Méthode de détection multi-modale de comportements anormaux en sécurité informatique

Baptiste Leterrier, Alexis Bitaillou, Benoît Parrein, Rémi Lehn

Méthode de détection multi-modale de comportements anormaux en sécurité informatique

Baptiste Leterrier
Silkke
contact : bl@silkke.com

Alexis Bitaillou
Polytech Nantes
dpt informatique

Benoît Parrein
Université de Nantes
LS2N

Remi Lehn
Université de Nantes
LS2N

***Résumé*—Le mot de passe est le moyen d'authentification le plus courant pour accéder à un système informatique. Malheureusement, il est vulnérable à différentes attaques, notamment par force brute. Afin d'augmenter la sécurité lors de l'authentification, nous cherchons dans par ces travaux un moyen d'authentification ne reposant pas uniquement sur le mot de passe, mais sur une combinaison d'éléments telle que la frappe clavier et les information biométrique et sur une validation en continue plutôt qu'une simple vérification à "l'entrée" dans le système.**

1. Introduction

La sécurité en informatique est le deuxième aspect le plus important après la disponibilité. Les logiciels de sécurité habituels (antivirus, pare-feu, etc.) ne suffisent plus pour détecter les menaces. Par exemple, une personne ayant un accès physique peut contourner le couple identifiant - mot de passe et ainsi dérober des données. La détection de comportements anormaux permet d'ajouter une strate supplémentaire de sécurité. La détection de comportements essaye de déterminer si les actions de l'utilisateur sont conformes à un profil établi. Le profil peut contenir par exemple les processus, l'activité réseau, les appels systèmes. La difficulté intervient lors de la création du profil. Si le profil est statique, le moindre comportement légèrement déviant peut déclencher une alarme et augmenter le nombre de faux-positifs. S'il est trop dynamique, il risque de rester dans une phase d'apprentissage indéfiniment et fera perdre en granularité de détection (seuls les événements vraiment déviant seront détectés). Dans tous les cas, la détection de comportement anormaux relève d'un problème d'apprentissage et de décision. Il convient aussi de définir une méthode la moins intrusive possible en terme de vie privée. Le détournement d'un systême de catpure des frappes de clavier peut être facilement détournée et une attention particulière sera prêtée à la méthode d'anonymisation de données récupérée lors des différentes expérimentations. Nous proposons tout d'abord de faire un rapide état de l'art sur le sujet. A partir d'une sélection de méthodes, nous allons réaliser les premières étapes une preuve de concept.

2. Présentation de la problématique

2.1. De la faiblesse des mots de passe

L'authentification sur un système informatique, se fait traditionnellement par un identifiant et un mot de passe. Ce mode d'authentification constitue le principal moyen d'authentification sur les systèmes informatiques. Sa simplicité a contribué à son adoption. Néanmoins, il est possible d'usurper l'identité de quelqu'un grâce à la connaissance de ces paramètres. En effet, certains identifiants sont standardisés comme les identifiants administrateurs, par exemple "root" sur les systèmes Unix. Malheureusement, le mot de passe est lui aussi attaquable. La puissance de calcul disponible a beaucoup augmenté. L'utilisation des cartes graphiques et de clusters pour la récupération de mot de passe a diminué sensiblement le temps des attaques avec des performances multipliées par 40 par rapport à la force brute sur processeur [1].

2.1.1. Les précédentes tentatives. Diverses alternatives ont été étudiées et développées. Ces alternatives sont axées sur l'identification biométrique, l'authentification matérielle et d'autres éléments mémorisables. Par exemple, l'empreinte digitale peut être utilisée pour identifier et authentifier une personne. Mais l'empreinte digitale est une donnée personnelles, son utilisation est soumise à restriction. Le résultat n'est pas exact. Comme décrit dans [2] , l'identification biométrique évalue la probabilité qu'un échantillon soit proche d'autres échantillons de référence et prend une décision. Ces alternatives sont potentiellement complexes à mettre en place et ne garantissent pas nécessairement de résultats. Un système d'authentification seul est insuffisant. Même les systèmes biométriques sont contournables comme exposé dans [3]. Un autre problème majeur de l'utilisation des sytèmes d'authentification biométrique est l'impossibilité de révoquer les accès car une personne ne dispose que d'un jeu limité d'empreinte digitale non remplaçable.

2.1.2. A la fusion des techniques. A défaut de créer de nouvelles techniques, les recherches se portent sur la fusion de techniques existantes. Par exemple, dans [4], la dynamique de la frappe de clavier et la reconnaissance faciale sont utilisées simultanément. Les deux techniques

ont des précisions différentes. La difficulté intervient lors de la fusion des données et de la prise de décision. Il faut pondérer le poids de chaque métrique pour aboutir à une tendance permettant d'identifier la personne.

3. Plan de l'étude

Il existe déjà des travaux sur les différentes méthodes d'authentification. Le premier objectif est d'établir un comparatif des techniques existantes. Notre comparatif s'effectuera sur différents critères. Par exemple, les moyens techniques nécessaires, les nombres d'échantillons nécessaires à l'apprentissage, la précision sont des critères utilisables. C'est cette voie qui sera suivie pour créer une solution hybride.

4. État de l'art

4.1. Dynamique des frappes de clavier

Le clavier est un des premiers périphérique d'un ordinateur. Son principe n'a pas évolué depuis des décennies. La frappe (dans son rythme, sa force) varie d'une personne à une autre, ce qui fait que notre frappe est unique, comme nos mains par exemple.

4.1.1. Présentation. Le clavier est présent sur quasiment sur tous les ordinateurs. De plus, la disposition des touches est relativement standardisée. L'utilisation de la frappe de clavier comme moyen d'identification n'est pas un concept "récent". Dès 1993, Brown et al. [5] utilisent ce concept pour ce qui semble la première fois en mesurant le temps de pression et le temps de relâche de chaque touche.

La dynamique des frappes n'est pas utilisable dès sa mise en place, un temps d'initialisation est requis. Un apprentissage est nécessaire pour utiliser la dynamique des frappes. La quantité d'échantillons dépend de la méthode de décision. La méthode de décision peut être statistique ou par réseau de neurones (comme dans [6]).

4.1.2. Intérêt de la proposition. L'identification par la dynamique des frappes est une technique intéressante. Le clavier est un périphérique répandu. Il est donc plus facile d'intégrer cette technique sur des équipements existants. Certaines métriques restent valables même sur des claviers virtuels.

Sur les systèmes dérivés d'Unix, la mise en œuvre semble relativement simple. Dans [5], ils utilisent une application pour d'intercepter les événements depuis le serveur X.org.

De nombreuses expérimentations ont déjà été effectué. La technique est donc relativement éprouvée. Les résultats sont généralement bons avec cette technique. La précision garantie est en moyenne proche des 80%.

4.1.3. Limites de la proposition. L'identification par la dynamique des frappes n'a pas que des avantages. Comme le résume [7], les résultats sont très variables. Le taux est influencé par la méthode de décision. Comme le montre [5], parfois, il n'y a pas de vraie différence entre une méthode de décision triviale comme la distance géométrique et une méthode de décision plus complexe comme un réseau de neurones avec rétro-propagation.

Le nombre d'échantillons a une grande importance et varie en fonction de la méthode de décision. Le nombre d'échantillons nécessaire n'est malheureusement pas prédéfini. De plus, la base d'apprentissage doit être régulièrement mise à jour. La dynamique de frappe étant une donnée biométrique, elle dépend de l'état biologique de l'utilisateur à un instant t. Ces particularités peuvent modifier cet état et donc la précision de l'identification. Les effets d'un changement de modèle de clavier n'ont pas été évalué.

Cette technique, si elle n'est pas encadrée, peut devenir dangereuse car il est possible de modifier le programme pour le transformer en enregistreur de frappes.

4.2. Son des frappes de clavier

Une variation de la dynamique de frappe consiste à utiliser le son de la frappe. L'objectif est donc d'utiliser le son du clavier pour identifier une personne. Comme pour la dynamique de frappe de clavier, le son de la frappe permet d'obtenir des métriques aux valeurs "uniques" pour chaque personne.

4.2.1. Présentation. Le clavier est souvent présent avec un ordinateur. Lorsqu'un microphone est présent, on peut enregistrer le bruit de la frappe sur le clavier. A. Peacocket al. [8] utilisent cette idée. La frappe au clavier est une caractéristique propre à chaque humain. Elle est fonction du corps et en partie du clavier. L'objectif est d'enregistrer le son de la frappe du clavier pour un mot donné. Cette technique n'est pas utilisable instamment. En effet, une durée d'initialisation est nécessaire. L'apprentissage dépend de la méthode de décision.

Le son des frappes de clavier n'a pas été beaucoup étudié. A notre connaissance, un seul article [8] traite du sujet.

4.2.2. Intérêt de la proposition. Le son de la frappe de clavier peut être facilement intégré aux ordinateurs portables. En effet, la plupart des ordinateurs portables sont doté d'un microphone intégré. Pour les ordinateurs fixes, la situation est complexe. Tous les ordinateurs fixes n'ont pas de microphone.

Les résultats des expérimentations sont bons. Dans l'article [8], le moins bon résultat atteint 88% de précision. Mais les tests sont effectués uniquement sur 4 groupes distincts.

4.2.3. Limites de la proposition. Cette technique a quelques inconvénients. La capture du son n'a pas été testé dans un milieu bruité. Dans un environnement type "open space", le micro pourrait capturer le bruit d'un autre clavier.

L'étude [8] a testé avec un seul et unique clavier. Dans le cas où l'utilisateur n'a pas d'ordinateur déterminé, l'impact du changement de clavier n'a pas été mesuré.

4.3. Mouvement de la souris

Bien que crée après le clavier, la souris s'est imposé comme outil d'interface homme-machine. Le nombre de touches est limité, mais la souris transmet ces déplacements.

4.3.1. Présentation. Après l'utilisation du clavier, la souris est utilisée pour l'identification d'utilisateur. L'objectif est d'utiliser les déplacements, les accélérations et les clics effectués avec la souris. Les accélérations sont calculées à partir des mouvements. Cette technique nécessite une durée d'initialisation. La technique est relativement jeune car elle est utilisée que depuis quelques années comme présentée dans l'article [9].

4.3.2. Intérêt de la proposition. La technique est conceptuellement relativement simple. Intuitivement, il semble que si l'utilisateur est une personne relativement âgée, les accélérations de la souris seront faibles et de courte durée. Le matériel nécessaire pour la mise en œuvre de cette technique est relativement abordable et/ou disponible.En utilisation supplémentaire/complémentaire, l'utilisation de la souris prend plus d'intérêt. Les articles [10] et [9] l'utilisent en complément du clavier pour affiner les résultats.

4.3.3. Limites de la proposition. Cette technique présentes quelques inconvénients. L'utilisation de la souris n'est disponible que dans un contexte d'identification continue. Il n'est pas pertinent de capturer des évènements sur une durée très courte. Au moins, le cas n'est pas évoqué dans [9].

Dans [9], les résultats sont relativement décevants, la précision est bornée entre 45% et 60%. Cela implique la nécessité de l'utiliser en tant que technique complémentaire et/ou supplémentaire.

5. Preuve de concept

Afin de tester nos hypothèses, une preuve de concept est en cours de développement. L'architecture foncionne en local dans un premier temps (capture et analyse sur la même machine). Dans le futur, l'architecture sera de type client-serveur mais conservera son aspect privé par une installation en réseau local. Un client autonome pourra aussi être envisagé dans le cas de postes nomades. Cette preuve de concept est composée de différents éléments. On peut le diviser en 3 parties (collecte, évaluation, exploitation). Un agent a été réalisé afin de collecter différentes métriques.

Il a été développé pour les environnements Microsoft Windows et Canonical Ubuntu 16.10. Il permet notamment de récupérer les différentes frappes au clavier et de les corréler avec l'activité système, les processus etc. Pour prendre une décision, nous développons une application permettant de réaliser un apprentissage. Cette approche vise à identifier une corrélation en le rythme de frappe et l'activité de l'utilisateur, dans le but de créer une tendance permettant de l'identifier dans le futur.

6. Conclusion

Les premiers résultats ont permis d'aboutir sur un ensemble de données d'apprentissage. Cet ensemble est actuellement utilisé dans un réseau de neurones qui permettra de potentiellement établir une tendance servant à l'identification d'un utilisateur.

La suite de l'experimentation consistera à exploiter les résultats pour rechercher la meilleure combinaison frappe/services. Cela ouvrira la voie à la création d'un ensemble de données permettant l'identification d'une personne.

D'autres méthodes sont aussi en cours d'expérimentation, comme l'utilisation des réseaux wifi pour l'identification des postures et comportements comme présenté dans l'article [11].

Références

[1] D. Apostal, K. Foerster, A. Chatterjee, and T. Desell, "Password recovery using mpi and cuda," in *2012 19th International Conference on High Performance Computing*, Dec 2012, pp. 1–9.

[2] K. Nandakumar and A. K. Jain, "Biometric authentication : System security and user privacy," *Computer*, vol. 45, no. undefined, pp. 87–92, 2012.

[3] N. M. Duc and B. Q. Minh, "Your face is not your password face authentication bypassing lenovo–asus–toshiba," *Black Hat Briefings*, 2009.

[4] A. Gupta, A. Khanna, A. Jagetia, D. Sharma, S. Alekh, and V. Choudhary, "Combining keystroke dynamics and face recognition for user verification," in *Computational Science and Engineering (CSE), 2015 IEEE 18th International Conference on*, Oct 2015, pp. 294–299.

[5] M. Brown and S. J. Rogers, "User identification via keystroke characteristics of typed names using neural networks," *International Journal of Man-Machine Studies*, vol. 39, no. 6, pp. 999–1014, 1993.

[6] S. Ravindran, C. Gautam, and A. Tiwari, "Keystroke user recognition through extreme learning machine and evolving cluster method," in *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Dec 2015, pp. 1–5.

[7] A. Peacock, X. Ke, and M. Wilkerson, "Typing patterns : a key to user identification," *IEEE Security Privacy*, vol. 2, no. 5, pp. 40–47, Sept 2004.

[8] M. Pleva, E. Kiktova, P. Viszlay, and P. Bours, "Acoustical keystroke analysis for user identification and authentication," in *2016 26th International Conference Radioelektronika (RADIOELEKTRONIKA)*, April 2016, pp. 386–389.

[9] S. Mondal and P. Bours, "Combining keystroke and mouse dynamics for continuous user authentication and identification," in *2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, Feb 2016, pp. 1–8.

[10] L. Fridman, A. Stolerman, S. Acharya, P. Brennan, P. Juola, R. Greenstadt, and M. Kam, "Multi-modal decision fusion for continuous authentication," *Computers & Electrical Engineering*, vol. 41, pp. 142–156, 2015.

[11] Z. W. M. L. Z. Y. Tong Xin, Bin Guo, "Freesense :indoor human identification with wifi signals," *arXiv preprint arXiv :1608.03430*, p. 6, 2016.