Specific Problems of User Identification Methods Implementation in Web-applications

Oldřich Horák

Abstract—This article describes problems of user identification methods implementation in web-applications. The implementation costs are accentuated as the condition of these methods comparison. At the beginning, some specific issues of user identification are discussed due to features of front-end applications and application protocols used in background connection. The identification method implementation costs are meant as a part of the TCO, and the term of “Addition Implementation Costs” is defined closely. Some tips and techniques are accentuated to use for undesirable multi-user issue detection and avoidance in the relation to the user policy and conditions of the given application. The usability is discussed across common web-application in comparison to special application such as public administration systems or geographic information systems.

Keywords—GeoWeb, Identification methods, Implementation costs, Information system, User identification, Web-application

I. INTRODUCTION

The usability of information systems depends on many features. One of these features is a user-friendly environment with a simple user interface. In present, modern systems use thin clients realized as web page or web-site presented via browser, and the robust system backend running on powerful server hardware. This configuration forms the “web-application” as a common information system design.

Most of the identification systems need a basic user access control subsystem at least. The access control subsystem can be designed for different security levels. It depends on the target application purpose and security measurements suitable for the owner of the information system.

The information system designed as the web-application deals with specific environment of the worldwide web rules and conditions. In general, the web is unsecured. Many issues and threats can encounter the information system in the form of the web-application. Designers are responsible for sufficient solution of such situations.

Some systems providing the user interface in the form of web-applications deal with the undesirable multi-user issue. Social networks or so-called browser games are sensitive to this problem. The only one user account per person is expected to use in these systems, and the ability to detect such rule violation is very important. No certain way to solve this task is used in general, but features of some identification methods are usable to support such solution.

Each one solution depends on the security demands of the information provided by the system. Some systems need to distinguish the separate identities of their users, the user’s accounts uniqueness is important for others. In general, each one needs to assign an access rights to provided information by the user or user-group identity. In the web environment, the process of user identification is more difficult.

II. WEB-APPLICATION ARCHITECTURE

A web-application consists of two basic part overall. As a network application it has a server part and one or more client parts (client/server architecture). “If we look at the client-server model in detail, we see that two processes are involved, one on the client machine and one on the server machine. [1]”

The server part is commonly found as a running service of a dedicated network server. The client part can be found on workstations as an application used to browse through the worldwide web service (web-browser), see Fig. 1.

Some types of web-application are based on frameworks providing component integration. Such components can be distributed and/or heterogeneous. Their services often act as a separate network application with their own client-to-server communication. [2]
Client-to-server connection is realized by unsecured (HTTP) or secured (HTTPS) protocol going over transport protocol (TCP). The common data format transmitted between server and client part is hypertext. Server part also provides other file types transmission including graphics and multimedia.

Other file type transports from client part to server is unhandy and only way to realize it is the interactive upload service provided by form HTML element. By this reason, any file transport from client to server is problematic, and it requires user’s effort. [3]

The capabilities of modern web-design techniques can be used for better presentation of the web content (i.e. Web 2.0 [4], etc.), but the client-to-server data transmission still remain limited.

Another way to transport more data from client to server is provided by integrated components in the form of embedded objects. But, such solutions lead to more security problems based on bad knowledge of the 3rd-party components internal design. [2]

A special purpose is the capability of disclosing false user accounts created by robots or viruses. The importance of this task increases nowadays. Such accounts are often used to illegal activities, i.e. social engineering.

Children and adolescents use the social networks most frequently, and the communication using Internet is their nature [5, 6]. Members of this social group are most endangered by some illegal activities in the Internet. The false identity created in the social network environment could be misused to some criminal acts. “For example, there are many cases where people take advantage of these social media to involve in activities that lead to crime cases. [7]” The right user identification can prevent this kind of acts.

Three groups of identification method exist. First, user can be identified by some unique knowledge. Typical case is using of a password. Next group includes methods based on ownership of some unique item, typically a token or a thing. The methods of the third group are based on some specific feature (mainly body feature) of given user as some unique biometrics parameters.

A. Knowledge Oriented Methods

These methods are based on user knowledge of given fact, i.e. username and the corresponding password. For user purposes are commonly used. The first one is used for monitoring of individual user’s activity, usually without user’s knowledge about the monitoring. Other one is used for authentication. In this case user’s cooperation with the system can be necessary, and user’s identification can be cognizant.

The usage of user identification is various, but two main purposes are commonly used. The first one is used for monitoring of individual user’s activity, usually without user’s knowledge about the monitoring. Other one is used for authentication. In this case user’s cooperation with the system can be necessary, and user’s identification can be cognizant.

A special purpose is the capability of disclosing false user accounts created by robots or viruses. The importance of this task increases nowadays. Such accounts are often used to illegal activities, i.e. social engineering.

Children and adolescents use the social networks most frequently, and the communication using Internet is their nature [5, 6]. Members of this social group are most endangered by some illegal activities in the Internet. The false identity created in the social network environment could be misused to some criminal acts. “For example, there are many cases where people take advantage of these social media to involve in activities that lead to crime cases. [7]” The right user identification can prevent this kind of acts.

Three groups of identification method exist. First, user can be identified by some unique knowledge. Typical case is using of a password. Next group includes methods based on ownership of some unique item, typically a token or a thing. The methods of the third group are based on some specific feature (mainly body feature) of given user as some unique biometrics parameters.

A. Knowledge Oriented Methods

These methods are based on user knowledge of given fact, i.e. username and the corresponding password. For user
identification by this method a unique username is required. But, the username and password can be accidentally disclosed, and the security of the identification fails.

Some form of shared or pre-shared password or pass-phrase without the username can’t be used for sufficient user identification, because more users could use the shared password to be accepted by a system. Such form of identification can be used for access into shared environment, i.e. a course of e-learning system, where is sufficient to distinguish between given groups of users only.

B. Ownership Oriented Methods

All the methods based on ownership of some item are inappropriate for secure user identification, when the owned item can be passed to other user or can be stolen. The improper user can be identified in this case, because of ownership of foreign identification item. Additional measures must be realized to avoid this situation.

C. Biometrics Methods

Specific parameters of human body or user’s specific behavior can be used for user identification. Many biometrics methods depend on special techniques or complicated equipment is needed for such body or behavior features extraction. However, some methods can be realized using common computer accessories or peripheral.

IV. IDENTIFICATION METHODS USABILITY

The usability of identification method depends on several basic conditions or features of the target system. Each method requires other type of features extraction, and each system requires other form of feature transmission across its infrastructure.

A. Network System Specific Conditions

The common feature of the network architecture designed systems is user remote access. An identification feature transport across the network can be slightly difficult there.

First, the amount of the information used for the identification can be bigger in many cases. A transport capability problem originates by this reason. The second problem springs from security issue, because to keep the identification feature safe is necessary. Third, the question of technological capability to gain the information by the client part from the user can be fundamental at several methods.

A qualified decision about whether to extract basic features from identification information by server or by client part must be done. It depends on used transport technology and on network application type.

The next problem originates from technical requirements. Some methods require sophisticated reading equipment that can be more expensive than others. It makes some methods unusable for network environment because each one client has to have its own expensive reader. It’s the disadvantage of some ownership-oriented and biometrics methods especially, so these can be used only by systems where the client count is limited.

B. Web-application Specific Conditions

The influence of conditions described above is stronger when using web-application environment. The transport is very limited by the protocol specification that is why only a small amount of data can be transmitted from the client to server part. The need of a quick identification is evident, so the transport speed has to be sufficient.

The basic application protocol used to hypertext transport (HTTP) is unsecured that is why it is not well usable for identification features transmission. The secured form (HTTPS) provides secured sockets layer (SSL) capability to encapsulate that transport into an encrypted stream, see Fig. 3. It increases the communication security and enables to keep the secret data safe. But, the encapsulation consumes some system and connection resources, such as computing time or transfer capacity.

Other security issue is connected with processing of the identification data obtaining or feature extraction. Basic technologies providing a scripting capability of the server or client part aren’t designed to be usable securely. The main problem rising from the insufficient security of scripting is on the client side.

The client part of the web-application is in general realized by web-browsers. Scripting support is limited to a few scripting languages only (JavaScript, VBScript) with many incompatibility issues between browsers. Along with them none security questions are solved, because the source code of the web page and other resources linked in are viewable and can be inspect by the browser additions as a plain text. Some compiled additions (i.e. Java-applets) can be decompiled easily by free utilities, so the security is out of the question. [8]

C. Identification Usage Specific Conditions

More conditions lead from the usage of user identification. If the identification is used for statistical monitoring of user behavior, the system needs to distinguish between users, but the knowledge of their identity is not necessary. The access control management system usually needs the user identity knowledge.

In the web-application environment the tracking cookie...
The main problem related to user identification in these applications is multi-account issue. The owning of more than one account by one player is mostly strictly prohibited, because it leads to undesirable advantage for the owner of more than one user profile account. Providers of these games spend many resources on the security measures against multi-account owners. The appropriate user identification methods can help them to avoid these issues.

D. Geographic Information Systems

The geographic information systems often provide the information with web-based interface as GeoWeb sites. The management of user access uses the same principles as described in previous paragraph. But, the GeoWeb often displays data from different servers with different access rules. The access management needs to be solved not only for the entire web-application, but for the individual layers of the map service.

VI. MOTIVATION FOR GEOWEB USER IDENTIFICATION

As it was described above, the GeoWeb sites can use data from more backend systems and servers, with various types of security level and access rules. It is the motivation for user identification. The GeoWeb site has to distinguish individual users or users group to allow or deny access to some specific map-layers or geo-information. It is important on systems used for tasks as Emergency action plan, Evacuation plan and procedure during floods, or Critical infrastructure and key resources.

A. Geographic Information Systems Architecture

If the GIS provide a GeoWeb interface, it has the form of the typical client/server application with the three basic layers:

**Data layer** can be provided by the given server, but some data can be dynamically loaded or linked from another server or public services across the local network or Internet.

**Application layer** has to process the user requests and provide appropriate responses. This task is typically divided to be processed partly by server and by client side of the system.

**Presentation layer** is the user interface used to collect the request and to visualize the response. [13]
For a better web-application implementation overview see Fig. 4. Client part is realized by web-browser. There the hypertext and images are presented to the user. A dynamic functionality at the client side is provided by client-side scripting tools and languages, i.e. JavaScript, Visual Basic Script, Java Applets, etc. These procedures fall under the application logic. Rest of the application layer is processed server-side, and provided by common web-server enhanced by scripting capabilities of PHP, ASP, JSP, or other programming languages and techniques. The server part provides the data layer functionality using simple filesystem and storage, or more sophisticated solution based on some intelligent database services. The object-oriented databases with many objects’ characteristics and relations are often used nowadays.

B. Geographic Information Systems User Groups
The GIS provided as web-application have some user groups with various relations to the presented data. Each one user group needs a different access to have granted. Because the GIS application can be able to analyze some information on demand, the user access has to be solved not only as the “data access”, but the solution has to be able to manage access to run these analyzes. There are these user groups:

High-end users – some specialists with access to run a high time or system resources consumption tasks, i.e. spatial analyses, etc.
Regular end-users – employees or regular customers, etc., with access to only several functions or data.
Casual end-users – citizens or common guests of the web-application with access to only some analyses results or public data. [14]

VII. USER IDENTIFICATION METHOD COSTS
Costs of the user identification method can be defined from two main points of view. First, we can express the costs of the solution implementation. It means all the resources spent on the changes extending the application to be able to provide the given user identification method. Second, we can evaluate the costs of the solution processing and providing. So it means the resources consumed in relation with the user identification method service running. The general cost value is expressed by some combination of these two viewpoints.

If we need to add the user identification functionality to web-application, the costs of this addition will be the matter of the user identification method choice. For this purpose the costs of addition needs to be evaluated. The cardinal evaluation is more difficult and can be used to express the absolute financial costs. In many cases the ordinal evaluation is sufficient. It provides only the pseudo-values usable for the methods implementation costs comparison. These additional system upgrade costs can be expressed as a total of these values:

\[ \text{Cost} = \sigma \cdot \text{server side addition implementation costs} + \kappa \cdot \text{client side addition implementation costs} + \tau \cdot \text{data transmission addition implementation costs} \]

Each one of these values consists of three parts. First, there are hardware costs. It means the value of the hardware components that are added to given system part (server, client or transmission circuit) for the user identification process needing. Most of the methods do not need additional server side hardware, but some client side hardware can be necessary.

Second part of the value represents the software costs, that is the value of additional software needed to user identification feature of the system. The maintenance of the software is included in this value. For details see Fig. 5:

![Fig. 5 Identification Method Implementation Costs](image-url)

Last value depends on the computing time when system processes the user identification. The costs of this time can be most important in some critical applications.

A. Server Side Costs
\[ \sigma_{\text{HW}} \] – server side hardware costs are low or none for most user identification methods.
\[ \sigma_{\text{SW}} \] – server side software costs are very dependent on type of method; if the data processes the server side, there is only one instance of the server software, but more data needs to be transmitted from clients.
\[ \sigma_{I} \] – server side identification costs depends on the identification process.

B. Client Side Costs
\[ \kappa_{\text{HW}} \] – client side hardware costs are low or none for some methods, but for others can be very high.
\[ \kappa_{\text{SW}} \] – client side software costs depends on the element processing collected data; if this process is provided by server part, no more client software addition needed, and this costs go to zero, but the client side data processing increases the costs with each one new client requesting to add some software instance in most cases licensed and expensive.
\[ \kappa_{I} \] – client side identification costs depends on the used method and the way of processing.
C. Data Transmission Costs

\( \tau_{HW} \) – data transmission hardware costs are none or very low, if the Internet is used as the transmission media; the system designer can not modify the infrastructure, excluding of usage in private network environment.

\( \tau_{SW} \) – data transmission software costs explanation is similar to above described \( \tau_{HW} \).

\( \tau_{I} \) – data transmission identification costs depend on the amount of data the system transmits from client to server side; if there are used some method with identification processing located in the server part, the amount of data to transmit can be high.

The generic evaluation of implementation costs can be expressed by equation (1) where the value \( TIC \) means Total Implementation Costs:

\[
TIC = N_{\sigma} \cdot \sigma + N_{k} \cdot \kappa + N_{\tau} \cdot \tau
\]  

(1)

The parameters \( N \) mean number of servers, clients and transmission links. So, if \( N_{\sigma} \) is equal to 3, it means there are three servers to implement the identification method addition. The \( N_{k} \) is the number of clients to be enhanced by the identification addition. Most of the systems are single-server. It means the value of \( N_{\sigma} \) is one. If the hardware costs needed to be evaluated, the expression (2) will be used. Similarly for the software costs (3) and per identification costs (4):

\[
TIC_{HW} = N_{\sigma} \cdot \sigma_{HW} + N_{k} \cdot \kappa_{HW} + N_{\tau} \cdot \tau_{HW}
\]  

(2)

\[
TIC_{SW} = N_{\sigma} \cdot \sigma_{SW} + N_{k} \cdot \kappa_{SW} + N_{\tau} \cdot \tau_{SW}
\]  

(3)

\[
TIC_{I} = N_{\sigma} \cdot \sigma_{I} + N_{k} \cdot \kappa_{I} + N_{\tau} \cdot \tau_{I}
\]  

(4)

We can use a matrix expression for the better form of evaluation. The unit costs matrix \( UC \) includes the nine values of the basic coefficient (5). The number of servers, clients, and transmission links to be adapted is expressed by the vector \( N \) in equation (6):

\[
UC = \begin{pmatrix}
\sigma_{HW} & \kappa_{HW} & \tau_{HW} \\
\sigma_{SW} & \kappa_{SW} & \tau_{SW} \\
\sigma_{I} & \kappa_{I} & \tau_{I}
\end{pmatrix}
\]  

(5)

\[
N = \begin{pmatrix}
N_{\sigma} \\
N_{k} \\
N_{\tau}
\end{pmatrix}
\]  

(6)

The values of \( TIC \) component (7) will be evaluated by expression (8):

\[
TIC = (TIC_{HW} \quad TIC_{SW} \quad TIC_{I})
\]  

(7)

\[
TIC = N \cdot UC^T
\]  

(8)

All the matrix expressions can be used for the better view of evaluation of the user identification methods implementation costs in future work.

VIII. IDENTIFICATION METHOD COSTS COMPARISON

The comparison by the costs is one of the important parts of the decision what user identification method use. In this point it is necessary to say that the weight of this criterion may be very different in dependency on the web-application usage. For the example of comparison see Table I. Some identification method implementation costs are compared by pseudo-values \textbf{none, low and high}, because the real costs are not evaluable directly in general case.

<table>
<thead>
<tr>
<th>Method</th>
<th>Server side</th>
<th>Client side</th>
<th>Data transmission</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_{HW} )</td>
<td>none</td>
<td>low</td>
<td>none</td>
</tr>
<tr>
<td>( \delta_{SW} )</td>
<td>none</td>
<td>low</td>
<td>none</td>
</tr>
<tr>
<td>( \rho_{HW} )</td>
<td>none</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>( \theta_{SW} )</td>
<td>none</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>( \varphi_{I} )</td>
<td>none</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>( \mu_{I} )</td>
<td>none</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>( \tau_{HW} )</td>
<td>none</td>
<td>low</td>
<td>none</td>
</tr>
<tr>
<td>( \tau_{SW} )</td>
<td>none</td>
<td>low</td>
<td>none</td>
</tr>
<tr>
<td>( \tau_{I} )</td>
<td>none</td>
<td>none</td>
<td>none</td>
</tr>
</tbody>
</table>

If we set a weight for these criteria, the result will be expressed as the sum of pseudo-costs (1), (2) to (4). The client side hardware and software costs have to be multiplied by the count of clients. All the per-identification costs must be multiplied by expected identification processes in the calculated period.

This example displays the password method as the low costs way. There are none client side and data transmission costs. The value of \( \kappa_{HW} \) is important for web-application, because of many clients in such system expected. Other method with none client side hardware costs is speaker recognition. Each one new personal computer includes sound interface with or without microphone (if none, it is cheap to buy it) today. It is sufficient hardware for speaker recognition processing. But, if the client side processing of the audio signal is needed, the client side software will be expensive. If the audio signal is transmitter to be processed server sides, more costs for the transmission will be expected.

The use of pseudo-values is not available in the case of general \( TIC \) evaluation, because there are not defined how many values “low” in total reach the value meant as “high”.

But, if we defined the weights of “low” and “high” in some units, the matrix evaluation would be usable very simply and correctly. The better way for this evaluation to be used is to express the coefficients in money. The approximate values are...
IX. IDENTIFICATION METHOD RELIABILITY PROBLEM

The reliability of some identification methods is the important problem of the general use. There are two weak points in the identification task processing. First, some methods enable the matter of identification to be stolen. It means the access can be granted to person who identifies him/herself by stolen password or token. The system evaluates it as the right identification in error. Second, some methods can not recognize the identification features with the success always. It leads to situation, when the right identification is evaluated as wrong and the access is not granted. Several biometrics methods have this problem. Repeated identification process can solve the issue in several cases. Some examples of identification methods weaknesses are introduced below in the Table II.

<table>
<thead>
<tr>
<th>Method</th>
<th>Weakness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Password</td>
<td>Can be forgotten or disclosed</td>
</tr>
<tr>
<td>ID Card</td>
<td>Can be stolen or unreadable</td>
</tr>
<tr>
<td>ID Token</td>
<td>Can be stolen</td>
</tr>
<tr>
<td>Fingerprint</td>
<td>Can be unreadable if injured</td>
</tr>
<tr>
<td>Eye identification</td>
<td>Can be unreadable in specific condition</td>
</tr>
<tr>
<td>Hand metrics</td>
<td>Can be inaccurate</td>
</tr>
<tr>
<td>Keyboard dynamics</td>
<td>Can be inaccurate or imitated</td>
</tr>
<tr>
<td>Speaker recognition</td>
<td>Can be inaccurate in specific condition</td>
</tr>
</tbody>
</table>

More identification methods can be combined to avoid the reliability inconvenience. It increases the user identification method implementation costs. These costs evaluation is rather difficult, and expects more specific values to be set. The identification method reliability has to be included in the criteria of methods comparison. Other way is to compare the more methods combination (with sufficient reliability) as the one option.

The weight of the reliability problem also depends on the purpose of the user identification. If the identification do not serve for granting of access, but for statistics or monitoring only, the decreased reliability can be covered by some statistic errors.

X. CONCLUSION AND FUTURE WORK

The usability of user identification in the web-application environment is the complex problem. The common task of user identification by information system is limited by the specific conditions and rules of the web environment and communication across the Internet.

The special focus of GeoWeb applications leads to more specific rules and needing given by various tasks of such systems. More users and user groups access levels and the security reasons of the provided information make the solution more difficult.

The decision what identification method select for concrete usage depends on the costs of that solution. At the other hand the usability of identification method depends on more characteristics such reliability and success ability. This problem can be eliminated by use of combination of two or more methods, but it may lead to more additional costs.

The future work is to elaborate more precisely the evaluation and comparison of identification methods. Inclusion of the reliability in described evaluation model is necessary, because the costs can not be the only criterion for the decision.

Some identification methods need to be tested in this specific environment with the focus given to the client side implementation. The server side can be modified or upgraded easily, but the thousands of user’s browsers used as client sides, can not be modified above the options of the web and hypertext standards.

ACKNOWLEDGMENT

This work was supported in part by the Grant Agency of the Czech Republic under Grant project no. 402/09/0219, and in part by the Student Grant Agency of University of Pardubice.

REFERENCES

Oldřich Horák was born in Vrchlabí, Czech Republic in 1973. His educational background: the bachelor degree (Bc.) in Applied Informatics (2004), the master degree (Ing.) in Information Management (2006), both at Faculty of Informatics and Management, Univerzity of Hradec Králové, Czech Republic; he is a student of doctoral study programme in Systems Engineering and Informatics, Informatics in Public Administration at Faculty of Economics and Public Administration, University of Pardubice, Czech Republic.

He worked in informatics and accounting department of house of correction “Husův domov” in Dvůr Králové nad Labem, Czech Republic (1993-1997), and as a secondary school teacher of computer networks, programming, and accounting at SPS in Dvůr Králové nad Labem, Czech Republic (1997-2004). Now he is a lecturer at Institute of System Engineering and Informatics, Faculty of Economics and Public Administration, University of Pardubice, Czech Republic.