Some Special Fields of Data Security

Keszthelyi András

Keszthelyi.Andras@kgk.bmf.hu

Our age is called information age which shows the importance of any kind of data. Because of this one should protect their (digital) data to prevent non-wanted entities accessing them. This is a permanent problem of enterprises and private persons as well. The technical prerequisites are given not only for eavesdropping the data of others but for securing our data as well. In this paper I show a possible and a low cost technical solution for a typical situation: how one can hide the digital footprints of his or her browsing the internet.

SSH, portforward, privacy, data security

1 Information Age

Our age, the present economic era is called Information Age. This name shows the global economy's movement in focus away from the *production of physical goods* (industrial age) toward the *manipulation of information*. This means somehow that information is the first-of-all basics of our everyday life. The words *data* and *information* are usually used as synonims. At this point it is not only interesting but very important as well do discuss what do these words mean: data, information, knowledge.

Data can be any kind of statements which can be interpreted. It can appear in very different forms such as text, pictures, numbers, sequence of characters and/or numbers, etc. and can be stored in/on any kind of data holder, written on paper, painted on canvas, etc., or in most cases digitally in a data storage equipment of a computer.

The *information* is none other than the result of the interpretation of some data, the meaning of the data for a given person. It is possible that different persons assign different meanings to the same data depending on their own knowledge and experiences they collected previously. The gathered information(s) may become *knowledge*. Data is quantitavie while information is qualitative.

Data security has at least two meanings. The first one is ensuring that our data is stored safe from corruption. The second one is that access to it is controlled by the

owner. Both fields are huge enough even to recite the possible problems and solutions.

Our data is an important part of our working or basic capital. According to the above it can be stated that we must protect our data first of all as we try to protect any part of our capital. Because we cannot control if anyone else extracts any information from our data after (s)he had got the data itself. We only can control the access to our data.

Data can be classified into two main groups. Once we can speak about *static* data which means that the data is stored locally under strict physical supervision. It means that nobody can get neither locally nor remote (i.e. *any*) access to our data storage equipments without our permission. On the other hand we can speak about *dynamic* data or better to say *data traffic* between different computers via the internet.

Below some special problems of controlling the access to our data traffic is discussed.

Possible Threatens

The internet and any parts of it should always be considered as an untrusted network. Data traffic can be eavesdropped at any router or gateway machine. This is possible because of all the traditional communication protocols (e.g. sending and receiving emails, ftp, web-browsing) are plaintext-based ones. So the technical possibility is given.

Who can be interested in eavesdropping against us? There are too many answers. We can be threatened by an individual spy or data phisher or cracker who is interested in getting explicitely our data. E.g. a released employer of us wants to make a revenge on us or a cracker decides that it is our webshop from where (s)he gets all the credit cards data or one of our business competitors tries some sniffing, etc. On the other hand our ISP is in the technical position to collect too many data about us, about our data traffic, e.g. about our whole email or web browsing traffic which can be serious even if only the address and url data is collected. Other service providers can be interested in collecting and analysing (some parts of) our traffic data. Lets think of displaying personalized advertisements on the monitor of the users on the basis of their previously recorded searching keywords. Last but not least governments and government agencies should be mentioned here. E.g. it is known but rarely written or spoken about that the USA tries to eavesdrop the whole electronic data traffic of the whole world. Unless I am not very much mistaken I think that each government wants to know the most possible about the enterprises located in its country and the worse: about its own citizens.

In Hungary according to EU rules it is controled by the law what types of data the different service providers (ISPs, cellphone providers, etc.) should log and store in case of a later investigation against a customer of them. I think it is very interesting that there are some cases in which it has become known that the service providers logged and stored user data against the law.[1]

I described some aspects and methods of securing the data traffic flowing via the internet in a paper at the conference of MEB 2007. Those methods were possible solutions for the very basic communication situations between employees and their enterprise(s). At this point we should face with more serious situatoins which affect even the privacy of each of us because everybody leave too many footprints in the internet and those footprints can be saved easily and retrieved in the worst moment. I think that from this point of view there is not much difference between the situation of a person and an enterprise.

Technical Background

Each computer which is connected to the internet must have a unique identification number. This is the so-called IP-address. It can be permanent or dynamically assigned. In the case of home computers and ADSL-like connections the latter is the usual. The IP-address identifies the computer in both cases as long as the dynamic IP-address assignements are logged for further use.

Data traffic between two different computers needs their IP-addresses so. One of the two computers is called client the other is called server. Client computers send "questions" and server computers sends "answers" backwards. These IP-addrresses can be monitored and logged at any gateway or router computer of the six or sixteen ones between the client and the server.

E.g. the client machine (IP-address: 91.139.35.50 on Oct 17, 2007 midday.) of the user sends the http request "GET index.php HTTP/1.0" to the server named www.bombagyar.hu (IP-address: 195.228.74.223). As an answer the server machine sends and html file to the client machine on which the running web browser displays the initial page of www.bombagyar.hu. In the present case the route was the following: 43-002.vivanet.hu (217.173.43.2); 43-014.vivanet.hu

[1] The so called Zsanett case in which the cell data of the cellphones of the accused policemen were available even after some weeks when cell data should only be logged in case of starting a call. The known blogger Tomcat was cleard up from an accusation on the basis of the cell data of the critical period even if he had not started any calls. In the case of a policeman accused by murder the court asks seven year old cell data. It is also known at least at some cases when Google and Yahoo collaborated with the Chinese government to identify chinese citizens. See e.g. the link below: http://www.sg.hu/cikkek/49877/google_yahoo_microsoft_kozosen_az_emberi_jogokert

(217.173.43.14); GE-1-5.core0.iszee.hu (88.151.88.13); 81.183.245.21 (81.183.245.21); 84.1.104.138 (84.1.104.138); sparta.freedom.hu (195.228.74.223). So at least six computers could have logged (or logged) the fact that on Oct. 17 somebody from the 91.139.35.50 IP-address red the initial page of Tomcat's blog. Among these six computers the first one is always the same: the gateway of the ISP the others may vary.

Of course theoretically it cannot be proved who was sitting in front of the computer at the given moment nor that the page was red nor even the page was displayed. Of course if all the http access log data of a one year long period is analysed it will be clear that the given computer (mine) asked for bombagyar.hu only once for the above example or regularly.

How to Secure Our Privacy

The most important and frequent activities in the internet are sending and receiving emails and web browsing. Both emailing and browsing provides very sensitive personal data if it can be logged. If one sends and receives only encrypted emails even in this case deductions can be made of his/her activities and preferences or at least his or her personal connections can be traced because only the email body can be encrypted the fields of the sender and of the addressee cannot. Based on the log of one's visited url-s and searched keywords a more complete personal profile can be described, including very sensitive information about one's political opinions, possible diseases, sexual interest, human relations, etc.

So what we needed is some kind of anonymity for web browsing. One may think that (s)he would like to hide his or her IP-address. It cannot be altered on the client machine by the user because if you do so your computer will not fit into the network and the client-server dialogue will not be able to be established. This conflict can be solved by using an http proxy server which acts as an interpreter between the original client and the original server. The connection between the client machine and the proxy server can, or better to say: must be encrypted. Of course the proxy server should be a trusted computer.

In this case the ISP (or the other intermediate computers) can only log the fact of the encrypted data transfer between the client and the proxy machine. In order to know more one should have both the ISP's logs and the logs of the outgoing data traffic of the proxy. It can be very hard if the proxy server is placed abroad or at least has a big data traffic on other demands or more than one proxy is used.

The basic rules are the following:

- the proxy must be trusted, i.e. the user can be sure that no unwanted logs are made or at least the possible logs are not in the reach of unwanted entities;
- the proxy must not be in the reach of the ISP of the user or in the reach of any entity which can get access to the ISP's logs.

Additional rules can be given:

- fake data traffic can be generated to different directions from the proxy server while fake data traffic is generated between the proxy and the user's computer;
- more than one user might use the proxy in order to make harder to any possible traffic matching experiments.

Technical Solution

An http proxy software must be installed on the proxy computer, e.g. Ffproxy. This software is responsible for the redirection of the http traffic of the browsing of the web. Ffproxy is a filtering HTTP/HTTPS proxy server. It is able to filter by host, URL, and header. Custom header entries can be filtered and added. Using another auxiliary proxy server is also supported.

OpenSSH is also required. It is a genuine software tool for improving the security of data transfer between computers. It identifies not only the communicating computers but the users as well.

SSH gives us three very important security services:

- **Authentication** based on two different methods which can be combined. Public key and password can be used. A public key is something which the user must *have*, a password is a phrase he or she must *know* to prove his or her identity digitally.
- **Encryption**. SSH ciphers the whole data stream by industry standard cryptographic algorithms such as Blowfish or AES (or some others).
- **Data integrity**. SSH guarantees the integrity of the data transferred over the insecure network by signing it digitally.

SSH makes it possible for users to:

- log in to a remote computer to run programs;
- transfer files between the local and remote computers in a secure way (scp);

- access remote network services in a secure way as if a VPN service were used (port forwarding).

The client configuration rules are the following. Web browser(s) must be configured to use a proxy server which the browser can connect to at a given *local* port, e.g. port 8888. In order to have the *remote* proxy at the *local* port 8888. SSH port forwarding is needed. SSH can redirect the data traffic between two computers. A command like the following

SSH -f -N -L 8888:proxy.server.valahol:8080 proxy.server.valahol

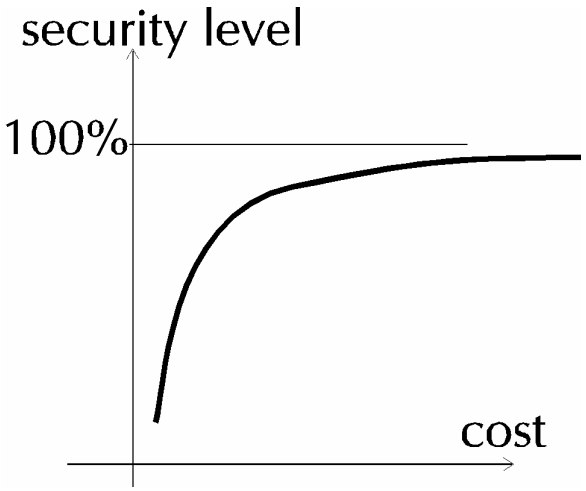
connects the local port 8888 with port 8080 of the proxy server. This connection is encoded by an algorythm which is considered strong. In the above example it is supposed that the usernames and user keys are correctly set on both machines.

Proxy server must be configured to listen to local (i.e. local to the proxy server itself) port 8080. Additional proxy configuration settings might be set in order to change some http request header elements, e.g.

Security Level

It is an important thing to know that a level of 100,00% of security can be reached in no circumstances. The security level of 100% can only be drawn nearer and nearer. The very first thing is to make a risk analyses. One must face the danger, one must determine the possible threatens as accurate as possible.

The other important thing to know is that the cost-securitylevel funcion is like in the figure below:



So an optimum is needed which depends on the special circumstances of the given person, of his/her position, and, of course, which depends on the (possible) interests of the other side.

And what about emailing? The simplest method is to use a free emailer from abroad via our http proxy, in addition we can use PGP or GPG to encrypt the email bodies if it is considered to be necessary.

The most important two things to know are the following: a) security level 100% cannot be reached; b) if a national security agency or if a mafia decides to get the knowlede of our browsing and emailing activities it will be able to get it in some way.

References

[1] Dwivedi, Hirmanshu: SSH a gyakorlatban. Módszerek biztonságos hálózati kapcsolatok kialakítására. Kiskapu, 2004.

[2] Gagné, Marcel: Linux-rendszerfelügyelet. Kiskapu, 2002.

[3] Flickenger, Rob: Linux Server Hacks. O'Reilly & Associates, Inc., 2003.

[4] Hagen, Bill von – Jones, Brian K.: Linux Server Hacks, Volume Two. O'Reilly Media, Inc., 2006.

[5] www.openSSH.com

[6] Nagy-Britannia: Átlátnának a járókelők ruháin. Minden emberről adatbank készül? http://www.mno.hu/index.mno?cikk=394378&rvt=3