Work in Progress: Uncovering the Privacy Implications of Web Usage [*]

Hendrik vom Lehn[†] Jó Ágila Bitsch Link[†] and Klaus Wehrle
Communication and Distributed Systems
RWTH Aachen University, Germany
{vomlehn, bitsch, wehrle}@comsys.rwth-aachen.de

Introduction

Internet users visit a considerable number of websites every day. This makes keeping track of which information is disclosed to which site increasingly difficult. To aggravate this situation, the inclusion of third-party services into websites is often hidden or easy to overlook. Information leakage can therefore become invisible and opaque.

Having incomplete knowledge about disclosed information is not a problem in itself, but can have implications for the users' privacy. Knowing about potential information leaks, on the other hand, allows users to assess and utilize websites in an appropriate manner.

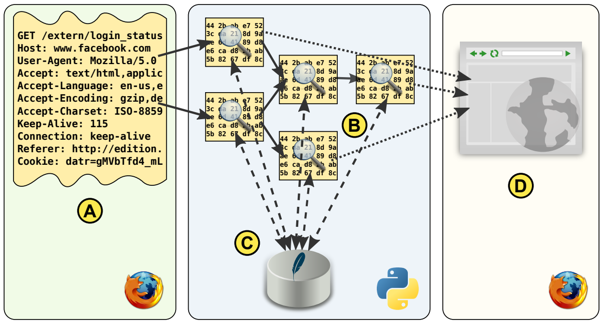
The poster introduces a novel approach that tries to fill this information gap: By locally monitoring and analyzing web traffic for disclosed information such as visited websites or geotags in pictures, it becomes possible to confront users with their digital personality. In combination with details how the information has been disclosed and advice how to avoid this kind of information disclosure, users get a tool at hand that helps them to keep an eye on their privacy while surfing the web.

Architectural Design

In order to realize the proposed concept, we present an architecture as depicted in Figure 1. This section roughly outlines this architecture and highlights selected key design aspects.

Our overall goal is to give users useful advice regarding the privacy implications of their web usage. Instead of interrupting a user while requesting websites, we perform the analysis in the background and store the extracted information. Later, the user can retrieve the information and draw her own conclusions from it. This allows to include information from multiple requests into the analysis and to provide the user with a complete picture.

An active intervention in critical situations is nevertheless a reasonable addition, as it can prevent privacy-breaches directly. We plan to add this functionality in the future and already have rudimentary support for this in our preliminary implementation.



```
GET /extern/login_status
Host: www.facebook.com
User-Agent: Mozilla/5.0
Accept: text/html,applic
Accept-Language: en-us,e
Accept-Encoding: gzip,de
Accept-Charset: ISO-8859
Keep-Alive: 115
Connection: keep-alive
Referer: http://edition.
Cookie: datr=gMVbTfd4_mL
```

Figure 1: Overview of the proposed architecture: A browser plugin intercepts the user's web traffic (A) and feeds it into a pipeline of analysis modules (B). Here, information is extracted, further processed, and stored in a database (C). Based on this information, we generate output on disclosed information as well as potential privacy implications and present it to the user (D).

Furthermore, we want to facilitate the extension of our framework with additional analysis capabilities. The core of our framework therefore consists of lightweight modules which make use of the infrastructure that the embedding framework provides. These modules can register for specific types of information and, after performing the analysis, emit extracted information to other modules. The modules thereby form dynamic chains – the analysis pipeline.

Together with the disclosed information itself, data to whom the information has been disclosed, a reference to the input information and advice on possible countermeasures is stored. This data can be used later-on to search and view the extracted information.

Finally, we want to present the extracted information in an appropriate manner. This may heavily depend on the type of acquired information. Generating meaningful summaries and detailed views is therefore also the task of the modules.

Conclusion

We anticipate that a full implementation of our architecture and the according modules will provide a mean for users to critically analyze their behavior with respect to privacy when browsing the web. The poster will provide more details on the proposed architecture, emerging difficulties and possible application scenarios.

[*] Poster abstract submitted for the poster session at NSDI'11, Boston, MA, March 30-April 1, 2011. No demo planned.

This work was supported in part by German State of NRW within the B-IT Research School.

[†] Graduate Students