Next generation of anonymous web browsers: a bit closer to democracy?

GULYÁS, Gábor György

e-mail: gulyasg@hit.bme.hu

Abstract

In our previous studies we analyzed several anonymous browsers in different aspects and we discovered that these services usually offered preset options in packages, with only a few available options. As a result, users were unable to have an influence on global options. Social networking, collaborative filtering might enlighten some new perspectives regarding enhanced options, increased role of users and might also open the possibility of integrating different services, such as anonymous web browsers and search engines. In this study we analyze the effects of these Web 2.0 compatible features on anonymous browsers and further possibilities.

Keywords: anonymity, privacy, anonymous browsers, social networking, collaborative filtering, Web 2.0

1. Anonymous web surfing

While surfing the web a lot of information is provided automatically about the user's computer, operating system and her web browser agent. Even crucial pieces of information can be revealed that might easily get the user tracked. Regarded as a puzzle, a group of cooperating user–tracking sites can put all the small pieces together and create a profile of the user's habits, containing accounts or other information valuable for marketing purposes.

Naturally, — as the title of this article suggests — there is a solution to this threat: *anonymity*. Being anonymous on the web seems to be natural, since compromising this property might never be revealed to the user herself. However, it happens many times.

Anonymous browsers offer complex preventive solutions. Using these services, the state of anonymity can be preserved, and in some cases certain undesirable contents, like advertisements, can be filtered out, too. We had analyzed and evaluated several services and technologies in [1].

Basically there are two types of anonymous browsers: web based and regular proxies. The main difference between them is that web based anonymous browsers can be reached through websites and their control panel is embedded into the viewed pages, while regular proxies do not have that much of transparency: a certain intermediary agent needs to be installed or settings have to be changed in the web browser agent. Both have special filtering systems to remove malicious code from the downloaded content, also narrowing the scope of revealed information about the user, although if someone is using regular proxies, in most cases her content filtering possibilities are more limited.

It is certainly important for these services to meet some theoretical criteria [2] as well, in order to preserve the possibility of anonymity. Accordingly, the use of anonymizing services (like Tor [4]) is required, in addition to other filtering features.

In our study we mostly discuss web based anonymous browsers, however, the discussed theories and methods can be easily applied to the other type as well. The reason of this partiality is simple: these web based applications have more deficiencies and consequently need to be fixed in more ways.

2. The age of Web 2.0 has begun: A brief review of technical features and possibilities

As an unusual aspect of design or purpose — keywords, such as these would describe Web 2.0 applications in a profusely simple way [3], and social networking, interactive websites are good examples. No technical advancements were necessary for these services to spread.

The basic protocol used to retrieve the content of web pages, called HTTP (Hypertext Transfer Protocol), follows a simple idea: the user defines a request, and the response is the content of a page (or the required file). As requests are generated by clicking on links or by submitting forms, this is a stateless protocol: the current state is defined by variables embedded into the links or hidden form inputs. In order to solve this problem, cookies (a special user side database storing snippets of information) were introduced in a later phase of developing the Web.

As the user roams the Web generating new requests, every time the request–response mechanism is invoked, the page is reloaded. Surfing this way is something like reading a book or a newspaper, where clicking on links corresponds to turning pages over.

The “breakthrough” technology is already a decade old; however, it has been applied only recently. AJAX (Asynchronous JavaScript and XML) allows web developers to create

interactive user interfaces on websites, increasing usability and functionality. As the dissolution of the abbreviation suggests, HTTP requests can be sent asynchronously without reloading the currently viewed page. Usually these request-response sessions contain small messages, like mere value modifications or requests. The use of AJAX technology attracts more embedded functionality.

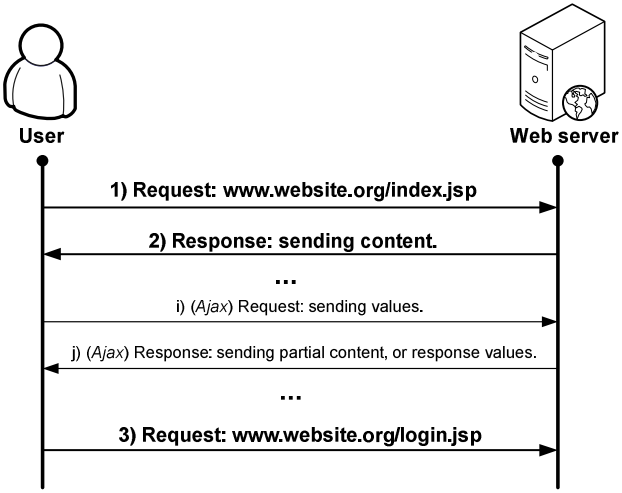

Figure 1. AJAX and regular HTTP requests.

On Figure 1 we illustrate the basic differences between conventional and Web 2.0 sites (using AJAX). After the first step the server side business logics generate the requested file. In case of conventional sites there would be no communication until another request, shown in step 3. If AJAX is used there can be several response–request sessions between step 2 and 3, user activity does not require reloading the whole page. From the viewpoint of the web server these requests are normal ones, however, special server–side scripts are being run for generating small partial output or response values.

These small protocol sessions guarantee the desktop-like application behaviour by enabling actions without affecting the whole page and no reload is required. AJAX creates a link between the front–end user interface and the back-end business logics and databases.

Websites embracing the principles of folksonomy[1] are often serving functions by using AJAX as the basic technology. Tagging, voting are the tools of folksonomy for the purpose of categorizing, classifying, raking, describing any type of content. These tools have become wide–spread today. Folksonomy helps creating systems based on semantic contents like search engines, related content commendatory services.

Instead of a static, preset central knowledge base a more democratic one can be created this way. Naturally, certain entities have to monitor malicious users and actions affecting global scope. Automated solutions (well defined protocols and cryptography) can be inserted to prevent frauds.

3. Anonymous web browsers and Web 2.0?

Essentially, anonymous web browsers are proxies. All required contents are relayed through them by filtering all incoming and outgoing information. Besides filtering, in order to preserve the users' anonymity, requests (and the respective responses) between the users and the proxy are encrypted and the system often uses MIX[2]-es.

Today's anonymous web browsers — these previously defined proxies — use preset features composed by the service provider these are being applied instantly for any part of any site the user visits. The user's options are usually limited and applied globally for any site and in any situation without exceptions. One can easily find examples to demonstrate the disadvantages of this approach, for example, filtering out all scripts on sites is a great and security–enhancing feature, however, some sites will be inaccessible this way. Such a plain functionality can be significantly improved by applying the collaborative filtering features, tagging and voting, following the principles of folksonomy. We recommend using these features jointly; users should add labels and should vote pro and contra for tags in order to increase their "entropy".

As a result, we gain a system with an open database created by a democratic society of users, instead of pre–written, ever–lasting rules from a company. Furthermore, one can also

[1] The practice of communities collaboratively creating and maintaining tags is called folksonomy, also known as collaborative filtering.

[2] MIX-es are used to obfuscate communication between the users and the proxy, and by applying MIX-es, ideally it would be impossible to pair incoming and outgoing messages at the proxy. For more information read the original proposal in [9].

customize her options according to her level of trust in the system and user managed datasets.

3.1. Tagging

Tagging[3] can be used for several purposes. Besides content tagging there are other uses available like tagging addresses of sites creating a *general description* including label groups such as subject tags, warning for adult content, indicating connections with other sites, etc. Certain supposedly cooperating sites grouped into a virtual network by tagging can be filtered out together or, similarly, parental locks can be applied to them.

Furthermore, tags might describe *security guidelines*: cookie policies, scripts management guidelines, labelling sites for phising[4], etc., making page–customized options available. In addition, the shared database is up–to–date anytime and accessible for anyone, we might say, it is "open source"!

Tagging *content* is also a great option, as we previously discussed. Different parts of web pages can be described as advertisements, containing disturbing material (text or images). In addition to visual content tagging script codes, web bugs, cookies and other types of malicious (semi–)hidden content can be marked for suggested filtering. The latter can be done on web sites very efficiently — we will show this later.

The smaller, the better — this seems to be the main philosophy of content tagging, though it is better otherwise. Pages on the web are created in HTML (Hypertext Markup Language) describing layout and design. The source code is stored in a plain text format represented by a modelling scheme named DOM (Document Object Model). DOM contains elements called tags (this might be a confusing name in this context, so let us call them DOM–tags) in a tree–hierarchy. Accordingly, DOM–tags can have child elements (depending on their type). In this context it might be worthy to label larger parts of the content, removing all its child elements as well.

---

[3] Tagging is the method of appending labels on different kind of subjects like pictures, web pages, books, etc. Tagging is used to enhance filtering, searching or simply to help categorizing content.

[4] Phishing stands for using camouflaged e-mails and web sites pretending to be authentic to gain passwords, credit card numbers, etc. The practice of phishing based on social engineering and on the technical subterfuge for stealing.

Analyzing the DOM structure is not necessary on the server side. Excellent developer tools are available in today's web browser agents which allow real time analysis on client side and also can block the download procedure of undesired parts.

3.2. Potentials in information– and identity management

Filter management is also easier with tagging. Tagged content and sites can be divided into groups, this way security settings are considerably easier to handle. Not to mention identity management possibilities: besides benefits of anonymity, sometimes it is also beneficial to manage different, independent accounts on a website, or on a group of sites. By managing independent accounts — or let us assume identities —, anonymity can be preserved despite the fact of using accounts. Technically speaking, this means that the anonymous web browsers have to offer the users an identity change function, and also during the change, all information that could be stored for tracking purposes have to be erased.

In our opinion the next generation of anonymous web browsers should support basic identity management features. If nothing else, the continuous evolution of e–commerce would be a single but powerful reason to support these functions. Even within these applications there might be several requirements which are not granted by the default protocols used for browsing web applications.

Filtering can also be extended by applying a special threshold function customized by the user herself. Not all the users have the same preferences or level of trust in collaborative filtering. To improve the suggested model all users should be able to define a threshold value to strain out presumably invalid, low rated tags.

3.3. Example: tagging and filtering with threshold values

In this simple example Alice visits an on-line gaming community and selects a game. Alice's goals are rather reasonable; she wants to get all the advertisements and sexual content filtered. Thus she uses a hypothetical next generation anonymous web browser with filtering enabled. This browser has a threshold-based filtering system that requires a threshold value set in order to indicate which tags having lower vote count will be strained out. In our example her value is 5.

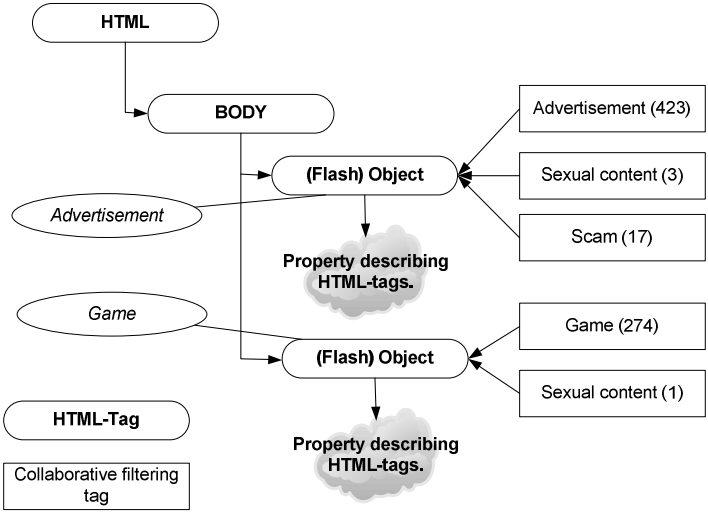

Figure 2. The sample page transformed into a tree hierarchy of HTML tags.

As Figure 2 illustrates, the page contains two flash objects: one of them is an advertisement and the other is actually the game. The HTML–tags can be seen with their respective vote counts. The first flash animation was definitely tagged as an advertisement by numerous users, and a few others tagged it as scam and sexual content.

It can be seen that a malicious or misbehaving user tagged the core services of the page, the game, to have sexual content. This would not matter for Alice. By her threshold setting the anonymous web browser would just ignore the tag. This method simply works just as well in the case of filtering the advertisement; however, the browser ignores the sexual content tag here, too.

Its simplicity makes the using of threshold values advantageous. Conversely, besides the fault tolerance property of the system it might be a weakness: important tags might be ignored. Sophisticated voting systems can improve this feature, for example, relying on user reputation; likelihood decisions might help as well.

4. Feasibility analysis

Anonymous browsers, like Privacy Enhancing Technologies (PET), serve both individual and democratic values and rights. In many cases these aims do not osculate with business

goals, and since it is hard to find financial support for these projects, the implementation phase cannot be initiated or, if initiated, it fails. In other cases financial problems, marketing purposes or other business objectives affect the result of the project.

In this chapter we discuss the feasibility of designing, implementing and supplying these services: in our belief these next generation systems require several magnitudes higher budget and time for designing, implementation, and their maintenance would also require a larger monthly expense. Who might ever want to build such a system, and for what purposes?

4.1. Basic differences

Considering the expected throughput of next generation anonymous browsers, a *sophisticated design* should be applied in order to create a scalable system. The databases and the back–end systems have to be distributed on a network of servers and the server application should also be implemented and designed in a complex manner.

Besides guaranteeing the integrity of the databases and user actions throughout the network, frauds should also be prevented. Since in next generation anonymous web browsers even a single user can affect global options by committing her tags, it is very important that a single user should not be able to enforce the options she likes. In ways like using automated bots to increase votes on tags or exploiting protocol flaws one could gain any kind of advantages.

Implementing the client–side application might also bring some troubles. According to the statistics in [5], even if it would be satisfactory to target the 88% of web users, the client–side application should be compatible for three different web browsing agents. In our opinion, the client–side application should be implemented for the five most famous browsers, targeting 97% percent of the browser market share. Implementing portable client–side applications such as discussed needs experienced professionals and developers as well.

Operational difficulties would occur as well as costs would be increased by several factors: the service should be run on a farm of servers, and it also requires wider bandwidth due to the nature of the application (using AJAX technology increases the number of response–request protocol runs between the server and the client).

Proper design can prevent frauds, however, moderating is also advised. Moderators should monitor malicious user behaviour by revising suggestions proposed by the system's fraud detection module. Paying a group of moderators seriously increases the operational budget.

Since filtering functions can also be implemented on the client side, there might be some performance expectations which might exclude the use of low performance hardware such as PDA-s, mobile phones or even old computers.

4.2. Possible investors and motives

Regarding the difficulties mentioned in the previous section, another question arises: who would invest in developing such a system? In this section we analyse different possible investors' motives and goals.

Anonymous web browsers serve democratic rights, so it seems to be natural to have *governments or other democratic organizations* taking over the responsibilities of developing these services. However, the number of democratic goals and possible achievements seems to be less then expected or enough to be a motivating factor. (In addition, even democratic governments are often counter–interested in providing such services to their citizens because governments claim control over the behaviour of their citizens even on the internet.)

In a recent Communication the Commission of the European Communities expressed its intention to support the development of PET technologies [6]. This point of view might change the current situation and similarly to the PRIME project [6] and it is not completely unlikely that once there will be a Europe-wide project supporting the designing and creating of a next generation anonymous web browser, or at least providing professional and financial support for a PET system including an integrated anonymous browser.

Online companies pursuing direct marketing, might also be interested in gathering personal data. The existence of these well–working companies reinforces the need for anonymous web browsers: utilizing the possibilities of the web they use a lot of tools for creating statistics and profiling individual users. For example, selling profiles for targeted advertising might be a good — and illegal — business in the future, too.

Using next generation anonymous web browsers this practice would be possible in a lawful way: all users should be informed and warned on what is observed and logged about them, and — similarly to other existing PET solutions — either they consent to forwarding their personal data, or only anonymous statistics would be created. No more tricks will be

required for hidden observation, and there will be a calculated (or hopefully growing) number of users whose activities can be lawfully monitored within the consented limitations.

Since all browsing information flows through the central server, monitoring becomes more effective: not only sites "wired with sensors", but any activity from the first visited site to the last one can be tailed. Some information which nowadays are be hard to collect, such as time spent on sites, might be easier to monitor. These features neccesitate a service that is compatible with all web services, and the user will never have to exclude the anonymizing service.

Another benefit of central filtering is advertisement removal; this change might attract a new type of advertising system. There is no need any more to sign contracts with sites for putting on advertisements and pay–per–click; the services on advertisements can be embedded into the user interfaces. Creating predictable advertising might be beneficial even for the users. Also, all concurrent companies' advertisements can be completely removed.

Collaborative filtering features attract the attention of *search service providers*. Experimental search engines gather results by interpreting the search query and the documents. This process can be helped by creating semantic description of sites, pages or even other kind of partial content built by tags added by users.

Creating a semantic web this way might raise a problem: subjective tags do not represent the content in a substantive way since users also vote on a subjective basis. For search engines understanding objective semantics is a better solution, although the main purpose is to find the content more relevant for the users' queries and perceptions. In this respect subjective tags might be better than objective ones, because of their relevancy for the users. (In smaller and specialized user communities tagging represents a more coherent approach.) Also a single user's tags can be used to understand queries or to build them.

In our opinion it is most likely that a company which is in the direct marketing business and also runs a web search service, like Google or Yahoo [8], whom we will refer to as The Company in the following, might profit the most of creating a next generation anonymous web browsing service. These companies have enough resources for the design, development and operational phases.

4.3. Moral considerations

It is worth mentioning that by building this system, The Company would be in an advantageous position compared to its competitors. The sooner it starts the system operating, the more advantages it will have, especially because of the co-evolutional business model between the two corporations mentioned in [8]. These huge corporations providing whole bunch of services are sometimes called informational superpowers. Introducing these services into their portfolio, or additionally integrating their services into the anonymous browser would exponentially increase their power regarding the possibilities mentioned in previous sections.

The basic features like creating unlimited (theoretically anonymous) statistics on user activities, filtering out their competitors' advertisements, restructuring the advertising policy raise the question: would it be right to let such a huge corporation extend its possibilities this way? These changes alone could kill smaller regional competitors.

In this context we should also consider our trust in these companies. Possibly they can not be trusted at all — strict business goals first, always — but we should be prepared for the worst case scenario.

4.4. Future work and possibilities

Besides using such a great anonymizing system for free, there would be other benefits, too, such as the possibility of integration with other services, for example on–line e–mail applications utilizing anonymous remailers. These possibilities improve usability and might affect users' willingness to use the service.

There might be even more possibilities. Nowadays' anonymous browsers remain on the lowest level of *identity management*: guaranteeing only complete (identity free) anonymity. However, with smart management of profiles and identities, identity management can also guarantee pseudonimity, while enabling accounted access for different services (even different accounts for different services).

In our future work we would like to analyze the possibilities of identity management on the web including e–mail privacy issues and the possibilities of PRIME integration as well.

5. Symbiosis of PET technologies and community based services

Generally, networking PET technologies are better on larger user bases that can be guaranteed by combining PETs with community based services. Greater user base enhances the chance of anonymity as in the case of using anonymizing networks, MIX-es.

Community based collaborative filtering provides a better way of database updating, refreshing, which might also be a relevant feature.

In our opinion, this symbiosis will hopefully strengthen the democratic nature of the internet and guarantee a level of anonymity and user-centric identity management that one should ever desire.

References

[1] Gulyás, G., Are the anonymous web browsers anonymous? Analysis of technologies and services, In: Alma Mater Series on Information and Knowledge Management No. 10, BME GTK ITM, Budapest, March 2006. pp. 9–30 [in Hungarian]

[2] Gulyás, G. and Schulcz, R., Next generation anonymous browsers. *Híradástechnika,* August 2007. pp. 24–27 [in Hungarian]

[3] Wikipedia's entries on Web 2.0, AJAX, Folksonomy, DOM: http://en.wikipedia.org/

[4] Tor http://tor.eff.org/

[5] TheCounter.com Global Statistics On Browsers Market Share in October 2007 http://www.thecounter.com/stats/2007/October/browser.php

[6] Commission of the European Communities: Communication from the Commission to the European Parliament and the Council on Promoting Data Protection by Privacy Enhancing Technologies (PETs), Brussels, 2.5.2007, COM(2007) 228 final http://eur-lex.europa.eu/LexUriServ/site/en/com/2007/com2007_0228en01.pdf

[7] Privacy and Identity Management for Europe (PRIME) project https://www.prime-project.eu/

[8] Google's and Yahoo's homepages: http://www.google.com http://www.yahoo.com

[9] Chaum, D. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM* 24, 2, 1981. pp. 84–88. Available at: http://www.freehaven.net/anonbib/cache/chaum-mix.pdf