October 2005

ETNO Expert Contribution on Data retention in e-communications - Council's Draft Framework Decision, Commission's Proposal for a Directive

INTRODUCTION

1.- Purpose of the document

This document analyses some of the **most relevant technical aspects** involving the impact that the implementation of a generalised data retention obligation would have for ETNO companies as e-communications services providers.

Although two parallel legislative initiatives[1] *from Council and from Commission exist,* the technical requirements for operators and service providers that both regulations raise are practically the same. Thus, this document would generally apply in both cases.

2.- References

Currently, last versions:

[1] "Draft Framework Decision on the retention of data processed and stored in connection with the provision of publicly available electronic communications services or data on public communications networks for the purpose of investigation, detection and prosecution of crime and criminal offences including terrorism".

Council of the European Union.

Brussels, 10th October 2005 (12894/1/05 COPEN 153 TELECOM 99).

[2] "Proposal for a Directive of the European Parliament and of the council on retention of data processed in connection with the provision of public electronic communication services". COM (2005) 438 final.

Commission of the European Communities.

Brussels, 21 September 2005

[1] Council's Draft Framework Decision and Commission's Proposal for a Directive of the EP and the Council.

1. GENERAL COMMENTS

- ETNO companies wish to reiterate their **commitment to work with the Law Enforcement Agencies (LEAs)** in the fight against terrorism and crime. Current co-operation ranges from the real time interception of phone conversations, data preservation to making available traffic data collected for legitimate business purposes.

- It must be pointed out that there is a major difference in storing data

  - voluntarily and doing so on a compulsory basis;

  - for own internal use and to do so for an external party

  - in bits and pieces in different systems and platforms and to do so in an aggregate manner

- As stored data under the new provisions may not be used for other purposes than fighting crime, operators will have to duplicate data for their own use (billing, network planning, etc.)

- There are many ambiguities in the proposed Directive's text and in its Annex. These must be clarified in order for the industry to provide authorities with an accurate feed-back.

- The **principle of proportionality** between the data retention obligation and the rights to secrecy of the communications, privacy and protection of data (Art. 52 of the Charter of Fundamental Rights of the European Union) must be respected. In this respect, the proposed measures should combine, in a balanced way:

  - the needs for security and the fight against crime and

  - respect for individual rights.

- Furthermore, any future regulation must have a level of confidentiality that prevents the serious and organised crime from having access to data and from becoming experts of the technical means available.

- In general, increasing the range of data to be retained will require major investment and increase costs for the industry. For certain types of data, serious doubts are raised about the technical, economic and administrative viability of collecting such data.

- Any future measure should introduce **harmonised criteria in the allocation of costs**. This would ensure fair competition among European operators and a level playing field for all operators and service providers across EU. Any regulation must target all actors equally in order to avoid distortion of competition. The battle against crime is first and foremost a public duty whose costs must not be imposed on the telecommunications industry. At the same time, **cost compensation** is an important factor that helps to restrain the data requests of the law enforcement authorities.

- Having said that, it must be avoided that the question of cost compensation supersedes the **more important discussion about the substance of the specific obligations in the proposals.**

2. SPECIFIC COMMENTS

Data storage: the basic prerequisites

- Successful collection of huge amounts of session data relating to mobile, fixed and Internet is a difficult task. Trying to make sense of the different data format and interpret them into something that makes sense and is of value for law enforcement agencies is even harder. The quality of the collected data is a major issue to avoid false positives or false negatives during the legal investigation.
- In addition, most of the data needed to fulfil the data retention requirements are raw data produced by the network. As most of this data is not needed for billing purposes, it is generally not stored or even captured. To store and retrieve this data is a critical issue that most certain require a standardized format. Above all, the main problem is to interpret and implement this data into a data call record and send it to the data processing centre, therefore costly technical upgrades being needed.
- In order to interpret and use the stored data it must be coded into a specific language, "format". Overtime, these formats change which in turn leads to the necessity to continuously update all stored data that one wants to use. This is a costly and complicated process, where the costs are proportionate to the amount of data stored – a rule of thumb is that the cost for updating data is in the order of 1/3 of the original investment.
- When the amount of retained and processed data becomes too voluminous, the data retention operation starts to interfere with network performance. It is difficult to say where the threshold is and at what point this interference will occur without detailed analysis, but it is important to highlight this risk with a view to the large volumes of retained ~~i~~nternet data which will be the direct consequence of the Commission's and/or the Council's proposals.
- At present, there is no search engine/analysing tool capable of analysing the amount of data to be retained under the new proposals. It can be expected that this would require a lot of manual work until industry and LEAs have developed such a search engine.
- When storage volume becomes an issue, operators may choose to clean out the raw information and only store processed data. This implies that it is no longer possible to revert to the raw data and carry out an analysis for errors. This further entails that operators would have to store both processed and raw data in order to carry out quick and in-depth analyses.
- When data is exchanged between two or more systems, great efforts must be made to ensure that the different sets of data are compatible to each other and belong together. This is due to the fact that different networks work with

different data clocks. This matching process is both complex, tedious and costly, and is of course directly proportionate to the amount of data being handled.

- Based on the volume of retained data the cost will increase due to the complexity of the systems involved in the management of data storage and recovery. The costs curve based on the volume of data to be retained can be considered linear for quantities close to 1 TeraByte[2] (TB). This situation would correspond to the needs of a small operator or Internet services provider. Bearing in mind the state of the art of storage technologies, higher volumes of data would mean incurring non-linear or even exponential costs since they involve a change in the design of the management systems, more powerful and sophisticated platforms, greater security measures, storage and support infrastructures as well as the necessary human resources to handle this type of systems. This is the case of a large operator or service provider.

'Generated and processed' data

Apart from the data volumes and retention periods, operators are especially concerned about the obligation to store new data types such as unsuccessful call attempts and location data at the end or even throughout a mobile call. **Operators do not currently store these data for business purposes.**

In any case, in order to comply with the proposed data retention measures, also the processing systems of all operators and service providers would have to be redesigned and/or upgraded in order to allow the exchange of such data.

As the extend of the technical effort necessary to retain a certain type of data is dependent on the question to which extend data is already 'available', **it is important that a clarification on the terms "generated" and "processed" is introduced in the final text:**

- It must be noticed that some data are not "**generated**" at all by the networks, that is, no relevant signal is available on the network (e.g.: MAC).

- Other data provide a relevant signal available on the network, but need to be interpreted and implemented into a data call record (for instance, as it is the case in most operators, it happens with "call attempts" or location data at the and or throughout a call), the network needs to be upgraded, implying costly investments.

- Finally, operators **process** data, that means the signal/information is already implemented into a data call record and sent to and stored in the processing/billing centre (this data is really 'available' and 'just' has to be stored for a longer period; however additional costs occur for update of storage and processing centres (storage and search capacity).

[2] 1 Terabyte equals 1024 Gigabytes

[4] A prior technical feasibility study should be carried out based on the various Internet equipment manufacturers.

- It should also be pointed out that any obligation to retain specific data referring to the **destination** of a communication (**e.g. name and address** of the subscriber or registered recipient, connection label or user ID of the intended recipients, IMSI, IMEI of the called party, internet connection label) is virtually impossible, whenever it involves different service providers. Furthermore, the handling and reshuffling of such enormous data volumes would overload the communications networks. Therefore, **any future measure must clarify that only the party offering the respective service can be subject to the data retention obligation. This party is the only one in a direct relationship with the customer and with sovereignty over the data.** As an example, those companies which act only as mere carriers cannot identify the he final subscriber of e.g. an email service, but only the service provider with a direct relationship with the final user is able to provide this information.

- This clarification needs the deletion of the obligation for service providers to retain data related to the destination of a communication involving different operators/service providers (with particular reference to Internet service access, Internet e-mail and Internet telephony: Annex of the proposed Directive, lett. b) point 3.) and a modification of current art. 3 of the Proposal for a Directive such as:

  Art. 3.2.

  Member States shall adopt measures to ensure that data retained in accordance with this Directive are only provided to the competent national authorities **BY THE PROVIDER THAT HAS OFFERED THE USED e-COMMUNICATION SERVICE** , in specific cases and in accordance with national legislation, for the purpose of the prevention, investigation, detection and prosecution of serious criminal offences, such as terrorism and organised crime.

  Fixed-line and mobile telephony

  - UNSUCCESSFUL CALLS.

  Currently Call Detail Recordings (CDR) corresponding to **unsuccessful calls** (e.g. called party does to take up, line busy) are not processed by most operators since they are not necessary in the billing process. This type of call involves a considerable percentage of the total number of calls made (in the case of **fixed and mobile telephony, it can mean** varying on different national situations from **around 40%** and up to 60% of the total number of calls).

  Complying with the current proposals and depending on the size and configuration of the network, it would require a big operator to make hardware and software upgrades of up to 1600 switches in order to interpret the signalling information into a data call record and send it to the data processing centre resulting in an additional investment.

Initial investment cost would be within a 3 digit million Euro range only for this data type. Additional costs occur for updating the data processing centre (storage and search capacity), additional security measures, more human resources.

The amount of data generated would require operators to change the Input Output-system. If operators are required to store raw signalling data, the volumes would be considerable.

Mobile telephony.

- LOCATION DATA AT THE END (or throughout) OF A MOBILE COMMUNICATION.

Currently, most operators process the location of a mobile call (Cell-ID) **only at the beginning** of an answered (successful) mobile call. In order to record the Cell-IDs also at the end of (or even throughout) a call and to transmit them to the data processing centre would require a costly network upgrade.

It is doubtful whether this requirement is proportionate since it is already possible to create a movement profile of the user by means of the Cell-IDs that have been stored at the start of each new call, and that there are additional systems in place (interception) to identify a current location, e.g. IMSI-catching.

As an example, the retention of unsuccessful calls and information about the cell for a period of six months would mean for a large operator a volume of 14,400 million call registrations.

- IMEI
Retaining the international number of mobile equipment (**IMEI**) will have limited use as proof of the user's identity since numbering can be manipulated and can be multiply assigned by manufacturers so that subscribers cannot be clearly identified. Several attempts in the past to implement a unique system have failed.

Operators usually do not transmit IMSI/IMEI numbers to each other, and only a small portion can therefore be registered. In addition, calls transmitted through the fixed network will be stripped of IMSI/IMEI numbers because the fixed network does not support forwarding.

Internet services.

Notions regarding data retention in traditional voice telephony services should be revised whenever they are transferred to Internet services. However, the proposed measures under discussion require the storage of data relating to e-mail services and "voice over IP" (VoIP).

Data for identifying the connection (i.e. "Internet Access Data" - IP address, log in and log time, User ID) have proven to be **the most important and useful tool for LEA's purposes.** Internet access data can be used to determine which IP address with which customer ID was connected to the internet, when, and for how long.

With regard to IP address, however, it must be noted that due to the design of certain Internet protocols, network infrastructures do not check the authenticity of the **IP origin address**. Computer programs exist, which are public and free of charge, that the user can employ to falsify the genuine IP address, static or dynamically assigned by the access service provider (technically known as IP spoofing).

Even when restricting the requested data to the Internet Services 'email' and 'VoIP' (some Member States even want Chat, http etc), huge additional costs would be incurred for the information systems necessary to answer the requests.

Concerning e-mail, it should also be noted that many operators only have a small market share in terms of e-mail services in the EU. The major mail providers – e.g. Hotmail, Yahoo and Google, are all located outside the EU and would not be subject to the proposed provisions.

Besides, reasonable doubt exists about whether the retention of Internet **service** data, such as in the case of the http (as some Member States want to introduce in the Council proposal), could be equivalent to interception of the communication, since the destination Web address can indicate the content that subscribers have accessed.

In addition to all the above-mentioned, specific services exist to confer anonymity to calls. Generally located in countries outside the scope of the European legislation these servers, which are public and free of charge, delete and filter the user's personal data and call identifiers, rendering any investigation impossible.

- MAC ADDRESS

Just like the IP call origin address, the **MAC address** (Media Access Control address) of the terminal equipment, assigned by the manufacturer, can be falsified (technically known as MAC spoofing) using programs available in the public domain. On the other hand, and in recent years, the uniqueness approach has not been maintained in the serialization by manufacturers, reutilising numbering systems for this type of devices.

The MAC is a piece of data used by the low-level protocols and amongst adjacent routers that is not spread either toward the Internet management systems or toward other remote routers (that means the MAC address is not even generated). The retention of the MAC address to include it in the registration of retained data associated with each communication, **in the event**

it were technically[4] viable, would force the network infrastructures to be updated, besides penalizing its performance.

These considerations about the ambiguity in the identification by MAC address diminish the importance of retaining this data associated with other Internet services different from that of the connection.
