

The hazards of technology-neutral policy: questioning lawful access to traffic data

By Alberto Escudero-Pascual and Ian Hosein

◆ ACM, 2004. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in Communications of the ACM, {VOL47, ISS3, (April 2004)}<http://doi.acm.org/10.1145/971617.971619>"

<ACM sub-line (below title)>

Policies are being updated to deal with new communications infrastructures; the path to policy renewal is fraught with danger.

</ACM sub-line>

After some successes and many miss-steps, the regulatory environment surrounding technology policy is transforming. Lessons taken from content, copyright, and cryptography policy processes, amongst others, resulted in the emergence of a number of technology policy innovations. Two particular innovations are the *internationalization* of policy-making, and *technology-neutral* policies. These innovations come with risks, however. The risks are particularly apparent when we look at policies on lawful access to traffic data.

Access to traffic data for law enforcement purposes is a traditional tool for investigation and intelligence gathering. *Traffic data* is an elusive term, due in part to technology variances. The policies regarding lawful access to traffic data, however, are increasingly set in technology-neutral language, while the language and policies are often negotiated at closed international fora.

Even while policy changes are argued as necessary due to international obligations and new technological realities, these policies tend to ignore technological details. Just as cryptography policies of key escrow were mis-interpreted by government as updates 'to maintain the status quo' [8]; updating legal definitions of traffic data while not acknowledging the increased 'sensitivity' of the data is problematic.

Historical Access

In the days of plain old telephone systems (POTS), after much legal debate, the content of communications was considered sensitive and therefore any interception required constraint, e.g. judicial warrants in the U.S., politician-authorized warrants in the United Kingdom. The same rule did not apply to *traffic data*: numbers called, calling numbers, etc. This data was considered less invasive, and therefore only required minimal constraint. An additional factor was that traffic data was stored by telephone companies and in turn was available for access by law enforcement agencies, while content was not: traffic data was available, legally less sensitive, and so, lawfully accessible.

The traffic data records collected by telephone companies are generally of a form similar to:

```
19991003070824 178 165 0187611205 46732112106 -----001----003sth 46 4673000---0013 14 10260
```

The most significant fields (emphasized) are: date and time of the start of the call, duration of the call and caller and receiver phone numbers.

Traditional investigative powers of access to traffic data were established with traditional technological environments in mind. Governments are now updating these policies to apply to modern communications infrastructures. If governments insist on applying traditional powers to these new infrastructures, the new policies must acknowledge that the data being collected now is separate from tradition.

The claim of technological neutrality

Many policy initiatives have involved articulations regarding the importance of being *technology-neutral*. When the Clinton Administration first announced its intention to update lawful access powers, they proposed to "update the statutes in outmoded language that are hardware specific so that they are technologically neutral" [7]. Meanwhile in the United Kingdom, it was noted in tempestuous debates in the House of Lords regarding the *Regulation of Investigatory Powers Act 2000* that:

The Earl of Northesk: "One of the many difficulties I have with the Bill is that, in its strident efforts to be technology neutral, it often conveys the impression that either it is ignorant of the way in which current technology operates, or pretends that there is no technology at all." [9]

Technology-neutral policy is seen as a way to deal with concerns of governments mandating a specific type of technology. While this is favorable in the case of some policies that affect market developments, technology-neutral lawful access policies may contain hazardous side-effects. Technology-neutral language may be used to ignore the challenges and risks to applying powers to different infrastructures.

Defining 'Traffic Data'

Inter-governmental organizations, particularly the Group of 8 (G8) and the Council of Europe (CoE), have been working for a number of years to ensure lawful access to traffic data. Both the G8 and the CoE have been criticized by industry and civil society because of their ambiguous and problematic approaches, and the closed nature of their processes. Yet policies continue to be decided there, and brought to national parliaments under the guise of 'harmonization' and 'international obligations'.

The G8 formed a senior experts group in 1995 to develop an international co-operation regime to address transnational organized crime. This 'Lyon Group' has since been active on high-technology surveillance-related policies, including three meetings with industry representatives throughout 2000

and 2001. Arising from that work, the G8 working-definition of traffic data is "non-content information recorded by network equipment concerning a specific communication or set of communications." [10]

Meanwhile, the CoE, a 45-member inter-governmental organization, convened closed meetings since 1997 to develop a treaty establishing lawful access powers across borders. The CoE Convention on Cybercrime defines traffic data as "any computer data relating to a communication by means of a computer system, generated by a computer system that formed a part in the chain of communication, indicating the communication's origin, destination, route, time, date, size, duration, or type of underlying service" [2]. In the convention's Explanatory Report [3], the CoE states that traffic data should be defined so as 'to not refer' to the content of a communication; but this is a non-binding interpretation.

One is left to wonder what is included and what is excluded by these vague definitions. While subject lines in emails may be content, uncertainty arises as to whether the name of files requested (e.g. HTTP requests), URLs (e.g. <http://www.computer.tld>), search parameters, TCP or IP headers, and other such data are considered content or traffic data. A report of a transaction by an individual with server 158.143.95.65 may be considered traffic data; but the name of the web site(s) run on a server may disclose more information (e.g. aidshelpline.org). Search parameters in the URLs and the name of files accessed may refer to the content of the communications. If we consider the next generation Internet, mobility bindings or routing information included in the IPv6 extended header will include absolute or relative location information. The location information is part of the mobility 'signaling' protocol and hence fits into the above definitions of traffic data.

Some states have tried to deal with this challenge in their legislative language. The UK's *Regulation of Investigatory Powers Act 2000* went through much iteration, particularly in the so-called 'Big Browser' debates, before the language was agreed upon. Traffic data is defined in theory as data about the

source and destination of a transaction, and data about the routing and the tying of separate packets together. This definition is complemented by the legal definition of 'communications data': data used by the network; or exists within logs; or other data that is collected by service providers. However the definitions are also quite clear about the extent of information that qualifies: this data does not include URLs per se, and may only include the name of the computers running a service, while the specific resource used qualifies as content, and accorded greater protection. Therefore, in the UK the IP address is traffic data, while `http://www.url.tld/file.html` is tantamount to content.

Other states have failed to respect this level of technological awareness. Previous U.S. policy differentiated between traffic data from cable and telephone communications. The *Cable Act* once protected traffic data to a greater degree than telephone traffic data, as viewing habits were considered sensitive. Now that cable infrastructure is also used for internet communications (which were previously used over telephone lines, and thus traditional laws applied), successive White House administrations worked to erase this cable traffic distinction, finally succeeding with the *USA-PATRIOT Act*. Rather than deal with the specifics of digital communications media and services, the changes in U.S. law reduces the protections of traffic data for cable internet communications to what had previously existed for telephone communications data.

This can be interpreted as a boon to law enforcement. According to Attorney General John Ashcroft:

Agents will be directed to take advantage of new, technologically neutral standards for intelligence gathering. (...) Investigators will be directed to pursue aggressively terrorists on the internet. New authority in the legislation permits the use of devices that capture senders and receivers addresses associated with communications on the internet [1].

Traffic data blurs with the content of communications as new communications infrastructures are encompassed under existing practices. The legal protection of this data is reduced as distinctions

applied are based on categorical decisions established under older technologies. The separation of content and traffic remains elusive, even in policy language.

Categorical Determinants: The technological

Traffic data under the plain old telephone system was considered derivative, and while informative, it did not necessarily disclose the sensitive details of an individual's life. Under the *Cable Act*, Congress accepted that the disclosure of individual viewing habits deserved greater protection; but such protections were later deemed unnecessary for the internet.

Traffic data's constitution differs by communications medium. Below we present dial-in records, wireless LANs, and search engines to preview what can be accessed by technology-neutral law enforcement powers. ²

Dial-In records

The Remote Authentication Dial-In User Service (RADIUS) is a client/server security protocol, designed to manage dispersed modem pools for large numbers of users. This tends to involve managing a single "database" of users, which allows for authentication (verifying user name and password) as well as configuration information detailing the type of service to deliver.

Many Internet Service Providers are outsourcing the access network to big operators that provide dial-up connectivity world-wide. Internet users dial into a modem pool attached to a Network Access Server (NAS) that operates as a client of RADIUS. The client is responsible for passing user information to designated RADIUS servers (managed by the ISP) and then acting on the response that is returned.

The RADIUS server stores usage information for dial-in users, often for billing purposes. When the user is authenticated and the session has been configured according to the authorization information,

an accounting start record is created. When the user's session is terminated, an accounting stop record is created.

The most significant fields of the "start/stop" records are:

Start and Stop Timestamps

Timestamp records the "Start" and "Stop" time on the RADIUS accounting host. The duration of a session "Acct-Session-Time" is computed by subtracting the "Start" and "Stop" timestamps.

Call(ed,ing)-Station-Id

Called-Station-Id records the telephone number called by the user. Calling-Station-Id records the number the user is calling from.

Part of the *Start and Stop RADIUS records* may look like:

```
Fri Oct 19 11:30:40 2001
  User-Name = "aep@somedomain.org"
  NAS-IP-Address = 62.188.74.4
  Acct-Status-Type = Start
  Acct-Session-Id = "324546354"
  Acct-Authentic = RADIUS
  Calling-Station-Id = "01223555111"
  Called-Station-Id = "02075551000"
  Framed-Protocol = PPP
  Framed-IP-Address = 62.188.17.227

Fri Oct 19 11:31:00 2001
  User-Name = "aep@somedomain.org"
  NAS-IP-Address = 62.188.74.4
  Acct-Status-Type = Stop
  Acct-Session-Id = "324546354"
  Acct-Authentic = RADIUS
  Acct-Session-Time = 21
  Acct-Input-Octets = 11567
  Acct-Output-Octets = 3115
  Acct-Input-Packets = 96
  Acct-Output-Packets = 74
  Calling-Station-Id = "01223555111"
  Called-Station-Id = "02075551000"
  Framed-Protocol = PPP
  Framed-IP-Address = 62.188.17.227
```

Figure 1. RADIUS Start and Stop records

From this log we can extract a limited amount of information regarding the content of the communications transactions that took place. The user has been identified (aep@somedomain.org), the number of the caller (01223555111, which is a Cambridge number) and the place being called (02075551000, London), IP address assigned (62.188.17.227), the duration (21s), number of bytes and packets sent and received, type of connection, date and time. The traffic data over time identifies

the change in location of a user despite the common dialed number. As users roam globally with different access telephone numbers, the user identification remains static. In this sense, the collected traffic data is mildly more sensitive than traditional telephone data: where POTS traffic data pivots around a given telephone/ID number, RADIUS data pivots around a user ID regardless of location; therefore disclosing location shifts. The ISP now knows everywhere their customers connect from; information that may be useful to other parties.

Wireless LAN association records

Such mobility becomes more problematic within wireless environments. In a standard wireless LAN environment using IEEE 802.11b, a radio cell size can vary from hundreds of meters in open air, to a small airport lounge. Before the mobile station (STA) is allowed to send a data message via an access point (AP), it must first become associated with the AP. The STA learns what APs are present and then sends a request to establish an association.

The significant records of a centralized association system log are:

time_GMT

Time when a mobile node associates with a base station

Cell_ID

Base station unique identifier in the LAN

MAC_ID

A unique identifier of a mobile device.

```
Time_GMT=20010810010852 Cell_ID=115 MAC_ID=00:02:2D:20:47:24
time_GMT=20010810010852 Cell_ID=115 MAC_ID=00:02:2D:04:29:30
[. . .]
time_GMT=20010810010854 Cell_ID=129 MAC_ID=00:02:2D:04:29:30
time_GMT=20010810010854 Cell_ID=129 MAC_ID=00:02:2D:20:47:24
[. . .]
time_GMT=20010810010856 Cell_ID=41 MAC_ID=00:02:2D:04:29:30
time_GMT=20010810010856 Cell_ID=41 MAC_ID=00:02:2D:20:47:24
[. . .]
time_GMT=20010810010900 Cell_ID=154 MAC_ID=00:02:2D:20:47:24
time_GMT=20010810010900 Cell_ID=154 MAC_ID=00:02:2D:04:29:30
```

Figure 2. Wireless LAN data logs.

In our analysis of collected logs we identified moments where two individuals were alone within a cell, and whether they arrived together. It is tempting to analyze these logs by drawing an analogy with the POTS, i.e. a registration of a mobile with an access point could be seen as the establishment of a phone call between both parties. This analogy is simplistic as it doesn't consider that the Cell_IDs represent places (airport, conference room, restaurants) and the registration timestamps can reveal if two nodes are (moving) together. Data mining of association records (registration and deregistration) can provide sufficient information to draw a map of human relationships. [4]

HTTP requests to a search engine

The above media may involve further traffic data in the form of internet protocols. The GET and POST methods in the Hypertext Transfer Protocol (HTTP) allow a web client to interact with a remote server. In the most common search engines, the keywords are included in the HTTP header as part of a GET method. All the web logs can be transformed to a W3C common log file format that contains the IP address of the client, the connection time, the object requested and its size.

```
295.47.63.8 - - [05/Mar/2002:15:19:34 +0000]
"GET /cgi-bin/htsearch?config=htdig&words=startrek HTTP/1.0" 200 2225
295.47.63.8 - - [05/Mar/2002:15:19:44 +0000]
"GET /cgi-bin/htsearch?config=htdig&words=startrek+avi HTTP/1.0" 200 2225
215.59.193.32 - - [05/Mar/2002:15:20:17 +0000]
"GET /cgi-bin/htsearch?config=htdig&words=Modem+HOWTO HTTP/1.1" 200 2045
192.77.63.8 - - [05/Mar/2002:15:20:35 +0000]
"GET /cgi-bin/htsearch?config=htdig&words=conflict+war HTTP/1.0" 200 2225
211.164.33.3 - - [05/Mar/2002:15:21:32 +0000]
"GET /cgi-bin/htsearch?config=htdig&words=railway+info HTTP/1.0" 200 2453
211.164.33.3 - - [05/Mar/2002:15:21:38 +0000]
"GET /cgi-bin/htsearch?config=htdig&words=tickets HTTP/1.0" 200 2453
211.164.33.3 - - [05/Mar/2002:15:22:05 +0000]
"GET /cgi-bin/htsearch?config=htdig&words=railway+info+London HTTP/1.0" 200 8341
212.164.33.3 - - [05/Mar/2002:15:22:35 +0000]
"GET /cgi-bin/htsearch?config=htdig&words=union+strike HTTP/1.0" 200 2009
82.24.237.98 - - [05/Mar/2002:15:25:29 +0000]
"GET /cgi-bin/htsearch?config=htdig&words=blind+date HTTP/1.0" 200 2024
```

Figure 3. Sample search engine traffic data.

If 'traffic data' residing in logs are analyzed, a great deal of intelligence can be derived. Observing the logs we can see for example, that 212.164.33.3 has requested (in a short period of time) information about "railway+info+London" and "union+strike" in two different requests. We may

identify not only the patterns of an individual's movements on-line, but also interpret an individual's intentions and plans. Or more dangerously one could derive false intentions ("child+pornography" may be a search for studies on the effects of pornography on children). Much more can be ascertained with some data-mining, even if IP addresses are assigned dynamically, allowing for traceability based on habits and interests; and compounded with location data, previous NAS data, etc., a comprehensive profile can be developed.

The shape of things...

Even the Council of Europe acknowledges, in passing, that the breadth of possible traffic data may be problematic.

"The collection of this data may, in some situations, permit the compilation of a profile of a person's interests, associates and social context. Accordingly Parties should bear such considerations in mind when establishing the appropriate safeguards and legal prerequisites for undertaking such measures. [3]"

No such safeguards or prerequisites are mandated nor discussed in detail, however. The convention text is mute on this matter.

Shifting between infrastructures gives different data; but converging infrastructures is even more worrisome. Mobile communications systems magnify the sensitivity of traffic data; wireless LANs were presented as an indication of the shape of things to come as we encounter new protocols and infrastructures, e.g. third generation wireless running IPv6.

Yet governments want access to this information. The collection and access methods currently under consideration are *preservation* (access to specified data of a specific user that are collected by service providers for business purposes), *retention* (requiring all logs for all users be stored beyond their business purpose), and *real-time* (government access to real-time data flows). These powers are enshrined in G8 and CoE agreements; and will be appearing in national laws near you, if they are not already there.

The national laws that enshrine these collection and access powers differ remarkably, despite being established under the umbrella/guidance of the G8 and the CoE. Neither organization places requirements on countries to require safeguards such as limitation on access, or specified purposes for collection, and authorizations through judicial warrants. We believe that national deliberation on these matters will be minimal because governments will claim, as they have, that proposed policies will be due to 'international obligations' and requirements for 'harmonization'. These international policy dynamics thus reduce deliberation and our ability to inform policy discourse.

The landscape for lawful access powers remains quite fragmented. UK law separates URLs from traffic data, and yet practices very weak access constraints (any number of government agencies may access this data); and later proposed retention regimes for periods ranging from 4 days (web cache), 6 months (RADIUS, SMTP, and IP logs), and 7 years [5]. The U.S. recently introduced technology-neutrality to its laws thus reducing earlier protections, but the U.S. has stronger access protections than the UK; and has no retention requirements.

With over 30 signatory states including the U.S., Canada, Japan, Romania, France, and Croatia, we can rest assured that there will be selective interpretation in implementation of the CoE cybercrime convention. Even among the G8 countries, the protections afforded to citizens' communications in Italy, Germany, the U.S. and Russia vary greatly.

While policies may vary, the sensitive nature of the data produced does not. 'Traffic data' analysis generates more and more sensitive profiles of an individual's actions and intentions, arguably more so than communications content. In a communication with another individual, we say what we choose to share; in a transaction with another device, e.g. search engines and cell stations, we are disclosing our actions, movements, and intentions. Technology-neutral policies continue to regard this transactional data as plain old telephone system 'traffic data', and accordingly apply inadequate protections.

Already the Canadian government, claiming its intention to ratify the CoE convention, has proposed to consider all telecommunications services as equivalent. “The standard for Internet traffic data should be more in line with that required for telephone records and dial number recorders in light of the lower expectation of privacy in a telephone number or Internet address, as opposed to the content of a communication.” [6] This is disingenuous: the civil servants who wrote this policy are the same who participated in drafting the CoE and the G8 agreements. Select components of these agreements are thus brought home for ratification while the recommended protections are not.

This is not faithful to the spirit of updating laws for new technology. We need to acknowledge that changing technological environments transform the policy itself. New policies need to reflect the totality of the new environment.

These technology policy innovations fail to do so. Governments seek technology-neutral policy, and are doing so at the international level. This appears to be to the advantage of policy-setters. New powers are granted through technological ambiguity rather than clear debate. International instruments, such as those from the Group of 8 and the Council of Europe, harmonize language in a closed way with little input and debate. This problem will grow as more countries feel compelled to ratify and adopt these instruments; or feel that it is in their interests to do so. Implementation, however, will still be fragmented and will likely be in the interests of increasing access powers of the state.

Attempts to innovate policy must be interrogated, lest we reduce democratic protections blindly.

About the Authors

Ian (Gus) Hosein is a fellow in the Department of Information Systems at the London School of Economics; and a fellow at Privacy International. For more information please see <http://is.lse.ac.uk/staff/hosein/>

Alberto Escudero Pascual is an Assistant Professor in the Department of Microelectronics and Information Technology at the Royal Institute of Technology (KTH) in the area of privacy in the next generation Internet. For more information please see <http://www.imit.kth.se/~aep/>

References

- [1] Ashcroft, J. Testimony of the Attorney General to the Senate Committee on the Judiciary. Washington D.C. September 25, 2001.
- [2] Council of Europe. Convention on Cybercrime, ETS no.185.
- [3] Council of Europe. Convention on Cybercrime Explanatory Report, adopted on November 8, 2001.
- [4] Escudero A. Contribution to the EU Forum on cybercrime. Location data and traffic data. Brussels. November 2001.
- [5] Gaspar, R. Looking to the Future: Clarity on Communications Data Retention Law: A National Criminal Intelligence Service submission to the Home Office for Legislation on Data Retention. Submitted on behalf of ACPO and ACPO(S); HM Customs & Excise; Security Service; Secret Intelligence Service; and GCHQ, August 2000.
- [6] Government of Canada. Lawful Access – Consultation Document. Department of Justice, Industry Canada, Solicitor General Canada. August 25, 2002.
- [7] Podesta, J. National Press Club Speech with (former) White House Chief of Staff John Podesta on "Cyber Security". Washington D.C. July 17, 2000.

[8] Reno, J. Law Enforcement in Cyberspace Address by The Honorable Janet Reno, (former) United States Attorney General. San Francisco: Presented to the Commonwealth Club of California, 1996.

[9] UK Hansard. "House of Lords 28th June, 2000 (Committee Stage)", Column 1012 (published by The Stationery Office Limited).

[10] U.S. Delegation to G8. Discussion Paper for Data Preservation Workshop. Tokyo, G8 Conference on High-Tech Crime. May 22-24 2001.

Endnote

1. The data presented has been obtained with permission from a telephone carrier, an internet service provider, and a large conference where wireless LAN access was provided. All transactions presented in this paper have been de-identified, and the time-logs were altered to reduce the risk of re-identification.