A Real-Time Authentication Method Based on Cursor-hidden Scene

Wei Zicheng
School of Computer Science
Beijing University of Posts and Telecommunications
Beijing, China
weizicheng@nelmail.iie.ac.cn

Chen Xiaojun
Engineering Laboratory for Information Security Technologies, IIE
Beijing, China
chenxiaojun@iie.ac.cn

Pu Yiguo
Engineering Laboratory for Information Security Technologies, IIE
Beijing, China
puyiguo@nelmail.iie.ac.cn

Xu rui
Institute of Computing Technology
Beihang University
Beijing, China
xurui@nelmail.iie.ac.cn

***Abstract*—Identity theft, as a common insider attack method, is difficult to detect because it is hard to distinguish the legal user from the masquerader with a stolen legal identity. Compared with traditional identity authentication method such as password or fingerprint, behavior biometrics based on HCI (Human-Computer Interaction) has been more useful and effective in real-time authentication. However, existed approaches either require longer authentication time or are designed for special scenario. In this paper, we propose a real-time authentication approach based on mouse-hidden scene to detect identify theft attack. When operations of the suspicious masquerader are detected, the cursor is hidden deliberately. Under the scene, we assume that mouse operators become anxious and show some unique and instinct mouse movement operations. Based on the movement traces, behavioral model is generated and used for detecting the suspicious masquerader. The experiments show the approach can achieve 2.6 percent of FAR and 3.3 percent of FRR while authentication time consumed is acceptable in practice.**

Keywords: insider threat, identify theft, behavior biometrics, mouse dynamic

I. INTRODUCTION

The threat of insider attacks to organizational security has been one of the most difficult challenges to address and it received increases attention in academic, commercial and government research communities. The organizations can often detect or control the outside attacks which try to access sensitive data either physically or electronically, and mitigate the threat of an outsider stealing company property. But the malicious insiders are more dangerous because they are much more familiar with the topology and security mechanism of the company than the outside attackers. Furthermore, they are difficult to identify because they typically use authorized access and regular business processes to commit crimes .The threat of attack from insiders is real and substantial .For example, WikiLeaks released a document set called “the Afghan War Diary” which includes over 91,000 reports about the detail information about the war in Afghanistan in 2010; Wen Chyu Liu, a retired research scientist was convicted in February 2011 of stealing trade secrets from his former employer and selling them to other companies, etc[16].

As one of the most common attack methods of inside threat, identify theft has been a hot area of research for several years. Identification and authentication are the major security services to recognize and remove this kind of attacks. And the technology of user authentication can be categorized into three classes as we show in figure 1.Knowledge based authentication is based on something one knows. The best examples are commonly known passwords and PIN codes. These systems do provide a strong frontline against the masqueraders if managed properly. However, there are several weaknesses, such as weak passwords or masqueraders obtain the password in some way like password files. Once authenticated, the adversary can easily abuse the victim’s account and the password will lose its function.Object based authentication relies on something one has and is characterized by possession, such as ID cards and tokens. Same with knowledge based authentication, it also face the once-authentication problem. The attackers can operate the victim’s computer to steal confidential business documents or personal privacy information when they leave without locking screen and tokens also can be forgotten, lost or stolen. The last class is biometrics based authentication which is related to what a person does, or how the person uses their body. It can overcome difficulties of knowledge based and object based approaches. There exist two categories of biometrics: physical biometrics and behavioral biometrics. Physical biometrics measures the physiological characteristics of a person, such as fingerprint, iris scan and hand shape geometry. These systems require special hardware and often only augment the logon process. Behavioral biometrics measure the behavior of a person, such as keystroke dynamics and mouse dynamics. The keystroke dynamic biometrics does not require special hardware. But in the Operating System with GUI, the keyboard operations is relatively less than the mouse operation. Existing researches have proved that mouse dynamics are more effective in authenticating users in real time. Existing researches about mouse dynamics have some common shortages: most of them need more than five minutes to authenticate, some of them can only get satisfactory experimental result in a fixed application and some others ask the users to do some operations for a long time like playing a memory games. They are all unpractical in face of the identity theft.

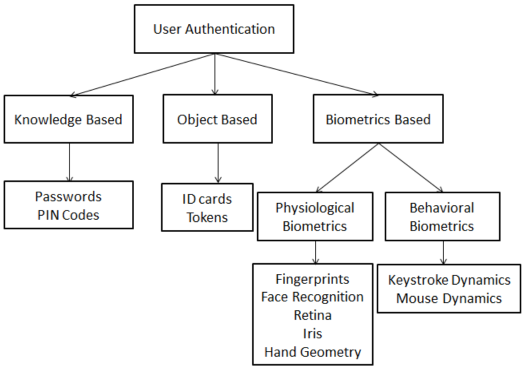


Figure 1. Existing user authentication methods

In this paper, a real-time authentication method based on cursor hidden scene is proposed. When operations from suspicious masqueraders are detected, the system hides the cursor and enforces the current user to enter into the presupposed cursor hidden scene. From the behavioral biometrics point of view, different people will show their unique behavioral features by instinct in the psychology of anxiety and thus the responding mouse movement features also will be identical for each person. Based on the analysis of user's unique mouse movement features, a real-time authentication method is proposed to generate the behavior model for users. By comparing the legal user's profile with the mouse movement trace captured in real time, the malicious insider masquerader can be detected. The experiment proves that our authentication method is feasible and efficient in detecting the attacks of masqueraders.

The following sections are organized as follow: Section II introduces related work of behavioral biometrics technology. Section III introduces our authentication method. In section IV, we describe the setup of experiment and analyze the result, and then conclude the paper and provide future work in section V.

II. RELATED WORK

As behavioral biometrics based on HCI can provide both static and dynamic authentication with no additional equipment than physical biometrics, it draws more attention of the scientists and has made considerable progress. We usually focus on two kinds of social behavioral biometrics technology: keystroke dynamics and mouse dynamics.

Authentication with keystroke dynamics has been studied extensively over the past three decades and their use for authentication has shown promising results [9, 10].Existing approaches always verify the identity of current users by analyzing the information related with individual's unique typing rhythms like keystroke dwell time[11], flight time [12] and latency[13], etc. With the popularity of graphical user interface, the proportion of mouse operations rising relative and the proportion of keyboard input show descending tendency, which gives existing keyboard dynamics method produces a corresponding impact and presents a great challenge. Meanwhile, mouse dynamic is more suitable than keystroke dynamic to verify users continuously for it need not the users to operate the computer in a predetermined pattern like most of the keystroke dynamic system. Thus the research focus is transferring to mouse dynamics.

As a new behavioral biometric [9], mouse dynamics was first investigated around 2003[1, 2] and developed rapidly during the past few years. The idea behind it is to collect all mouse actions when the user interact with a graphical user interface and extract behavior profiles and model for the user, aiming at verifying current user ultimately. Ahmed and Traor conducted series of experiments to prove their methods [3].In main experiment, they collect 284 hours of raw mouse data over 998 sessions of 22 participants and analysis the captured data, extracting features like MSD (movement speed compared to traveled distance), MDA (average movement speed per movement direction) and ATA (average movement speed per types of actions) from it. Then they modeled the behavioral characteristic with artificial neural networks and achieved an average EER of around 2.46 percent. The two other smaller experiments, involving seven participants, showed relatively good results by tiny-controlling the environment variables. But the re-authentication time of this method ranges from 17 minutes to 30 minutes, which is impractical in attack detection. Pusara and Brodley [4] attempted to distinguish users according to extract mouse features such as angle, speed and distance in a defined window. They used C5.0 decision tree algorithm as their classification algorithm and achieved an average FAR of 0.43 percent and an average FRR of 1.75 percent. The detection time of their experiment depends on the parameters set by the different users. Meanwhile, their detection model is application dependent, making it impossible to be used in practice. Gamboa and Fred [5, 6] studied the possibility of user authentication based on mouse dynamics in a web memory game. They found that sequences of 50 strokes yield an ERR of 2 percent. Nan et al. [7] presented a method based on fine-grained angle-based metrics they defined, which is able to re-authenticate a user to a high accuracy. But the authentication time that needs to achieve the accuracy is more than 37 minutes. Chien-Cheng Lin et al. [8] presented an approach utilizing the file-related operations via a mouse. For best, he achieved a FAR 4.9 and a FRR of 5.2 with the data of 20 users accessing files using Explore. Their method can be only used under special environment.

III. CURSOR HIDDEN BASED REAL-TIME AUTHENTICATION

In this section, we mainly discuss and validate the feasibility of the method.

A. Cursor-hidden Scene

It's a great challenge to select an appropriate scene. For mouse dynamics, many factors can affect the interactive features of person's behavior during a session, which will decrease the accuracy of the results. Subjective factors such as user’s emotion can affect the accuracy of modeling and judgments. For example, feelings of anxiety may accelerate user’s interaction frequency. We cannot rule out this kind of errors. Objective factors such as the screen resolution and dots per inch of the mouse can also affect the behavior model without data normalization. But we can do the corresponding control.

As we know, the reaction of different person is various in the face of unexpected situations. Similarly, the behavior of different users also exist some differences when the cursor suddenly disappears. Some people will move their mouse rapidly in random directions while some others’ mouse events concentrated in some fixed directions. All of them will show their own characteristic like the trajectory of movement, the frequency of the mouse direction changes or the average movement speed. So we propose the cursor-hidden scene.

In order to prove the feasibility of our theory more clearly, keeping the factors as follows the same in our experiments is necessary: the type of the mouse and screen, the dpi setting and the screen resolution of different users. Thus, we put forward the following hypothesis:

- Hypothesis I: Assuming that the emotion of a user in the mouse-hidden scenarios is consistent. It ensures the user’s behavior do not produce large fluctuations.
- Hypothesis II: Under normal circumstances, a malicious user does not change the mouse when he uses legal user's computer.

B. *Feature Selection Under Cursor Hidden Scene*

In our approach, firstly, we collect and analyze the information of mouse movement generated by each user within 5 seconds after the disappearance of the cursor, extracting the behavior features of them. Secondly, we build the behavior model base on the feature vector. Finally, we detect the identity theft attack with the model we build. To prove the feasibility of this method, we randomly select two users' data and extract some of their characteristics. We also make some corresponding contrast with the data and the comparative results will later show up in the form of figures. Meanwhile, the mouse click events generated in few seconds are so little that this kind of features can’t represent the users’ behavior characteristics effectively, thus we don’t analyze and discuss them.

We divide the directions of mouse movement into eight separate parts and each part contains 45 degrees. The mouse moves continuously in the same direction over a certain distance is recorded as a valid movement behavior in our experiment. The features below are calculated based on the valid movement records.

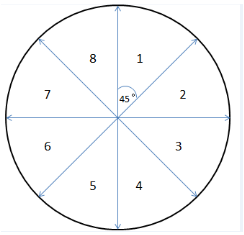
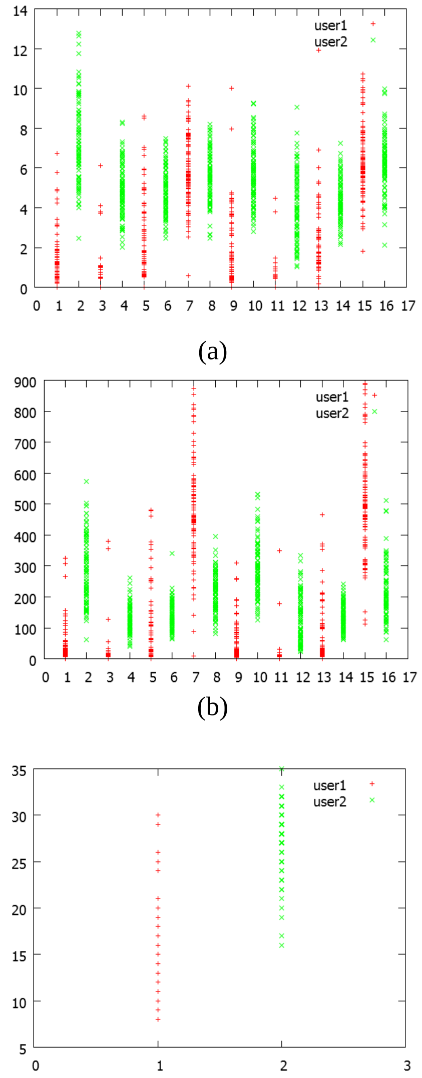


Figure 2. Movement direction of the mouse

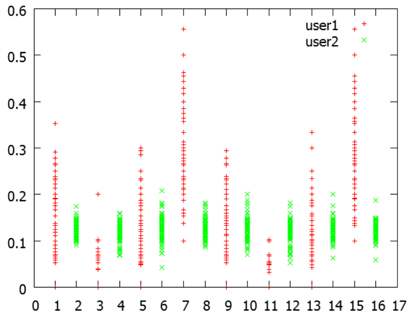
The figure 3 shows the comparative results of two users. Red and green represent two different users we randomly chosen. We calculated traveled speed, the average movement distance and features about numbers of movement events in different directions.



(a)

(b)

(c)



(d)

Figure 3. (a) The average movement speed of eight directions(b)The average traveled distance of eight directions(c) The overall number of movement events (d) The number of movement events in eight directions

From Figure 3, we can see that different users show obvious diversity in the behavior model during the few seconds after the mouse disappears. The distribution of user1's mouse events is more stochastic than user2's. The distribution of different users' features also has their own range. According to the result of experiment above, we can classify the users both effectively and accurately and detect identity theft attack.

C. *Real-Time Authentication Method*

In the approach, the authentication system can be divided into three components as it is shown in figure 4.They are data collection component, data analysis and feature extract component and classifier component.

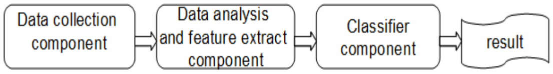

Data collection component
Data analysis and feature extract component
Classifier component
result

Figure 4. Experimental flow

- *Data Collection Component*

This component is mainly used to collect raw mouse information. Now we set the cursor disappears for five seconds every once in a while, and use windows hook mechanism to obtain and record the mouse movement and click information during this period in a binary stream. The contents of our records are: the timestamp, the horizontal axis, the vertical axis, the type of mouse event, the ID and Command Line Arguments of current foreground process and the title of current window. We primarily use the first four items of records to build the model, and the rest are used for computer forensics and attack intention analysis in the future.

Compare with the data collection component of other real-time authentication system, we use much less disk storage space as well as much lower CPU usage for the reason that we collect data only when it is necessary.

- *Data Analysis and Feature Extract Component*

In this component, we analysis the collected data and extract the feature vector from them. We set the time threshold and the distance threshold to filter the original data and remove some of the noise data, avoiding the influence of events like traveled distance less than few pixels. Then we calculate the average, variance and the frequency of occurrences of the features. The result will be stored in the feature vector used for modeling users' behavior characteristic with SVM. The features we choose will be discussed in IV.

- *Classifier Component*

In this part, we use the model generated from the features to verify the current user's identity and make a reasonable judgment. It uses the movement information within 5 seconds after the cursor disappears to authenticate the identity of current user.

We select support vector machine (SVM, also support vector networks) [14] as our classification algorithm. In machine learning, it is supervised learning model with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier.But when it was used for classification and regression analysis, current academic community has not yet formed a unified model of the choice of kernel function and the parameters. It is said that the optimal SVM algorithm parameter selection can only be selected by virtue of experience and a wide range of search or use the interactive cross validation functions provided by the software package.

Libsvm[15] is a software package which is designed and implemented by Chih-Jen Lin of National Taiwan University. It is a simple and easy-to-use SVMs tool for classification, regression, and distribution estimation. After experimental comparison, we finally set the parameter s as 0 and set the parameter t as 3.

IV. Experments And Result Analysis

A. *Experiment Environment*

As we described above, in order to limit the unintended influence of objective factors on the experimental results as low as possible, we tiny control the hardware environment. The experimental terminals are Lensovo personal computers with a Core 2 Duo 2.8 GHz processor and 2GB of RAM. The monitors are 17" LCD monitors (Lenovo ThinkVision series) and the resolution is set at 1449 ×900. We equip the computer with a Shuangfeiyan optical mouse, running the Windows 7 operating system. We make the data acquisition environment as consistent as possible for the users.

In the experiments, we collect the mouse data when the cursor is hidden, on average, 20 times for every user. Five seconds per time. We extracted one features record from the data per second. The training data set contains 150 records, which include 100 normal user's records and 50 abnormal user's records. It greatly reduce the time needed for data acquisition and training. At the same time, we carry on different dimensions of feature vector extraction respectively, calculate the corresponding FRR and FAR and investigate the effects of feature vector selection on the experimental results. The features we chose in different experiment are as follows:

- Experiment 1(17 dimensions totally): the average traveled distance and the percentage of movement events in eight directions; the sum of mouse events.
- Experiment 2(25 dimensions totally): the average traveled distance, average movement speed and the percentage of movement events in eight directions; the sum of mouse events.
- Experiment 3(49 dimensions totally)：the average and variance of traveled distance, movement speed in eight directions; the percentage of movement events in eight directions; the sum of mouse events.

B. *Experiment And Result Analysis*

TABLE I. RESULTS OF THE EXPERIMENTS.

Dimension	FAR	FRR
17	**17.3%**	**5.3%**
25	**6%**	**5.3%**
49	**2.6%**	**3.3%**

Table I gives the result of our experiment. We can see from it that the accuracy constantly improves with the characteristic dimensions increases. We achieved a satisfactory recognition rate when characteristic vector is 49 dimensions, FRR and FAR reaches the 3.3 percent and 2.6 percent respectively. Compared with the existing method, on the premise of guarantee high accuracy, our method needs less time to collect enough data for training(around two minutes)and the authentication time (few seconds)is greatly shorter than other systems, thus greatly improves the availability in real-time authentication system. At the same time, we can still optimize the selection of features vectors properly, improving the accuracy.

V. CONCLUSION AND FUTURE WORK

In this paper, we propose a new approach of user re-authentication based on cursor-hidden scene. It is suitable for combining with existing host intrusion detection system to provide users with secondary verification, further enhancing the accuracy of identity theft detection. Under the condition of keeping accuracy, our method has the following features compared with conventional methods :1)During the training phase, we need less original training data and generate the user model in a short period of time .2)In the identification stage, you can utilize the data of cursor hidden once(usually few seconds) to make accurate judgments and speculation.3)We need not deploy and keep special software environment, and we also do not make any assumptions on the user's behavior.4)The characteristics of frequent mouse-behavior without a fixed trajectory in a short time is hard to imitate, which makes the robustness of our methods better. Finally, we have verified our theory through experiment, achieving a FAR 2.6 percent and a FRR of 3.3 percent.

In the future, we will focus on the following aspects: selecting and optimizing the feature vectors and the classifier parameters, further improving the accuracy of prediction; the influencing factors of the screen resolution and mouse DPI will be added to existing system, expanding the scope of our system; combining the collected application information to do the research on computer forensics and attack intention analysis, etc.

ACKNOWLEDGMENT

This work is supported by National High Technology Research and Development Program of China, 863 Program (Grant No. 2012AA013101).

REFERENCES

[1] R. Everitt and P. W. McOwan, "Java-based internet biometric authentication system," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 25, no. 9, pp. 1166–1172, 2003.

[2] A. Ahmed and I. Traore, "Anomaly intrusion detection based on biometrics," in Information Assurance Workshop, 2005. IAW '05. Proceedings from the Sixth Annual IEEE SMC , 2005, pp. 452–453.

[3] A. Ahmed and I. Traore, "A new biometric technology based on mouse dynamics," Dependable and Secure Computing, IEEE Transactions on, vol. 4, no. 3, pp. 165–179, 2007.

[4] M.Pusara and C.E.Brodley, "User re-authentication via mouse movements,"in Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security,ser. VizSEC/DMSEC '04. New York, NY, USA: ACM, 2004, pp. 1–8.

[5] H. Gamboa and A. Fred, "A behavioral biometric system based on human-computer interaction," pp. 381–392, 2004.

[6] H. Gamboa and A. Fred. An identity authentication system based on human computer interaction behaviour. In Proceedings of 3rd International Workshop on Pattern Recognition on Information Systems, 2003: 46-55.

[7] N. Zheng, A. Paloski, and H. Wang, "An efficient user verification system via mouse movements," in Proceedings of the 18th ACM conference on Computer and communications security, ser. CCS '11. New York, NY, USA: ACM, 2011, pp. 139–150.

[8] Lin, Chien-Cheng, Chin-Chun Chang, and Deron Liang. "A New Non-intrusive Authentication Approach for Data Protection Based on Mouse Dynamics." Biometrics and Security Technologies (ISBAST), 2012 International Symposium on. IEEE, 2012.

[9] F. Monrose and A.D.Rubin, "Keystroke dynamics as a biometric for authentication," Future Generation Computer Systems, vol.16, no. 4, pp. 351–359, 2000.

[10] M. Karnan, M.Akila, and N. Krishnaraj, "Biometric personal authentication using keystroke dynamics: A review," Applied Soft Computing , vol. 11, no.2, pp.1565–1573, 2011.

[11] K. Xi, Y. Tang, and J. Hu, "Correlation keystroke verification scheme for user access control in cloud computing environment," The Computer Journal, vol. 54, no. 10, pp. 1632–1644, 2011.

[12] S. Singh and K. V. Arya, "Key classification: A new approach in free text keystroke authentication system," in Circuits, Communications and System (PACCS), 2011 Third Pacific-Asia Conference on, 2011, pp. 1–5.

[13] F. W. M. H. Wong, A. Supian, A. Ismail, L. W. Kin, and O. C. Soon, " Enhanced user authentication through typing biometrics with artificial neural networks and k-nearest neighbor algorithm," in Signals, Systems and Computers, 2001. Conference Record of the Thirty-Fifth Asilomar Conference on, vol. 2, 2001, pp. 911–915 vol.2.

[14] support vector machine. http://en.wikipedia.org/wiki/Support_vector_machine.

[15] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[16] insider theft cases,http://www.fbi.gov/about-us/investigate/counterintelligence/the-insider-threat.