Advances in Computer Science Research (ACSR), volume 72
IV International Research Conference "Information Technologies in Science, Management, Social Sphere and Medicine" (ITSMSSM 2017)

Increase in security of authentication services through additional identification using optimal feature space

Andrey Iskhakov
Faculty of Security
Tomsk State University of Control Systems and Radioelectronics
Tomsk, Russia
iskhakovandrey@gmail.com

Roman Meshcheryakov
Faculty of Security
Tomsk State University of Control Systems and Radioelectronics
Tomsk, Russia
mrv@security.tomsk.ru

Sergey Iskhakov
Faculty of Security
Tomsk State University of Control Systems and Radioelectronics
Tomsk, Russia
iskhakov.sy@gmail.com

Alexey Krainov
Faculty of Physics and Engineering
Tomsk State University
Tomsk, Russia
akrainov@ftf.tsu.ru

***Abstract*—The research focuses on topical issue of the Internet security. In particular, the issue of level increase in security of authentication services through additional identification using optimal feature space is being considered. This article is devoted to the practical application of additional identification technologies in authentication services. The paper presents sets of informative features characterizing the access subject. A classification of methods for identifying the user's work environment is proposed. The article presents the experimental results of intercomparison between scientifically-grounded methods and technologies for identifying the user's work environment.**

Keywords—authentication; identification; information security; IoT; identification methods; attribute; user; browser fingerprint; cookies

I. INTRODUCTION

The global network of Internet is one of the main instruments of mass communication today. Its fast-growing development is inseparably linked with new scientific discoveries and technological innovations in different areas of the IT-industry. These factors promote the evolution of information systems enabling remote communication with users. There is an increase in the amount of user data in the network because today almost everything can be done online: from payment for utility services to buying air tickets. Obviously, at the same time, there is a growing number of cyber attacks [1].

This problem in particular is specific to the Internet of things (IoT) infrastructure [2, 3] – the concept of the network of physical objects ("things") equipped with integrated technologies for interaction with each other or with the environment. The importance of this information exchange sector lies in the ability to rebuild economic and social processes, avoiding the necessity for a human participation in some of the actions and operations [4]. Furthermore, the emphasis in this concept is given to information collected by smart devices, turning the data into a guide for action for individual proprietors or groups of people, is of value in this concept [5].

One of the important tasks in the security of network system elements such as IoT is the implementation of an effective auditing and access control subsystem for users who establish remote connections to network elements and smart devices [6]. In particular, within the scope of this study, issues of enhancing the security of authentication services through additional identification of access subjects will be considered.

II. NECESSITY OF THE ADDITIONAL IDENTIFICATION

As an example, consider the infrastructure of cloud-based video surveillance system in the IoT concept. Such systems long ago stopped being systems merely streaming video from camera onto the device for viewing. The era of cloud-based video surveillance came today: the era of the smart remote monitoring which does not need constant control of the operator. Such video surveillance is called smart because its cloud computing architecture is capable not only of storing and reading out video information, but also realizing certain algorithms of the video analysis. The standard system of local video surveillance can be called passive and dependent, and cloud-based video surveillance can be described as active video surveillance capable of analyzing what it "sees" and reporting to users important events by means of notification messages.

Restriction of access to a control bar by a similar information system through standard authentication methods (login / password) is often not enough. It can be connected to business-processes of application object operation. In such cases information security administrators are stimulated to integrate additional mechanisms of visitor identification. However, the objective look at similar mechanisms helps to reveal imperfection of the existing approaches and technologies of user identification. For example, the widespread methods based on logging of visitors' IP addresses and storage of control footing on client's device (cookies technology) are not capable of resisting the usage of network address translation mechanisms (NAT) [7, 8], the proxy services, anonymizers and dynamic addressing. Besides, in such cases the user exercises complete control over cookies contents (including legitimate opportunities of destruction and change of data).

At the same time, there are data acquisition methods characterizing user operating environment. The user operating environment is data about operating system, fonts, screen parameters, installed plug-ins, visited links, etc. Attempts of use of the listed data as identification signs are known [9-11].

This work was supported by the Ministry of Education and Science of the Russian Federation within 1.3 federal program «Research and development in priority areas of scientific-technological complex of Russia for 2014-2020» (grant agreement № 14.577.21.0172 on October 27, 2015; identifier RFMEFI57715X0172).

However, using such a technology involves increase in traffic volume that is not acceptable for many network devices, including, for example, representatives of the IoT infrastructure.

Thus, the primary objective of the research is defining optimal feature space and an additional identification method allowing to increase reliability of identification, applied in authentication service, of users with the available entries in the database of an information system.

The most common and popular option for creating a convenient and cross-platform interface for interacting with users of the IoT infrastructure is using web-oriented technologies operating such languages as HTML, CSS, JavaScript, etc [12]. Usually, web pages are viewed using special software - browsers, which are widely used today. Authentication methods for such applications are divided according to the resource type, the structure of the network organization and the technology used in the matching process. Table 1 presents the most common authentication methods.

TABLE I. AUTHENTICATION METHODS IN THE WEB APPLICATIONS

Method	Implementation / protocols	Purpose
By password	HTTP authentication (Basic, Digest, NTLM, Forms)	Users authentication
By one-time password (OTP)	Forms	Enhanced user authentication
By certificate	SSL/TLS	Strong user authentication in secure applications; service authentication
By access keys	–	Authentication of services and applications
By “tokens”	SAML, WS-Federation, OAuth, OpenID Connect	Delegated user authentication; delegated application authorization

The variety of these methods makes it possible to apply a differentiated approach to the construction of the authentication services depending on the tasks and the resources. The performance features of some systems necessitate the development of a method for the additional access subject identification, which allows detecting suspicious user sessions. Example of such cases is the local IoT infrastructure of cloud-based video surveillance, limited by computing resources, considered by the authors. In such system there is no possibility of installing a full-fledged Intrusion Detection System (IDS) [13, 14], capable of detecting abnormal user activity.

III. AUTHENTICATION FEATURE SPACE

In this research following classification method for identifying a user operating environment [15] was formed by the authors:

A. Setting a unique user ID

The methods are based on the use of the following featutes (characteristics):

- Cookies – a small data fragment stored on the user's computer;
- Local Shared Objects (LSO) – the type of metadata that is stored as files on each user's computer; today all versions of Flash Player use LSO;
- Isolated Storage – isolated Silverlight storage; as with LSO, from a technical point of view, there are no barriers to storing the session identifiers;
- HTML5-repositories (localStorage, File API and IndexedDB) are intended for maintenance of constant storage of the arbitrary portions of the binary data corresponding to a specific resource;
- Browser cache objects. This mechanism was not intended to be used as random access storage. But if the service returns a JavaScript to the user document with a unique identifier inside its body and sets the value of headers “Expires / max-age” as distant future, then the identification script will be stored in the browser's cache. After such a manipulation it is possible to access this script from any page in a network, simply requesting the script download from the known URL;
- abstract identifier ETag (tag of the cached document version);
- The Last-Modified header (date of the cached document version);
- Application cache (HTML5) - a set of functions that provides advanced caching of web application resources;
- SDHC dictionaries. This method is a compression algorithm developed by Google, which is based on the use of the dictionaries provided by the special server. The client receives a dictionary file containing the lines that may appear in subsequent replies. After that the server can simply refer to these elements inside the dictionary, and the client will independently generate a page on their basis;
- Use of the internal DNS browser cache;
- Other storage mechanisms (window.name or session.storage) which allow to store and request an unique identifier in such a way that it remains even after deleting all browsing history and site data;
- Use of the protocol features. Origin Bound Certificates (persistent self-signed certificates that identify the client for an HTTPS server) - as a unique identifier, it’s possible to take a cryptographic certificate hash, provided by the client as part of a legitimate SSL handshake. TLS also has “session identifiers” and “session tickets” mechanisms that allow clients to resume interrupted HTTPS connections without performing a full handshake.

B. Use of the calculated features of the user automated system.

- browser “stamps” that are based on creating identifiers by combining a set of parameters available in the browser environment. Each of the identifiers separately

is of little interest, but their combination forms a unique value for each machine. For example, User-Agent string, system clock deviation, CPU and GPU information, lists of installed fonts and plug-ins, information about extensions and additional software;

- "network stamps" – the values of external and local IP addresses, outgoing TCP / IP connections port numbers, information about the used proxy server, etc.

C. *Analysis of dynamic identification features (behavioral analysis)*

This method allows to identify clients among different browser sessions, profiles and in case of private browsing. For example, the following can be used:

- mouse gestures characteristics;
- frequency and duration of keystrokes;
- data from the accelerometer;
- zoom level, use of special features.

IV. THE EXPERIMENT PROGRESS

The following scientific methods and technologies for identifying the user's working environment identifying were used in the study:

1) Identification method using the component profile, which is a tuple of the most informative user's working environment data [16];

2) Cross-Browser Fingerprinting method (CBS), based on computer profiling according to the time of various graphical operations execution (per minute) [17];

3) Panopticlick Fingerprints technology;

4) Evercookie technology [18], which combines HTTP cookie, Flash cookie, Silverlight Isolated Storage, PNG and canvas cookie, session storage, local storage, Indexed DB, ETag, Java Persistence;

5) FingerprintJS technology.

According to [19], these methods find wide practical application in tasks of users identification on the Internet. For their comparative analysis an experiment was conducted. It included calculation of the identification reliability and the assessment of the computational resources requirements (run-time value). In order to obtain statistical characteristics, experimental data was collected on the basis of 3 cloud based video surveillance systems observing public places for 3 months. The selected systems used basic password authentication.

Fig. 1 shows the result of the experiment (evaluating the time of performance).

The data was obtained by averaging the estimate of 15,000 requests. The average value of the first request to the main content of the page takes the same time.

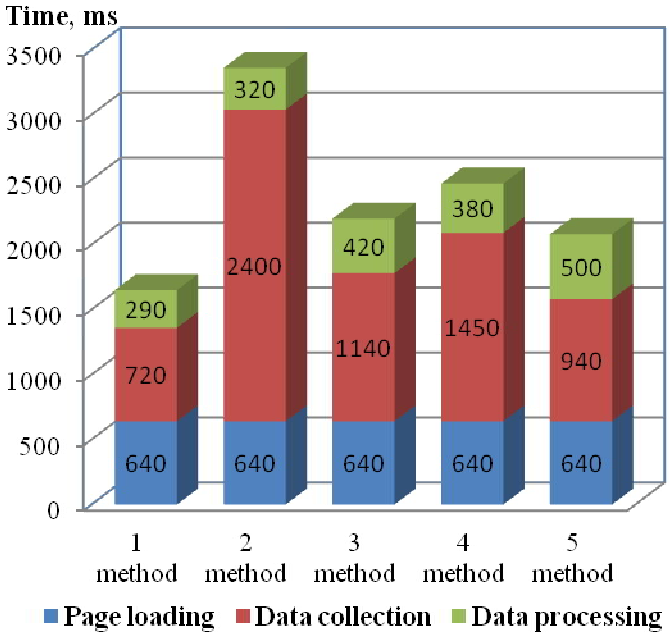

Fig. 1. Results of testing the methods speed

The process of data collection of the 2nd method takes much more time than the others. This is due to the need to perform a variety of graphical operations (for example, drawing a raster image on the face of a cube using WebGL with hardware video acceleration).

The next step was evaluating the dependence of the identified users number on the noise level (fig. 2). By noise in this work we mean deliberately or accidentally distorted data that cannot serve as basis for identification.

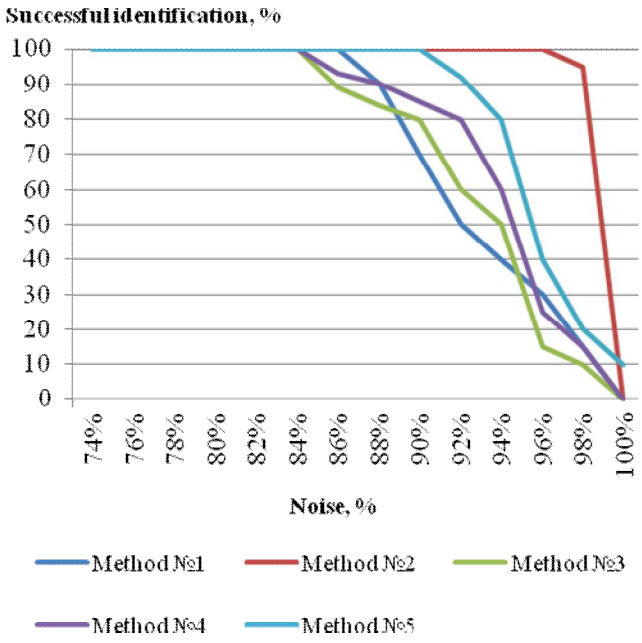

Fig. 2. Dependence of the methods effectiveness on the degree of the sampling noisiness

The experiment results reveal that all the methods show maximum efficiency when the sampling noise is up to 84% inclusive. With more than 90% noise introduced, all methods except the second and the fifth considerably worsen their results. With 95% data distortion only the Cross-Browser Fingerprinting can be attributed as robust. For example, with 96% noise level the percentage of correct identification in methods 1, 3-5 does not reach 40%.

So the second method showed the best result of user identification reliability, and while the first one – in terms of speed. However, it is necessary to take into account that in the above-mentioned case the identification accuracy affects the rate of content transfer. Therefore, the practical application of the second method in such services as cloud-based video surveillance will lead to significant delays in users' work. So, authors consider it appropriate to combine the first and second methods, taking into account the peculiarities of business processes of the particular information system.

In the process of studying these technologies, the following massive deficiencies were revealed:

1) The change in the new browser versions release policy has recently reduced the effectiveness of using the UserAgent attribute;

2) Apple products (IPhone, IPad and others) are characterized by a high degree of hardware unification. This means that the byte array obtained from the Canvas Fingerprint will be the same for the IPhone (with the IOS operating system up to version 8.1). This leads to a decrease in the identification accuracy.

3) A large number of computers that are still using older versions of Internet Explorer do not allow to get a list of installed plug-ins.

CONCLUSION

Previously used technologies [20] make it possible to identify users of one browser with an acceptable accuracy, but modern methods allow to identify users that are intentionally using several different browsers.

Of course, such anonymization methods as the use of Tor networks allow to bypass similar checks of additional identification. However, the list of output nodes of this technology is published and is constantly updated. The developer of authentication services should automatize the update of this registry and set an appropriate lock.

The experiment results show that today there is no universal tool that allows for reliable additional identification of users with minimal labor input. As with authentication technology, it is necessary to apply a differentiated approach to choosing an optimal feature space. It allows to increase the reliability of user identification with existing entries in the information system database used in the authentication service.

REFERENCES

[1] S. H. Y. Hsu and S. J. Dick, "Information sharing & cyber threats", 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), 2017, pp. 89 - 94.

[2] K. Gupta and S. Shukla, "Internet of Things: Security challenges for next generation networks", 2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH), 2016, pp. 315 - 318.

[3] K. Zhao and L. Ge "A Survey on the Internet of Things Security", 2013 Ninth International Conference on Computational Intelligence and Security, 2013, pp. 663 – 667.

[4] S. N. Shukla and T. A. Champaneria, "Survey of various data collection ways for smart transportation domain of smart city", 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2017, pp. 681 – 685.

[5] D. B. Kochieva, Iot devices secrity. All ingenious is simple [Electronic resource]. URL: https://www.atlex.ru/wp-content/uploads/2017/05/ kochieva-zaschita-iot.pdf (access date: 12.10.2017) (in Russian).

[6] A. Iskhakov, R. Meshcheryakov and Yu. Ekhlakov, "The Internet of Things in the security industry", INTERACTIVE SYSTEMS: Problems of Human - Computer Interaction. - Collection of scientific papers, 2017, pp. 161 - 168.

[7] V. Olifer, N. Olifer A. Computer Networks: Principles, Technologies and Protocols for Network Design. John Wiley & Sons, 2005, P. 1000.

[8] S. Iskhakov, A. Shelupanov, R. Meshcheryakov, "Simulation modelling as a tool to diagnose the complex networks of security systems", Journal of Physics: Conference Series, 2017.

[9] E. E. Bessonova, I. A. Zikratov, Yu. L. Kolesnikov, V.Yu. Roskow, "Method of user identification in the internet network", Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2012, pp. 134 - 138 (in Russian).

[10] I. Cantor, Methods of identification on the Internet [Electronic resource]. URL: http://javascript.ru/unsorted/id (access date: 12.10.2017) (in Russian).

[11] P. Eckersley, How Unique Is Your Web Browser? [Electronic resource] URL: https://panopticlick.eff.org/browser-uniqueness.pdf (access date: 03.10.2017).

[12] Find and compare Javascript IoT projects and hardware. Nodejs on your Pi, Arduino or custom Bluetooth or Wifi board [Electronic resource] URL: https://www.postscapes.com/javascript-and-the-internet-of-things/ (access date: 03.10.2017).

[13] J. N. Davies, P. Comerford, M.V. Verovko, I.S. Skiter, I.S. Posadska, "Intrusion prevention within a SDN environment", MMC, 2017, vol. 1, pp. 39-48.

[14] A. A. Gendreau and M. Moorman, "Survey of Intrusion Detection Systems towards an End to End Secure Internet of Things", 2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud), 2016, pp. 84 – 90.

[15] A. Zhukov, Fingerprinting the browser. How users track the web [Electronic resourse] URL: https://xakep.ru/2015/01/30/user-web-tracking-howto/ (access date: 10.10.2017) (in Russian).

[16] E. E. Bessonova, Method for identifying users on the Internet using a component profile (dissertation) [Electronic resourse] URL: https://isu.ifmo.ru/index/B996F9609F0750E3BBDF52445A22CFC1 (access date: 14.10.2017) (in Russian).

[17] Y. Cao, S. Li, E. Wijmans, "(Cross-)Browser Fingerprinting via OS and Hardware Level Features", Conference: Network and Distributed System Security Symposium, 2017 [Electronic resourse] URL: http://yinzhicao.org/TrackingFree/crossbrowsertracking_NDSS17.pdf (access date: 14.10.2017).

[18] G. Fleishman, How to kill the evercookie and supercookie, the cockroaches of tracking, 2017 [Electronic resourse] URL: https://www.macworld.com/article/3152056/privacy/how-to-kill-the-evercookie-and-supercookie-the-cockroaches-of-tracking.html (access date: 14.10.2017).

[19] Methods for identifying users on the Internet, 2017 [Electronic resourse] URL: https://serfmoney.ru/cpa/metody-identifikatsii-polzovatelya-v-internete/ (access date: 15.10.2017).

[20] Methods for identifying on the Internet, 2017 [Electronic resourse] URL: http://javascript.ru/unsorted/id (access date: 15.10.2017).