# Estimation of the time for calculating the attributes of browser fingerprints in the user authentication task

*A.Y.* Iskhakov[1,*] and *A.A.* Salomatin[1]

[1]V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, 65, Profsoyuznaya, Moscow, 117997, Russia

**Abstract.** This paper presents an overview of the essence of web space which may contain information required to identify and authenticate users on the Internet in order to prevent attempts at malicious acts. Essences with a common content or the same detection method are structured into groups. The greatest attention is paid to one of these groups of entities: «the browser fingerprint» group. An approach and software have been proposed that allow for automated search of the values of entities belonging to this group and for estimating the time of this search for a specific infrastructure in order to develop adaptive authentication mechanisms. Moreover, an analysis of the average time to calculate the values of attributes has been carried out for the most informative browser fingerprints.

## 1 Introduction

The need to develop scientific and technical solutions in the area of adaptive user authentication is beyond doubt today both among representatives of the scientific community and major developers of information systems. At the same time, the implementation of algorithms in the field of risk-oriented authentication with dynamic configuration of verification factors [1] requires the use of a large number of sources of information that allow you to decide on the legitimacy of the performance of a particular operation by the user. In connection with the widespread introduction of web-oriented technologies, the authors are working on the formation of methodological and algorithmic support which makes it possible to build personalised models of web services users [2] by analyzing the various entities of the web space.

It should be noted that this study is part of the authors' development of a dynamic adaptive authentication system. This technology implies that any authentication factor has a level of trust, and the choice of an authentication factor depends on the risk level of a particular operation. Such decisions are undoubtedly extremely important and relevant. On the one hand, they provide an opportunity to protect the client from compromising the account by analysing the activity of its profile for anomalous characteristics, and on the other hand, to strike a balance between convenience and reliability, providing in some cases the opportunity to reduce the number of authentication procedures. It is therefore necessary to build

* Corresponding author: karateka30@mail.ru

personalised models as part of building each user's security profile. To this end, it is proposed to analyse the various web essentials associated with their virtual account.

## 2 Classification of entities and their descriptions

The abstract object relating to the client-server interaction business logic, where the client is a real user and the server is a technical platform containing a web-oriented service is called the essence of the web space in this paper. The purpose of this study is to improve the accuracy of correct user authentication. In order to achieve this goal, two tasks need to be solved: to study the types of entities and their descriptions, and to conduct an experiment that will lead to conclusions about the effectiveness of the selected entities and the approach applied to them.

In the course of the study and taking into account [3, 4], the structure of the entity groups that can act as sources of a digital user trail has been drawn up. Figure 1 shows the main blocks (source categories) as well as several individual examples for each source.

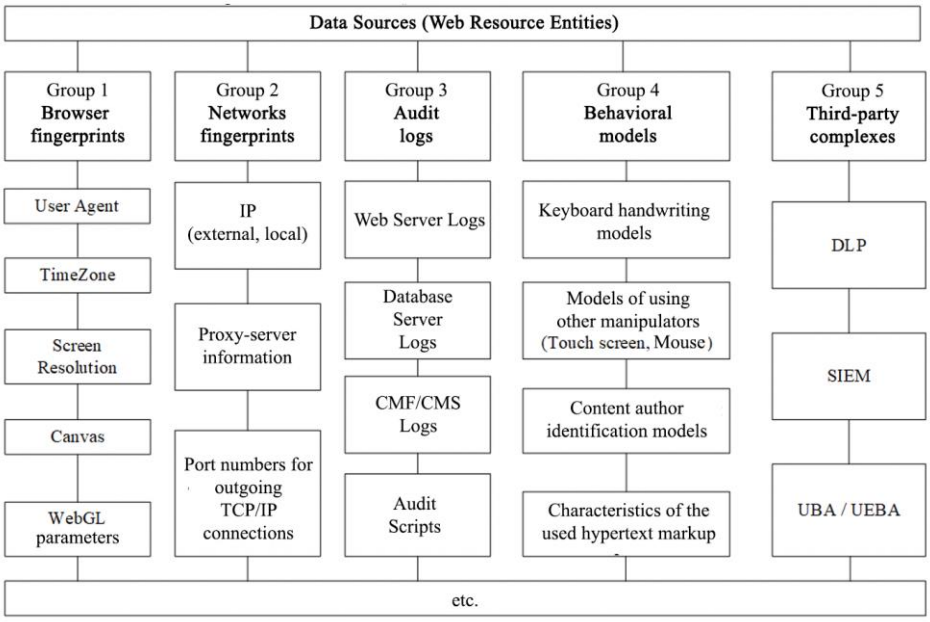


**Fig. 1.** Data sources allowing to obtain a personalised user model for adaptive authentication.

The first group of entities combines informative, unique attributes called browser fingerprints. The group can include the following identifiers: User Agent, TimeZone, Screen Resolution, Canvas, WebGL parameters.

- UserAgent is an identifier string that contains data about the operating system (OS) and the user's browser, as well as the version of the browser's assembly. Due to the fact that updates for modern browsers are issued with a regular frequency, the assembly number cannot serve as an informative attribute that identifies the user in the web space.

- TimeZone. The name of the time zone and a numeric offset relative to 0 zone are usually used as identifying attributes.

- ScreenResolution. It is a set of two numbers: the length and width of the screen, expressed in pixels.

- Canvas is the feature that characterise the ability of a particular device to generate a two-dimensional raster image. Two parameters can be obtained which identify this process. The

first parameter is canvas winding. It is binary and defines the ability to use this functionality. For most browsers the value of this parameter is "true". The second parameter is the unique canvas identifier string, which is formed by converting a gradient colour object using the base64 encoding.

- WebGL is related to image construction by using shaders and other WebGL functions (e.g. anti-aliasing, rendering, filtering). One of the attributes defining this essence can be a unique identifier string, which is obtained by converting a gradient object with shaders using the base64 encoding, taking into account all exceptions and WebGL features. However, the marked WebGL capabilities can be obtained and investigated separately.

The second group of entities is network fingerprints. These include IP addresses (local and external), information on using Proxy servers, and port numbers for outgoing TCP/IP connections.

- An external IP address is understood to mean an IPV4 or IPV6 address that characterises the connection device globally, within the international Internet network. This type of IP address is not always an accurate indicator that identifies the user, as providers often provide customers with dynamic IP addresses. The use of a local IP address as one of the identifying attributes is not reasonable because ranges of private addresses are very limited.

- The Proxy server is an intermediary between the client and the server. In this case, the web server does not accept data from the user's device, but from a proxy server acting as a certain transit point. Moreover, data exchange may also take place in the opposite direction. In this case, the proxy server may be used not only by scammers to hide their address and intercept user requests, but also by ordinary users, for example, to anonymise and protect against certain network attacks.

- Port numbers for outgoing TCP/IP connections are numerical values that servers use as access points to IP addresses. When exchanging data, two IP addresses and two port numbers for a particular protocol are used. In turn, the port is used to send data in the process and have it confirmed by the recipient's operating system.

The third group of entities is the audit logs. They are text files in which information about the user and his actions is collected. Audit logs include web server logs, database server logs, high-level CMF/CMS logs and special software and audit scripts developed by the resource owner.

- Web server logs are the text files that store the information of queries received by the server, including the IP address, the requested page, the status of the response, information about the referring page (referrer), the number of data transferred, user agent, date and time.

- Database server logs contain information about user requests to server databases. It is possible to analyse the codes of the queries themselves, the time of execution and the date of the query.

- High level CMF/CMS logs are site content management logs. CMS is a content management system that facilitates the management of a website, its support and the creation of standard websites. CMF - framework management system is a project written on the basis of Model-View-Controller (MVC). MVC is a scheme for independent management of server data by dividing it into three components: model, presentation and controller.

- Special software/audit scripts. These are software scripts connected to a web resource that allow collecting additional information about the user.

The fourth group of entities is behavioural models. These include such models as keyboard handwriting models, computer mouse usage models, TouchScreen models, text author identification models and characteristics of the hypertext markup used.

- Keyboard handwriting is a type of behavioural biometric characteristic. The parameters evaluated by this complex characteristic may include input speed, input dynamics, input error rate, stylistics of capital letters (e.g. the Shift key or CapsLock key is held down), etc.

- The models for using other manipulators (TouchScreen, Mouse) are similar to the previous model, but describe screen gestures or mouse uses. For example, which mouse buttons are pressed, how often the user moves the cursor.

- Models that identify the author of a text represent a semantic-syntactic analysis of the text and the definition of basic descriptive attributes for the author of the text.

- The characteristics of the hypertext markup used are the characteristics of the HTML page. For example, font styles, font size, order of text output on the screen, etc.

The last group of entities are third-party complexes. These include such popular complexes as DLP, SIEM and UBA/UEBA.

- DLP is the technology for preventing leaks of confidential information from the information system outside, as well as a technical device for such prevention.

- SIEM makes it possible to manage security information and security events.

- UBA/UEBA contains information about security systems that perform intelligent analysis of user behaviour.

## 3 Estimation of the time for calculation of the browser fingerprints

It is clear that the application of certain authentication factors [5-7], as well as the use of certain entities, must be determined adaptively based on the specific nature of the protected resource. Suppose that the information required for adaptive authentication of the user on the site should be quickly collected. It has been suggested that mechanisms for collecting browser fingerprints should be investigated as identifying attributes.

The collection of information characterising browser fingerprints can be done using software code written in the JavaScript language. A ready-to-use library has been developed in this language to obtain such attributes is called «fingerpring.js». It is popular and is regulary being updated in order to increase the accuracy of identifying parameters. The current version is Fingerprintjs 2 2.1.4.

The attributes that can be obtained from this library are shown in Figure 2.

The information on the selected attributes was implemented individually and the average time required to calculate each attribute was measured. Let us consider the parameters already mentioned such as: UserAgent, TimeZone, ScreenResolution, Canvas, WebGL.

The value of UserAgent is "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/85.0.4183.102 Safari/537.36". It is possible to extract information from the proposed line and to claim that the operating system is Windows 10.0 and the browser is Chrome.

Information about Timezone attributes is presented by two parameters: timezoneOffset and timezone: timezoneOffset = -180 and timezone = "Europe/Moscow". The example can be interpreted as follows: time zone - Europe/Moscow, where the time is 3 hours (180 minutes) ahead of the world UTC (UTC +3).

The value of the ScreenResolution parameter is an array [1920, 1080]. This means that the length of the screen is 1920 pixels and the width is 1080 pixels.

Canvas and WebGL identification strings were obtained. The identity string for Canvas looked like this: "canvas fp:..." where by "..." the encoded continuation of the string is meant. Matches with the canvas identifier were noticed in the WebGL identifier line, but the difference was that the line was supplemented by a WebGL component.


**Fig. 2.** Browser identification attributes defined using JavaScript in the Chrome browser.

WebGL parameters were also obtained individually, which were in an array of 65 elements (see Figure 3).

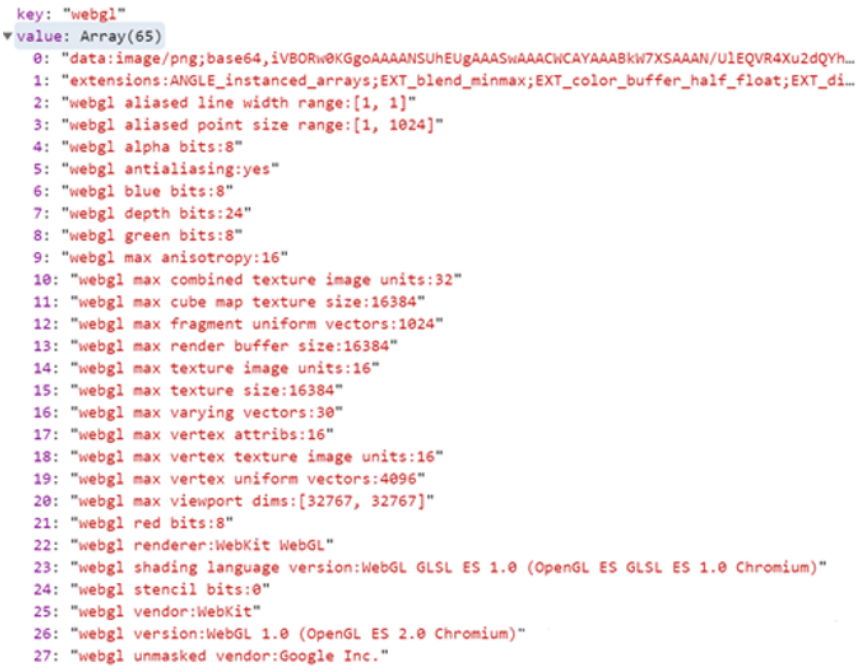


**Fig. 3.** Some WebGL parameters.

These parameters included WebGL exceptions (extensions), characteristics and language for shaders (e.g. webgl fragment shader medium int precision rangeMax: 30), vendor characteristics (webgl unmasked vendor: Google Inc. webgl vendor:WebKit), rendering

characteristics (rendering names - ANGLE (Intel(R) HD Graphics 630 Direct3D11 vs_5_0 ps_5_0 and maximum rendering buffer size 16384). Moreover, the bit characteristics of WebGL were obtained.

"webgl alpha bits:8".It is the number of bits defining alpha - the channel (transparency) for one pixel.

"webgl blue bits:8". It is the number of bits defining blue - channel (blue) for one pixel.

"webgl depth bits:24". It is the number of bits defining the colour depth for one pixel.

"webgl green bits:8". It is the number of bits defining green - channel (green) for one pixel.

"webgl red bits:8". It is the number of bits defining red - channel (red) for one pixel.

"webgl stencil bits:0" - number of bits defined for stencil buffer of the image (stencil buffer is used when creating various special effects, the most popular use is adding shadows to the image).

The results of the average computation time measurements of the above attributes individually and together are shown in Table 1. The values of the entities were rounded to tenths and the average time to determine these values was the mathematical average and also rounded to the first decimal point.

**Table 1.** Time table for calculating the main browser fingerprints.

| | UserAgent | TimeZone | **ScreenResolution** | Canvas | WebGL parameters | Main browser fingerprints |
|---|---|---|---|---|---|---|
| Time to calculate the indicator. ms | 0.7 | 0.8 | 0.7 | 10.8 | 42.2 | 51.4 |
| | 0.7 | 0.8 | 0.7 | 16.1 | 38.6 | 62.3 |
| | 1.8 | 1.5 | 1.3 | 19.7 | 41 | 48.5 |
| | 1.4 | 0.9 | 0.7 | 11 | 39.6 | 57.1 |
| | 0.7 | 0.9 | 0.7 | 14.1 | 41.3 | 47.9 |
| | 1.7 | 1.9 | 1.4 | 21.7 | 39.4 | 50 |
| | 0.7 | 0.9 | 0.7 | 11.1 | 40.3 | 47 |
| | 1.3 | 0.9 | 0.7 | 10.9 | 38.3 | 49.5 |
| | 0.7 | 0.9 | 0.7 | 10.3 | 41.6 | 48.5 |
| | 0.7 | 1.7 | 0.7 | 12.4 | 40.7 | 49.5 |
| | 0.7 | 0.9 | 0.7 | 9.6 | 38.7 | 50.6 |
| | 0.7 | 0.9 | 0.7 | 11.8 | 40.7 | 47.9 |
| Average time. ms | 1.0 | 1.1 | 0.8 | 13.3 | 40.2 | 50.9 |

Thus, the following conclusions can be drawn from Table 1:

1. The average time to detect the informative attribute "ScreenResolution", which is used for adaptive user authentication, is less than 0.001 seconds, which is a good indicator.
2. The average time to identify WebGL parameters required for user authentication is the most time-consuming among the main browser fingerprints and takes approximately 0.04 seconds, which is due to the fact that it contains a large amount of data and requires work to build an image processed using shaders and other WebGL functions.
3. The average time it takes to collect information about the main noted browser fingerprints characterising a user in a web space is approximately 0.05 seconds, which is effective.

It is possible to identify other informative browser fingerprints [8-9] than those shown in Figure 1, using fingerprints. Such fingerprints can include the following attributes: adBlock, deviceMemory, hardwareConcurrency, touchSupport.

- "adBlock" is the binary parameter that determines whether or not Adblock is enabled in the browser, which allows you to block ads and tracking and protects against malware.
- "deviceMemoru" is the parameter that describes the size of RAM on the device. In the task to be solved, the value of this parameter is 8, which means the size of RAM is 8 Gb.
- "hardwareConcurrency" means the maximum number of threads involved, which usually depends on the number of cores on the computer. In the case under study, this value is 4.
- "touchSupport" is the parameter that describes the characteristics for touchscreen. It is an array of three values: the maximum number of touch points on the screen, whether the base touch event (touchevent) has been successfully completed and whether the "ontouchstart" property is available. In the task under consideration, the array was as follows: [0, false, false], which meant that there was no touchscreen support at all.

The results of the average time measurements of these four informative attributes are shown in Table 2.

**Table 2.** Time table for calculating additional browser prints.

| | adBlock | deviceMemory | hardwareConcurrency | touchSupport |
|---|---|---|---|---|
| Time to calculate the indicator. ms | 0.9 | 0.7 | 0.7 | 0.9 |
| | 0.9 | 0.7 | 0.7 | 0.9 |
| | 0.9 | 0.7 | 0.7 | 0.9 |
| | 0.9 | 0.7 | 0.7 | 0.9 |
| | 0.9 | 0.7 | 0.7 | 0.9 |
| | 0.9 | 0.7 | 0.7 | 0.9 |
| | 0.9 | 0.7 | 0.7 | 1 |
| | 0.9 | 0.7 | 0.7 | 0.9 |
| | 0.9 | 0.7 | 0.7 | 0.8 |
| | 0.8 | 0.7 | 0.8 | 0.9 |
| | 0.9 | 0.7 | 0.8 | 0.9 |
| | 1.0 | 0.7 | 0.7 | 0.8 |
| Average time. ms | 0.9 | 0.7 | 0.7 | 0.9 |

The average time to identify additional browser fingerprints that can be used for adaptive user authentication is less than 0.001 seconds, which is insignificant compared to attributes such as WebGL or Canvas parameters.

One of the results of the study of informative attributes and average time estimates for calculating the value of a feature is that for the selected browser fingerprints there is an inverse relationship between the amount of information on the feature and the time it takes to obtain the value of the feature. This is schematically illustrated in Figure 4.

Based on the basic and additional browser prints examined, it can be concluded that in most cases it does not take a long time (about 1 millisecond) to obtain informative attributes of this group. Moreover, the entity group contains characteristics such as Canvas and WebGL string identifiers, which, although time-consuming to calculate, are unique in their length and character set and can be informative with sufficient accuracy, in adaptive user authentication.

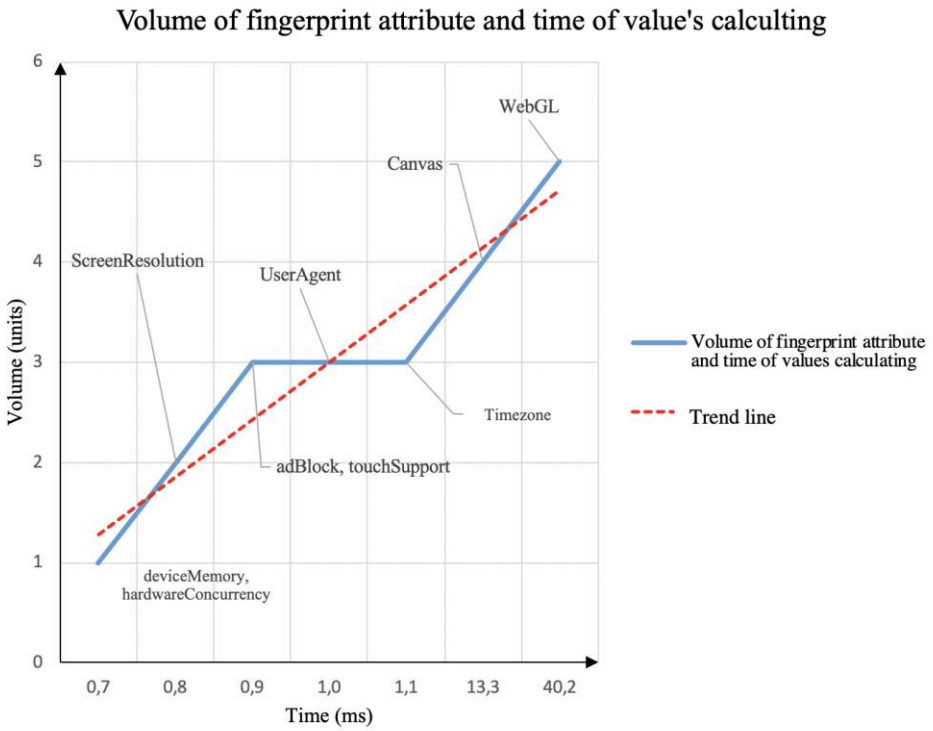


**Fig. 4.** Dependence of the volume of information in the browser fingerprint on the time required to determine the value of the attribute.

## 4 Conclusion

In order to realise the possibility of rapid detection of information security incidents in web resources, an analysis of the essence of both individual and complex, united groups was carried out. Special attention was paid to such a group of entities as browser fingerprints. An approach was proposed that allows for the pre-screening of informative browser prints and the calculation of the average time to obtain them using a specific example. Knowledge about the time of obtaining the attribute gives the possibility to choose the most effective attribute among the considered ones.

## Acknowledgements

This work was partially funded by Russian Federation President Grant for the young scientists [MK-2421.2020.9].

## References

1. Iskhakov A Y and Meshcheryakov R V 2017 *J. Phys.: Conf. Ser.* **803** 012056
2. Iskhakov A Y, Iskhakova A O, Meshcheryakov R V, Bendraou R and Melekhova O 2018 *SPIIRAS Proceedings* **6(61)** 147
3. Hu S, Bai J, Liu H, Wang C and Wang B 2017 *10th International Symposium on Computational Intelligence and Design (ISCID)* 482

4. Pilankar P S, Padiya P *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)* 1947
5. Chen X, Xu F, Xu R, Yiu S M and Shi J 2014 *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* 121
6. Kaminsky R, Enev M and Andersen E 2008 *University of Washington. Technical Report*
7. Feher C, Elovici Y, Moskovitch R, Rokach L and Schclar A 2012 *Information Sciences* **201** 19
8. Lewis A, Li Y and Xie M 2016 2016 *IEEE Conference on Communications and Network Security (CNS)* 380
9. Wu Q, Zeng Y, Lin Z, Wang X and Yan B 2017 2017 *8th International IEEE/EMBS Conference on Neural Engineering (NER)* 564