Tapping and Data Retention in Ultrafast Communication Networks[*]

Bart Custers

Research fellow
Tilburg Institute for Law, Technology and Society
Tilburg University, 5000 LE Tilburg, Netherlands
Senior Consultant
Capgemini Consulting Services, 3500 GN Utrecht, Netherlands
B.H.M.Custers@uvt.nl

Abstract. In the fight against crime and terrorism, many governments are gathering communication data in order to gain insight into methods and activities of suspects and potential suspects. Tapping, or wiretapping, has been used for a long time and nowadays most countries are extending this to data retention, i.e., large-scale storage of various kinds of data available on communications. At the same time, however, efforts are being made in the field of technology to develop a new generation of communication networks, based on ultrafast optical and wireless communication. This is likely to result in a significant increase in the speed and volume of information transfer on communication networks such as the Internet. These increasing amounts of information require increasing storage and analysis capacity, for which automated solutions are being developed. In this contribution, the way in which these technological developments influence the possibilities of tapping and data retention is discussed and some suggestions are made on how to deal with this.

1. Introduction

Government organizations, particularly those engaged in fighting crime and terrorism, have a particular need for personal information. Personal data may help to find out who a person is, whether he poses a risk to society and with what people he is in contact. Such information may play an important role in preventing crime and terrorism, as well as in solving or reconstructing in retrospect any such cases that have taken place.

In recent years, crime and terrorism have become more organized. As a result, it is no longer sufficient to investigate and profile individual suspects. There is also now a need to reveal the networks of people where these suspects operate. Gaining insight in who is communicating with whom may bring other suspects into scope, particularly *first offenders*, who were hitherto unknown to the authorities. In a way, finding out who knows who has become easier in the information age, as communication increasingly takes place via such information and communication networks as phones and the Internet. Tapping these communication lines is technologically straightforward. *Telephone tapping* is almost as old as the telephone itself. Apart from the term telephone tapping, the term *wiretapping* is often used to include tapping Internet communication. Nowadays, the term *tapping* is becoming more common, since this also includes various types of wireless communication networks. In this contribution, the common term tapping is used to indicate the tapping of all forms of electronic and/or digital communication. Communication without any tools such as phones or Internet can also be overheard and or recorded, but is beyond the scope of this contribution.

Data retention is a more recent form of investigating who knows who. Because of the ever-growing storage capacity of information systems, it has become technologically possible to store *all* communication that takes place over these networks. It is important to note that this is not currently happening, but it can be done. Nevertheless, the secret services in the United States are building large databases with communication data. Echelon is a global electronics communications surveillance system that gathers and processes vast amounts of communication data (Hagar, 1997, Madsen, 1998). Furthermore, in 2002, the US Department of Defense was planning a project known as Total Information Awareness (TIA). This project envisioned the creation of a gigantic government database of personal information, including communication data, to be analyzed under various models to detect patterns and profiles for terrorist activities (Markoff, 2002). When a major news story about TIA broke, civil liberties groups, commentators and politicians voiced criticism. In 2003 the program was renamed Terrorism Information Awareness and it was stated that privacy would be protected, though without specifying how. However, that same year, the US Senate stopped funding TIA (see also Solove, 2004).

In March 2006, the European Union adopted a directive that requires telecom operators and Internet providers in all member states to implement data retention systems for both telephone and Internet traffic. It is

[*] This paper was first published in Kierkegaard, S.(2007) Cyberlaw, Security and Privacy , pp. 289-300

significant to note that this EU Directive does not require or allow the retention of the *contents* of any communication. This contrasts to tapping, which focuses on the content of any form of communication. Data retention focuses on the storage of call detail records of telephony and Internet traffic and transaction data. Basically this concerns phone calls made and received, emails sent and received and web sites visited. These data provide an idea of who is in contact with whom, when, and how frequently. When possible, further identifying information may be added, as well as location data.

With the rise of new technologies and the ever-increasing volumes of information being transferred, new security issues arise regarding tapping and data retention. In the past decades there has been a significant increase in communication between people. At the same time there has been a significant increase in data storage and analysis relating to this communication. This raises the question of how these technological developments influence the potential of tapping and data retention.

This contribution will provide a brief overview of current use of tapping and data retention that will provide an answer to the question above and make some suggestions on how to deal with these new developments. In Section 2, the technological developments regarding ultrafast communication networks will be discussed briefly. In Sections 3 and 4, tapping and data retention will be discussed, respectively. How tapping and data retention works in ultrafast communication networks and what effects this may have will be explained in more detail in these sections. In Section 5, (European) privacy legislation will be discussed briefly. In Section 6, conclusions will be drawn and some suggestions will be made on how to deal with these effects. The focus will be on the developments in the EU and US.

2. Ultrafast Communication Networks

Technological change is exponential. According to Moore's Law, the number of transistors on an integrated circuit (a 'chip' or 'microchip') for minimum component cost doubles every 24 months (Schaller, 1997). This, more or less, implies that storage capacity doubles every two years or that data storage costs are reduced by fifty per cent every two years. Gordon Moore's empirical observation was made in 1965; by now, this doubling speed is approximately 18 months. Moore's Law deals with storage capacities, but similar observations have been made for communication speed and volume. According to Gilder's Law, the total bandwidth availability of US communication systems has tripled every twelve months since the 1980s and will expand at the same rate for the next 30 years to come (Raessens, 2001).

Moore's Law is not only about making existing technologies more efficient. It also takes into account the new ideas and inventions in the field of information technology. The latest developments to increase the speed and volume of transferring information on communication networks are focused on changing from electronic communication to optical communication. This is likely to result in a significant increase in the speed and volume of information transfer on communication networks. This new type of communication is referred to as *ultrafast communication* (Miller, 2004). In order to achieve all-optical networks, efforts are being made to develop and introduce optical communication hubs. Many optical fibers are already used for communicating optical signals over longer distances, but there are currently no optical alternatives for many electronic building blocks, such as flips flops, gates, buffers, memories, shift registers, and transistors.

Optical communication is not the only method for ultrafast communication. *Wireless communication*, using electromagnetic waves, is also considerably faster than electronic communication systems. The speed of wireless networks is often slowed down because wireless networks may involve electronic transmission at both ends of a data transmission. The development of all-optical building blocks will overcome the limitations for ultrafast communication systems.

3. Tapping

Wiretapping, or tapping for short, focuses on the contents of communication. The information communicated between persons may be very useful in criminal investigations. Obviously tapping is considered a serious infringement of some basic human rights, such as privacy and freedom of expression. Most legal systems explicitly mention privacy of letters and privacy of phone calls [1], as there are many different types of privacy. Those most commonly distinguished are spatial, relational and communicational privacy, all of which have physical and informational aspects (Blok, 2002).

Since tapping violates basic human rights, it is generally not allowed, but there are exceptions. In Western countries, such exceptions are strictly controlled and often concern (serious) suspicions of crime or terrorism or both. The police are often the executing authority. In most countries, tapping needs to be authorized by a court. Permission to tap is only provided under strict conditions (Koops, 1999). The crimes have to be severe; tapping is not allowed for minor offenses. Usually this means that the case must involve an offense punishable with, for instance, at least four years of imprisonment. Another condition is that there is a reasonable expectation that the suspect will participate in the conversation. Not only suspects' phones can be tapped, but also, for instance, relatives' phones. When the communication involves people with professional rights of non-disclosure, such as lawyers and doctors, monitoring the conversation is not allowed. Tapping is only allowed for public networks; private networks are beyond scrutiny.

The rapid developments in communication infrastructure are posing major challenges to the technical feasibility of tapping. Tapping regular (electronic) communication systems usually works by clicking a tapping device on the wire and copying (or interfering with) the signal. From a technological perspective optical tapping is quite different. The socket of the cable has to be removed to get to the fiber. When bending the cable, part of the optical data stream will no longer follow the path of the fiber but will go straight ahead [2] . This sub-stream can be read when using the proper devices. When a lot of light is being tapped, this may cause the quality of the signal at the receiving end of cable to decrease. This could indicate to users that a cable is being tapped.

Obviously this way of tapping signals is rather complicated, even when cryptography is not used to encrypt the information. Since most communication is at least partially wireless (usually the parts of the communication closest to the end users), it is much easier to tap wireless information. The advantage of tapping wireless communication is that the signal is transmitted in all directions and can therefore be easily intercepted. The major disadvantage of tapping wireless signals is that it requires being physically present at the place where the signal is being transmitted wireless. Generally speaking, the wireless part of a communication covers only a very small distance compared to the complete distance over which the communication takes place. For instance, a phone call between Europe and the United States is transmitted via long-distance wires on the bottom of in the ocean (covering thousands of kilometers) and only via short wireless distances close to the users (covering a few kilometers). Furthermore, ensuring that the largest part of the communication takes place via optical transmission has the advantage of higher quality and less use of energy [3]. Since wireless signals constitute the weakest link when it comes to tapping, it is likely that this is where actual tapping will take place.

Users may protect themselves from content tapping by using encryption. This is why governments seem to prefer encryption methods with *trapdoors*, i.e., possibilities that allow a person with additional information to tap encrypted data flows (Van der Lubbe, 1997, Schneier, 2000, Abelson et al., 1998, Akdeniz, 1998). Government institutions may then use such additional information for criminal investigations. Companies like Microsoft, Netscape, and Lotus have implemented trapdoors in their software (Leprovost, 1999).

4. Data Retention

In March of 2006, the European Union adopted Directive 2006/24/EC on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks. This European Directive requires all EU member-states to implement national legislation to ensure that communications providers retain particular data, for a period of between six months and two years. The providers are mainly Internet Service Providers (ISPs) and telecommunication companies. As indicated above, in contrast with tapping, data retention does not focus on the contents of the communication, but rather on the storage of call detail records of telephony and Internet traffic and transaction data.

Obviously these data are stored with a purpose. According to Article 1 of the Directive, the aim is to use the data for "investigation, detection and prosecution of serious crime". Considering that a lot of communication takes place, data retention involves building vast databases. The costs for this are for the ISPs and telecom operators, who are subjected to fines if they fail to comply. In the light of the developments regarding ultrafast communication systems, it is expected that the volumes of communication will significantly increase. Hence, the capacity needed for storage and analysis of the data will increase accordingly. This will result in an overload of information. Therefore, the ability to distill useful information from these large amounts of data is becoming more and more important. Technologies are being developed for this and one of the most promising is *data mining*, which provides an automated analysis of data in order to find patterns and relations (Adriaans and Van Zantinge, 1996, Fayyad et al., 1996). Such an automated analysis may result in revealing networks around particular people and it may result in risk profiles of both individuals and groups (Custers, 2004).

The use of such risk profiles may have some typical advantages and disadvantages. The main advantages concern efficacy and (cost) efficiency, as it may be easier to find the individuals or target groups that are being looked for. A particular advantage is the possibility of finding so-called *first offenders*. This works as follows: if the characteristics of a particular individual in the database are very similar to the characteristics of some known individuals, it is assumed that such a person poses an increased risk. Obviously this does not mean the person actually is a terrorist or is planning a terrorist attack. However, there are also some disadvantages of using such risk profiles. One of the main problems is that the profiles may not always be accurate (Custers, 2003). There may be false-positives (i.e., innocent people who also fit the profile) and false-negatives (i.e., terrorists and criminals that do not fit within the profile and are hence out of scope). The false-positives may result in arrests of innocent persons; the false-negatives may result in missing the persons who need to be identified. When particular criteria, such as ethnic origin and religion, are used to create a profile, this may result in unjustified discrimination[4]. If such profiles become known publicly, this may also result in stigmatization of particular groups (Harvey, 1990).

People who want to be protected against data retention have several options to avoid leaving traces of their communication or to ensure that traces lead to other persons. These methods cause decreased reliability of the data retained. Basically identifying persons communicating via a network means establishing a link between the user and the network. Furthermore the user needs to be identified, often by traditional methods, using identity documents, face recognition, passwords, keys, etc. The methods of identification are easily tampered with, rendering the user anonymous. Tampering with the link between the network and the user is a typical form of *identity fraud* and is often straightforward. The Internet can be used anonymously by walking into an Internet café somewhere in the world. When registration is required, it is usually easy to provide a fake name. People who want to disseminate computer viruses often use this method in order to be untraceable.

The user access point to a network is usually indicated by an address. For cell phones, this is on the SIM card (Subscriber Identity Module), a removable smartcard for cell phones. On the Internet, IP addresses (Internet Protocol addresses) are used, which are numbers that locate someone's computer on the Internet. Some IP addresses are static, i.e., they do not change every time a user logs on to the Internet. If a user has a dial-up connection to the Internet or is using a computer that is connected to the Internet intermittently, it is most likely picking up a dynamic IP address from a pool of possible IP addresses at the Internet service provider's network during each login. Obviously, dynamic IP addresses may be used to tamper with the link between the network and the user, since this is no longer a uniquely identifying link. It might be easy to retrieve the IP number used at a particular point in time, but it could prove difficult to build a dossier on a particular IP number when there are different users. At one moment the IP number could be used by a terrorist suspect; at another moment, it could be used by someone else who has nothing to do with this suspect. There are many technological applications that can be used for accessing and using the Internet anonymously. Using telecommunication networks anonymously is simple as well. For instance, anonymous phone calls can be made by buying a prepaid cell phone in a supermarket. The phone can be thrown away afterwards. At the moment, this may not be really cheap, but prices are decreasing.

5. Legal Protection of Privacy

When confronted with the developments described in the previous sections, many people ask whether there is any legal protection of privacy that may prevent the effects of tapping and data retention. In this section I will indicate that there are cases where these (mainly European) privacy laws fall short. The main reason for this is that tapping and data retention legislation usually overrules privacy legislation. Privacy is often regarded as a hindrance in fighting crime and terrorism, although views on this are changing. Analyzing all available data is not always as effective as a targeted (and more privacy-preserving) approach.

However, even when privacy legislation is not overruled by tapping and data retention legislation, there are some difficulties with the legal protection of privacy. A brief explanation of how (European) privacy legislation works is required. In Europe, the collection and use of personal data is protected by a European Directive (the so-called 'privacy directive'), which has been implemented in national law in the member countries of the European Union[5] . Privacy principles that are safeguarded in the European privacy directive correspond to the principles in the Organization for Economic Co-operation and Development (OECD) guidelines, [6] which were also included in the Council of Europe Treaty of Strasbourg [7]. (For a more detailed account, see Bygrave, 2002.)

These principles are:

- the *collection limitation principle*, stating that "[t]here should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject";
- the *data quality principle*, stating that "[p]ersonal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up to date";
- the *purpose specification principle*, stating that "[t]he purposes for which personal data are collected should be specified and that the data may only be used for these purposes";
- the *use limitation principle*, stating that "[p]ersonal data should not be disclosed, made available or otherwise used for purposes other than those specified, except a) with the consent of the data subject; or b) by the authority of law";
- the *security safeguards principle*, stating that reasonable precautions should be taken against risks of loss, unauthorized access, destruction, etc., of personal data;
- the *openness principle*, stating that the subject should be able to know about the existence and nature of personal data, its purpose and the identity of the data controller;

- the *individual participation principle*, stating, among others, that the data subject should have the right to have his personal data erased, rectified, completed or amended;
- the *accountability principle*, stating that the data controller should be accountable for complying with measures supporting the above principles.

These privacy principles for fair information practices are based on the concept of *personal data*, which is described in article 2 sub a of the European privacy directive as ‘data concerning an identified or identifiable natural person’, a definition that also stems from the OECD guidelines. Personal risk profiles contain personal data and are therefore protected by the (national implementation of the) directive, but group risk profiles do not necessarily contain personal data and may therefore lack this protection. Particularly in the case of data retention, the use of group risk profiles is useful to avoid privacy legislation. A group profile is a property or a collection of properties of a particular group of people. Group risk profiles may contain information that is already known, for instance people who smoke live on average a few years less than people who do not smoke. But group risk profiles may also show new facts, such as, for instance, people living in zip code area 8391 are (significantly) more often terrorists. Group profiles do not necessarily describe a causal relation. For instance, people driving a red car may show (significantly) more criminal behavior than people driving a blue car. As was already indicated in the previous section, group profiles, though useful, may result in stigmatization and errors. (For a more detailed discussion, see (Custers, 2004)).

6. Conclusions

Currently, there is a lot of information on communication being collected and processed to support the fight against crime and terrorism. As a result of new technological developments, such as the development of ultrafast communication networks, it is expected that the amount of communication data will continue to increase. This new generation of networks is likely to be a combination of optical and wireless devices. The former are relatively hard to tap; the latter are relatively easy to tap. Tapping should only be possible if the prescribed conditions allow it. It is therefore particularly recommended that cryptography is used to prevent unauthorized tapping for the wireless parts of ultrafast communication. This cryptography should not be too strong to be deciphered in cases in which tapping is allowed. The use of trapdoors and technologies such as key recovery systems, key escrow systems and trusted third-party encryption may be useful to achieve this.

Whereas tapping concerns the contents of the communication, data retention focuses on storing and analyzing communication data, particularly call detail records regarding phone calls and Internet traffic. Ultrafast networks will require larger capacities for storing and analyzing data. The former is relatively easy, since storage capacity continues to grow (though the costs involved are the subject of a major discussion); the latter is a significant problem. Analyzing vast amounts of data needs to be automated, for example, by means of data mining. However, most data mining technologies are not yet sophisticated enough for large-scale use. Furthermore, a major disadvantage is that the risk profiles resulting from the automated analyses may not be accurate. False-positives may result in investigating and even arresting innocent people. False-negatives may result in criminals and terrorists being out of scope.

When risk profiles have limited accuracy, they should only be used with the utmost care, in order to prevent investigating and arresting innocent people. It is recommended to always perform double checks on existing risk profiles and not to merely rely on data in databases, but to also conduct significant fieldwork. In order to prevent the worst forms of unjustified discrimination and social polarization, it is recommended not to include sensitive personal data, such as religion and ethnic background, in the risk profiles.

It is important to note that the increasing speed of network communication on tapping and data retention does not present much of a difference to civil liberties issues. Issues like privacy, guilt by association and wrongful arrest do not present much of a difference if a fast or slow network is presented. However, the combined effects of new technologies and new powers for government organizations have far-reaching consequences for the constitutional rights and privacy of individuals. Recent research in the Netherlands shows that the over the past few years, the Dutch government has approved numerous laws that have drastically increased the intelligence-gathering powers of the police, judiciary and intelligence services (Vedder et al. 2007).

Summarizing, tapping and data retention in the age of ultrafast communication networks may be very useful to reveal criminal and terrorist networks and to find first offenders. Both aspects are increasingly needed in the fight against crime and terrorism. However, because of the increasing amounts of data that are communicated over ultrafast networks, it is vital to start by determining which data should be collected. Even though all data can be stored, it is not recommendable to do so because the overview will be lost. It is better to make a selection of the data that may be useful. This will make the approach better targeted and effective than storing and analyzing all available data.

References

1. Abelson, H., Anderson, R., Bellovin, S.M., Benaloh, J., Blaze, M., Diffie, W., Gilmore, J., Neumann, P.G., Rivest, R.L., Schiller, J.I., and Schneier, B. (1998) *The Risks of Key Recovery, Key Escrow & Trusted Third Party Encryption*; A report by an ad hoc group of cryptographers and computer scientists. <www.cdt.org/crypto/risks>.
2. Adriaans, P., and Zantinge, D. (1996) *Data mining*, Harlow, England: Addison Wesley Longman
3. Akdeniz, Y. (1998) No Chance for Key Recovery: Encryption and International Principles of Human and Political Rights. In: *4th International Conference on Ethical Issues of Information Technologies*, Ethicomp 98, Rotterdam. Zielinski, C. (1998) The Ethics of Encryption, In: *4th International Conference on Ethical Issues of Information Technologies*, Ethicomp 98, Rotterdam.
4. Blok (2002) *Het recht op privacy* [The right to privacy]. The Hague: Boom Juridische Uitgevers.
5. Bygrave, L.A. (2002) *Data Protection Law; approaching its rationale, logic and limits*, Information Law Series 10, Den Haag: Kluwer Law International.
6. Custers, B.H.M. (2003) Effects of Unreliable Group Profiling by Means of Data Mining. In: G. Grieser, Y. Tanaka and A. Yamamoto (eds.), *Lecture Notes in Artificial Intelligence*, Proceedings of the 6th International Conference on Discovery Science (DS 2003) Sapporo, Japan. Berlin, Heidelberg, New York: Springer-Verlag, Vol. 2843, pp. 290-295.
7. Custers, B.H.M. (2004) *The Power of Knowledge*, Tilburg: Wolf Legal Publishers.
8. Fayyad, U.M., Piatetsky-Shapiro, G., and Smyth, P. (1996) From Data Mining to Knowledge Discovery: An Overview. In: U.M. Fayyad G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy (eds.) *Advances in knowledge discovery and data mining*. Menlo Park, California: AAAI Press / The MIT Press.
9. Hagar, N. (1997) Exposing the Global Surveillance System, *Covert Action Quarterly*, Winter 1997.
10. Harvey, J. (1990) Stereotypes and Group-Claims; Epistemological and Moral Issues, and their Implications for Multi-Culturalism in Education, *Journal of Philosophy of Education*, Vol. 24, No. 1, pp. 39-50.
11. Koops, B.J. (1999) *The Crypto Controversy; A Key Conflict in the Information Society*, The Hague, Netherlands: Kluwer Law International.
12. Leprovost, F. (1999) Encryption and Cryptosystems in Electronic Surveillance: A Survey of the Technology Assessment Issues. In: *Development of Surveillance Technology and Risk of Abuse of Economic Information; An Appraisal of Technologies of Political Control.* In: D. Holdsworth (ed.) European Parliament, Directorate General for Research, Scientific Technological Options Assessment (STOA).
13. Madsen, W. (1998) Crypto AG: The NSA's Trojan Whore? *Covert Action Quarterly*, Winter 1998.
14. Markoff, J. (2002) Pentagon Plans a Computer System That Would Peek at Personal Data of Americans, *New York Times*, November 9, 2002.
15. Miller, D.A.B. (2004) Ultrafast Digital Processing, In: A. Miller, D.M. Finlayson, D.T. Reid (eds.) *Ultrafast Photonics*, Bristol, Philadelphia: Institute of Physics Publishing.
16. Raessens, B. (2001) *E-business, Your Business*. Utrecht: Lemma.
17. Schaller, R.R. (1997) Moore's Law: Past, Present and Future, *Spectrum*, IEEE, Volume 34, June 1997, pp. 52-59.
18. Schneier, B. (2000) *Secrets and Lies; Digital Security in a Networked World*, New York: Wiley Computer Publishing, p. 241.
19. Solove, D. (2004) *The Digital Person; Technology and Privacy in the Information Age*, New York: New York University Press.
20. Van der Lubbe, J.C.A. (1997) *Basismethoden cryptografie*, Delft: Delftse Universitaire Pers. p. 143. Systems with exceptional access include *key recovery systems*, *key escrow systems*, or *trusted third-party encryption*.
21. Vedder, A. Vedder, A.H., Wees, L. van der, Koops, B.J., & Hert, P.J.A. de (2007). *Van privacyparadijs tot controlestaat? Misdaad- en terreurbestrijding in Nederland aan het begin van de 21ste eeuw*. Den Haag: Rathenau Instituut. (Studies Rathenau Instituut, 49).

Notes

[1] See for instance Article 12 of the Universal Declaration of Human Rights and Article 8 of the European Convention on Human Rights.

[2] Note that actually cutting the cable is not necessary to tap the information flow.

[3] Generally speaking, optical transmission is in only one direction, whereas wireless transmission is usually in all directions. As a result the strength of a wireless signal decreases with a factor of $1/r^2$ over a distance r.

[4] Note that telephone and Internet data do not include such characteristics; however, they may be derived with some accuracy from location data, since particular locations may be indicators for characteristics like ethnic background and religion.

[5] European Directive 95/46/EG of the European Parliament and the Council of 24th October 1995, [1995] OJ L281/31.

[6] See <http://s3-hq.oecd.org/scripts/pwv3/pwhome.htm>.

[7] See <http:/www.coe.fr/dataprotection/Treaties/Convention%20108%20E.htm>.