Research on Mobile Terminal User Identification Based on Big Data

Zhenhua Wang [1,a,*], Yangsen Yu [1,b] Tao Xue [2,c]

[1,3] School of Information Engineering, China University of Geosciences, Xueyuan Road, Beijing, China

[2] Geological Survey Institute, China University of Geosciences, Xueyuan Road, Beijing, China

[a] wangzh@cugb.edu.cn, [b] 61412129@qq.com, [c] xuetao@cugb.edu.cn

Keywords: Mobile terminal, User identification, Big data analysis.

Abstract: Mobile user identity is an important part of mobile Internet security research. In this paper, it introduces the development status of user identification and authentication technology in "Internet +" era, and briefly describes their implementation plan and evolution trend. the security problem of each general authentication technology is analyzed. This paper describes the combination of mobile device identification, user behavior analysis and forecasting based on big data, and routine identification, and realizes mobile user identification to improve the accuracy of mobile user identification.

1. Introduction

The definition of mobile terminal is narrow and broad. The narrow definition of the mobile terminal generally refers to the mobile phone, tablet computer, mobile terminal general definition is connected to the Internet can move the device, such as notebook computers. User identification refers to the process of confirming the identity of an operator in a mobile Internet system to determine whether the user has access to and use of a resource and to enable the access policy of the mobile device and the network system to be performed reliably and efficiently, to prevent the attacker to fake legitimate users access to resources, to ensure system and data security, and authorized visitors to the legitimate interests. Mobile terminal user identification in the computer network security, social security, precision advertising, etc., has a wide range of applications.

2. Conventional User Identification Technology

Generally, user identification is performed by techniques such as passwords or biometrics on a mobile device. Comparison of various cryptographic authentication methods lists in Table 1.The user's password is set by the user. In the network login, one enters the correct password, the system thinks the operator of mobile terminal is a legitimate user. In fact, because many users in order to prevent forgetting the password, often use such as birthdays, phone numbers and other easy to guess the string as a password, it is likely to cause a password leak. If the password is static data, the verification process is required in the computer memory and the transmission process may be intercepted by the Trojan programs or malicious programs. Therefore, the static password mechanism, whether it is used or deployed are very simple, but from the security point of view, the user name and password is a way of unsafe authentication.

SMS password requests the dynamic password containing 6 random numbers in the form of mobile phone SMS, and the authentication system sends the random 6-bit password to the customer's mobile phone in the form of short message. The user enters this dynamic password during login or transaction authentication so as to ensure the security of the system identity authentication.

Dynamic token is a dynamic password, that is, some unpredictable random number combinations based on a specific algorithm, randomly generated at a specific interval, and each password can only be used once. Mobile terminal token is used to generate dynamic password of the mobile applications, in the process of generating a dynamic password, will not produce any communication and cost. Mainstream is based on the time synchronization mode, every 60 or 30 seconds to change a dynamic password, the password is valid once, and it produces 6 or 8 dynamic digits for a one-time authentication. There are now dynamic, token-based, two-way authentication. Dynamic synchronization based on event synchronization is a synchronization principle triggered by user action. It is really a secret, and because it is bidirectional authentication, that is, the server verifies the client and the client needs to authenticate the server.

Biometrics is a discipline in computer science that uses biometrics to identify people and conduct access control. A technique for identity authentication through biometrics such as measurable body or behavior. Biological characteristics are the only physiological features or behavior that can be measured or can be automatically identified and validated. Biological characteristics are divided into two categories: physical characteristics and behavioral characteristics. Body features include: fingerprints, palm type, retina, iris, body odor, face, hand blood vessels and DNA; behavioral characteristics include: signature, voice, walking gait and so on.

Table 1 Comparison of various cryptographic authentication methods

Cost and use Type	Hardware cost	network requirements	storage location	operability	security
Static password	Low	No	mobile terminal	medium	Weak
SMS password	Low	Yes	Server	poor	Strong
Dynamic token	Low	Yes	Server	poor	Strong
Biometrics	High	Yes	mobile terminal	good	Strong

Traditional mobile terminal user authentication mainly uses the password technology and the biometric technology, these technologies have some certain limitation and the unreliability. Once the mobile device is lost or modified, or the user password is cracked, the user identity is used by others, which will produce a series of troublesome things. In order to improve the reliability and accuracy of mobile user identification, we use large data technology and related algorithms to carry out mobile user identification. This technology needs to be considered from three aspects: first, to identify the use of equipment that information is static attributes; Second, the analysis of a series of user behavior, that is, dynamic attributes; Finally, combined with conventional user identification method for mobile users Identification.

2.1 Static Attribute Analysis of Mobile Terminal

For mobile devices, under normal circumstances, belong to a person's special equipment. Mobile devices on the Internet, through the motherboard, CPU, memory, network card and other information to determine a complete device. At present, these hardware information is generally stored in the mobile device operating system, the conventional software to read information, only read the hardware information, only read the data stored in the operating system files, and this information may be changed. In order to prevent the hardware system is changed, each time the hardware information is confirmed, the hardware information is read directly. There is also a case where the mobile device is disassembled or reassembled, so that it can cause trouble to confirm a mobile device, and that a mobile device integrity can be identified as including the motherboard, the processor, and the storage device.

The location information of the mobile user is an important part of the user data, mainly from the mobile device. Generally, a mobile device belongs to a person's special equipment. Before you can collect information about your mobile device, make sure that the basic information for your mobile device is accurate. Mobile devices on the Internet, through the motherboard, CPU, memory, network card and other information to determine a complete device. At present, these hardware information is generally stored in the mobile device operating system, the conventional software to read hardware information, only read the data stored in the operating system files, and this information may be changed. In order to prevent the hardware system is changed, each time the hardware information is confirmed, the hardware information is read directly. There is also a case where the mobile device is disassembled or reassembled, thus causing trouble to confirm a mobile device, and that a mobile device integrity can be identified as including the motherboard, the processor, and the storage device.

2.2 Dynamic Behavior Analysis of Mobile Terminal Users

The dynamic behavior data of mobile users is the most important source of large data. By use of the applications on the user's mobile devices, the user dynamic behavior data are collected within the permission of the law.it is necessary to ensure that users of data security and confidentiality. The dynamic behavior of mobile end users generally includes: moving routes, network behavior and social behavior.

Based on the big data of the mobile trajectory analysis, through the mobile device application positioning system, through a certain technology to obtain the location of mobile devices, generally through the IP address, MAC address, Wi-Fi information, GPS information, tracking user location information, to analyze the user's daily walking route, often go to the place, stay time, frequency and other data. Using large data clustering methods and correlation analysis, once you want to identify the user identity, can be used as an important basis for identification of reference.

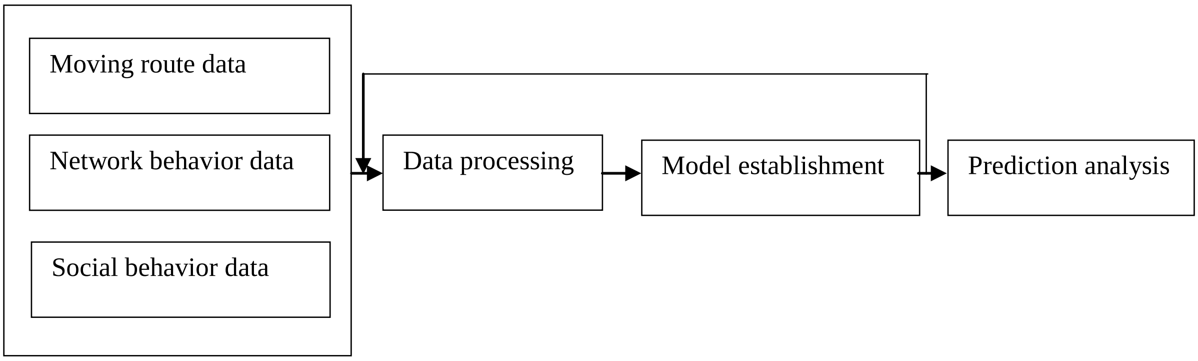

Fig. 1 Steps to process big data

Moving route is user’s daily trajectory. Through the mobile device application positioning system, or a certain technology to obtain the location of mobile devices, the position data are collected through the IP address, MAC address, Wi-Fi information, GPS information, etc. By tracking user location information and analyzing the user's daily walking routes, we can know the place where a person often goes and the time when he stays there, and the frequency and so on. Using big data clustering method and correlation analysis, the classification of different users and single user behavior clustering analysis can be used as an important basis for mobile user identification reference.

Network behavior generally refers to the daily use of the Internet according to the behavior, such as: browse the site information, search behavior, shopping APP, pay APP and so on. By analyzing the user's online behavior information, we can get the following information: user interests, games, chat, community usage, stock, reading, music, integrated portal access, browsing information, and so on. In addition, by the use of IM APP users, we can get the trend analysis information of the users: day Frequency, week Frequency, business Class, access Time, and so on.

Social behavior data are a series of valuable data that a person will produce in the social activities. Common social behavior such as shopping behavior, consumer behavior, travel behavior, social behavior and so on. Through the comprehensive analysis of the social behavior data, they help us determine whether a person's social behavior is his own in the mobile user identification.

Users’ big data collected by mobile devices are preprocessed, and models are built by use of machine learning or deep learning. In the process of user identification, the model algorithm is used to determine whether the user's identity is consistent with some of its regular activity attributes. Finally, use the application's authentication to confirm user information.

3. Conclusion

The use of big data technology for personal behavior analysis, combined with the traditional identification technology, provides more reliable mobile user identification accuracy.

References

[1] LI Zhong-miao, WANG Yu-xiang, TAO Xiao-long, CAO Wen-jie.Is a Method of Identity and Correlation for Security Domain [J]. Software Journal, 2016, (Issue 2): 170-174.

[2] DONG Ji, GUO Wei, DU Yun.Study on the Problem of Identity and Authentication Security in "Internet +" [J]. Internet World, 2015, (Issue 10) .9-14.

[3] Nan Wang;Qindong Sun;Yadong Zhou;Si Shen.A Study on Influential User Identification in Online Social Networks[J].Chinese Journal of Electronics,2016,Vol.25(3): 467-473.

[4] Reza Zafarani;Lei Tang;Huan Liu.User Identification Across Social Media[J].ACM Transactions on Knowledge Discovery from Data,2015,Vol.10(2).

[5] Sanjin Pajo;Dennis Vandevenne;Joost R. Duflou.Automated feature extraction from social media for systematic lead user identification[J].Technology Analysis & Strategic Management,2017,Vol.29(6): 642-654.

[6] Xiaoping Zhou;Xun Liang;Haiyan Zhang;Yuefeng Ma.Cross-Platform Identification of Anonymous Identical Users in Multiple Social Media Networks[J].IEEE Transactions on Knowledge & Data Engineering,2016,Vol.28(2): 411-424