7. Herrmann, D., Banse, C., Federrath, H.: Behavior-based tracking: Exploiting characteristic patterns in dns traffic. Computer & Security 1-17 (2013)
8. Stockinger, T.: Implicit authentication on mobile devices. The Media Informatics Advanced Seminar on Ubiquitous Computing (2011)
9. Micallef, N., Just, M., Baillie, L., Kayacik, G.: Towards an app-driven mobile authentication model. In: Proceedings of the 9th Symposium on Usable Privacy and Security, UK (2013)
10. Shi, E., Niu, Y., Jakobsson, M., Chow, R.: Implicit authentication through learning user behavior. In: Burmester, M., Tsudik, G., Magliveras, S., Ilić, I. (eds.) ISC 2010. LNCS, vol. 6531, pp. 99–113. Springer, Heidelberg (2011)