Privacy Protection Without Impairing Personalization by Using the Extended System MASKS and the Extended Contextualized P3P Privacy Policies

Robson Eduardo De Grande
Computing Department
Federal University of São Carlos
Zip Code 676 - 13565-905
São Carlos, Brazil
robson_grande@dc.ufscar.br

Sérgio Donizetti Zorzo
Computing Department
Federal University of São Carlos
Zip Code 676 - 13565-905
São Carlos, Brazil
zorzo@dc.ufscar.br

ABSTRACT

Current privacy protection mechanisms are limited to provide privacy guarantees for Web data gathering. MASKS system provides user's anonymity and allowance for implicit data collection, although it lacks security for collecting explicit data. It does not allow browsing session by dividing cookies into different groups of interest. Project 3P introduces a mechanism to keep users aware of privacy policies while navigating through the sites. It offers low reliability for implicit data gathering. Thus, developing a system to combine an extended MASKS system and an extended P3P ensures existing qualities of both mechanisms and supplies each other's limitations. The extended MASKS system includes sessions in the masking proxy to enable browsing session creation. The extended P3P incorporates information about user's benefits by increasing his/her understanding on privacy practices. Running comparative tests with users allows an evaluation on advantages by using this new combined system. The results showed that the combined system provides users' navigation with higher reliability. Thus, the implementation of this system proves useful for providing privacy during Web browsing, without impeding data gathering.

Categories and Subject Descriptors

H.4.3 [**Information Systems Applications**]: Information browsers; H.5.2 [**Information Interfaces and Presentation**]: Evaluation/methodology; K.4.1 [**Computers and Society**]: Privacy

General Terms

Design, Reliability, Experimentation, Security, Verification

Keywords

Privacy, personalization, Web, user, navigation

1. INTRODUCTION

It is introduced a mechanism that guarantees privacy for implicit and explicit data gathering to increase user trust in his/her browsing. Besides privacy warranty, this system allows information obtainment for creation of Web services, such as personalization. The system joins two tools: one, by inserting anonymity in the Web navigation, the other, disclosing data to the user about privacy policies on accessed sites.

Privacy and personalization are mostly important to online services. Both are fundamental requirements to provide attractiveness to site. Business popularity acquired from the Web relies on provided user's security for gathering information and how it presents its services through the web pages. On a research by Teltzrow [20], 64% of Web users stopped accessing some site or buying something through the Web because they did not know how provided information was going to be used after gathered. On another research, Kobsa [12] shows that sites offering personalized services achieved an increase on the number of clients up to 47%.

Recognized as a valuable item that can lead an on-line marketing to success, user trust relies on offered privacy. Risk concerns of unwarranted disclosure or inappropriate use of gathered information induces the user not to access certain services or to make a choice for a more reliable one. On a research, Jutla states that 53% of users do not trust e-commerce sites which collect data, 66% of them do not register in Web sites by fearing the improper use of their information, and 4% of them provide false data to online registration [11].

Furthermore, privacy is intrinsically related to the control that an individual possesses over his/her personal information [1]. Gathering or use of information without Web user knowledge or permission leads to a loss of control. Privacy maintenance causes user confidence to a site, and gives the site a good reputation within a proper time. Thus, user's increase perception of control over his/her data reflects the growth on accessing Web services and products.

Otherwise, user personalization makes navigation easier and turns it more according to the supplied content related to

[1] http://www.privacilla.org

his/her interests. A visitor's profile identification allows displaying site pages in a format that guides to an easier and faster search, besides services presentation is molded according to receiver preferences. Based on Kobsa [12], clients need to feel they possess a personal and unique relationship with an enterprise.

Some services to be personalized demand a minimal knowledge on who will receive it. Personalizing is turning something closer to someone's characteristics. Add to that, Jutla [11] evidences that great number of users would disclose information to receive some benefit in exchange. Thus, Web sites have necessarily to collect information from their visitants to apply some personalization.

Web data gathering can be accomplished in an implicit or explicit way [5]. In the first case, obtaining data succeeds without the user knowledge or consent. This lack of knowledge by an individual regarding what occurs with his/her data can lead to a loss of control of his information. In order to accomplish collecting data, sites use cookies and observation of page requisitions. This browsing observation captures control data, which is related to the utilized communication protocols. In this manner, through the gathered data analysis, it is possible to delimit a profile to those who frequent a site.

Explicit gathering requires the user to expose his/her information in evident manner. In this case, the collected data is not captured by an automatic mechanism. It is from content, related to personal information of an individual. In this type of data acquisition, the visitor is aware of its existence and he/she has the choice to consent it. Thus, in this gathering process is harder to occur some type of privacy impair. However, an invasion may succeed regarding to site attitude when using collected information.

There are several techniques of privacy protection in order to avoid privacy invasion in the data gathering by increasing user control over his/her information. Pseudonym creation, use of privacy contracts, masking of requisitions, contextualized disclosure of privacy practices and personalization benefits and the anonymity introduction to the navigation are examples of these techniques.

The anonymity prevent from any information disclosure. The user maintains a total control over his/her information, since no data is available. Meanwhile, it becomes impossible to create personalization, due to impediment of data obtainment that impairs provided services that need browsing identification.

The information acquired by the implicit method is fundamental to the e-commerce interaction and to the clickstream data capture [3], [15]. It makes possible creation of users profile based on their interests, navigation patterns, preferences and others. Due to its importance, anonymous navigation must not obstruct this type of data gathering.

A taxonomy [9], [10] is utilized to classify these tools and techniques into different privacy layers. According to this classification, no tool approaches all the protection layers. In this way, MASKS and P3P are joined into a unique system to cover a bigger number of this classification layers. Both are tools that provide privacy and trust on information disclosure and allow data gathering for personalization purpose.

MASKS system provides privacy to the navigation and it allows the implicit information acquisition. This anonymity server makes possible the identification of profiles in groups of interest without user is recognized. Meanwhile, by the proxy property to divide cookies into different groups of interest, it restricts the profile delimitation and makes session management impossible.

Project 3P discloses information about the site privacy practices to the users' data. It introduces a mechanism of automatic reading and evaluation of privacy contracts. For that purpose, it is necessary a specific format to the policies creation and to the user agent construction. Although these privacy contracts inform about site practices, they cannot bring any guarantee to the user, mainly by the implicit gathering.

With the purpose of increasing the user control over his/her information, a new privacy protection system is built. Through the combination between MASKS server extended and Project 3P extended, the system adds anonymity to the implicit gathering and bigger trust in the explicit sending of user information. Navigation sessions are included in the MASKS proxy to extend it. The P3P is expanding by the addition of information about the user benefits in the data disclosure.

Outlining paper sections are as follows. Second section presents implicit and explicit data gathering methods. In third section, existing privacy protection mechanisms and their respective limitations are exhibited. In fourth section, the MASKS server functioning, architecture and limitations are exposed. The fifth section describes the Project 3P and its limitations. In sixth section, anonymous browsing system and site privacy practices disclosure are defined. In seventh section, it is described the methodology of the tests accomplished. Eighth section presets results. Finally, results, implication of modifications accomplished and future work are discussed.

2. IMPLICIT AND EXPLICIT DATA GATHERING

Implicit data is obtained in a passive manner. It results from observing the navigation of an individual. This way, a user is not capable to identify when noticing browsing occurs. Meanwhile, it is possible to recognize the product generated through this collected information by offered service modifications. Depending on how data capture occurs, a person's privacy can be impaired, mainly if these data is associated to his/her real identity.

This data is captured in order to evidence a person's interests when accessing the Web services. A visitor profile draft lies on his/her interests, it is used with the purpose of personalization application. The collection of this information uses cookies and clickstream capture techniques.

2.1 Clickstream

Clickstream term, also known as clicks paths, denotes the path or route that a user accomplishes through one or more Web sites [3]. The route passed by discloses a sequence of choices offered on a site, the pages visited, the page through which a person arrived at the site, the access time in the site, the accomplishment of an online purchase and the other sites visited through it.

Clickstream data is a natural product of Web navigation, generated automatically without the need to interact with the user. These data can contain information about objectives, knowledge and interests. According to the marketing and e-commerce perspective, there are interests in using knowledge acquisition techniques to these data for propaganda application to improve the understanding and forecast the user choice behavior.

Through the clickstream data, it is possible to implement interactive propaganda systems in a dynamic way, to deduce individually an objective, helping users that do not keep in mind what they want. This inference is achieved by recommendation systems [3].

An identification of user's computer must occur in order to receive an answer to every request. Navigation trough a visited site server or the computers wherever the communication packages passed through can register this identification. The following techniques can be used to collect clickstream data without refinement [15]. Server log files are registers of visitor access in sites. Panels of access capture are mechanisms of requested observation inserted by third parts in the access site interface. Internet service providers (ISP) are able to make a complete navigation register of their users, so that they are the communication canal to the Web.

Statistical and data mining methods are used to extract knowledge of data sets captured by these methods. Knowledge mining mechanisms can generate results such as: consumers' demographic characteristics, predictions, patterns of malicious use, identification of consumer loyalty, patterns of behavior and others [4].

However, anonymous navigation techniques make this data gathering impossible. Anonymity tools are capable of removing any information from communication headers that identify the user computer. Even though, sites are able to recognize the access of a determined visitor through the cookies.

2.2 Cookies

Cookie is a small information fragment referred to a user. Its content, generated by a Web server, consists of a characters string contained in the user browser memory. Although its content is only text, it can contain information to identify a user.

The initial objective in creating cookies was to introduce a memory in the HTTP protocol, which does not store navigation status. The status is necessary to distinguish transactions. Thus, through them, it is possible to store the last browsing situation before finishing the connection for the sake of future navigation.

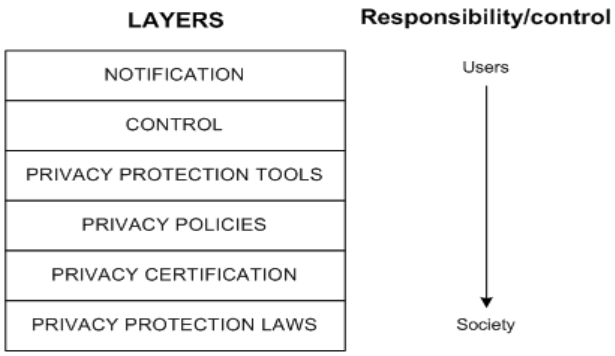

Figure 1: Privacy Protection Layers.

According to its specification [14], a cookie has syntactic properties involving attribute and value pairs. It contains a name, a content and a validity period. The Web server creates the cookie content, and it is stored in the user computer by the browser. Whenever a user requests a page, the browser looks for all the cookies that refer to the page and send them to the site.

The large use of cookies is mainly to create personalized services and to maintain communication sessions between the user and the site. Maintaining the navigation present status, it records the visitor session situation in a particular moment. Besides, a cookie contains information that can evidence the user requested history, applied to determine a profile and to generate personalization. Thus, blocking cookie storage prevents providing several Web services, turning them nearly impossible to be provided.

3. PRIVACY PROTECTION MECHANISMS

Several specific tools maintain the user navigation privacy. Their primary objective is to introduce larger information control to the user when accessing the Web services. According to the approaches of these mechanisms to the content or control data gathering, there is a division in different categories, privacy contracts, pseudonyms, anonymity and navigation masks. Ishitani classifies these categories in different approaches layers.

Concept of privacy protection layers was introduced by Ishitani [9], [10] as a taxonomy of all user privacy protection context. This classification facilitates the identification of the differences among each approach of existing mechanisms.

In accordance with Ishitani, this taxonomy is divided in 6 distinct layers, which can be visualized in the figure 1 [9].

Notification informs the users about the risks that they can suffer during their navigation by the Web.

Control is relative to the mechanisms that allow users have a better control over their information in avoiding the data obtaining tools.

The privacy protection tools mask the virtual or real user identity to protect him/her. This layer owns anonymity,

pseudonymous and requisition masking mechanisms.

The privacy policies present site descriptions about its privacy practices.

The privacy certificates comprise the regulation services of sites privacy practices.

Privacy protection laws regulate the privacy protection.

Privacy policies are assertions provided by sites to inform user about their privacy practices. Warning user about the gathering practices and the data dealing, his/her acceptance of privacy contract terms occurs by continuing the access to the site.

The policies add trust to the user on information disclosure. A research showed that about 90% of users desire that sites look for permission before to use the personal information throughout marketing [16]. Another research evidenced that 76% of users consider very important privacy policies [6] and that 55% of them assert that a privacy policy comforts the information disclosure [2]. However, the policies are written in a complex language and they are hardly read by the users [19]. Thus, there are methods to simplify and automate their analysis.

Platform for Privacy Preferences Project [2] (P3P) [5] establishes an automatic mechanism for privacy policy analysis. It introduces a standard format, XML, to construct policies and a tool to examine them automatically. This way, the user has larger control over his/her data through the management of his/her navigation.

Kobsa [13] presents a framework for the disclosure contextualized of privacy practices and personalization benefits. This method notifies the user according to the context he/she is.

For all methods of this approach, user needs trusting the policies and he/she believes that the site follows them correctly. Besides, the privacy contracts offer few guarantees on implicit data gathering.

Another approach is the pseudonyms creation for the purpose of users access the Web services without presenting their real identities. It is inserted an anonymity orientated to the content information, and, essentially, it does not deal with the implicit data gathering and neither the user profile identification. Janus Personalized Web Anonymizer (JPWA) [7] is an example of this type of mechanism. It operates as a proxy that generates automatically nicknames in the authentication process. However, supposing it is possible relating a pseudonym to a user, it will expose all his/her actions already accomplished.

For anonymous navigation purposes, using a server or a set of servers removes any information that identifies the user browsing. Using anonymity does not allow any type of data implicit collection, and consequently it becomes impossible to apply personalization. Anonymizer, Onion Routing and Crowds are typical mechanisms of this approach.

[2]http://www.w3.org/P3P/

Anonymizer [1] is a Web proxy that forwards the user requests and applies certain methods to mask them as belonging to him. Thus, Web sites cannot distinguish users, identify the access source and associate a request to a determined user. Several routers constitute Onion Routing [8], each one works as a proxy that applies methods to improve privacy. This routers network is dynamically built, and it is fault tolerant and avoids traffic espionage by creating random paths.

The Crowds system [17] works with the cooperation of members from an anonymity group. Each user contributes by hiding the real source of a request. A request is forwarded random and dynamically to a member from a Crowd group or to the destination site. Due to the routing shift in the requests, it is not possible to determine the original user.

MASKS system [9] is another anonymous navigation approach. Nevertheless, it allows personalization through cookies. It guarantees privacy to the control data with no denigrating the provision of personalized services.

4. MASKS SYSTEM

MASKS (Managing Anonymity while Sharing Knowledge to Servers) system introduces anonymity to the user browsing without blocking information to the sites. Its objective is to allow some personalization, even without the user identification. The system presents some important privacy characteristics, it is efficient and adaptable, it avoids the user data storage, it makes flexible the amount of information that user desires to disclose, it accepts cookies and it does not need modifications in the existent communication protocols [18].

To bring privacy, the MASKS proxy removes any communication control information that identifies a user. This anonymity is viable because large portion of navigation is summed up in searches and visualization of documents, without the explicit information exchange with Web sites [18]. Besides, data sent by the user does not cause impair to his/her privacy, since the explicit information acquisition is known and consented. Thus, the MASKS server inserts privacy guarantees to the implicit data gathering.

Cookies are essential for providing personalization to the MASKS system. The system supplies them as the only possible way to enable sites to accomplish implicit collection. The anonymity proxy inserts the cookies in the communication according to the user interest profile. So, site server does not hold user private data and a Web site cannot realize if the access was performed by a group of similar interests instead of a unique user [9].

Requisition anonymity proxy and a user PSA agent (Privacy and Security Agent) compose MASKS system. PSA is a plug-in added to browser acting as a system configuration and as an user access interface, allowing him/her to be informed about his/her masks and to alter them [18]. It removes data from HTTP requested headers and, according to user, redirects the browser communication traffic to the proxy. Figure 2 presents the system architecture and possible situations that a requisition can be masked [18].

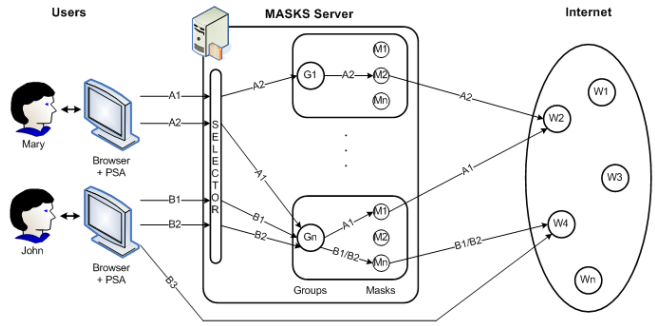

Figure 2: MASKS System achitecture and functioning.

MASKS server provides privacy by the exchange of IP of all HTTP requisition that passes through it by the masking process with cookies insertion. Cookies inclusion occurs according to the interest profile presented in each URL requested by the user. An interest group selector, a semantic tree, groups and masks manager compose MASKS system [9].

Due to the difficulty on detecting user profile in his/her navigation, group selector analyses each URL requested and assigns to it an interest group. This process uses a semantic tree; it classifies the profile groups in categories. A group designation process must be effective and semantically significant. Designated requisitions for a group need to be lead to same subject pages [18]. Thus, the URL analysis and group choice consists in to select previously the group from the semantic tree that indexes the URL. If it isn't found, the group is specified by the query terms from the URL. In negative case, the group is indicated according to the URL terms. Finally, the Root group is select.

A category tree is used to construct semantic tree. It is defined by the Open Directory Project [3]. A volunteer subject classification accomplishes this category definition. This method is used because the object semantic classification is complex for automatic techniques application [9]. The classification consists of a Web site list organized by categories and it is constantly updated.

Each tree knot is an interest group, and each group contains a set of masks, that correspond to user requisitions. A group can possess several masks, one for each page of same interest. In such a manner, a user can receive several masks during the browsing, one for each different interest presented while navigating.

Masks correspond to cookies sent to a site through the HTTP protocol. All proxy users make use of them, useful for characterizing requisitions. Sites modify them as a navigation consequence and their insertion in the system masks updates them as well as the semantic tree.

Although maintaining the cookies is used for personalization, the anonymity proxy impedes the access to the cookie main feature; it impedes the storage of navigation status. Without this resource, sites are not able to maintain navigation sessions. The grouping of masks in different interest groups disables login process, purchases and others, which basis is in the cookies.

Besides, two different requisitions can use the same mask whenever they relate to the same page or to pages of same category of interest. In this case, it is possible for user to access the account of others when he requests the same page through the MASKS system. Thus, the cookies lose their memory and they only maintain the feature in labeling a requisition profile.

5. PLATFORM FOR PRIVACY PREFERENCES PROJECT

P3P, Platform for Privacy Preferences Project, created by the Wide World Web Consortium [4], intends providing the users to keep their privacy while accessing Web site services. For that, it specifies criteria to constructing P3P privacy policies and user agents. The objective of this platform is to automate reading and analysis of privacy policies by introducing a pattern to construct it.

Project 3P presents a standard based on the P3P vocabulary and it enables Web sites to express their privacy practices. Privacy policies are defined by using a XML codification with name spaces, which assert the information collected, how to obtain it, place and time of storage, the person responsible and purpose for the gathering. The purpose of P3P vocabulary is not just an indicator of obedience to laws or conduct codes, but it is to describe a site's behavior.

Privacy preferences, privacy policy reference, P3P privacy policies and a user agent compose the Platform. Preferences are configurations established by users determining which information can be disclose. These privacy options determine the user agent behavior when analyzing the site's P3P privacy policies.

A reference file links the accessed page to one or more P3P policies. The references allow determining if policy covers certain site region. The assertion codification is in XML with name spaces that can delimit one or more policies to an entire site, to site parts or to a sole Web document. They indicate the location of privacy policies file. There are three manners to locate reference file.

A place previously known can establish a file location. In this case, the user agent may obtain the policy references file previously to any request for resource. This standard place is the path "*/w3c/p3p.xml*" added to the site server address.

HTTP headers can contain information that indicates the policy reference file position. Creating a P3P field in a new answer HTTP header may inform its location. This field is named *policyref*.

A link tag in the HTML code is used to obtain the position of policy reference file. The href attribute indicates the file place. A XHTML link tag can also be used to determine

[3]http://dmoz.org

[4]http://www.w3.org

the localization of this file.

P3P policies are a translation of site privacy policies into a standard format. The P3P vocabulary is the basis for indicating them. To a set of information, these policies delimit gathering purpose, storage time, place, and the use given to the data collected.

User agent is a software that acts together with browser to assist user navigation. Building of this agent can be made directly in the navigator; it can be a plug-in or a proxy server. Initially, it looks for references file to obtain the policies related to requested page. Through these policies, it accomplishes a comparative analysis between the site practices and the privacy preferences delimited by the user. At last, it signalizes to the user and it informs him about this privacy situation.

The platform only informs the user about site practices. The P3P policy is only informative and it does not provide guarantees regarding true site attitudes. In the case of implicit gathering, the P3P offers few guarantees, since the user cannot detect this obtaining data procedure.

6. SYSTEM OF ANONIMOUS BROWSING AND DISCLOSURE OF SITE'S PRIVACY PRACTICES

The system objectifies to protect the user privacy throughout data gathering in Web navigation. The MASKS system and the Project 3P compose the two modules. The combination of these mechanisms conciliates the qualities of each one and it supplies their limitations. Therefore, implicit data gathering offers privacy guarantees by the anonymous navigation and, privacy contracts inserts security into explicit collection.

The constructed mechanism has a user agent and a server that masks requisitions. The agent redirects the user navigation to the anonymity proxy, it removes information from the HTTP headers and it looks for P3P policies extended to signalize to the user. This agent is composed by a PSA, MASKS agent, and by an extended PPA, a P3P user agent. Figure 3 shows the architecture and the functioning of all developed system.

The Module that masks requisitions is the system MASKS extended. This modification introduces sessions in the masking proxy to make possible the creation of navigation sessions, which are necessary to access the Web services and to provide security to the information disclosed explicitly. Besides, this added extension makes cooperation between both modules more practical.

The Module of privacy policies and personalization benefits disclosure is the extended Project 3P. This extension incorporates the communication of information about user benefits in data sending, considering the users' interests in providing data to receive some benefit in exchange [11].

6.1 MASKS Server Extention

The MASKS proxy extension restores the cookie memory to the user navigation and maintains his/her browsing anonymity.

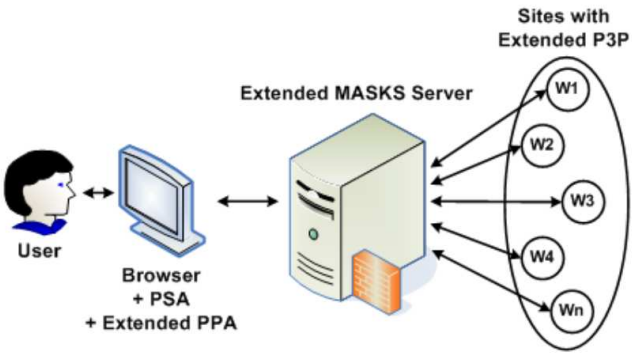

Figure 3: Functioning arquitecture of system of anonimous browsing and disclosure of site's privacy practices.

Therefore, there is insertion of sessions in the MASKS server. These sessions allow an identification of user navigation status, which is lost afterwards. This loss is resultant from cookies mixing among masking groups and users. In the extended system, every one has access to the group cookies, except to those related to user session in a site.

The period spent to accomplish certain activity or part of it defines a session. In the Web, one considers session the time in which a transaction occurs between user and site; identifying and determining its existence is a cookie function. Thus, cookie validity determines a session existence period.

Sessions of long term identify user, and short term ones are used to make logins in systems, purchases, and other services wherever there is a great amount of interaction with user. Web sites normally utilize cookies with no validity, without the field "*expires*", to maintain these short-term sessions. Session insertion into proxy allows sites to create these short-term sessions.

In the server, session creation is on interest groups level, each group receives a validity time to continue active in a cache area. When a group session expires, the group returns to the semantic tree and updates it with new cookies. Therefore, groups from semantic tree remain updated with users' navigation. It occurs when the user loses his/her interest in some topic during determined period.

An addition to the architecture permits to extend MASKS system, a session manager, a mask searcher and a cache area. These new modules operate with the user agent, selector and group manager, without modifying the original parts. See figure 4.

The cache area contains mask session repository and group copies. User sessions divide this area. Each session possesses groups and masks utilized by some user. Groups' copies maintain reference with the original groups. Each user receives a masks repository that has all the cookies applied by it in his/her navigation.

In order to update according to user browsing, the repository maintains semantic tree groups masks copies. Figure 5

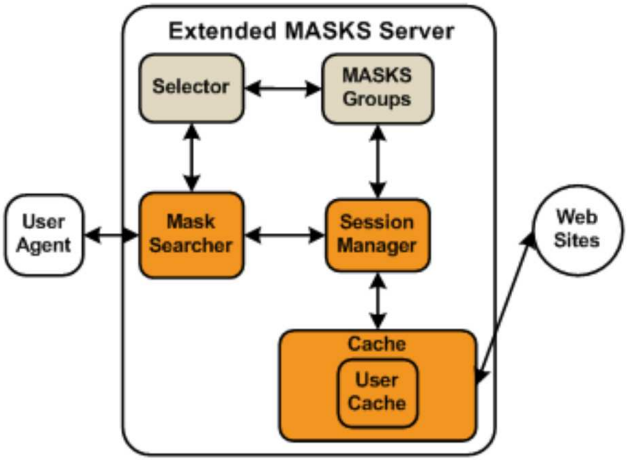

Figure 4: Extended MASKS System achitecture.

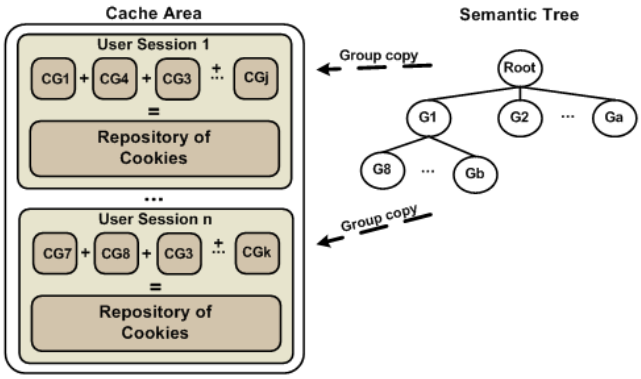

Figure 5: Cache Area and its user sessions.

displays its contents and behavior. Its purpose is to avoid interruption on user navigation session and to facilitate applying cookies treatment criteria, pursuant their specification [14]. It is necessary that the newer cookies overwrite the older ones to maintain an order in the session. It must occur in the cache session on a global way, and on a specific way in its groups in order to keep updated. Thus, repository contains every session cookies plus is responsible to actual status maintenance for all user sessions.

The mask searcher, with the requested URL, selects the group through the MASKS selector. After that, it searches by the group copy in the user session. If the session does not have a copy, it creates a new copy there, and, if there are cookies in the copied group, it processes transference to the repository. Finally, the requisition receives its respective cookies and forwards to the site.

The session manager is responsible for the creation, maintenance and elimination of user and group sessions in the anonymity proxy. A user session holds one or more group sessions. The user makes a user session whenever he first requests through the MASKS server.

The user, navigating, can have his/her requisition assigned to a new group session or to one that had expired. There is the generation of a group session, when a group does not exist in the user session, copying the group to the user cache area and there is the insertion of group cookies into the repository. This insertion occurs with the cookies copy that did not exist in the repository; it is to maintain the consistence of sessions in the sites. In this way, the extended system utilizes the existent cookies to insert interest group characteristics at the communication beginning with a site with no impair to existent navigation sessions.

The manager maintenance is within valid period of a group session, in repository updating as well as on the group copy cookies updating. A requisition forwarded to a session, updates the group validity. Updating of cookies occurs in all the answers to requisitions.

Terminating a group session, the manager removes it from the user session cache and inserts the group copy masks in the respective original group. This insertion selects the cookies presenting a defined validity. The user session must be finished when cookie session validity expires, or whenever the cookie does not have validity and the browser is closed [14]. Maintaining cookies updated to other user's access and freeing cookies to someone else identity is this copy responsibility.

The manager closes a user session when it has no one in the group session. It eliminates the user session from the cache area, as well as its cookies repository.

Such as in MASKS, the added extension to it continues incapacitating the user computer identification. However, the user session maintenance in the anonymity server makes possible the creation of navigation sessions and it allows making personalization closer to the user for the generated session.

6.2 P3P Extention

Expanding Project 3P improves the user understanding about his/her information disclosure in transactions through the Web. The user accesses Web services to receive some of his/her interest advantage in exchange. The user interest in accessing some service depends on knowing about the benefits he can receive. As though privacy policy instructs about privacy practices, it can also inform about the benefits received through a service access.

Developing project presented by Kobsa [13] applies when orienting privacy policies construction and increasing information disclosed is necessary. For a page or a set of pages, there is contextual and specifically built P3P privacy policy. Thus, the user visualizes only the policies pertinent to the site region, which he accesses. Besides, P3P policies specification is extended to hold data about the user benefits.

The constructed P3P policies present a hierarchic organization to disclose privacy practices in a general and contextualized way. Common privacy statements are inherited among policies. General policies to a set of pages have their statements inherited by policies that are more specific. Figure 6 exemplifies that. So, delimitates specific and general contexts to all site.

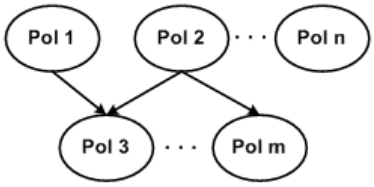

Figure 6: Hierarchical building of extended P3P privacy policies.

Specific user benefits are on each element "*PURPOSE*" on policy. To justify sending user information, data gathering purpose must result in some advantage to him. Add to that, in order to delimit user advantages, clearly defining purpose can be better understood making more conscious decisions about the information disclosure to the site [13].

Since benefits are specific for each application, creating a standard classification to the disclosure of these advantages is complex. In this way, each site must describe its services such as user benefits in a contextualized manner.

The container "*STATEMENT*" delimits elements of "*PURPOSE*", expanded to hold a user benefits description. Modified purposes are as follows:

<current> </current>,
<current> </current>,
<admin> </admin>,
<develop> </develop>,
<tailoring> </tailoring>,
<pseudo-analysis> </pseudo-analysis>,
<pseudo-decision> </pseudo-decision>,
<individual-analysis> </individual-analysis>,
<individual-decision> </individual-decision>,
<contact> </contact>,
<historical> </historical>,
<telemarketing> </telemarketing>,
<other-purpose value=""> </other-purpose>.

In order to describe the user advantages for each purpose, it is created a new element: <user-benefits>PCDATA</user-benefits>.

An extended P3P agent must be built to find meaning in this new specification benefits. For each element of "PURPOSE", agent must look for respective user advantage description by permitting this data obtaining. At the end, it must present, along with the warning of the site's privacy practices, information about this new element. Therefore, the user is aware of data gathering, site policy and benefits he can receive.

7. TESTS

Submission of 15 people is accomplished to developed tool qualitative tests in order to verify the modifications added to user agent. Information about benefits proved to increase user trust, as well as anonymous navigation. The criteria for analysis were user understanding and security, in browsing and data sending. One accomplishes comparative evaluation between P3P user agent and its extended version, and between global and contextualized privacy policy disclosure. Besides, anonymity proxy is tested in order to measure the user trust augmentation in his/her navigation.

The test participants utilize three user agents, which inform about site privacy policies. They simulate the browsing by the same site with three different versions of P3P privacy policies: one contextualized, the other extended and a third global. At the end, these participants answer a comparative questionnaire asking about which methods brought more security to them.

The user evaluation was by three P3P privacy agents: Privacy Bird [5], PPA (Privacy Police Agent) and PPA extended. The Privacy Bird originally was created by the AT&T Corp. to attend the P3P specifications. To execute the test in a standard way the interfaces of Privacy Bird are used as basis to construct the PPA, which is according to the P3P, and the PPA extended, which is create according to the P3P modifications.

To the test, the user preferences were configured to the maximum privacy level. In this way, the user visualizes the checking report in a standard way, and then he has access to all possible warning messages.

In the first part of the evaluation, the user familiarizes with each agent that he will use. This familiarization consists in presenting the tool basic functioning. At practice, after each page access, the user must consult the checking report that the agent makes available. At the end, one has executed of one-fourth of navigation at the site version that has global privacy policies.

After the navigation simulation, there is a presentation of a comparative questionnaire to the user. In this set of questions, the user informs subjectively which agent presents more trust in its privacy reports, which agent and P3P policy construction maintain the user better informed and how much an anonymous navigation mechanism that allows implicit data gathering increases his/her trust.

8. RESULTS

Two different contexts describe the results in this work. In one of them, it is presented information obtained through the application of comparative tests. This subjective data evidences the user opinion about the use of system introduced in this paper. In the other context, it is described the result in the sessions added in the MASKS proxy. There is a demonstration of efficiency in maintaining sessions during the anonymous browsing.

8.1 System MASKS Extended Efficiency Evaluating

The insertion of sessions in the anonymity server allowed the user to receive Web services that utilize navigation sessions. The implemented extension impedes identifying any information by the Web sites. Thus, the extended system

[5]http://www.privacybird.com

offers privacy by the anonymous navigation and it provides knowledge discovery and browsing sessions by the cookies.

There was a benefit to implicit data gathering with the sessions in the server. Cookies session can identify user navigation, but the session closure in server brings anonymity again to the user. Thus, the extended server provides a momentary identification of the user browsing, which one remains while the session in the system is valid.

MASKS server, without the modifications to insert sessions, removes the cookie memorization feature. A user that navigates by a site receives a semi-static personalization; an interest group confines it. Add to this, it does not allow services that utilize navigation sessions, and it presents insecurity in the user navigation sessions, they are public domain in the system.

According to Ishitani [9], without the MASKS use, a user requests a page sequence, p1→p2→p3→p4→p5→p6. In this sequence, pages p1 and p2 are from sports, p3 and p4 are from business, p5 and p6 are from tourism, and the site observes it as a sequence of a unique user.

Using the MASKS system to mask the same user requisitions, dividing the sequence in 3 interest groups in the server, p1→p2, p3→p4 and p5→p6. By the cookies inserted according to the semantic of each requisition, the site identifies through them 3 different sequences, as p1→p2, p3→p4 and p5→p6. In conformity to the cookies there is no correlation between requisitions from different groups, what limits the personalization and makes creation of navigation sessions impossible.

The system extension organizes the user navigation sequence. The page requisition sequence, presented previously, executed in the extended system is not broken either loses its order. The site identifies the sequence as from a unique user, and it is able to identify the transition between two pages from different interest groups.

Creating sessions in the server, the user browsing does not invade other's navigation session. With session finishing, user identity ends as well. So, there is a better produced personalization, since the implicit data gathering is dynamic at the session level in the extended MASKS proxy, and user privacy is kept.

8.2 Comparative Tests Results

Once tests were over, all participants agreed that anonymity proxy brings more security in their navigation. However, 84% of them find it more convenient not to have their navigation identified. In this way, although the anonymity is not a necessary requisite to all of them, it brings more security in the maintenance of browsing information.

During evaluation, 84% of users consider pleasant receiving personalized services while navigating anonymously. The others 16% that would not like to access these services in this context also would not appreciate to receive any type of personalization. Thus, to those who have interest in receiving personalized services, the extended MASKS system is ideal.

In the tests, the extended agent brought more trust to the user and satisfaction in disclosing his/her information than other agents. All test participants concluded that the extended agent details more data gathering purpose, it presents more information and approaches closely to user understanding.

In the situation of explicit information sending, 34% of research participants accepted to send their data only by knowing their benefits. The other 66% accepted or rejected to disclose their information anyway. However, 100% of them affirmed that the extended agent advises them more clearly and makes them more aware.

The 87% of users preferred the contextualization of P3P privacy policies. They concluded that it is more objective, it does not compete with navigation and it relates to the accessed page context. The others 13% chose the global construction of policies because it allows them to make a verification of whole site in just one time.

9. CONCLUSION

The presented system provides privacy to the user navigation in the implicit and explicit information gathering. The proxy extension that masks requisitions guarantees the anonymity without impairing the data obtaining in implicit way. The extended Project 3P inserts privacy contracts that assure more trust in the explicit information sending.

The MASKS system extension is efficient in providing privacy to the user navigation, in allowing information discovery to the personalization application and in maintaining user navigation sessions. The session in the anonymity server simulates the user browser management. In this manner, it is possible to maintain navigation sessions without linking them directly to a browser.

Explicit data gathering is the only way to identify user real identity. By the extended system, a user, connecting to a site, is dependent having his/her navigation observed by the cookies. However, user identifies explicitly, agreeing with site privacy policies impeding this process.

The insertion of sessions in server allowed increasing the user access to Web services that utilize the cookie memory feature. The system finishes user browsing sessions if they use cookies without defined validity.

The insertion of user benefits proved to increase the user knowledge to make decisions. By the benefits, he/she is more capable to choose to access a determined service. Besides, disclosure of this data allowed to focalize in the application context and to direct the advice to a more understandable speech to the user, without technical terms.

Mostly preferred by evaluation participants was the contextualized privacy policy building. It brings more objectivity disclosing it to the user. Thus, site's privacy practices presentation is more comprehensible.

The extended system MASKS use offers trust to the user in his/her browsing and it allows the implicit data gathering. According to the accomplished research, users consider more

reliable to use a system that masks requisitions.

For future work, it is necessary to improve the access interface of system user agent by considering usability criteria to provide to the user a more intuitive navigability by the system and a better understanding of information presented. Add to that, the insertion in the system of a mechanism of privacy seal conference to bring larger privacy guarantees to the P3P policies.

In the extended MASKS server, using a 20 minutes validity period to maintain group session, it expires within this period with no user requisition related to it. Nevertheless, it is necessary to accomplish more tests in order to adjust this session validity. The adjustment of this time is critical for user navigation sessions maintenance.

10. REFERENCES

[1] Anonymizer. Anonymizer enterprise network privacy/security appliance, 2004. Anonymizer Inc.

[2] L. Behrens. Privacy and security: The hidden growth strategy, August 2001.

[3] R. E. Bucklin, J. M. Lattin, A. Ansari, D. Bell, E. Coupey, S. Gupta, J. D. C. Little, C. Mela, A. Montgomery, and J. Steckel. Choice and the internet: from clickstream to research stream. In *U.C. Berkeley 5th Invitational Choice Symposium*, pages 245 –258. Mareting Letters, February 2002.

[4] A. Cavoukian. Data mining: Staking a claim on your privacy, 1998. Technical report, Information and Privacy Commissioner.

[5] L. F. Cranor. 'i didn't buy it for myself' privacy and ecommerce personalization. In *WPES?03*, Washington DC, USA, October 2003. AT&T Labs-Research.

[6] DTI. Informing consumers about e-commerce, September 2001. Department for Trade and Industry. Conducted by MORI.

[7] E. Gabber, P. B. Gibbons, Y. Matias, and Y. Mayer. How to make personalized web browsing simple, secure, and anonymous. In *Proceedings of Financial Cryptography'97*, page 17?31. Springer-Verlag LNCS 1318, February 1997.

[8] D. Goldschlag, M. Reedy, and P. Syversony. Onion routing for anonymous and private internet connections. *Communications of the ACM*, 42(2):39–41, 1999.

[9] L. Ishitani. *Uma Arquitetura para Controle de Privacidade na Web*. PhD thesis, Dept. of Computer Science Universidade Federal de Minas Gerais, Minas Gerais, Brazil, 2003.

[10] L. Ishitani, V. Almeida, and W. Meira. Masks: Bringing anonymity and personalization together. *IEEE Privacy&Security*, 1(3):18–23, May 2003.

[11] D. Jutla and P. Bodorik. A client-side business model for electronic privacy. In *16th Bled eCommerce Conference and Transformation*, pages 463–479, Bled, Slovenia, June 2003.

[12] A. Kobsa. Tailoring privacy to users' needs. In *8th International Conference in User Modeling*, pages 303–313, Berlin, Germany, 2001.

[13] A. Kobsa and M. Teltzrow. Contextualized communication of privacy practices and personalization benefits: Impacts on users' data sharing and purchase behavior. In *Privacy Enhancing Technologies: Fourth International Workshop*, pages 329–343, Toronto, Canada, 2004. Springer LNCS.

[14] D. Kristol and L. Montulli. *HTTP State Management Mechanism*. Bell Laboratories, Lucent Technologies, October 2000.

[15] A. L. Montgomery, S. Li, K. Srinivasan, and J. C. Liechty. Modeling online browsing and path analysis using clickstream data. *Marketing Science*, 23(4):579–595, 2004.

[16] R. Morgan. Community attitudes towards privacy, June 2004. Technology Overview. Roy Morgan Research.

[17] M. K. Reitter and A. D. Rubin. Crowds: anonymity for web transactions. *ACM Transactions on Information and System Security (TISSEC)*, 1(1):66–92, November 1999.

[18] B. G. Rocha, V. A. F. Almeida, L. Ishitani, and W. Meira. Disclosing users' data in an environment that preserves privacy. In *Workshop On Privacy In The Electronic Society*, pages 71–80, Minas Gerais, Brazil, 2002. Dept. of Computer Science Universidade Federal de Minas Gerais.

[19] M. Teltzrow and A. Kobsa. Impacts of user privacy preferences on personalized systems: a comparative study. In *In CHI-2003 Workshop: Designing Personalized User Experiences for eCommerce*, pages 315–332, Dordrecht, Netherlands, 2003. Kluwer Academic Publishers.

[20] M. Teltzrow and A. Kobsa. Communication of privacy and personalization in e-business. proceedings of the workshop. In *WHOLES: A Multiple View of Individual Privacy in a Networked World*, Stockholm, Sweden, 2004.