Anonymity in Campus Networks

Best Practice Document

Produced by CESNET led working group
on Service support and CSIRT
(CBPD140)

Aleš Padrta, CESNET
March 2012

Document No: GN3-NA3-T4-CBPD140
Version / date: March 26, 2012
Original language : EN
Original title: “Anonymity in Campus Network”
Original version / date: V1.0 of March 26, 2012
Contact: apadrta@civ.zcu.cz

CESNET bears responsibility for the content of this document. The work has been carried out by a CESNET led working group on Service support and CSIRT as part of a joint-venture project within the HE sector in Czech Republic.

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 238875, relating to the project 'Multi-Gigabit European Research and Education Network and Associated Services (GN3)'.

Table of Contents

Executive Summary

The article is focused on the anonymity of university computer networks. Following a general introduction we will discuss the negative aspects of anonymity for the university's reputation, the solving of security incidents and fulfilment of some obligations. The primary problem with anonymity is when some offence leaves footprints on the internet leading back to the university. We will also offer some recommendations how to set up anonymity in the university's computer network from a technical point of view, but also legislative, and how to best deliver such information to the user.

1 Introduction

Anonymity conceals the identity of the person, institution or other entity connected to a network. The ability to hide an entity's identity is facilitated by pooling them with other identities within a network so that the individual entities are not distinguishable from one another.

On closer examination it is apparent that anonymity does not always need to have the same degree of strength, because the more entities that are pooled together, the stronger is the overall anonymity. An entity loses its anonymity the moment they become distinguishable from the others - meaning they can be identified.

Every subject leaves some traces by its presence and movement within a defined environment, be it physical or virtual, or on the internet. And the longer this entity operates within that environment the more traces (or "footprints") they leave behind them, revealing more information about them. Such revealed information can be used to differentiate the individual entities and subsequently weaken overall anonymity. In some cases the traces can be analysed to exactly identify the entity.

2 Digital Footprints and Anonymity

Every use of information technology creates so-called digital footprints. Some of these footprints arise directly due to the technical nature how a network and its individual services function, while many other footprints are left behind by the user voluntarily [1].

Contrary to traces or marks left behind in the physical world, digital footprints pose several unpleasant characteristics when trying to stay anonymous. One major problem is their durability: such footprints can be easily saved in digital format and it is not possible to guarantee they will be erased after a certain period. Even after a footprint has been erased from its original location, it can remain within a browser's cache, a backup media or other location. Another problem is the relative ease with which some information can be accessed by the greater public, or indirectly through other means. Searching and comparing information in digital format is a simple matter, because pooling together the dataset can portray a general image of a group of entities [2].

2.1 Connection to the Internet

The first area where valuable information is revealed and which can compromise anonymity is the connection to the internet and the network communication. The problem is that every network operator requires a certain amount of network data in order to monitor the system's proper functionality (or to optimise it) and assure its usability, including security incidents mitigation. The collection of such data can significantly compromise anonymity [2].

One fundamental clue which can identify an entity is its IP address. As is generally understood, IP addresses are assigned across the global network in a hierarchal manner, where individual IP blocks are first assigned among RIR (Regional Internet Registries) and then distributed among LIR (Local Internet Registries), who in turn assign IP addresses to their individual clients (typically ISPs or larger organisations). Information concerning the assigning of IP blocks to organisations are publicly accessible and anyone can determine the owner (refer to Table 1).

Regional Internet Registry	Webpage where to find detailed information
AfriNIC	http://www.afrinic.net/whoisdb/whois-afr.htm
APNIC	http://wq.apnic.net/apnic-bin/whois.pl
ARIN	http://whois.arin.net/ui
LACNIC	http://lacnic.net/cgi-bin/lacnic/whois
RIPE NCC	https://apps.db.ripe.net/search/query.html

Table 1: RIR overview with their public registers of assigned IP address blocks.

Each organisation then assign individual IP addresses to their end users, although sometimes blocks of IP addresses are first divided among different departments or branches etc. of the organisation, which could also be assigned direct central authority. Besides the assigning of an address it is also standard procedure to

maintain detailed information concerning the use of an IP address over time. Each organisation is theoretically capable of determining which exact device used a specific IP address at any moment. It is then a simple matter to determine who the local user is, based on their IP address, be it due to their request to register their equipment and consequent assignment of an IP address, or by direct authentication via 802.1X protocol.

Another source of valuable information, which may compromise anonymity, is user communication. Even if we exclude the transferred content, the source and target IP address, the time of communication, and the amount of transferred data can be used to identify the users. Such as their specific behaviour, the usage of different web services, etc. Furthermore, such information is accessible not only to organisations the end points of communication belong to, but to all owners of the infrastructure involved in the data transfer concerned.

The primary sources of operational data are data (net)flows, IDS (Intrusion Detection Systems) reports on forbidden or abnormal activities, and access to interesting or important services at the transport layer of ISO/OSI model.

Excluding statistical reasons, most operating data are usable for a relatively short period of time (days or weeks), because their use for resolving current problems or for planning decreases over time. Secondly, the longer such data are stored the more space they require to store. Another aspect is legislative, as different countries have different minimum times for storing operating data (in the Czech Republic the period is 6 months for internet service providers [ISPs]). So, as you can see, these digital footprints can remain for quite a while once they have been created.

2.2 Use of Services

Basically, all internet users use many of the services it offers. This may concern access to web pages, information system servers, terminal services and many others. Accessing these external services gives rise to number of digital footprints which may compromise anonymity, if only for the reason that such services require a certain amount of user information in order to function properly.

First of all, the user's IP address is always known, which, as explained in the previous section, automatically reveals who exactly is using the service. The use of a third party service often also requires that the user is authenticated, forcing the user to reveal its identity right then and there.

Furthermore, user's requests and actions can also be stored, such as any comments they might add to a forum or what kinds of files they download. There are also audit records form important information systems which allow user activities reconstruction, which consequently lead to lower anonymity.

Less known digital footprints are supplementary information which the client automatically sends to the servers. Originally, such information was meant exclusively to optimise the provided service or fine-tune problematic versions of the client, for which reason email clients and web browsers always send information relating to their version, the operating system version, resolution of the monitor and so on. The website http://browserspy.dk provides a relatively detailed picture of how different web browsers behave and exactly what information each browser automatically provides. Not all users use the same web browser and same configuration, in which case such information may be used to differentiate users and subsequently compromise anonymity.

2.3 Information Made Public Voluntarily

The digital footprints described in the previous section arise due to the technical aspects of how computer networks operate, and are therefore very difficult to conceal. But much of such data are not public, while the internet providers store them for some time and may release them to third parties by court order.

However, users often voluntarily leave behind digital footprints that are much more accessible to the public and that result in an accordingly greater compromise of anonymity [1][3]. Considering the personal information that users regularly post on blogs, chat sessions, dating websites, discussion forums and social networks, we can see quite a lot is being revealed. Starting with their full name, interests and opinions, right up to their daily

schedule, where they live and details of their work. In this way they divulge valuable personal information to almost anyone, as such revealing their identity. Such quality information as who they know within social networks, photographs about themselves, and other personal information.

Typical users is not aware of the fact, that personal information is made public, who all will have access to it and to whom else such information may be automatically sent. Such publicly accessible information is usually enough to identify who is behind a nickname used on blogs or social networks.

If someone were able to obtain information from all the mentioned parts, meaning from the internet connection provider (ISP), the services provider and from the individual user, it would not be a problem to identify all their activities, or the user behind any activity.

2.4 Pseudonymity

While surfing through cyberspace, although the digital identity of a particular user may be known, their mapping to an actual person is not. An example would be a user with a particular nickname on a discussion forum or social network. The other users know they are corresponding with the same person (or group of users in some cases), but they do not know exactly who that is. Such a user creates its own history and reputation, contrary to an entirely anonymous contributor. When a user hides behind a nickname (pseudonym) it is referred to as pseudonymity [4]. One can think of it as half way to anonymity.

One's pseudonymity automatically surfaces during the first phase of identifying the actual user when available digital footprints are selectively filtered to match a particular interesting entity. In other words, corresponding digital footprints can be compiled from different sources in cyberspace to give an indication of the true identity behind a pseudonym. At present social networks divulge such information the most.

2.5 Relation to Anonymity

The need or appropriateness of one's anonymity depends on the user's role in cyberspace. A regular user will have different requirements than, for example, an ISP or bodies involved in criminal proceedings.

A regular user would prefer total anonymity (not only within a computer network), because he gets the access, but cannot be prosecuted for breach of any of the site's regulations. From a psychological perspective an anonymous user is less hindered in their cyber activities than one who is aware that their identity may be revealed.

It is in the interest of commercial ISPs to make sure they are charging the right services to the right users. Other information requires additional effort to collect and store them, but do not pose any gain to the provider. On the other hand, the legislation can define the mandatory set of operating data and the duration of its storage. For this reasons a commercial provider is indifferent to the anonymity of paid-for web services.

Universities are not usually understood as public ISPs, for which reasons less restrictions are imposed on them, i.e. there are no duties on storage of operating data (this is at least true in the Czech Republic). On the other hand, this status can lead to unpleasant consequences. While commercial providers can easily distance themselves from any user involved in illegal activities, this becomes more sensitive for universities because it can shed poor light on their reputation. Furthermore, since there is a high fluctuation in the user base, in particular among the student body, it is difficult to increase security awareness within the academic environment. This results in irresponsible behaviour among some of the users, illegal activities, and security incidents. To maintain the university's good reputation it becomes important to properly investigate serious security incidences and implement corresponding disciplinary action. This is simply impossible without user identification or in other words the total dismantling of anonymity. From this perspective user anonymity within a campus network is not perceived negative [5].

3 Recommendations for a Campus Network

As it was explained in the previous section, on a global level the university should eliminate anonymity from its campus network. The main reason for this is to preserve a good reputation since security incidents incurred by its own users could negatively affect it, in particular if they are publicly visible. The absence of anonymity helps two primary areas [5]:

- Users who are aware that their actions are not anonymous will tend to behave more responsibly. Simply because they are aware that they could suffer consequences for their actions. On the flip side of the coin, an anonymous user is less concerned about violating the law or the university's internal regulations.
- If security incidents are to be fully resolved, interaction will be required with the responsible user. Either to find and reprimand the culprit, or to educate the user how to avoid such an incident in the future. Either way, the identity of the user must be determined.

For this to be possible the university must become involved in three fundamental areas. The first is to communicate with the users in order to inform them about the anonymity issue and explain it. The second is to adopt certain technical measures, such as to collect the operating data required to identify a user. And third to convey the regulative end of it, such as the internal guidelines and operating codes that all should be aware of, at least on a basic level.

3.1 Communication with the User

The university's stance against anonymity should be made known and regularly reminded to all users in an appropriate manner. One typical method is through IT training, such as an introductory course when first arriving to the university, regular seminars focusing on IT services or, ideally, regular required training on how to use the university's computer system. These seminars should include a presentation of anonymity within a computer network, instruct users as to what information on what activities are readily available to the university and, if possible, mention a case whereby culprits were pursued. Table 2 shows the primary presentation subjects and a brief description of them.

Such training sessions should improve user awareness concerning their anonymity within the campus network and concerning the internet and computer networks in general. This will make them act more responsibly within the campus network, which is the primary goal, but it could also help them behave better in their personal lives, at least what concerns the making public of their personal information and their trust in the internet's anonymity.

Subject	More detailed description
The concept of anonymity	Defining the concept and emphasising the different strengths of anonymity.
The concept of digital footprints	Defining the concept and its problematic characteristics relating to anonymity (accessibility, duration, ability to aggregate)
Anonymity of internet connection	Relative anonymity of IP address – assigning of IP address, publicly accessible information and IP address anonymity within the campus network

Anonymity of connection to the campus network	Anonymity of university-assigned IP address – show what type of operating data are stored and how they can be used to determine user identity behind an IP address
Anonymity of services	Operating data of services provider – information required for operation, additional information of clients and practical data (such as web requests, email headers and information provided by the browser)
P2P Anonymity	The principle of P2P operation with reference to public knowledge of individual node IP addresses and the previously mentioned anonymity of connection to the campus network. Determining the source of copyright material is not difficult.
Anonymizers	The principle of functioning – masking IP addresses, warning of output node problem (providing a third party with access to the network is usually forbidden), other persons' problems accounted to a user
Information Made Public Voluntarily	Users can be exhibitionists, show examples of what they reveal on their blogs or social networks, mapping digital identities to real users, the easy access of such information and the length of time it is stored, and the correlation among them.
Conclusion	The internet is not anonymous – depending on the ISP, provider of services and individual users, i.e. nobody can be relied upon anonymity. It is better to act on the internet as if you were fully identifiable.

Table 2: Presentation topics.

3.2 Nonanonymous Assignment of IP Addresses

The basic technical measures related to the assigning of IP addresses from the campus network and whose purpose is to determine who was using which campus IP address at any given time. By knowing these three pieces of data – IP address, time and user – the necessary information can be compiled [6].

To acquire this information it is first necessary to have the connection of all equipment to the network properly under control. From a practical perspective, it is possible to allow connection access by two methods – based on the user's digital identity, or registration of the equipment by the user.

The use of the 802.1X protocol is the easiest method of user identification, as it allows unlimited communication after authentication only. In the academic environment, the eduroam wireless network works this way. To reverse search the user identity for a specific IP address the logs from the radius server, DHCP server and user register are required. With this information the user name (including the home organisation if using an eduroam network) assigned to a particular user can be determined.

Another reasonable option is to assign IP addresses by central authority based on request, whereby the user and the computer's MAC address were submitted, often together with the required domain name. If the request is submitted on behalf of a third person – such as when dedicated personnel are assigned to administer the IT of particular departments and who register on behalf of computer illiterate employees – the nature of the request should be mentioned. Ideally, all these requests should be recorded and serviced within a system of administering requests (such as Request Tracker or OTRS) and where the required information could be easily determined.

It is not advisable to allow the user to connect itself to the computer network or to choose its own, actually available IP address. This would not be ideal for other reasons than anonymity – in particular with a conflict of IP addresses which may result if at some point the central authority assigns the already "squatted" address. It is also recommended that the used IP addresses have direct and reverse records within the DNS. This works by assigning IP addresses from DHCP while using Dynamic ARP Inspection and DHCP snooping, which only permit connections to equipment assigned an IP address from a specific DHCP server. For older equipment,

which does not support DHCP, must be set with the relevant exceptions on the relating network components. The process of registering equipment into the network also depends on reconfiguring the DNS and DHCP servers and the possible reconfiguring of some network components.

To assign IP addresses nonanonymously the network must be set up accordingly while operating data from the DHCP server, radius server registration system and user administration system (Identity Management – which maps the login to a physical person) should be stored for a sufficiently long period. Table 3 shows an overview of the systems with the data they provide and their recommended storage times.

Information source	Provided data	Comment
DHCP server	Time – IP – MAC	Each assigned IP address is stored
Administration of registrations	User – MAC (+ assigned IP)	One-off record in system of administering requests
Radius server	Time – MAC/IP – login	Each assigned IP address is stored
Identity Management	User – login	One time storage of each user

Table 3: Overview of data sources for the nonanonymous assigning of IP addresses

3.3 Network Traffic Monitoring

Monitoring the network's operation provides relatively important log data (data flow). These are not records of all transferred data, which would not be technically feasible considering its volume, while some degree of privacy should be applied to the transferred information in general. The transferred content is not entirely stored but only that information relating to the flow, such as the source or end IP address, the source or end port, the protocol, volume of transferred data and flags. This information does not serve to directly identify the user but can complete a picture of its behaviour and help to verify past events [6].

As was stated earlier, even while excluding the transferred data itself, the remaining volume of data is still quite large, making its storage in entirety problematic. For this reason all the information is stored on a short-term basis, to the order of several days, after which the less important data is gradually erased from the system. Some records can be aggregated when erasing differentiated data, which greatly reduces the required storage volume. Table 4 shows an appropriate setting when erasing data.

Storage period	IP addresses	Volume of data	Protocol	Ports	TCP flags
3 days	YES	YES	YES	YES	YES
2 weeks	YES	YES	YES	YES	
1-6 months	YES	YES			

Table 4: Aggregation of data flows

3.4 Nonanonymous Access to Services

Another area which should be examined is access to services, in particular those which may cause problems to the university.

At first, we should mention those services which enable third party access to IP addresses, typical access to terminal servers, or public computer rooms on campus. This is similar to the previous section, only that the user does not use its own equipment but is able to perform the same functions as if so.

There also exist services which may be abused to create problems for the university. Examples of this are data storage locations (by uploading copyrighted material) or discussion forums on the university's official pages (by publicly posting illegal comments). Internal services could also be used inappropriately, such as critical

information systems (the curriculum or financial matters), whether it be intentional or not. Even in this case do we recommend being able to identify the source of inappropriate activity.

To deanonymise provided services, additional operating data, which create the service-time-user connection, must be stored. For more detailed logging a more detailed event-time-user level can be used. Every organisation has its own infrastructure specifics, for which reason it is not possible to describe in detail from where exactly the required data may be drawn. Nevertheless, Table 5 shows an overview of which information should be stored.

Information source	Provided information	Comment
Authenticating system	service – time – login	Each connection to the service
IS auditing records	activity – time – login	Every activity within the system

Table 5: Overview of information sources for nonanonymous use of services

3.5 Operating Data and its Storage

What concerns the time period for storing data, if the minimum time is not strictly dictated by law, regular operating data should be stored over a period of several months. Deeper history is rarely required. An exception to this rule might be selected audit records of key information systems which may need to be stored over a longer period. Such as a financial audit which occurs in a period of one year, where an overview of user activities would be stored over the entire year in order to understand the source of any discrepancies. In the case of an educational institution the duration of storage would relate to the length of study. Table 6 shows an overview of recommended storage times for different types of data.

Information source	Storage period
DHCP server	6 months
Radius server	6 months
Administration of registrations	6 months (considering the frequent connection of the system to request management, the storage period is virtually unlimited)
Identity Management	The duration of the identity's existence + 6 months
Authenticating system	6 months
IS auditing records	6 months, or longer for critical IS, depending on the nature of stored data
Data flows within the network	6 months (one week for all and then only aggregated selections)

Table 6: Recommended period for storing data

Local legislation should be considered in this case as well. For example, in the Czech Republic, public providers are required to store operating data for at least six months. Even though universities do not fall under this category, it is still recommended to strive for these requirements voluntarily.

If the storage room is available there is no need to quickly delete older data as they can always be used for statistical purposes. In these matters we should also mention legislative pressure to store personal data only as long as is required, for which reason the user's personal information (name, residence, date of birth, residence etc.) cannot be kept indefinitely.

3.6 Stance to Anonymous Networks

As was previously stated, users prefer to maintain anonymity, for which reason they may resort to such anonymous networks as TOR, I2P or freenet [7]. Such networks are created by volunteers whose equipment form a group of nodes used to make the network communication anonymous. Data always transfers across several nodes, whereby each one only knows the previous or subsequent node. This makes it longer and more

difficult to identify the user, because the opposite direction is used for searching the source of communication, and if the IP addresses are drawn from all over the world, the delays may result in the loss of necessary operating data.

The university may experience core problems if a user with IP address falling within the university's dedicated range connects to one of the anonymous networks, such that the connection functions as the last communication node (referred to as output node). In such a case the system's IP address is used by an entirely unknown person, making them anonymous and unidentifiable in case of need.

These arguments prove that an anonymous network is not ideal for universities, although those for scientific or study purposes might make an exception. Furthermore, the general operation of an anonymous network is often in breach of the ISP contract because the campus network is actually provided to a third party.

3.7 Legislation

Besides internal incentives like protection of one's name or reputation, a reduction in anonymity may also be required due to external sources. In most cases it is not required directly, but arises from other requirements specified under law or AUP of ISPs – this would concern an interest group such as CESNET or JANET within a campus environment. A typical example would be the requirement "not to provide connection access to the network to another legal entity or natural person", which is rather difficult to fulfil if nobody knows who is connected.

The organisation should issue internal guidelines which would explain to users exactly how they should use their services – define their rights and obligations. In this way the user would get an idea of set limits and avoid potential conflicts. This is usually accomplished with a document titled Rules for Network Use and should always include the following information:

- The user should be aware and approve that the network will be monitored in order to resolve security incidents. One such formulation might be: "The IT department reserves the right to monitor network traffic. In the event that it becomes necessary to check access rights to the network's sources or if the guidelines are breached in another manner, based on the IT head's decision an authorized employee could be assigned to monitor the behaviour of a particular user."
- The obligation of users not to provide connection to the network to other legal entities or natural persons, which is usually required as per the agreement with the ISP and which deal with the problems of an anonymous network or of providing an output node. This may be formulated as follows: "When operating within the network you must not provide access to the network or other services to legal entities or natural persons."
- The user's obligation to always operate with its own identity and not to try and hide it – so that users are responsible for their own actions. This may be formulated as follows “While operating within the network you must not use any software which would provide you with anonymity or use such software which could provide you with an alternate identity."

Such formulations are better served if also worded in a more laic manner than to only repeat the guidelines. This is best delivered through the already mentioned user training sessions, informative posters and other popular media channels.

4 Conclusion

The anonymity of a computer network is relative and can be gradually compromised by the electronic footprints which are left within it. Both ISPs and service providers have access to technical data, which could determine a user's identity. Furthermore, much personal information is often voluntarily provided by the user. By collecting enough of such information a user's identity may be revealed.

Network anonymity is not advisable for a university because it puts its reputation at high risk. It is caused by fact, that an anonymous user cannot be reprimanded for its actions. A second argument is that security incidents cannot be properly resolved without interaction with the user, whether it was a victim of attack or the source of the problem. The requirements of the ISP or the law, which cannot be fulfilled if the network allows the anonymity, should be also considered.

The basic recommendation is that the university apply technical measures preventing anonymous use of the campus network and web services provided. Emphasis should be on network security and the collection of necessary operating data which must be stored for a certain length of time. It will also be necessary to properly formulate the rules in the internal guidelines.

The guidelines must then be addressed to all users, as their awareness of the risks alone will cause them to act more responsibly. Some appropriate approaches include introductory user training, IT seminars, and other popular teaching tools. This form of education is not only in the university's interests but also those of the user, who will learn about the risks of internet anonymity and benefit from them beyond university life.

References

[1] Padrta, A.: “Anonymous me”, 2009, in Czech language
http://download.zcu.cz/public/Prezentace/seminar%20bezpecnost%20091202/ja_anonym.pdf

[2] Kácha, P.: “Anonymity in the Internet”, 2011, in Czech language
https://www.vsb.cz/share/uploadedfiles/public/9872/seminare/20110414/anonymita_v_internetu.pdf

[3] Schneier, B.: “Identifying People using Anonymous Social Networking Data“, 2009
http://www.schneier.com/blog/archives/2009/04/identifying_peo.html

[4] Schneier, B.: "Pseudonymity”, Schneier on Security, 2011
http://www.schneier.com/blog/archives/2011/08/pseudonymity.html

[5] Padrta, A.: “Campus Network Anonymity”, 2011, in Czech language
http://www.cesnet.cz/akce/2011/monitorovani-kampusovych-siti/p/padrta-anonymita.pdf

[6] Padrta, A., Orkáč, R., Mach, J.: “Log – Best Administrator Friend“, 2011
http://www.cesnet.cz/akce/2011/monitorovani-kampusovych-siti/p/padrta-logging.pdf

[7] Satrapa, P.: “Anonymous P2P networks - users strike back”, 2006, in Czech language
http://www.lupa.cz/clanky/anonymni-peer-to-peer-site/

Glossary

RIR	Regional Internet Registry
LIR	Local Internet Registry
ISP	Internet Service Provider
IDS	Intrusion Detection System
DHCP	Dynamic Host Configuration Protocol
DNS	Domain Name Service
MAC	Media Access Control
ARP	Address Resolution Protocol
TOR	The Onion Routing