Bilateral Anonymity and Prevention of Abusing Logged Web Addresses

Thomas Demuth
Department of Communication Systems
Universität Hagen

Andreas Rieke
ISL Internet Sicherheitslösungen GmbH
Germany

- Overview -

- **Introduction and motivation**
  - Why providing anonymity for providers and users of web pages?
- **Situation/Fundamentals**
  - Hypertext Transfer Protocol/HyperText Markup Language
  - Weak points concerning anonymity
- **Providing anonymity for web servers**
  - Traitorous Data
  - How to hide a web page
- **Providing anonymity for web surfers**
  - Privacy attacks
  - Solutions

- Introduction and motivation -

- The internet is an open network.
- Usage of the WWW increased in the last years.
- In WWW, both participants (client and web server) know each other (at least via IP number).
- Client (web browser) sends a request to the server using an address in a special format.
- Server responds with the corresponding web object (e.g. HTML page, video, or executable program).

→ Access of web pages leads to protocol/communication data

- Introduction and motivation -

- Data can be and is collected (by companies, governments, etc. observing networks and providing web pages).
- In most cases, the transferred information is user/ institution/company related.
- Data pools can be exchanged and matched/ synchronised.

→ Location of web pages are publicly known by address.

→ Users of web pages *think*, they are anonymous in WWW.

- Situation/Fundamentals -

The WWW represents a giant hypertext document; web client and browser communicate with each other in a standardised form:

HTTP - Hypertext Transfer Protocol describes:

1. Syntax of addressing

By URL (Uniform Resource Locator):

[scheme]://[server].[domain]/[path]/[object]

Elements reveal information about server (place, affiliation) or author (of an object).

[server] and [domain] are the most sensitive information.

- Situation/Fundamentals -

2. Structure of messages

Header of a message can contain (traitorous) meta information (fields) like:

- email address of the client
- type of web browser or server
- preferred language

Example:

```
GET http://www.milcom2000.org/
Connection: Close
Accept: image/gif, image/x-xbitmap, image/jpeg,
image/pjpeg, image/png, */*
Accept-Charset: iso-8859-1,*,utf-8
Accept-Encoding: gzip
Accept-Language: en,de,de-DE,de-AU,de-CH,en-US
Host: www.milcom2000.org
User-Agent: Mozilla/4.74 [en] (Win98; U)
```

- Situation/Fundamentals -

Most web pages are created using a description language:

HTML – HyperText Markup Language

A HTML page contains

- textual information and
- instructions (“tags”).

By tags, text can be attributed, pictures can be displayed, or tables can be created.

But: Tags can also refer to another web page.

- Weak points regarding anonymity -

Meta/Administrative information in

- HTTP header

URLs (references in clear text) in

- HTTP header
- HTML text (HTTP body)

Active Attacks

- Why server anonymity? -

To hide the identity, place, or affiliation to an organisation of the sender in a relation of communication:

- Anonymous contribution of a paper to a conference.
- Surveying of a company's new product which should not be associated with the company.
- Anonymous blackboard containing the advertisements of employees.
- Realisation of the anonymity of speech in totalitarian countries.

- How to hide a web page -

- Use of a proxy to prohibit direct communication between client and server.
- The proxy uses an asymmetrical crypto system (like RSA) and offers a form to encrypt URLs.
- Encrypted URLs can be published and used like normal URLs.
- Only the proxy is able to decrypt the URL.
- The proxy also handles the transport of the request to the web server and the return transport of the requested web page

- JANUS/Rewebber -

JANUS/Rewebber is a service offering the use of anonymous URLs.

An URL is encrypted and published:

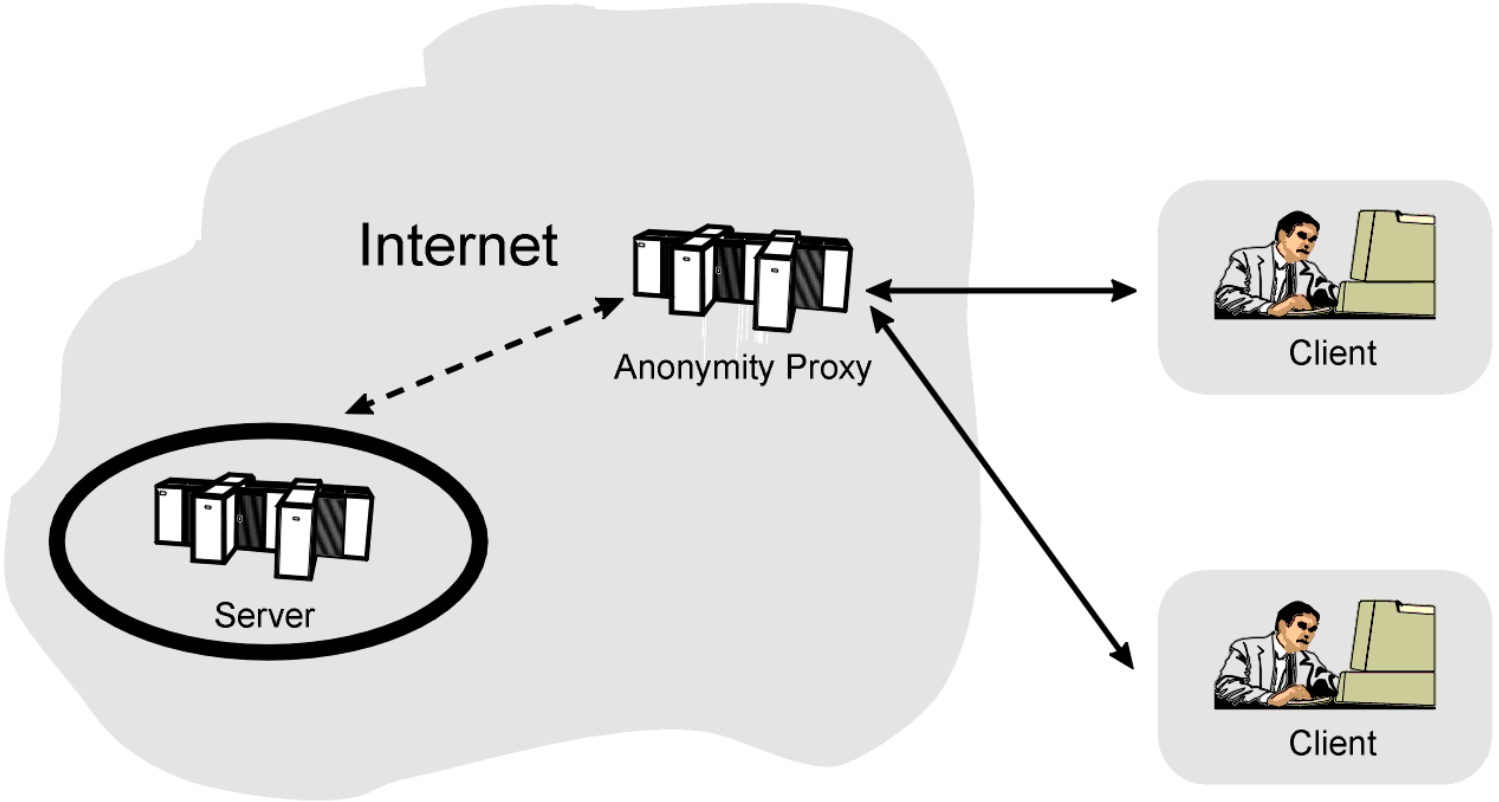
http://www.milcom2000.org/

becomes

http://www.rewebber.com/surf_encrypted/MTBuln5EFN8u$EFEnK68VF898GjCmhmsjYvIwvAndRGHFBF$KpVEVXUXzHC5ezz0hAmzSSEH2gRh4N6Iy4ifTXxs4lmbWK94ERRUUuauT8C6RyF+sN8KyQUR0BT1Vv9UX5s=

and can be used with a regular web browser.

→ A web server/page can not be identified by its URL.


Internet
Anonymity Proxy
Server
Client
Client

- JANUS/Rewebber -

JANUS/Rewebber eliminates or modifies sensitive header fields in a server's response.

JANUS/Rewebber offers to contact a web server by an anonymous URL.

Further references in the transported web page are handled automatically by the system.

→ Server anonymity

- Why client anonymity? -

Profiling of visitors of web pages

(„Mr. Smith's hobbies are ..., his incoming is ...“)

→ Loss of privacy, spying of personal tendencies

Undesired identification of institutions/ companies

(„Why is institution/company x interested in such a way in our product line y/ our institution structure?“)

→ Backtracking of web page users

Client anonymity ↔ User privacy

- Attacks at client anonymity -

- Collecting and analysing meta data of requests for web pages
- Caching of web pages
- Cookies
- Reuse of logged URLs

- Attacks at client anonymity: data collection -

Problem

- Unwanted collection and analysis of meta data of requests of web pages

Solution

Using an anonymising proxy:

- The proxy removes or modifies meta data and acts like and for the client.
- The web server contacted notices only one client.
- References in the web page are modified such that their usage automatically involves the proxy

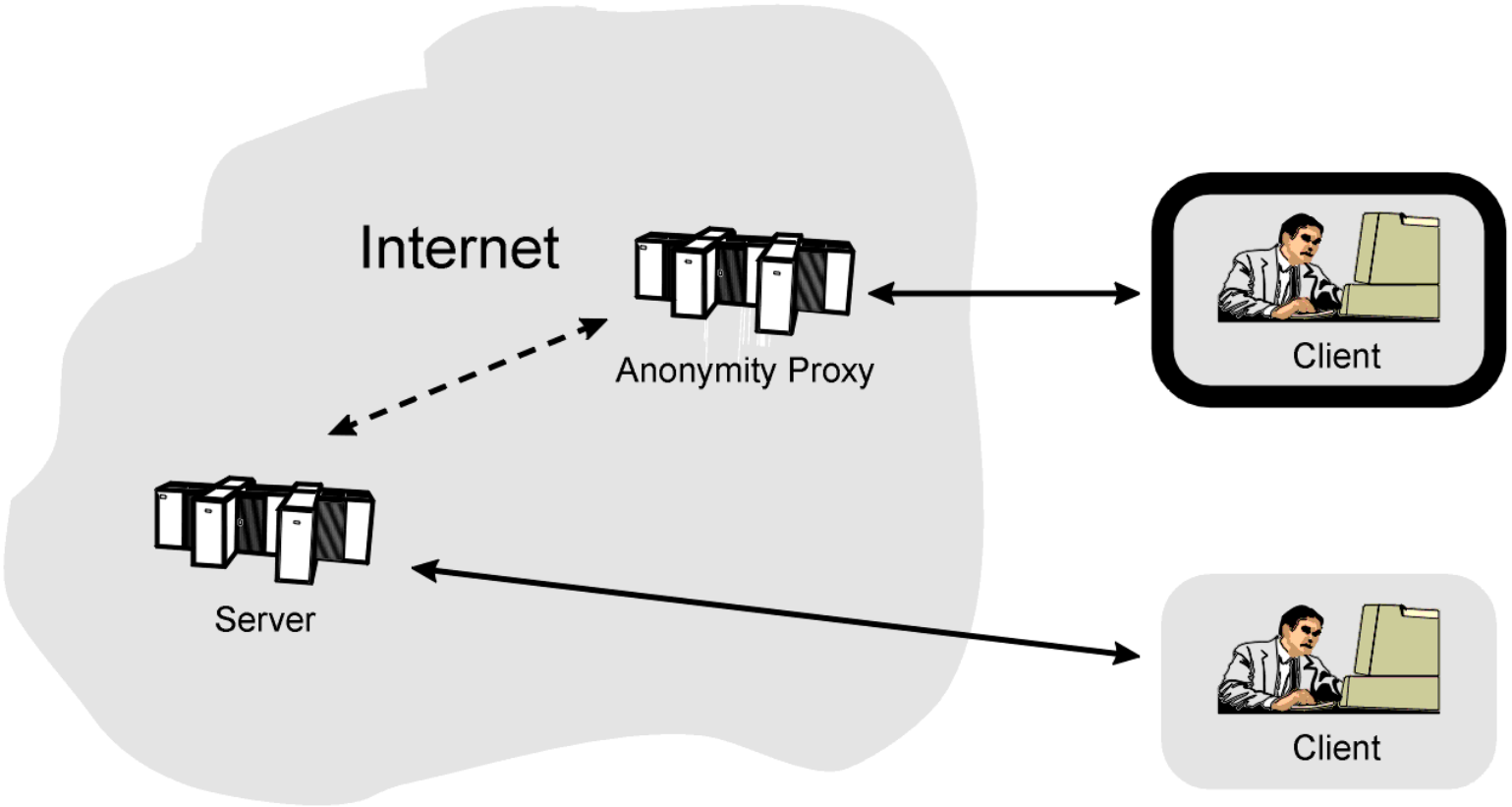
- Client anonymity -

Services for client anonymity already exist:

- The Anonymizer
- Crowds
- Lucent Personalized Web Assistant
- Onion Routing

General idea:

- Web pages are accessed via one or more proxies


Internet
Anonymity Proxy
Client
Server
Client

- Attacks at client anonymity: caching -

Problem

- Web server sets a coded HTTP header „**`Last-Modified`**“ to mark the user/client individually (meantime attack).

Solution

- Filtering of „**`Last-Modified`**“ headers
- Drawback: Each request is transported to the web server; for each request a web page is transported back to the client

- Attacks at client anonymity: cookies-

Problem

- Web server sets cookies to recognize the user/client later on

Solution

Using an anonymising proxy:

- The proxy removes/blocks cookies set by a server
- Drawback: Sites that really need status management can not be used.

- Attacks at client anonymity: cookies -

Alternative

- The proxy offers a „cookie jar“:
- Cookies are stored at the proxy individually for each proxy user.
- Drawback: The user has to authenticate to the proxy
- Advantage: User can not be identified when switching to non-anonymous mode (without proxy)

- Attacks at client anonymity: logged URLs -

Problems

- Logged URLs can be used by ISPs/observers to recover user's visited web pages
- Also subject to PCs accessed by more than one user: offices, internet cafes

Solution

Combination of

- encrypted URLs

and

- time stamps

- Attacks at client anonymity: logged URLs -

- Introduction of a time stamp field to an anonymous URL
- User can determine the (relative) time, the URL will be valid
- The time stamp consists of two elements: Absolute and relative date/time
- The proxy is able to determine the validity of an URL
- References in the transported web page are converted to time valid URLs based on the known time stamp
- In case of URLs not longer valid, an error message is displayed by the proxy

- Attacks at client anonymity: caching -

Problem

- Caching of web pages at instances between client and server („feature" of web servers and web proxies to enhance performance)

Solution

- Using a feature of HTTP

  The command „**`pragma:no-cache`**" instructs all transporting instances no to cache the corresponding page.

- Attacks at client anonymity: Summary -

- Accessing of web pages indirectly (via proxy)
- Filtering of meta data
- Avoidance of web page caching
- Removing/handling of cookies
- Temporarily valid URLs
- Further automatic involvement of the proxy

→ Client anonymity

- JANUS/Rewebber in the World Wide Web -

JANUS/Rewebber

is accessible via

http://www.rewebber.com/

and

https://www.rewebber.com/

- Handling of Misuse -

JANUS/Rewebber keeps a black list of URLs about content providers not admitted.

Trying to access such a web page via JANUS/Rewebber, a client gets a denying reaction: