

# What do they know about me?

[Ius mentis](#)

**ABSTRACT:**The Internet is quite anonymous. There are so many people, that one particular person can hardly be noticed between all the visitors, and most providers take adequate measures to protect their customer's personal data, so that an outsider can't even retrieve their real name. Great for the user who wants to surf some sites nice and quietly, but a big problem for advertisers. How can you advertise if you don't know who your target group is?

## Cookies

The first problem that occurs is that visitors of a web site are unrecognizable. All the server knows is that a particular computer requested a particular page and then a number of images belonging to that page. It is not possible, however, to connect the address of the computer to one particular visitor. Every time the user dials in to his provider he receives a different IP address.

Cookies are a way to solve this problem. A cookie is a piece of information that a server sends to a browser in response to the first request. The browser sends the cookie back with every subsequent request. The server can then make use of the information in the cookie, and optionally send back a new cookie. If the server puts a number in the cookie that is unique for every visitor, the server can then check all page request for that particular number and thereby verify when that particular visitor came back. This way the visitor's steps can be tracked and a profile can be built.

Cookies can have certain advantages. For example a bookstore can connect its customers' address information to particular cookies, so that the customers no longer have to reenter this data with every new order. A news site can determine using a cookie which articles a particular visitor would like to read or not.

Cookies also have their disadvantages. If several sites work together, they can combine their databases and construct a more extensive profile regarding their common visitors. There is a partial protection against this: a cookie can only be sent back to a server within the same domain as the server that originally sent it to the browser. However, if the cookie is sent along with an advertising banner, it is the advertiser who gets the cookie back, and so he is able to construct a profile based on the Web sites visited by these users (if these sites feature his advertisements). *bijstaan*).

In most browsers you can disable accepting cookies, or indicate on a per-cookie basis to accept or reject a cookie. Another solution is to install a firewall or a dedicated "cookie cutter" program that blocks all cookies.

# Scripts

After the introduction of JavaScript it became possible to add interactivity to a web page or to make it do certain things itself. The page can now for example verify whether a form has been filled in completely (EG by checking whether the data entered in the phone number field does not contain any letters), but also automatically submit a form. It is also possible to have a form on a web page which upon submission is sent to a particular e-mail address. These two features in combination comprise a large security hole: if a page directly and automatically submits an (empty) form to a particular e-mail address, the recipient of the form then knows the e-mail address of every visitor. He can then send advertisements to those addresses.

While this bug was fixed quickly, many variations and other tricks to obtain visitor e-mail addresses were devised. Additionally, a security bug made it possible to read particular files on the hard disk, assuming their path and filename were known in advance. This way files in a fixed location (like the Netscape configuration file, or the password file for UNIX systems) could thus automatically be sent by e-mail to any particular destination.

Using JavaScript one can also find out what screen resolution a visitor uses, how many colors are supported and other system information. This can be useful when trying to create an adaptive web page, but this information can also be transmitted to a server to extend the profile of the visitor.

There is not much one can do against such malicious scripts. The only real solution is to disable JavaScript or VBScript in the browser. This does mean that a number of sites becomes less usable, because those sites rely on JavaScripts for navigation or required functionality.

# Headers

When the user enters a URL or clicks on a hyperlink, the browser connects to the indicated server and requests a copy of the selected page. In addition, the browser transmits a collection of other information, such as the name and version number of the browser, the operating system on which it is running and the page in which the hyperlink was present. This way to server already has some nice pieces of information regarding the user. Combined with cookies for user tracking this can result in a reasonably accurate profile for the user.

This can be prevented by using a proxy. A proxy is a program that requests pages and passes them on to the browser. The proxy can be configured in such a way that it does not pass on any information other than what is strictly necessary. A server can then only learn that a particular proxy requested a particular page, but it receives no cookies or information regarding the user of a browser behind the proxy. The Anonymizer is the most well-known proxy. It is also possible to install a proxy on your own PC. This is usually faster than a proxy located in an entirely different location. A disadvantage is that the server now still learns the IP address of the user.

Many providers also offer their own proxy server. This way they prevent that all users have to individually request the same information from the same server, which wastes time and bandwidth. The provider proxy retrieves the page once and then passes it on to every user who needs it. Such proxies are usually only meant to save time and bandwidth, and they do not filter any information.

## Spam

Anyone who has ever posted something on Usenet, or who listed his e-mail address on his web page, will quickly receive unwanted commercial e-mail. Usually they concern things like "Get rich in three weeks!!!!" or making cheap long distance calls in the United States (very useful for Dutch recipients). The senders of such electronic junk mail or spam messages construct their address lists by simply collecting all e-mail addresses they come across. Retrieving e-mail addresses from postings on Usenet is very simple. One can also simply download a number of web pages and extract e-mail addresses therefrom.

Traditional direct marketing tries to contact people who might be interested in a particular product or service, based on a profile. Sending a paper brochure is quite expensive, so it makes sense to first evaluate whether a particular recipient is a potential customer. E-mail on the other hand is free, so there is no reason why you shouldn't simply send your e-mail brochure to every e-mail address. This is the main reason why "spamming" has become so popular. A spammer rarely worries about the profile of the recipients, despite subject lines such as "I saw your webpage and I thought..." that seems to suggest the opposite. These texts are only meant to trick you into reading the message.

Some people add a "spam block" to their e-mail address, which is a piece of text that renders the address invalid. Humans can remove this text and e-mail them in the usual way, but the automatic programs that harvest e-mail addresses are not that smart. This also has some disadvantages:

somebody else can't just reply to a message, but always has to manually edit the e-mail address. Additionally messages sent to the invalid address are undeliverable and have to be handled as such by the mail system. This needlessly puts a higher load on mail servers.

A warning to Linux-users: by default many Linux installations start a mail server in the background. This makes it possible for third parties to relay e-mail via this server. The final recipient of the e-mail then thinks that the message came from the relay. Complaints regarding such spam then end up with the administrator of the relay, rather than with the spammer. To prevent this, relaying should be disabled in the mail server - or even better, the mail server should be disabled completely.