UDC 004. 9 **doi: 10.26906/SUNZ.2023.3.103**

A. Kapiton[1], O. Dziuban[1], R. Talybov[1], B. Kanivets[1]

[1] National University «Yuri Kondratyuk Poltava Polytechnic», Poltava, Ukraine

COMBINED METHOD FOR PROTECTING USER PRIVACY WHEN USING A WEB BROWSER

Abstract. The article examines the issue of digital tracking, which allows users to be identified using cookies and digital fingerprints. A comparative analysis of current web tracking methods shows that approaches based on detecting and blocking web trackers are less effective than methods using value substitution. These approaches do not protect against passive fingerprinting techniques and are easily detected, making them vulnerable to user identification. To solve this problem, a hybrid method of forging transmitted data is proposed, including random generation and data fusion, as well as detecting trackers for immediate protection. Attributes with a wide range of possible values and complex compatibility are proposed to be replaced by real values with minor distortions.

Keywords: security, confidentiality, web browser, digital fingerprint, cookie, fingerprint, web tracker.

Introduction

One of the most common privacy vulnerabilities on the Internet is digital tracking, the slow, constant, and relentless accumulation of user data [1]. The components of this vulnerability are browsing history, application usage, online purchase information, and geolocation information. These insignificant individual data can be combined into a significant whole, which allows you to ‘identify the user. Trackers collect data on clicks, views, clicks and transitions and create extensive behavioral profiles, for example, to show targeted ads or optimize website performance. However, this information can reveal political preferences, religious beliefs, race and ethnicity, education level, monthly income, consumer habits, as well as physical and mental health, and in the hands of malefactors can be used for discrimination, blackmail or propaganda [2]. For this reason, the issue of protecting one's own privacy is particularly acute. To increase the confidentiality of user data, various legal, organizational and technical measures are taken. These methods include detection and blocking the execution of the web tracker program code, as well as the substitution of data transmitted to the web resource in order to create a false user profile.

Analysis of recent research and publications. Analysis of the works of leading scientists who expressed their opinions in materials published in magazines, collections of reports of conferences held in our country and abroad, gives confidence in the relevance of the researched issues. I. Sanchez-Rola, I. Santos investigate the problems of automatic analysis of web tracking. C. Giuffrida, S. Bardin, G. Blanc consider features of detection of intrusions and malicious software and vulnerability assessment [1]. G. Gebhart, B. Cypher study aspects of the use of information technologies for corporate tracking [2]. R. Laperdrix, W. Rudametkin, B. Baudry analyze the method of redirecting modern web browsers to create unique browser fingerprints [3].

Main part

Web tracking mechanisms mainly include two methods of tracking: the use of COO files [3, 4] and the creation of a digital fingerprint of the device (browser). Sometimes they can be used together to increase the accuracy of identification. A cookie is a piece of data that websites store in the user's browser and process each time they connect to the site. Cookies often store a unique user identifier assigned to him by a web resource, which can later be used to personally identify the user on other web resources. Cookies are classified according to their expiration date: persistent - such cookies can be stored from several months to several years and session cookies, which are deleted after the browser is closed. A cookie-based tracking mechanism consists in hosting a third-party server script on a website by adding a piece of code. Such a script will call JavaScript code hosted on a third-party web server. This code can perform various functions, for example, download and display an advertising banner of an affiliate advertising network or upload a special software library for developers to the site. It can also contain code that allows this server to install third-party files in the user's browser and track them when this script is loaded. Thus, the tracking web server can place its tracker on different websites and receive data about the user's browsing history, time spent on certain pages, and other data that is provided by a website with an embedded tracking script (Fig. 1). The disadvantage of this tracking method is that the files, and therefore the identification information about the user, are stored on the user's side. This feature leaves the user the option to delete coom files, thereby temporarily stopping digital tracking.

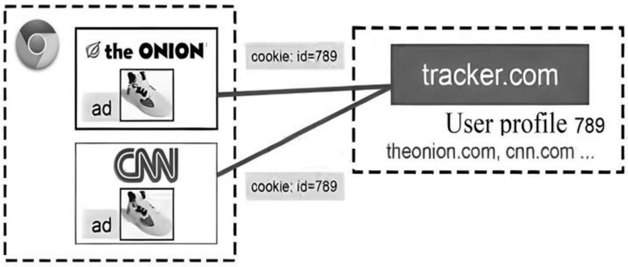

Fig. 1. Cookie-based tracking

Therefore, the mechanism of taking a digital fingerprint or fingerprinting is often used as a web tracker. A digital fingerprint is a set of information about the software and hardware-software configuration of a client device connected to a web resource. The basic concept of digital fingerprinting is the initial and re-identification of a user based on device-specific information. Sources of information for the formation of digital fingerprint are: data about user settings; equipment characteristics; operating system characteristics; browser characteristics; user behavior.

Digital fingerprints are classified by source and are divided into: passive based on information automatically transmitted by the browser in the content of the HTTP header and active - collected when executing JavaScript code on the client side, to obtain information provided by various web browser programming interfaces (API). In addition, there are modern complex techniques for the formation of digital prints. They are the most efficient, because the result of their work depends on the hardware configuration parameters of the device, for example, on the graphics accelerator or sound card used in the device. These techniques include Canvas Fingerprint, WebGL Fingerprint, and Audio Fingerprint. software configuration of a client device connected to a web resource. The basic concept of digital fingerprinting is the initial and re-identification of a user based on device-specific information.

Canvas Fingerprint is a method for getting device characteristics that uses the HTML 5 Web Technology Canvas element. This element is a surface capable of displaying 2D graphics in a browser using client-side code. The essence of the method is that some graphic elements and text are drawn on the element hidden from the user. The result of rendering the same code in this case may differ, depending on the operating system, font library, video card, driver version, browser engine, and so on. The resulting image is represented as a string encoded in Base64 (Fig. 2).

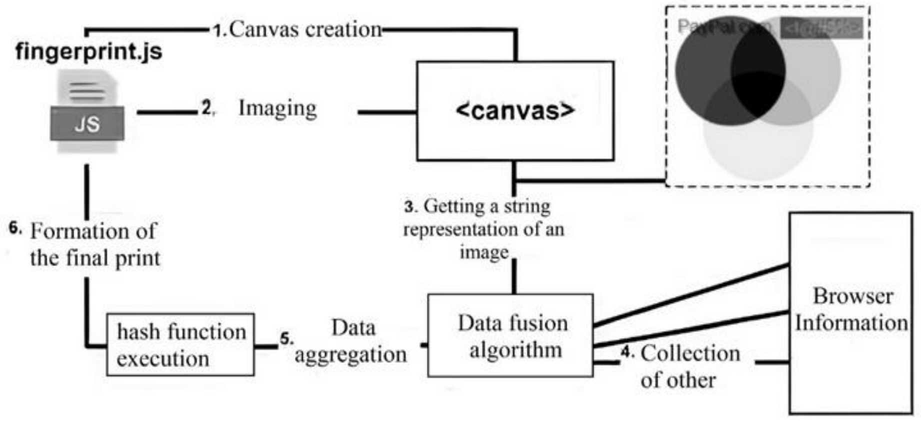

Fig. 2. Canvas Fingerprint

WebGL Fingerprint is an evolution of Canvas Fingerprint. In this case, the web tracker code causes the browser to render scenes using 3D graphics technology, which are also superimposed with various graphic effects. The resulting image is translated into the resulting byte array, which will differ depending on the hardware and software configuration of the device. Audio Fingerprint is to generate audio signals. The characteristics of such signals may also differ due to differences in the hardware or software used. This method is implemented using the Web Audio API web browser interface. In general, this fingerprinting technique is a sequential chain of nodes for working with audio in the browser and occurs in three stages: generating an audio signal, analyzing the characteristics of an audio signal, and hiding an audio signal.

Most browsers do not protect user information by default. This means that they allow you to store cookies, including those received from third parties, and also do not limit in any way the existing interfaces used in the formation of digital fingerprints. Blocking third party cookies will prevent the user from being uniquely tracked by third party websites. The task of increasing the user's protection against fingerprinting is much more difficult, since in this case the countermeasures provided by the browser are very limited and ineffective.

The essence of the method lies in the fact that third parties rely on the stability of digital fingerprints to associate them with one device. By sending random values instead of real ones, the collected fingerprints become so different and unstable that the tracker cannot identify devices on the network. The downside is that randomizing inappropriate combinations of certain parameters can result in lower user privacy.

Unification assumes that the transmitted data is brought to uniformity for each user. The problem of this method is the inability to change some attributes, for example, the characteristics of the hardware configuration of the device. The presence of unmodified unique attributes becomes a serious security vulnerability for the user, especially if they are incompatible with unified data. To solve the problem of counteracting digital tracking, a variant of a hybrid approach to replacing transmitted data is proposed, including both randomization and unification of some data. The choice of the data spoofing method is determined by the fact that it provides protection against all the considered methods of forming digital fingerprints, and it also lacks the stage of detecting a tracker, which allows for instant protection. With the proposed varnant, characteristics that have a wider range of possible values and are difficult to unify (Canvas, WebGL, Audio) will be replaced by real ones with the addition of small distortions. In turn, the attributes that must comply with the value adequacy property and whose change may disrupt the site will be subject to unification [3].

At the same time, it is planned to take into account the connectivity of the parameters with each other, as well as the adequacy of their values. It is proposed to unify the following parameters: HTTP headers (User-Agent, Accept-Language, Content-Language, DNT) and global JavaScript browser settings (navigator userAgent, navigator.appVersion, navigator.userAgentData, navigator.vendor, navigator.platform, navigator. language, navigator.doNotTrack, screen.height, screen.width, screen pixelDepth). First of all, before sending an HTTP request with the required headers to the requested web server, the value of the described headers will be replaced and transmitted already in a modified form. Thus, Protection against passive digital fingerprints will be provided.

Then, after the page starts loading, but before any other scripts are loaded and executed, a script will be injected and executed that will replace the values of the globally available JavaScript parameters listed earlier. After that, all other scripts on the site will be loaded. Thus, all web trackers, when trying to get the values of the characteristics through the browser API, will receive already changed values.

The next step is to provide protection against the well-known fingerprinting techniques described earlier. The general principle of protection operation for each method of hiding is the same. The security script will override the JavaScript implementation of the methods used in fingerprint generation. In the future, when a website tries to generate a digital fingerprint, when using methods, the overridden code will be executed.

The algorithm of fingerprinting technology using WebGL is similar to the previous one. However, in this case, noise will be added to the rendered scene data buffer, which is subsequently used to form a fingerprint.

The data buffer can contain the values of the coordinates of the vertices on which the drawing takes place, or the values of the colors. For Audio Fingerprint, it is proposed to replace the software implementation of some nodes that receive the characteristics of the generated audio signals. In particular, it is proposed to change the analyzer node (AnalyserNode) and the output buffer node (AudioBuffer) (Fig. 3).

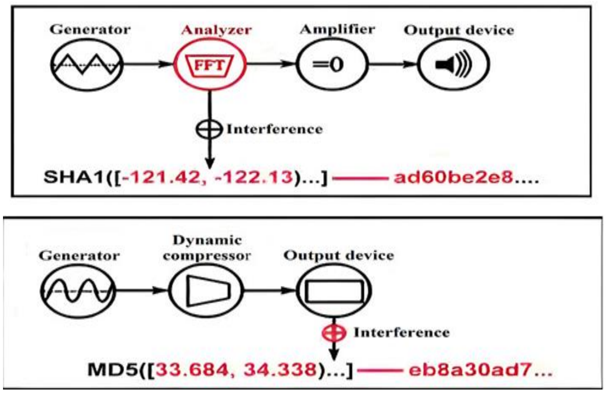

Fig. 3. Audio Fingerprint protection

Conclusions

The possibility of violating the user's privacy when working with a web browser using web trackers is shown. The main mechanisms of digital tracking are considered, the principles of their operation, varieties and modern implementations are described. It was revealed that tracking using digital fingerprints pose the greatest danger, since they cannot be effectively blocked by the browser. A hybrid method for counteracting fingerprint tracking is proposed, which consists in unifying simple characteristics while observing the relationship and adequacy of values, as well as randomizing complex characteristics that depend on the hardware configuration of the device and are not subject to unification.

References

1. Sanchez-Rola I., Santos I. Knockin' on Trackers' Door: LargeScale Automatic Analysis of Web Tracking. In: Giuffrida, C., Bardin, S., Blanc, G. (eds) Detection of Intrusions and Malware, and Vulnerability Assessment. DIMVA 2018. Lecture Notes in Computer Science, vol. 10885. Springer. doi:10.1007/978-3-319-93411-2_13
2. Gebhart G., Cyphers B. Behind the One-Way Mirror: A Deep Dive Into the Technology of Corporate Surveillance / Electronic Frontier Foundation. URL: https://www.eff.org/wp/behind-the-one-waymirror
3. Laperdrix P., Rudametkin W., Baudry B. Beauty and the Beast: Diverting Modern Web Browsers to Build Unique Browser Fingerprints // Materials 2016 IEEE Symposium on Security and Privacy (SP). С. 878-894, doi:10.1109/SP.2016.57

Received (Надійшла) 23.07.2023
Accepted for publication (Прийнята до друку) 13.09.2023

Комбінований метод захисту конфіденційності користувача під час використання веб-браузера

А. Капітон, О. Дзюбан, Р. Талибов, Б. Канівець

Анотація. У статті розглядається проблема цифрового відстеження, яка дозволяє ідентифікувати користувачів за допомогою файлів cookie та цифрових відбитків пальців. Порівняльний аналіз сучасних методів веб-відстеження показує, що підходи, засновані на виявленні та блокуванні веб-трекерів, є менш ефективними, ніж методи, що використовують заміну значень. Ці підходи не забезпечують захист від пасивних методів зняття відбитків пальців і легко виявляються, що робить їх вразливими для ідентифікації користувача. Для вирішення цієї проблеми пропонується гібридний метод підробки переданих даних, включаючи випадкову генерацію та злиття даних, а також виявлення трекерів для негайного захисту. Атрибути з широким діапазоном можливих значень і комплексною сумісністю пропонується замінити реальними значеннями з незначними спотвореннями.

Ключові слова: безпека, конфіденційність, веб-браузер, цифровий відбиток, cookie, відбиток, веб-трекер.