Towards Privacy-Preserving Knowledge-based Authentication: A Bayesian Network Approach

Tahani Alsubait
tmsubait@uqu.edu.sa
College of Computer and Information Systems, Umm Al-Qura University, P.O.Box: 715, Makkah, Saudi Arabia

Abstract
Authentication is a cornerstone in secure systems aiming to restrict access to legitimate claimants only. Authentication systems can be generally classified into knowledge-based (e.g., passwords), token-based (e.g., credit cards), or biometric-based (e.g., fingerprints). In this paper, we discuss the strengths and weaknesses of each class of authentication approaches with an emphasis on privacy related issues. We survey and present the related literature showing a gap on addressing users' privacy concerns. We propose a Bayesian network approach for addressing and modelling privacy factors. We discuss the preliminary evaluation of the proposed approach. Recommendations for making privacy features more tangible and suggestions for future research directions are discussed.

Key words:
Privacy, Security, Knowledge-based authentication, Bayesian networks

1. Introduction

Authentication is the process carried out by secure systems to ensure authorized logins to the system and knowledge of the identity of the person logging in [1]. Different classifications have been suggested for authentication systems. For example. the National Institute of Standards and Technology (NIST) [2] identifies three main factors for distinguishing different authentication schemes: knowledge-based (i.e., what the user knows) such as using passwords, token-based (i.e., what the user owns) such as using credit cards, or biometric-based (i.e., what the user is) such as using fingerprints for authentication. Knowledge-based, token-based and biometric-based authentication are sometimes referred to as secret-based, object-based and inherence- or characteristic-based authentication, respectively. Moreover, knowledge-based authentication can be further classified into text-based, question-based and image-based systems.

Remote or on-line authentications are considered more challenging than face-to-face authentication as the former is not supervised and is largely uncontrolled [3]. With this respect, authentication plays a major role in online systems, as they help in establishing higher degrees of trust and credibility. Consequently, authentication is a cornerstone in many applications such as educational, medical, financial, and retail systems, to name a few.

Recently, privacy-preserving security systems have become trendy with an aim to seclude sensitive information about users. In this paper, we propose a Bayesian Network approach for privacy-preserving knowledge-based authentication. We first explain the main considerations of authentication systems, with privacy being one of them. Then we provide a motivational discussion of the importance of privacy-preserving authentication systems. We then present a brief review of related studies and then present the proposed approach. We conclude the paper with a discussion of the proposed approach's evaluation and suggest some future work directions.

2. Theoretical considerations of authentication schemes

There are many factors and considerations that impact our evaluation of authentication systems. Below, we discuss the three main considerations, namely, usability, security and privacy. These factors can overlap with one another and improving one factor may positively or negatively affects the other.

2.1 Usability

Multiple factors affect the usability of authentication approaches. For example, token-based authentication approaches which authenticate the identity of users based on owning a physical object may remove the cognitive burden on users as they are not required to make mental efforts to memorize a shared secret. In contrast, password-based authentication has several shortcomings in terms of usability such as the high cognitive load for memorizing different passwords [4]. On the one hand, both token-based and biometric-based approaches require the use of special equipment for identity identification, making them less usable compared to knowledge-based authentication. On the other hand, usability in knowledge-based authentication, described as the ability of a legitimate claimant to provide the required knowledge for authenticating their identity, might decrease as security measures increase [5].

Manuscript received April 5, 2020
Manuscript revised April 20, 2020

2.2 Security

Users’ concerns regarding the cognitive load associated with password memorizations may lead to security issues represented by bad practices in choosing weak passwords or making them accidentally accessible to undesired parties. A trade-of between security and usability in authentication systems is usually acknowledged [6]. Security aspects of authentication are concerned with denying access to non-legitimate claimants or attackers. Examples of security attacks include shoulder surfing attacks, social engineering attacks, phishing attacks, malware attacks, brute-force attacks, and dictionary attacks to name a few.

Security can be regarded as the most mature aspect of authentication systems. Many mechanisms have been suggested to overcome security problems related to authentication such as cryptographic hashes and captcha [7, 8]. However, password-based authentication is still vulnerable to many guessing attacks [5], therefore strict password policies are highly recommended.

2.3 Privacy

The domain of privacy can be looked at from different angles and it can partially overlap with the domain of security. Considerations related to privacy include but are not limited to protecting personal information used for authentication from theft or misuse, protecting users from being observed, preventing/limiting information gathering about users. In this paper, we put an emphasis on privacy, as concerns about it increase with the development of authentication systems relying heavily on knowledge about users.

For example, Ratha et al. [9] demonstrated that biometric-based approaches suffer from the issue of database cross-matching where users can potentially be tracked from one application to the next due to the uniqueness of biometrics.

Privacy aspects of authentication are concerned with making legitimate claimants the only ones who know or have access to the information required for authentication.

3. Why privacy-preserving authentication?

Previous studies have shown that users have multiple concerns regarding privacy in authentication systems. However, usability and security seem to be more cared about and addressed [10], compared to addressing users’ privacy concerns. In fact, privacy can shape user attitudes and intentions to use information systems [11]. Fortunately, there is an increase in the number of research articles related to privacy in authentication systems. Figure 1 shows the number of publications in the last 10 years as they appear when searching for privacy and authentication using IEEE Xplore digital library.

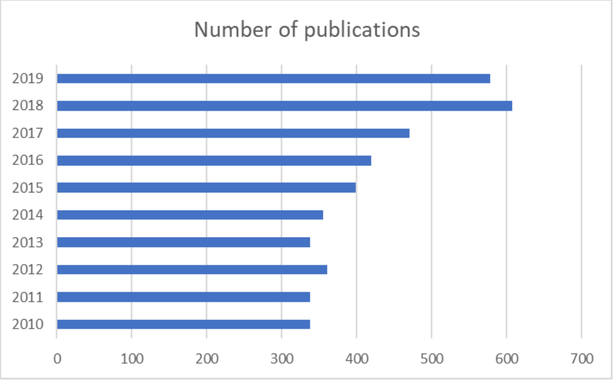

Fig. 1 Number of publications related to privacy in authentication systems in the period 2010-2019

4. Related work

A large volume of research was devoted for studying and developing user authentication approaches and techniques. Different studies have focused on different aspects of authentication with great interest in understating human factors that can have an impact on the effectiveness of authentication.

Ullah et al. [12] surveyed multiple users to investigate their privacy concerns regarding authentication in online examination environments. They proposed and implemented a Profile Based Authentication Framework (PBAF) for online examination. Online questionnaires were employed to gather participants' feedback on privacy and usability concerns. The results show that participants have a degree of concern regarding sharing personal and academic information with little or no privacy concern on using favorite questions.

Zimmermanna and Gerber [1] performed a laboratory study allowing 41 participants to interact with 12 different authentication systems to examine their perceptions. Their research findings show that users’ preference correlates positively with usability whereas there was no correlation between preference and privacy. A multiphase approach of using password followed by fingerprint authentication scored highest in terms of intention to use and lowest in terms of concerns regarding expected problems.

Hang et al. [13] present an evaluation of dynamic security questions for fallback authentication. Among many factors, they discuss privacy factors concerning the type to questions that can be presented to users without raising their privacy concerns. For example, they show that users do not prefer questions based on personal photos. They suggest that challenging questions presented to users during authentication should be about personal data, but not data that users would mind to expose to others.

Rabkin [14] examined security questions for a number of personal banking systems and outlines that they rely in part on the hardness of the information retrieval process which would diminish as personal information becomes ubiquitously available online.

Chen and Liginlal [15] state that knowledge-based authentication is a prominent approach for user authentication and propose a Bayesian network approach for this class of authentication. They define factoids as pieces of information provided by the claimant to authenticate themselves. They classify factoids to personal (e.g., favorite subject or sports team) or nonpersonal (e.g., job title), static (e.g., date of birth or mother's maiden name) or dynamic (e.g., last call or current employer). Their proposed approach utilizes maximum likelihood estimator (MLE) method for parameter estimation. They evaluated two Bayesian network structures for KBA, namely, the Naive Bayes (NB) and the Tree Augmented Naive Bayes (TAN), and show that the TAN achieves better authentication accuracy. The model focuses on memorability and guessability metrics of KBA, which are associated with usability and security respectively. This research proposes a Bayesian network approach that addresses the privacy metric of KBAs.

5. The proposed Bayesian network approach

This research is motivated by the need to quantify the degree of privacy factors when dealing with information used in the process of knowledge-based authentication. We tackle the problem by shedding light on different perspectives in the following subsections.

5.1 Knowledge sources

There are multiple data sources that can be utilized for KBA including student information, human resources information, credit information, demographics, national information databases, insurance information, mobile usage databases, airline databases and property ownership information. Other emerging knowledge sources include the use of information available on social media portals or smart phone applications' usage data. Moreover, specific system information such as previous logins and transactions may be utilized for authentication purposes. Relying on such knowledge sources has the advantage of making use of prior knowledge of the user, hence enhancing users' convenience and avoiding the inherent shortcomings of password-based authentication and the need for prior registration.

5.2 Factoids selection

Compared to the presumably ad-hoc ordering and selecting mechanisms for security questions presented in a number of exiting question-based authentication systems, in this research we propose a probabilistic method for question selection. The motivation behind this is that the selection mechanism impacts the three main factors of authentication schemes, namely, security, usability and privacy. For example, Chokhani [16] has suggested a random selection and ordering mechanism, lacking theoretical backing for the proposed design. In addition, choosing between static vs. dynamic questions may have an impact on the resulting levels of security, usability and privacy.

5.3 Network architecture

A Bayesian Network (NB) is a directed acyclic graph (DAG) which is composed of a set of vertices representing factors under consideration and a set of directed edges representing dependency relationships among factors [15]. BNs can be seen as generative probabilistic models which employs interdependent factoids taken from different knowledge sources resulting in an authentication decision ("true" for successfully authenticated users and "false" for unauthenticated users). An authentication decision can be made through Bayesian inference of a generated hypothesis.

Chen and Liginlal [15] suggest a BN model of KBA consisting of the following components:

The hypothesis related to the authenticity of a claimed identity represented as a class variable, with possible outcomes {true and false}

A subset of factoids constituting a challenge-response session represented as a vector, with possible outcomes {correct and wrong}.

A vector representing knowledge sources on which a KBA system relies, with possible outcomes {high, medium, and low} indicating the trustworthiness of data taken from the knowledge source.

A set of three directed edges representing dependency relationships between all factoids on the class variable y, knowledge sources on the associated factoids, and interdependencies among factoids.

We suggest to extend the model with two vectors: a vector representing the privacy class of each factoid with possible values {high, medium, and low} and a vector of schemas representing different groups of people with similar profiles (e.g., males, females, children, adults) with possible values of {true and false} indicating the pertinence to a given group. The motivation behind the latter is that certain groups of people are expected to have similar attitudes to privacy concerns. In addition, we add a dependency edge for the privacy class of each factoid and a given schema.

5.4 Model architecture

The architecture of the proposed KBA model is shown in Figure 2 below. It is composed of the following modules: Decision maker, model moderator, history keeper, and question generator. The decision maker evaluates user responses to challenge questions from different knowledge sources and estimate an authenticity variable. The model moderator is the core module of the model which estimates and updates the model parameters such as factoid status and privacy level. The history keeper stores login attempts and hand them later to the moderator module to perform necessary updates. The question generator composes questions based on knowledge taken from the different knowledge sources utilized by the system. It also respects user privacy preferences.

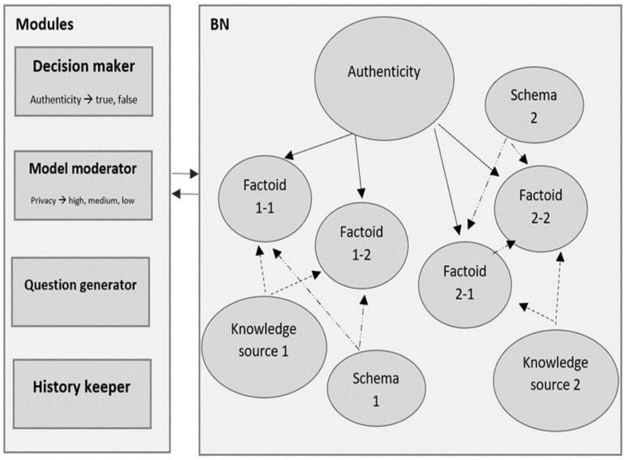

Fig. 2 Proposed model architecture

5.5 Authentication decisions

To make a decision of whether to authenticate a user or not, the probability of authentication is estimated using Bayes' rule then it is transformed into a true or false value by comparing it against a threshold. The threshold can be determined based on the preferred security level.

Other parameters under consideration such as the privacy parameter can be estimated using standard statistical learning methods, e.g., the maximum likelihood estimator.

The privacy associated with a given schema is calculated based on the number of times a factoid was refused to be answered by legitimate users of that schema and requested to be replaced by another question. To make sure that this is not due to difficulties in recalling the answer, users are asked to choose between two options when they ask to replace a question, namely, do not know the answer and refuse to answer. Values are then transformed into one of the following categories: high, medium, and low. Low values indicate that this factoid does not adhere to the minimum requirements of privacy in authentication systems. The model moderator module may set initial privacy values based on expert opinions and update them later with subsequent authentication events. An alternative option is to assume uniform distribution of privacy classes among factoids and assign privacy classes according to this distribution. Alternatively, a fixed privacy value may be initially assigned to all factoids, with the intention to update it later.

6. Preliminary evaluation and discussion

To examine users' preferences and attitudes towards privacy-preserving authentication systems, we conducted a questionnaire-based survey. We recruited 50 participants who were female adults, with age ranging between 20 and 45 with the majority being in their twenties. The participants were asked to try a knowledge-based authentication system utilizing questions generated from university students' database regarding courses in which the participants were enrolled. The majority of the participants (80%) expressed that they prefer to have the ability to choose the category of questions to be challenged with. However, 10% had neutral preference and 10% disagreed. We closely examined two generated questions to investigate students concerns about them. For the first question, 66% of participants indicated that they prefer to change the question. It was about their personal academic performance in courses they were enrolled in. For the other inspected question, only 34% of participants preferred to change the question. The question was about the specific semester the participants took the course in. This might be due to participants' preference towards less personal questions with concerns on protecting their privacy. Regarding the preferred question categories, 38% of them stated that they prefer questions about their academic experience such as grades, GPA, instructors names, year of enrollment, and current timetables. In addition, 26% of participants indicated that they prefer personal questions such as birth of date, mother name, and national id. Finally, 14% of them indicated that they prefer to be asked about their favorite stuff such as their favorite hobby, sport, book, team, color, food or subject.

These preliminary results motivate the next steps of this on-going research. In particular, there is a need to implement the proposed model to be able to measure and control users' preferences towards knowledge-based authentication, specifically those related to privacy. Also, it can be seen that users' attitudes are not shaped based on question categories only. Questions within each category may vary in terms of how difficult and how private the users consider it to be. This can be seen in the findings of our inspection to the two questions presented above.

7. Conclusion and future work directions

The main point of this study is to establish a new probabilistic model for user authentication by considering privacy metrics.

The preliminary effectiveness of the proposed method has been confirmed by our questionnaire-based survey. However, this study is part of an on-going research and there are several problems that need to be solved our future work such as:

(i) applying the proposed approach to various authentication environments with different knowledge sources and user considerations,
(ii) finding a reasonable method to determine an optimal order for presenting challenge questions in a given session.
(iii) develop a method to estimate the difficulty of each factoids: i.e., how difficult it is to answer the question regarding it.

References

[1] V. Zimmermann and N. Gerber, "The password is dead, long live the password – A laboratory study on user perceptions of authentication schemes," International Journal of Human-Computer Studies, vol. 133, 2020, pp. 26-44.

[2] W.E. Burr, D.F. Dodson, and W.T. Polk, “Electronic Authentication Guideline: Recommendations of the National Institute of Standards and Technology,” Nat’l Inst. of Standards and Technology (NIST), NIST Special Publication 800-63, version 1.0.2, 2006.

[3] A. Ullah, H. Xiao and M. Lilley, "Profile based student authentication in online examination," International Conference on Information Society (i-Society 2012), London, 2012, pp. 109-113.

[4] R. Biddle, S. Chiasson, and P. C. Van Oorschot, 2012. Graphical passwords: Learning from the first twelve years. ACM Computing Surveys (CSUR), 44(4), pp.1-41.

[5] S. Komanduri, R. Shay, P.G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor and S. Egelman,. 2011. Of passwords and people: measuring the effect of password composition policies. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). ACM, New York, NY, USA, 2595-2604.

[6] L. F. Cranor and S. Garfinkel, “Secure or Usable?”, IEEE Privacy and Security, Vol. 2, 2004, pp. 16-18.

[7] L. Gong, M. A. Lomas, R. M. Needham, and J. H. Saltzer, 1993. Protecting poorly chosen secrets from guessing attacks. IEEE journal on Selected Areas in Communications, 11(5), 648-656.

[8] P. Pinkas, and T. Sander, 2002. Securing passwords against dictionary attacks. In Proceedings of the 9th ACM conference on Computer and communications security (CCS '02), Vijay Atluri (Ed.). ACM, NY, USA, 161-170.

[9] N. K. Ratha, S. Chikkerur, J. H. Connell and R. M. Bolle, "Generating Cancelable Fingerprint Templates," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 4, pp. 561-572, April 2007.

[10] C. Katsini, M. Belk, C. Fidas, N. Avouris, and G. Samaras, “Security and usability in knowledge-based user authentication: A review,” in Proceedings of the 20th Pan-Hellenic Conference on Informatics (PCI ’16), ACM, New York, NY, USA, 2016, pages 63:1–63:6

[11] C. Morosan, “Voluntary steps toward air travel security: an examination of travelers’ attitudes and intentions to use biometric systems,” J. Travel Res. 51 (4), 2012, 436–450

[12] A. Ullah, H. Xiao, M. Lilley and T. Barker, "Privacy and usability of image and text based challenge questions authentication in online examination," 2014 International Conference on Education Technologies and Computers (ICETC), Lodz, 2014, pp. 24-29.

[13] A. Hang, A. De Luca and H. Hussmann, “I Know What You Did Last Week! Do You?: Dynamic Security Questions for Fallback Authentication on Smartphones,” CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, April 2015, 1383–1392, https://doi.org/10.1145/2702123.2702131

[14] A. Rabkin, “Personal knowledge questions for fallback authentication: security questions in the era of Facebook,” J. Travel Res. 51 (4), 2012, 436–450

[15] Y. Chen and D. Liginlal, “Bayesian Networks for Knowledge-Based Authentication,” IEEE Transactions on Knowledge and Data Engineering. 2007. vol. 19, issue: 5

[16] S. Chokhani, “Knowledge Based Authentication (KBA) Metrics,” Proc. KBA Symposium. Knowledge Based Authentication: Is It Quantifiable? 2004