

Traffic Data Retention

Impact on civil society organizations

George Danezis
University of Cambridge, Computer Laboratory

During the last two years a series of national and international legislative measures have been put in place to regulate access to traffic data by Law Enforcement Authorities and requiring telecommunications and Internet service providers to retain the patterns of communications going through the networks they control [1]. While the primary reasons given for such measures is the investigation of serious crime or terrorist activity partly perpetuated on-line, recently proposed legislation (later withdrawn) in the U.K. would have allowed a wide variety of government bodies access to this information [2]. Usually legislation, such as the U.K. RIP Act 2000 [3], is drafted in a way that implies that traffic data can be seized for any criminal investigation if it is “proportional” and “necessary”. Such data is primarily going to be used to trace communications, and accesses to on-line services, back to private individuals.

Traffic data generally fall into four main categories.

- *Subscriber data* provide a link between communication identifiers and a physical person. An LEA can use them as a phone book: given a network address or a phone number, the individual that owns it or the home it is connected to, along with other details that a provider might have can be retrieved.
- *Communication data* provide a trace of who has talked to whom. Usually they link together communication identifiers, along with additional information such as the time and length of the connection or its status.
- *Location data* provide information about where the two (or more) ends of a communication are physically located. This might include the GSM cell of a handset, or the base station of a wireless network connection.
- Finally, *Service or Usage data*, contain information about the use that has been made of the network. While not logging the content of communications, information about the pages accessed on a web server, or the addresses to which an email was sent, can be considered as usage data.

Traffic data is, from a technical point of view, an ill-defined term. Lobbying efforts in the U.K. have restricted the definition of traffic data included in the R.I.P. Act, only to include information identifying the physical piece of hardware used. Therefore only information about which physical server a user is accessing can be requested by LEAs. On the other hand, a recent document, leaked from EUROPOL, contains a wish list of traffic data to be seizable that encompasses all the log information that a web server would generate, down to the granularity of individual web pages [4]. The issue of granularity is therefore important.

While the difference of granularity of the information between the R.I.P. definition and the EUROPOL wish list can be seen as the difference between communication data and usage data, other issues are not so easy to distinguish. For example, the granularity with which one logs the size of the data transmitted during a connection makes a huge difference to the information that can be later extracted. If the size information is aggregated over a large period of time, very little information can be extracted. On the other hand, if the granularity of this information is fine enough, one can extract from the patterns of traffic, the type of service used, and even in some cases the exact web pages browsed [5]. This is a typical example illustrating that simple communications data, with appropriate processing can reveal information that would otherwise be classified as usage data.

The main questions about traffic data retention are:

- Is it going to be effective in achieving its goals of catching hard-core criminals?
- What are going to be the effects on the rights of other users?
- Is it worth it?

Technical comments on the first two questions are contained in my paper entitled “Comments on the E.U. Cyber-crime forum” [6]. The main line of argument in the paper is that the data retention measures will have limited effectiveness against people who are ready to commit illegal acts, such as stealing mobile phones or hacking routers. On the other hand, requiring operators to retain the data and serve requests, without full compensation, will create a private, highly efficient and effective infrastructure of surveillance. Additionally the measures cannot be effectively implemented on new communication paradigms such as wireless ad-hoc community networks, or peer-to-peer overlay networks. Either these technologies will render traceability impossible or they will have to be restricted.

Because comments about privacy as a communication right in the paper are vague, it is worth while as my contribution to the World Civil Society Forum to highlight concrete threats that civil society might face if the traffic data intercepted is misused. These threats come from misuse by the private holders of the data, rogue or corrupt elements in the LEAs or Security Services, or institutional abuse. While each of these threats out of context is dismissed as purely imaginary, or as showing bad faith, by the proponents of blanket traffic data retention, we will see that there are well documented precedents that should cause concern.

Data Protection legislation was brought into effect in order to limit in breadth and time the volume and processing that personal data, and in particular sensitive personal data such as religion, political opinions and sexual orientation, can be subject to. It can be argued that traffic data fall, in many cases, under the category of sensitive personal data since they can reveal information about the health condition or sexual orientation of an individual. In particular, usage data, such as access to on-line health information, or dating services are particularly revealing.

Furthermore traffic data are privacy sensitive because they can reveal private information that a user might not be aware of. The user is usually aware of the sensitive information contained and disclosed by content data, since it is explicitly communicated. The same is not true for traffic data, since the information contained in them is extracted by a context that the initial user might not be aware of. For example, while it is obvious to most users who someone that intercepts a mobile phone call can overhear all conversation, it is not obvious that access to traffic data also reveals the location of the caller. It is even less obvious that the times, duration and frequency of calls can reveal data about the relationship of the communicating parties, although nothing about it has been explicitly revealed during the conversation.

Traffic data of all sorts, already have a great amount of value because of the use one can make of them for direct marketing. Traffic data disclosure can also lead to individual embarrassment if they reveal practices that are not accepted as socially mainstream. The inability for the average person to model why their personal data is sensitive is often due to the fact that they do not perceive the threats to be more serious than leading to hassle or embarrassment. This inability is often due to the fact that there is a lack of perception of oneself as an actor in society with plans and hopes that, for tactical reasons, might not be appropriate to be disclosed, in particular to persons that might not share them.

From the point of view of the organized civil society the threats linked to marketing and minor embarrassment, that are often cited as the most important reason to protect one's private data, are only of secondary concern. By default, civil society organizations bring individuals together, along with their plans and hopes, with the clear aim to act upon society. Because it takes place in the social

sphere, their work can be broadly defined as “political” and the main concern that they have should be the usage of traffic data as a means of political surveillance and control.

Unfortunately the terms political surveillance and control do not seem less vague than the original term “privacy”. They are also overloaded by politically motivated tales of repression and half-accurate and unofficial accounts that make it difficult to assess the real threat to civil society. The difficulty in assessing the real threat leads to a near impossibility in protecting against them, without resorting to complete paranoia that would be more harmful than helpful to the social work of an organization. In order to remain as exact as possible, one can refer to the official U.S. Senate Church Committee Reports, in respect to domestic surveillance with political aims [7].

A typical example of surveillance for political aims is provided in the section C.2.(a) [7]:

The "Women's Liberation Movement" was infiltrated by informants who collected material about the movement's policies, leaders, and individual members. One report included the name of every woman who attended meetings, and another stated that each woman at a meeting had described "how she felt oppressed, sexually or otherwise". Another report concluded that the movement's purpose was to "free women from the humdrum existence of being only a wife and mother", but still recommended that the intelligence investigation should be continued.

It is clear that the information the FBI was after was lists of members and their degree of involvement in the organization. In 1976, when the final report was written, informants were necessary to perform this task (indeed 83% of the operations involved informants). Nowadays, thanks to the ubiquitous use of technology, access to traffic data would most probably be sufficient. If the announcement was sent by email list, the recipients could be traced. If the discussion was taking place on-line as a discussion group, the logs of the service could be requested, and finally, even if the meeting is taking place physically, 3rd generation mobile phones would give accurate enough location information to deduce who is present. In less than 25 years a few hours of work of an informant have been reduced to a few requests for traffic data. Furthermore, the request can be filed after the event takes place, for as long as the data is kept.

The question remains: what political use can be made of intelligence material, and by extension traffic data? According to the same U.S. Senate report the adverse impact of improper intelligence activity can be categorized as follows:

1. General efforts to discredit
2. Media Manipulation
3. Distorting Data to Influence Government Policy and Public Perceptions
4. “Chilling” First Amendment Rights
5. Preventing the Free Exchange of Ideas

Usual methods to perform the above were anonymous accusation letters, aiming at discrediting someone's personal, political, professional or academic life. Planted articles in the media have also been used. Interviews were conducted to enhance the feeling of paranoia and to induce the feeling that legitimate political dissent was questionable and being monitored. Meetings were prevented from taking place by lobbying the venues.

In order for the above operations to be successful in disrupting the work of the monitored organizations, the recipients of the letters or the targets of the accusations have to be appropriate. It is worrying that the traffic analysis of communication data can be used to perform the target selection in such a way that operations such as the above have maximal impact. Indeed, traffic data contain both the social and organizational blueprint of any association, including members, friends, leaders, hubs of information, and their relations. They also provide a complete blueprint of the relations between civil society organizations, and between particular organizations and the rest of

society.

It is tempting to dismiss the belief that such practices are still taking place today, less than 25 years after the final Church committee report, or even to believe that the will to disrupt civil society organizations is not anymore present in modern security services. Unfortunately, intelligence and secret services in many countries have not been the subject of such scrutiny, and no data is available to give an authoritative picture of their activities. While no certainties can be given, some reports that appear in the media a few times a year seem worrying.

“The same day [June 18, 1998] more than 10.000 people caused an estimated £2 million damage. They gained a reputation for violence which climaxed in reports, flatly denied, that RTS is stockpiling weapons in bunkers around the capital.” - Observer, Sunday November 28, 1999 (Sara Ryle, Nick Paton Walsh and Tony Thomson)

“Specialist firearms teams are being drafted in to police this year's May Day demonstrations in the City of London over fears that rioters armed with samurai swords and machetes will infiltrate the protest” - Observer, Sunday April 22, 2001 (Martin Bright and Frank Kane)

“In Manchester, we were booked to speak at the university. A few weeks ago, it canceled the booking. Having spoken to the police, it had decided we were “a potential security risk”. So the organizers hired a hall in the Co-op's headquarters instead. A fortnight ago, the security company running the hall annulled the contract on the advise of the police. [...]” - George Monbiot on the difficulties of organizing the “Globalise Resistance” tour 2001, Guardian, Thursday February 1, 2001

“Widespread collusion between the security forces and loyalist paramilitaries in Northern Ireland continued unchecked for years because a culture of “gross unprofessionalism and irresponsibility” allowed officers to create a climate in which Catholics could be murdered with near impunity, a comprehensive investigation has found.”- Rosie Cowan and Nick Hopkins, Guardian, Friday June 14, 2002

“Three weeks after they ordered Oakland police and the FBI to pay Earth First organizers \$4.4 million, jurors were allowed to speak for the first time Tuesday, and one of them said “investigators were lying so much it was insulting.” [...] “Bari, who died of cancer in 1997, and Cherney argued that the investigation, which has never cleared them as suspects, had undermined their credibility and hurt their ability to promote forest preservation.” - Cops, FBI lied about probe, juror says. Woman speaks out on Earth First trial after gag order lifted, Jim Herron Zamora, Chronicle Staff Writer, Wednesday, July 3, 2002, ©2002 San Francisco Chronicle.

The above do not provide any certainty that systematic abuses are taking place, or that traffic data is going to be a tool for political surveillance, but give rise to a legitimate need to protect sensitive aspects of an organization's structure. From a security engineering point of view, it is important to realize that too much protection, or even near paranoia, is by itself disruptive for an organization. If flows of information are arbitrarily restricted, “just in case” they are intercepted, or if people are not allowed to communicate, or when strict protocols have to be applied, the effectiveness of an organization is already severely diminished.

Therefore, before even considering imposing any restrictions, it is important to understand the environment and the nature of surveillance and then only act appropriately and in a proportionate fashion. Many organizations, such as the Association for Progressive Communications, have issued manuals about how to deal with surveillance [8]. Most of the measures proposed are concerned with surveillance awareness and safe information practices, and involve procedural, not technological solutions. Technological solutions, such as throwaway accounts, encryption, steganography and anonymous communications, can only be used to support and complement them.

Evading and managing traffic data surveillance

As highlighted above, by far the most important countermeasures against surveillance are of an organizational nature. The first priority is to educate all users of technology to be aware of the side information that is leaked by the technology they use. GSM phones leave billing information, lists of calls, links between the handset numbers and phone numbers, and the cells the phone was active in. Networked computers leave traces on each service they access on-line, and through the network they use. Even devices that are not obviously communicating can leave traces: automatic toll payment tags and other automobile identification schemes, credit card usage includes location information, etc. Attention should also be taken about the link between the traces left. While using a phone card and public phones might hide the identity of the caller, using it repetitively by phoning many numbers or from many locations reveals a great lot more information. As a general rule, a collection of data on a subject is worth more than the sum of its parts.

As soon as a consciousness of the dangers associated with technology exists, it is important for an organization to assess what effects they can have on its work. In particular, it is important to render explicit the nature and the significance of the high-level information leaked by traffic data. For example, could an analysis of phone logs reveal the plans to establish a new office in a particular town? Or can the phone logs reveal who is right now present in a particular region, if they are used for both business and work purposes? A careful assessment has to be made about the operational significance of revealing the information leaked. It is important not to adopt an over-zealous hiding policy since not using phones, cars, credit cards, computers could impede the effectiveness of an organization's operations. Instead it is more helpful to understand for each item of information that could be leaked what the risks are, and to follow a risk management rather than a risk avoidance attitude.

Once the assessment of the risks has been made, given a technological environment, some choices present themselves. Either one can abstain from leaking the information by not using a technology (which comes at a price), one can use a different technology that offers better protecting or is adapted not to leak any information, or one can choose to use a technology despite the risks because the benefits are worth it.

While not using a technology or simply accepting the risks can be a practical solution to many real life situations, sometimes it is not possible. Trying to select a substitute technology or protecting one requires again serious consideration of the technology and the organization's needs. Usually, public access phones, or terminals in libraries and cyber-cafes, offer an easy and relatively anonymous way of communicating. Traditional mail systems also provide less traffic data than its technological counterparts. Another generic solution is to use anonymizing proxies to access services. These can be used for web-browsing [9], email, and can be implemented for mail by first sending it to the central offices of an organization before it is dispatched to its final recipient. Steganography can be used to hide information in files, such as images, that do not give rise to suspicion. Encryption can be used to hide some usage data, but not to generally protect traffic data.

While protecting one's organization is necessary, in the long run the best protection against political surveillance is making sure that any existing surveillance regime is limited and has effective juridical, political and public oversight. It is therefore necessary for civil society organizations to organize themselves both as lobbying groups, to promote education and privacy legislation and practices, but also as a watchdog to collect, monitor and expose possible political use of surveillance powers.

- [1] EPIC Data Retention Page
http://www.epic.org/privacy/intl/data_retention.html
- [2] FIPR Press Release - FIPR Appalled by Huge Increase in Government Snooping.
<http://www.fipr.org/press/020610snooping.html>
- [3] FIPR Regulation of Investigatory Powers Information Centre.
<http://www.fipr.org/rip/>
- [4] Expert meeting on Cybercrime: Data Retention
<http://www.gilc.org/europol.pdf>
- [5] George Danezis, Traffic Analysis of the HTTP Protocol over TLS
<http://www.cl.cam.ac.uk/~gd216/TLSanon.pdf>
- [6] George Danezis, Comments on the E.U. Cybercrime forum
<http://www.cl.cam.ac.uk/~gd216/RetentionComments.pdf>
- [7] Intelligence Activities and the Rights of Americans, Book II, Part I Introduction and summary
<http://www.derechos.net/paulwolf/cointelpro/churchfinalreportIIa.htm>
- [8] Paul Mobbs, Living Under Surveillance
<http://secdocs.net/manual/lp-sec/scb7.pdf>
- [9] The Anonymizer
<http://www.anonymizer.com>