Chapter

User Authentication Based on Knowledge of Their Work on the Internet

Pavel B. Khorev

Abstract

This chapter analyzes existing user authentication methods for remote access to information systems and disadvantages of these methods. The method of multifactor authentication of users when they are accessing remote information systems, combining validation of knowledge on secret password and verification of conformity of the habits and preferences of Internet user's interests, is defined by registration in the system. Using the history of Web pages, the Internet browser creates a list of Web pages the user has visited in the past period of time. It is proposed to use the Bayesian classification for user's knowledge based on the analysis of information about Web pages visited by the user. For user authorization from someone else's computer, the user is invited to ask for additional questions to test knowledge of subject areas, which they selected during registration in the information system. This chapter defines the language and tools for implementation of the proposed authentication algorithm: the programming language PHP and the MySQL database management system (to create a database of registered users), Web-based open source application phpMyAdmin (to create and administer MySQL database management system), and the JavaScript programming language and HTML (for creating extensions for browsers receiving a list of the addresses of the Web pages visited by the user).

Keywords: user authentication, remote access, the document object model, classification of text documents, Bayesian method, PHP programming language, MySQL database management system

1. Introduction

Many Internet sites and portals (including educational institutions) should limit access to their content (commercial secrecy, personal data, intellectual property, and other sensitive information) for unauthorized users. For example, universities of distance education must provide reliable authentication of students in carrying out evaluation tasks. Financial institutions (banks) should provide access to customer accounts only after credible evidence of their authenticity.

A security user login procedure largely determines the security of an information system as a whole (and in the case of distance learning systems and the reliability of the results of implementation of the students of educational tasks). Authenticating the name of the logged in user is one of the steps in the logon process.

To authenticate users of information systems, there are well-known techniques. The first group of such methods is based on checking the knowledge of the user based on some memorized secrets (e.g., secret combination password). In the authentication process, knowledge of this secret is validated. The second group of authentication methods is based on checking the user ownership of certain hardware which becomes the subject (device), such as a smart card or USB token. The device contains the base secret of the user (e.g., its private key digital signature), which does not need to remember. Third party authentication methods are based on checking whether a user has characteristics that could not be separated from it. These characteristics, for example, can be printing finger, face, or voice. Such methods are related to biometric authentications.

But all these methods are as effective as possible in local authentication, when there is no doubt about the source of information for user authentication. If remote authentication has no such confidence, because data for verification can be provided to outsider, you want to create new authentication methods, suitable for use in remote access to information systems. These methods can be used in addition to the existing methods of authentication. Additional authentication methods, for example, are to verify a user based on the knowledge test about his preferences and competencies.

Data on user knowledge can be collected using the analysis of the content the user has visited Internet resources. Such an analysis could be based on the methods of classification of text documents.

2. Analysis of existing methods

Traditional methods of authentication (verification of secret knowledge-based reusable password and verification of biometric characteristics and devices) have a common disadvantage: the ability to intercept confirming the authenticity of the user information with its subsequent playback of the infringer to perform any action against a target system on behalf of a registered user. To improve protection against unauthorized access to sensitive information, it is usually recommended to use the two or multifactor authentication.

Consider the drawbacks of traditional methods of authenticating users of information systems in the case of remote access systems.

2.1 User authentication based on validation of secret knowledge

The main advantage of validation of knowledge-based authentication of user secret reusable password is the ease of its implementation and use. At the same time, the password authentication has many drawbacks:

- many users choose passwords that are easy enough to pick up due to lack of password length, their simplicity, and repetitiveness;
- the possibility of using the violator of readily available software tools for picking passwords;
- the ability to use social engineering techniques by the infringer (obtain the password by tricking the user); and
- the ability to “steal” the password as you type with the keyboard or intercept the password when it is sent over a computer network.

2.2 User authentication using devices

User name authentication using authentication devices is based on the uniqueness and the confidentiality of the information contained in the memory of the device. As such, information, for example, the private (secret) key of the user's electronic signature could be used. In the process of authenticating, the correctness of such key is validated using the user's public key certificate issued by a trusted certificate authority and is stored in an information system in which the user registered.

Most often the following devices are used for authentication:

- tokens that require connecting to your computer using a USB port and constituting in fact microcomputer;
- smart cards also constituting a microcomputer, but additionally requiring the use of card readers;
- passive devices (e.g., iButton or Touch Memory), which can only store information to authenticate the device owner.

For active USB devices, added protection from theft applies reusable passwords (so-called PIN-codes), the knowledge of which confirms the use of authentication devices to its rightful owner. Other advantages of authentication devices are no limits in the length and complexity of storage in the device memory and the ability to detect the fact that the device is lost or stolen and lock it in this case.

Authentication procedure using active devices may include the generation and verification of one-time passwords or occasional request response calculation (model "handshake").

But the use of authentication devices also has a number of disadvantages:

- the possibility of device failure or accidental damage;
- the additional cost issued by the registered users of the devices and their readers;
- the need for a free USB-port or additional equipment to connect your device to your computer;
- the possibility to manufacture copies of analog devices or wrongdoing or creating his software emulator; and
- the need to deploy a public key infrastructure (PKI) when using the private key of the user as electronic signature stored on his secret device [1].

2.3 Biometric user authentication

The biometric authentication checks that the user is unique and inseparable from his personality characteristics shared by physical or static (patterns of papillary lines or fingerprints, hand shape, iris and the retina of the eyes, face shape, etc.) and behavioral or dynamic (timbre, handwritten signature, tempo text input with the keyboard or keyboard "handwriting," etc.).

The advantages of biometric authentication refers the validity of authentication, user friendliness (it does not need to remember long and complex passwords or

permanently carry the device authentication), and the complexity of the falsification of biometric characteristics of the offender.

The disadvantages of biometric authentication are:

- the additional cost of the equipment to read the biometric characteristics;
- storage standards of biometric characteristics in plaintext, resulting in risk of violation of the privacy of the user;
- the possibility of failure to a registered user due to an accidental large deviation of his scanned characteristic from the reference value;
- the possibility of interception of biometric characteristics when it is sent over the network.

For protection against interception of biometric characteristics and its subsequent reproduction and when the violator tries to log on to the system on behalf of others, cryptographic methods and tools can be applied. However, the use of encryption when data are transferred across the network assumes the task of managing the encryption keys. The use of a digital signature to confirm the source of biometric data requires the solution of the problem of public key certificate management devices to read such data. These causes reduced the effectiveness of the use of biometric authentication for remote user access.

Biometric authentication in Russia, now, has been started to be used to authenticate clients during their remote access to their accounts [2]. In this case, users must first register with the bank on the list, which is set by the Central Bank of Russia. To authenticate the user, the following actions are then performed:

1. entering their login and password set during registration;
2. photographing their face using camera notebook or other devices (e.g., tablet, smartphone, etc.); and
3. using a microphone, the computer speaks the text received from the authentication server and displays on the screen.

This method of authentication refers to multifactor authentication. It combines checking the knowledge secret password and authentication based on static (face) and dynamic (voice) biometric characteristics. The use of this method does not impose additional requirements to the equipment of users' computers. Specific technical solutions to this project refer to the trade secrets of its developer and financial organizations. Therefore, the effectiveness of addressing the shortcomings of biometric authentication, mentioned above, is difficult to assess.

For this reason, use a similar solution for remote user authentication information systems (e.g., universities of distance education) which appears to be unfounded so far.

2.4 The use of traditional methods of authentication for remote user access

Overall lack of traditional authentication methods for remote user access is the lack of reliable evidence of the source of data for authentication. These data can be reproduced after their "sniffing." One solution to this problem might be to establish a secure connection between a client and a server using SSL/TLS. Such a decision requires the establishment of a public key infrastructure (PKI) and certificate

management [1]. This places additional requirements on the information systems and their owners. Such requirements may be redundant for distance education universities and other organizations with limited budgets.

Another possible solution would be to use the USB device of the remote users to generate one-time passwords for authentication procedure. One-time password intercepts the violator as it will not give the possibility of unauthorized access to information system resources. The use of this decision will complicate the administration of information system and will require additional expenses. Therefore, such a decision is also uncomfortable for distance education universities.

Authentication based on user testing of knowledge collected during his work on the Internet is free from these deficiencies.

3. User authentication method based on knowledge

Knowledge-based authentication often is used as a second authentication factor when using a password or user password recovery in case of loss. In this authentication scheme, the user is prompted to answer at least one additional "secret" question. But this simple schema is not free from following flaws:

- the user can forget his answer to the question;
- user response can be guessed by the infringer;
- the number of supplementary questions may not be very large; and
- answers to additional questions contain only a very small part of the knowledge of the user.

So an authentication method should be developed, which involves the collection of sufficient information. The information collected should be unique for each user, registered in the information system. It is also advisable to use the developed method that does not require any additional user action.

When designing a remote authentication method based on the knowledge of the user on the Internet, you must ensure the collection, accumulation, and use of information about the habits and preferences of the user global network. Analyzing the data of interest, habits, and preferences of Internet user may apply the analysis log of visited Web pages by this user. Among the functions of Web browsers is a function of preserving the history of visited sites and portals by the user in the appropriate journal. This function does not need to include especially constant collection of data on Internet user has visited, in the visit log saved addresses and titles of visited Web pages, as well as the date and time when they were.

Getting the user browsing history of Internet resources (scanned documents) is possible through the development of special extensions (Add-ons) for Internet Explorer and other browsers [3, 4]. However, browser manufacturers can set restrictions on the use of extensions; for example, Google Chrome, which allows installing extensions only from the shop, Chrome Web Store. Enable developer mode gives you the ability to install extensions from an arbitrary location (e.g., from the selected developer folder).

Using a known document object model (DOM) [5], it is possible to present the contents of the document (e.g., a user visited the Web page) in the form of a set of objects with certain properties. Support for this model is obligatory for all Web browsers.

In the DOM, document is presented in a tree structure. It provides a unified way to navigate through the document. This tree structure is called a node tree. Access to all the nodes can be accessed through this tree.

Using the document object model in the analysis of any user visited the Web page allows you to retrieve the value of the properties, which contains the keywords, description of the document, its title, and a list of all its internal headers list captions to the pictures (if they are available in the document). Obtaining these data provides an opportunity to analyze the document and determine the:

- set of key words;
- the number of occurrences of each of these keywords (phrases) into a document; and
- the position of the occurrences of keywords in a document.

Additionally, the results of the analysis provide an opportunity to offer the user a list of keywords (phrases) that best characterizes his interests.

For automatic document classification, visited by the user during his work on the Internet (its inclusion in one or more thematic rubrics), further analysis of the content of the document is required. You can use the following methods of classifying [6]:

- **Method of support-vector machines (SVM):** this method solves the problem by constructing a nonlinear plane separating the decision. Due to the peculiarities of nature space signs, the border decision method of supporting vectors was built, which has a high degree of flexibility in solving problems of classification of various levels of complexity.
- **K-nearest neighbors method (K-NN):** the method is based on memory usage and, unlike other statistical methods, does not require prior training, designed for classification. This method provides high efficiency, but demanding to computing resources in the stage classification.
- **Bayesian method:** this method is based on the theorem stating that if the densities of the distribution of each of the classes are known, then the search algorithm can be written in an explicit analytic form. This algorithm is optimal and has minimal error probability. In practice, the distributions of classes typically are not known. They have to be assessed (restore) on training samples. As a result, Bayesian algorithm ceases to be optimal; so as to restore the sample density is possible only with some margin of error.
- **Decision tree method:** decision tree-based classifier for category is a tree whose nodes are the terms; each edge is a labeled condition, and leaves are marked. In practice, use the binary decision trees, in which the decision of moving on the ribs is done with a simple check for terms in the document.
- **Method of neural networks:** artificial neural network is a collection of interconnected neurons. Each neuron is an elemental converter input signals at output signals. Passing on a specific set of network input signals, we get a certain set of signals to the output. A text categorizer based on neural networks is a network of elements which forms input elements presented by the terms of

the document output items submitted by categories, and links between the elements that define a dependency relationship and are marked with weights.

For remote user authentication, documents are classified and analyzed to identify those subject areas that are of interest to the user. To store the collected information, the database (DB) is used. The database will then be used for the remote user authentication.

Bayesian method is used as the proposed method of authentication. This method of classification is based on the theorem stating that if the densities of the distribution of each of the classes are known, then the classification algorithm with the minimal probability of errors can be specified explicitly. In practice, the distribution density classes not known. These probabilities has to evaluate (restore) on training sample.

Bayesian method is used when solving different tasks of information security: when spam is detected in an e-mail message, when evaluating the security of information systems, and others.

In our case, the classified document is rich in properties whose order is not important. The submission of the document was obtained by its previous analysis.

The advantages of Bayesian classification method include:

- this method allows the relatively quick classification of Web pages that must be specified when the user is authenticated;

- this method is characterized by the ease of programming;

A database containing information on behalf of the user, whose authenticity is confirmed, includes the following tables [7]:

- In the table "users," set attributes (columns) as id of the user, his name ("login"), the hash value of the password, the e-mail address of the user, sign mandatory password change at next logon, and date and time of the last logon user. It contains information about the users registered in the information system.

- In the table "interests," set the id attributes of interest and its name. It contains information on those subject areas that represent the interests of the user.

- In the table "users_interests," set columns as the id of connection user and interest, user id, id of interest. Information from this table links a user and his interests, identified by the analysis log of visited Web pages of the user's Internet browser.

- In the table "keywords," set the columns id keyword, keyword. Keywords are stored here, and they let you associate a document with a specific subject area.

- In the table "questions," set the columns id question, id of interest, which includes the question, the text of the question, and the user's response. The questions stored here will be asked to the user for authentication when it is not possible to analyze the history of the Web pages they have visited.

Relational database model consisting of the specified tables is presented in **Figure 1**.

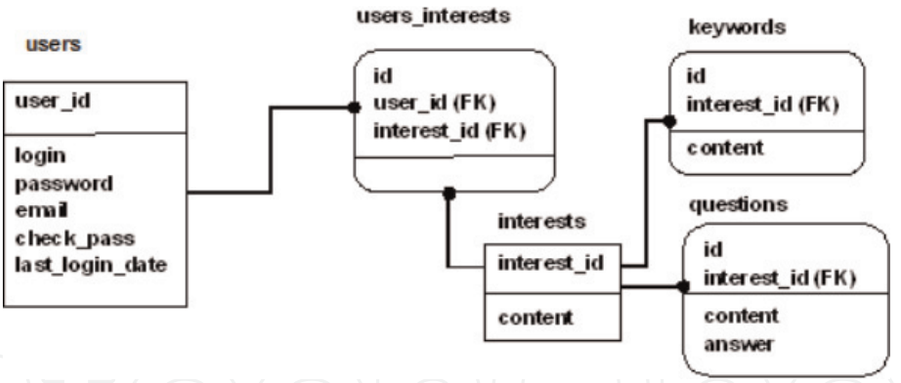

Figure 1.
The relational model DB.

When registering, the user specifies his username and password. Hidden to the user happens an analysis of visited Web pages using browser extensions. Further defines the user's interests. All data received are stored in the database. List of interests will be stored in the table "interests," associated with the "users" table—list of users of the system. Each user has an individual set of interests, so the "users" table one column will be "interest_id," which will store a list of interests of each user. The table "interests" will need at least three columns: "id" (number of entries in the table), "interest_id" (the number of the record interest in the table), and "content" (the name of interest).

In order to verify the conformity of the contents of your browser's browsing history, interests of user authentication need to be somehow mapped. Each html page can have a set of keywords, description, and header (title). After receiving a list of interests on the basis of the last visited URLs, the received data must be compared with user data stored in the database of the interests that have been entered into it after registration.

4. User authentication algorithm based on checking his knowledge

If the user tries log in into the information system from his device, it can authenticate using the following algorithm. In this algorithm, the user's browser history and his interests are identified (specified) and compared with the data received when you register a user in the system (see **Figure 2**):

1. Get a list of URLS of the Web pages contained in the log visits.

2. Retrieve tags for each Web pages from the list (its title, list of keywords, description).

3. Analysis of the information obtained to determine the user's interests.

4. Retrieving information about the interests of the user from the database that was created when its registration in the system.

5. Comparison of two sets of interests for checking user knowledge.

This method retrieves the last 1000 entries from the user's browsing history over the past 30 days. if the number of records for this period is smaller than 1000 analyses of all the log entries for the specified period of time.

Each time a remote user is authorized, the login and password are checked, as well as data analysis on its work on the Internet. Unbeknown to the user, the list of

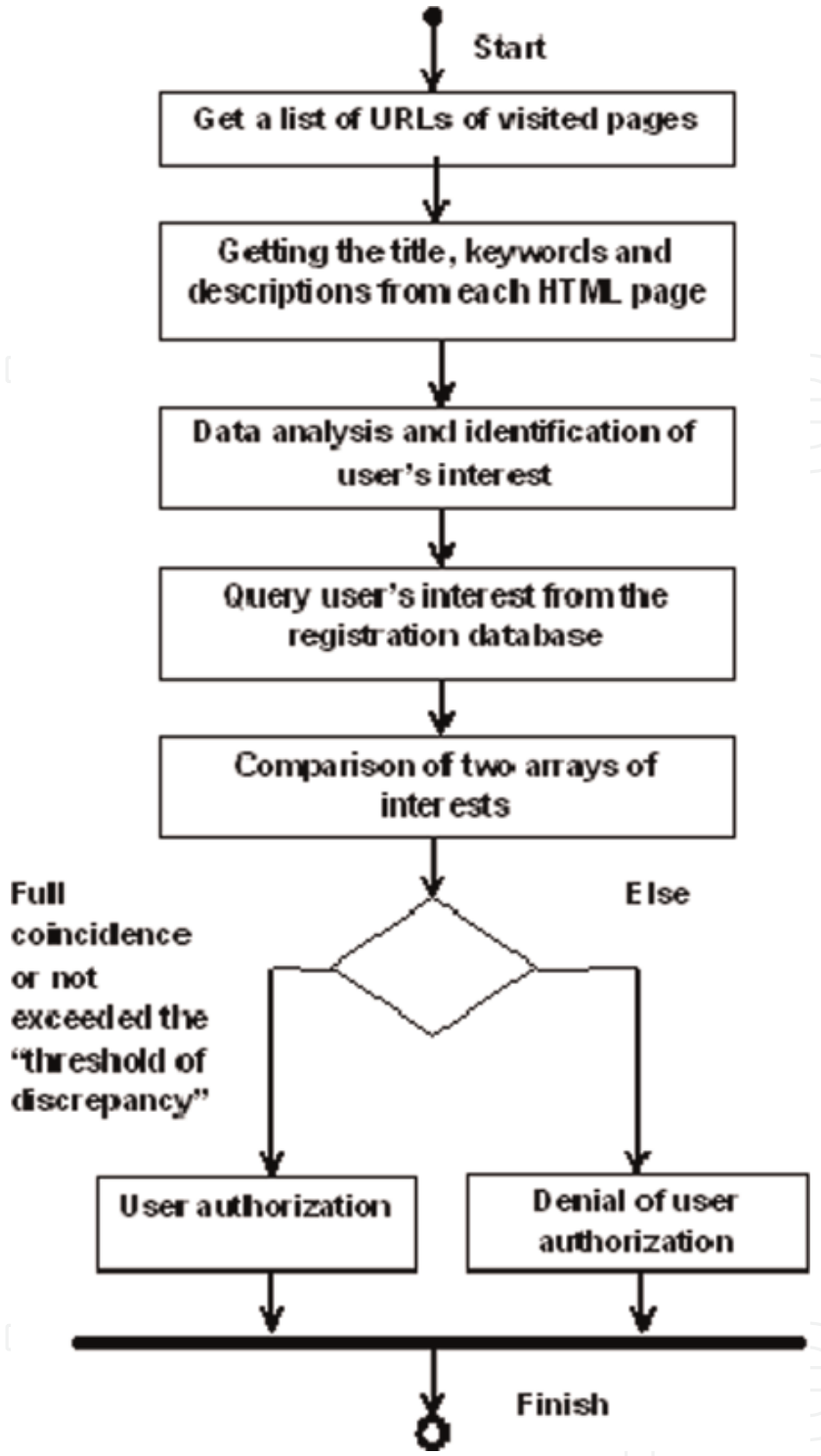

Figure 2.
The algorithm for checking the conformity of the contents of the history of the user's browser to its interests.

recent addresses of visited Web pages using Internet Explorer extensions is accessed. Further defines the user interests based on the information about the documents in history. Then the interests of the user who is authenticated are compared with those that are stored in the database. Authorization will be considered successful based on the two conditions:

- if "login" and the hash value of the password match; and
- if the difference between the interests of the user defined when authentication and retrieved from the database does not exceed the so-called "threshold of discrepancy" [8].

Let us say that from the moment of registration of the user prior to its authorization in the system, the user has actively worked on the Internet and visited the new Web pages that are not reflected in the browser's history of the Internet when registering. Then, the set of the interests of the user defined with its authorization may not match the set, which is stored in the database. Hence the use of "threshold of discrepancy" sets the maximum allowable difference between the two sets of interests. This threshold can be set by the administrator of the information system, where registering the user.

Let us say that when authorizing a user, the "inconsistency threshold" is not exceeded, and the set of certain user interests is less than his set of interests from the database. In this case, during the authorization process, the user will have to answer the questions of those subject areas that are not in the set, a specific authorization. The user is given a limited time to each response. One interest from DB corresponds to one question, and the user's incorrect answers are fixed.

Next, the user will be asked new questions of those substantive areas, the questions of which he gave incorrect answers. The maximum number of issues relevant to each interest is too limited. If you then remain relevant to the interests of user domains, the questions of which he was unable to give the correct answers, then the user authorization will not be available.

If in a set of interests that are stored in the database for each registered user, no interests, which were identified as a result of his successful authorization, these new interests are added to the database.

Application of the developed method of authentication will increase the validity of this procedure when providing remote user access to information system resources. This will reduce the potential damage from thefts of valuable information. For universities of distance education, possible loss may be associated with damage to the business reputation of the extradition documents on education for student evaluations that were falsified.

5. Methods and means of implementation

To create DB registered users, apply the programming language PHP and the database management system (DBMS) MySQL as well as Web-based open source application phpMyAdmin, designed to create and administer MySQL DBMS. phpMyAdmin allows you to administer a MySQL server, which can execute SQL-queries and view the contents of database tables.

Using phpMyAdmin, create a new database and add 5 new tables: "users," "interests," "users_interests," "keywords," and "questions."

Web browser extensions (such as Google Chrome and Mozilla Firefox) can be created using programming language (such as JavaScript) and hypertext markup language (HTML). This expansion will be used when authenticating for getting address list of Web pages viewed by the user.

Let us look at how to create extensions, for example, for Google Chrome browser. The file was originally created with the obligatory ".JSON" manifest, which contains information about the extension: extension name, version, description, version, and the location of the manifest icon in the browser address bar approx.

Example manifest file:

```
{
 "name": "Typed URL History", //the name of the extension
 "version": "1.2", //version of the extension
```

```
  "description": "Reads your history, and shows the top thousand pages you go to by
typing the URL.", //description
  "permissions": [
        "history",
        "tabs"
        ],
  "browser_action": {// the extension will have an icon next to your address bar
  "default_popup": "typedUrls.html", //the title of the html page that will be
              //displayed when clicking on the icon extension
  "default_icon": "url.png" //the name of the image that will be used as the icon
      },
  "manifest_version": 2 //manifest version
}
```

After you create a manifest, they are created with HTML and JS-files: "typedUrls.html" (HTML page that describes the type of window that is displayed after clicking on the icon extension) and "typedUrls.js" (the file that implements the collection of information about the user's browser log). For the implementation of the algorithm, the following functions were created:

- function showURLs(historyItems) (gathers a list of URLs from the user's browser history); and
- function showHistory() (displays the collected history pages for a specified period of time).

To invoke the necessary functions, event handler "addEventListener" is used.

Creating such extensions when using the proposed method of user authentication allows you to automate the process of analyzing your browser history at the time of registration and authorization of the user. Users will not be required to enter any additional information for its authentication (it introduces only the "login" and password). The extension generates a list of Web page addresses. This list is passed on to the authorization service. Then this list is parsed to determine the set of user's interests (using Bayesian method) and decision on user authentication or deny his access to the system.

When implementing user authentication algorithm, two Web pages are created:

- a page with a form for data input by the user's authorization; and
- a page with a form for user registration.

The master page is considered to be an authorization form. Here you can log in if already registered, or register by clicking on a hyperlink.

On the logon page, the user is allowed to go through the procedure of authorization. If the user has not yet logged in, you can go to the registration page. On this page, the user specifies the user name ("login"), as well as an e-mail address, which will be sent with a random initial password, that will be created by the service registration. If the user has entered valid data that satisfy the conditions (the login name should be between 5 and 15 characters, containing only letters of Latin alphabet, digits, and the characters ' _ ' and '-', and e-mail address must be valid and cannot be used twice), then the user will be registered.

When a new user is authorized for the first time, it will need to change the initial password. Without changing the initial password, the user is not authorized and will be accessible only to change password page.

If a user has forgotten his or her password, he or she may recover it by using the function "forgot password?"

To exclude threats of kidnapping registered users passwords directly from a database on a Web server, passwords should be stored in a database in a hashed form. To do this in PHP, there are special functions, e.g., md5 (MD5 hashing algorithm that produces a hash value with a length of 128 bits) [9]. This function returns the result string with hexadecimal hash value.

It is possible to crack a user's password by using a special dictionary. To protect you from this attack, the password is hashed together with a random number (salt). This salt can be calculated using the PHP function uniqid. This function uses the system timer and pseudorandom number generator for maximum uniqueness and unpredictability of salt.

Impurity is added to the password when it uses concatenation operation (.) and stored in an additional field "uniqid" database table of registered users.

When registering the user list of URLs stored in an array, the Next array analysis and Bayesian classification occur in subject areas that you are interested. After their definitions, user interests are recorded in the database as follows:

1. Access the table "interest" to obtain a unique number (the "interests_id") of each user's interests.

2. Next all the unique rooms of interest are assigned to a user in table "users_interests."

3. In the field "user_id," a unique number of the user; and in the field "interest_id," a unique room of interest are recorded.

4. Thus, in table "users_interests," rows as many as the user's interests will be stored exactly, as it was determined.

When you try to log in, user input verification occurs with those that are stored in the database (the password is hashed first, and then compared with the one stored in the database password). Username and password must match exactly with those that are stored in the database.

When authorizing, a user browser extension should get a list of thousands of URLs, which he attended in the last 30 days. If their number is less than 1000, the extension will keep all the available URLs for these 30 days. Next to each URL is determined by its area of expertise (there may be several).

After categorization, the entire list of URLs of interest of the user is compared with its interests in the DB. If the difference exceeds the threshold of discrepancy, the authorization will be refused.

If the user is authenticated from someone else's computer, to authenticate it gets a list of interests. In this list, user must select the interests of the subject areas that have been identified during registration. Then the user specifies additional questions—one for each subject area (as described above).

If the remote user session duration exceeds the maximum possible period, to continue the work he would have to pass reauthorization. After a specified period of time to a user, that is, on any Web pages, page opens instead of "authorization." Such a modification is introduced in order to enhance the security of user in the system, because while you are out of the workplace, an attacker could gain access to confidential data, posing as the owner of the account records.

6. Conclusions

The principles of authentication of users based on their knowledge of their work on the Internet are identified, as well as analyzed by means of collecting such knowledge. The methods to gather and compile information about users of the Internet are analyzed, including browser history log and the DOM of an html page. The methods for solving classification tasks in relation to the interests of Internet users are also analyzed. Their advantages and disadvantages are revealed. In order to accomplish the above objective, Bayesian method was selected.

Also, authentication algorithms are developed and implemented for:

- checking the conformity of the contents of the user's browser history and its previously defined interests; and

- calculating the level of "inconsistency" and the decision to authorize a user.

Extensions for browsers such as Google Chrome and Mozilla Firefox, allowing receiving log information browser visits within a specified time period are developed as well.

Thus, the work examines the shortcomings of existing methods of authentication when accessing remote information system. The method of multi-factor user authentication does not require the user to commit additional action during authorization. This method increases the reliability of the user's authorization results compared to the password authentication.

Compared to the use of the device-based authentication method, this method does not require extra costs and does not complicate the administration of information systems due to the need for programming and accounting for issuance of authentication devices.

The application of the method described does not require creating a cryptographically secured connection between a remote user and server information system. Setting a connection involves the creation of a public key infrastructure that also complicates the administration of information system.

Application of the developed method of authentication increases the security of your information systems without the need to increase the cost of its administration. This is especially important for organizations with limited budgets, which include distance education universities.

Acknowledgements

The author expresses sincere gratitude for student E.V. Mazaeva, for making software implementation of the proposed method.

Author details

Pavel B. Khorev
National Research University "Moscow Power Engineering Institute", Moscow, Russia

*Address all correspondence to: pbkh@yandex.ru

References

[1] Smith RE. Authentication: From Passwords to Public Keys. Boston: Addison-Wesley Publishing Company; 2002. 576p. ISBN-13: 978-0201615999. ISBN-10: 0201615991

[2] Remote Identification [Internet]. 2018. Available from: https://www.cbr.ru/fintech/remote_authentication/ [Accessed: 29 March 2019]

[3] How to Add Add-ons in Internet Explorer [Internet]. 2018. Available from: https://www.wikihow.com/Add-Addons-in-Internet-Explorer [Accessed: 29 March 2019]

[4] Create Your Own Browser Extensions, Part 1. Extend Your Reach Into Chrome [Internet]. 2013. Available from: https://www.ibm.com/developerworks/library/os-extendchrome/ [Accessed: 29 March 2019]

[5] Document Object Model (DOM) [Internet]. 2009. Available from: http://www.w3.org/DOM/ [Accessed: 29 March 2019]

[6] Kuznetsov RF. Web Page Classifier Based on SVM-Multiclass [Internet]. 2006. Available from: http://romip.ru/romip2006/10_kuznecov.pdf [Accessed: 29 March 2019]

[7] Khorev PB. Authenticate users with their work on the Internet. In: 2018 IV International Conference on Information Technologies in Engineering Education (Inforino); Moscow, Russia; 2018. pp. 1-4

[8] Mazaeva EV. Method of two-factor authentication of users based on knowledge of their work in the Internet. In: Science and Education: Materials of the XII International Research and Practice Conference; Munich; November 2–3, 2016. Munich, Germany: Vela Verlag Waldkraiburg; 2016. pp. 67-69

[9] The MD5 Message-Digest Algorithm [Internet]. 1992. Available from: https://www.ietf.org/rfc/rfc1321.txt [Accessed: 29 March 2019]