Written evidence on Investigatory Powers Bill: Technology Issues submitted by IT-Political Association of Denmark

Introduction

1. IT-Political Association of Denmark (IT-Pol) is a Danish digital rights organisation that works to promote privacy and freedom in the information society.[1] The activities of IT-Pol are funded entirely by membership contributions. IT-Pol is regularly consulted by news media and politicians in Denmark about the technical and privacy aspects of data retention.

2. IT-Pol is a member of European Digital Rights (EDRi), and we cooperate on an ongoing basis with digital rights organisations in other countries. Data retention in Europa is a key element in our international cooperation.

3. This submission contains comments on the proposal for internet connection records (ICRs) in the draft Investigatory Powers Bill. The motivation for ICRs in the Draft Bill and the outlined retention requirements are very similar to the session logging data retention scheme which was used in Denmark from 2007 until 2014 when it was repealed for lack of effectiveness.

4. For proper context, this submission will contain a brief description of session logging and a summary of the self-evaluation report published by the Danish Ministry of Justice in December 2012. This will be followed by our comments on the proposed ICR scheme and recommendations for the British Parliament.

Session logging in Denmark

5. Danish data retention took effect in September 2007 with ”logningsbekendtgørelsen”[2], and consisted of two parts: the data retention requirements for telephony and internet access in the now annulled European Data Retention Directive (2006/24/EC) and the special session logging requirements for internet traffic, described in paragraph 6-7.

6. Under session logging, the following information about internet packets must be retained: source and destination internet protocol (IP) address, source and destination port number, transmission protocol (like TCP and UDP), and timestamp.[3]

7. The main rule was to retain this information for the first and last packet of an internet session, which is not precisely defined. Alternatively, an internet service provider (ISP) could retain information about every 500th packet at the boundary of their network, and this rule was used by most ISPs in practice. Internal traffic within the ISP's own network, such as DNS lookups,[4] was excluded from the session logging requirements.

8. The specific session logging requirements were negotiated between the Ministry of Justice and the Danish ISPs for a couple of years. The rules reflect a compromise of what the Ministry of Justice wanted to achieve with session logging and what the ISP industry

[1] Website of IT-Pol Denmark: https://itpol.dk/

[2] Logningsbekendtgørelsen (Danish administrative order for data retention). The original order is available at https://www.retsinformation.dk/forms/r0710.aspx?id=2445. An English translation done by the Ministry of Justice is available at https://itpol.dk/sites/itpol.dk/files/TFR40200.pdf

[3] See section 5(1) in the administrative order https://itpol.dk/sites/itpol.dk/files/TFR40200.pdf

[4] DNS lookups translate domain names, such as www.parliament.uk, to IP addresses.

viewed as technically and economically feasible. For large ISPs, it was important from a cost perspective to limit the collection points to the boundary of their networks where traffic is exchanged with other ISPs. In Denmark, the equipment cost of data retention systems is borne solely by the telecommunication companies (access to the data is billed to the police).

9. The purpose of session logging was to retain information about all types of communication between two individuals, whether by telephone or the internet. The main argument for session logging by the Ministry of Justice was that the destination IP address and port number could identify the communication service being used.

10. In December 2012, the Ministry of Justice published a self-evaluation report about Danish data retention.[5] According to the report, communication data from session logging had only been used in a limited number of cases. The only example given in the self-evaluation report was a case involving online banking fraud on a minor scale, and even here session logging played a minor role. The Danish Security and Intelligence Service (PET), which is responsible for domestic counter-terrorism in Denmark, stated in the report that it had only been relevant to request session logging information in a very limited number of investigations by the service.

11. The report also mentioned technical difficulties by the Danish police in handling the massive amount of data available through session logging. In 2013, about 3500 billion records about telecommunication were retained in Denmark (620000 per citizen), of which more than 90 percent was due to session logging.

12. In June 2014, the Danish government decided to repeal session logging.[6] The decision coincided with the Danish response to the data retention judgment from the Court of Justice of the European Union (CJEU) on 8 April 2014. However, the Ministry of Justice emphasised that session logging was not repealed because of the CJEU judgment, but because the session logging rules were unable to achieve the stated objective (investigation and prosecution of crime).

13. The Ministry of Justice has indicated that session logging could be re-introduced in Denmark if the technical problems could be resolved, and the Ministry of Justice would discuss this with the ISP industry. To date, no proposal for a revised session logging scheme has been put forward.

Comments about internet connection records in the draft Investigatory Powers Bill

14. The following remarks about ICRs are based on our ongoing analysis of session logging in Denmark, which we have used for consultation responses, media contact and other advocacy work.

15. The focus in the following is on technical issues, as requested by the Science and Technology Committee. However, we would like to briefly point out that collection of ICR

[5] The data retention self-evaluation report from the Ministry of Justice is available at http://www.ft.dk/samling/20121/almdel/reu/bilag/125/1200765.pdf. There is no English translation. The article ”In Denmark, Online Tracking of Citizens is an Unwieldy Failure” in TechPresident, 22 May 2013, covers the report. Available at http://techpresident.com/news/wegov/23918/denmark-government-will-not-allow-ordinary-citizens-have-digital-privacy

[6] Press release: The Ministry of Justice repeals the rules about session logging, 2 June 2014 http://www.justitsministeriet.dk/nyt-og-presse/pressemeddelelser/2014/justitsministeren-oph%C3%A6ver-reglerne-om-sessionslogning (in Danish)

information will be very intrusive to the private lives of British citizens. The destination IP addresses will, in some cases, contain sensitive information about political and religious preferences of citizens through their choices of online news media, visits to websites of political parties and candidates as well as religious groups and societies.

Definition of Internet Connection Records

16. It is natural to compare ICRs to call detail records (CDRs) from telephony which contain information about the calling party (A-number), called party (B-number), starting time and the duration of the call. CDRs are collected because they are needed to billing in most cases. Even if CDRs are not needed for billing because of flatrate subscriptions, CDRs are easy to create because telephone calls are highly structured.

17. ICRs on the other hand do not naturally exist in the technical infrastructure of an ISP. The information contained in ICRs is used by the ISP for routing internet traffic, but the information is not regularly retained because it is not used for billing. Customers may pay for data volumes, but almost never based on the destination of the internet traffic. This implies that ICRs will have to be created by the ISP in order to be retained.

18. Second, and more importantly, it is inherently difficult to define ICRs in a meaningful way. Unlike the telephone system, the Internet Protocol is stateless. A telephone call has a certain duration, which the telecommunication company can keep track of because a telephone line is in use. Internet traffic is divided into packets with are, in principle, routed independently of each other between the end-points (source and destination IP addresses), and only the end-points can associate the individual internet packets with a particular mode of communication.

19. Retaining information about every internet packet, or even every 500th packet as in the Danish session logging scheme, will generate a lot of data. The typical maximum size of an internet packet is 1500 bytes, and many packets with interactive traffic such as instant messaging or online gaming are smaller. Therefore, the number of packets generated from ordinary internet traffic will be quite substantial.

20. Something as simple as a reading an article from an online newspaper will typically generate thousands of packets, not just to the web server for the newspaper itself, but also to the many third-party elements that are often included on web pages, for example social-media elements and online advertising. Furthermore, lots of internet packets will be generated automatically by services running in the background on most internet-connected devices, for example social media apps like Twitter and Facebook on smartphones.

21. The destination address in an ICR could be an IP address, as in the Danish session logging scheme, or a domain (server) name. The wording in section 71(9)(f) of the draft Investigatory Powers Bill covers both possibilities. When the internet communication takes place, a domain name is translated to an IP address through a DNS lookup. However, this information is not necessarily available to the ISP in a form whereby it can be readily associated with the ICR. Among other things, the customer could use a different DNS server than the one provided by the ISP, for example Google DNS. Therefore, a requirement to retain information about domain names will require some form of deep packet inspection (DPI) on the internet traffic, and with the increasing use of encryption on the internet, this task becomes more complex. If the destination IP address is sufficient for ICR retention, DPI is not required.

22. ISPs will have to purchase special equipment in order to be able to retain ICR information about their customers since this is not a regular task for an ISP. Depending on how ICRs are defined, it could even be the case that ISPs would have to make changes to their technical infrastructure in order to satisfy specific ICR retention requirements. This is something that all British ISPs must prepare for, even if they have not yet been served with an ICR retention notice by the Secretary of State.

23. One of the objectives of ICRs in the Investigatory Powers Bill is to identify the individual device that has sent a communication online.[7] This includes cases where the police has obtained a number of IP addresses and wants to know if they were used by a single individual, and cases where a source IP address may have been used by a large number of individuals at the same time because of carrier-grade network address translation (CG-NAT). In both cases, these types of requests from the police will require the ISP to search the entire collection of ICRs from all customers which, as mentioned in paragraphs 19-20, can be quite substantial. Most likely, ISPs would have to build a dedicated database infrastructure for ICR data in order to support the requirements of the request filter in the Draft Bill.

24. The cost of building systems to retain ICRs and responding to requests from law enforcement under the "request filter" system could be quite substantial. Since these systems are not necessarily trivial to implement, ISPs may have to acquire these systems before they are served with a retention notice for ICRs. This of course depends on the deadline given in a retention notice. These costs, if not fully covered by the British government, are likely to have a substantial fixed element which would effectively discriminate against smaller ISPs and new companies that consider entering the ISP business. That would have negative consequences for the competition landscape and consumer choice for internet access services.

Technical limitations of using ICRs in police investigations

25. Device identification seems to be the primary objective of ICRs, but there are limits as to what devices an ISP can actually identify. In general, the ISP can only identify devices that are connected directly to the ISP. A smartphone can be identified when it uses mobile data for its internet connection as this involves a direct connection between the smartphone and the ISP. However, if the smartphone connects to the internet through a WiFi access point (for example a WiFi hotspot in a hotel or pub), the ISP serving that access point only sees connections coming from the access point device itself. This means that the ISP is unable to distinguish between the individual devices (for example smartphones and laptops) that may be connected to the internet through the access point.

26. In paragraph 21, we pointed out that requiring domain (server) names in the ICR would add considerably to the complexity, and most likely also the cost, of retaining ICRs. On the other hand, ICRs that only contain destination IP addresses may be of limited use for the police in some cases. First, multiple websites are often hosted on the same server with the same IP address due to the scarcity of IPv4 addresses. A so-called reverse DNS lookup, which returns a domain name from an IP address, will not be unique in this case and could point to the wrong server (than the one actually accessed). Secondly, IP addresses of some servers change over time, so the DNS information may have changed between the time of the actual communication and the police investigation, which could be up to 12 months later. The underlying problem here is that the internet is much more dynamic than the

[7] Purpose 1 on page 14-16 of "Operational Case for the Retention of Internet Connection Records", part of the overarching documents for the draft Investigatory Powers Bill.

telephone system.

27. The ICR data set for an individual will contain a complete profile of the behaviour on the internet of that individual, subject to the limitation mentioned in paragraph 28 below. This includes all communication services accessed, at least to the extent that they can be determined from destination IP addresses. However, the number of records in this data set will be really large. Many individuals use file sharing software or online gaming that tend to generate lots of internet packets to unknown IP addresses. This is how peer-to-peer systems function. Looking for traffic to illegal websites or communication services could be a “needle in the haystack” problem. In Denmark, this has been a practical problem for the police when analysing data from session logging.

28. It is very easy for an individual to hide the final destination of the internet traffic and make the ICR data useless from the viewpoint of law enforcement. If the individual uses a VPN connection, the destination address in the ICR will be that of the VPN server, not the real destination of the traffic. VPN providers in the United Kingdom could also be subjected to ICR retention requirements, but foreign VPN providers will not, in general, be subject to similar requirements. Another possibility is to use Tor (a well-known anonymisation network), in which case the ICR will contain information about random Tor entry nodes, not the final destination of the traffic. It is very likely that the use of VPN and Tor will increase when the public becomes aware of ICR data retention.

Recommendations for the British Parliament

29. First of all, IT-Pol would encourage the British Parliament to carefully consider whether data retention with ICR is really a feasible project. The Danish experience with session logging, as documented in this submission, has been strongly negative, even from the viewpoint of law enforcement (that is, the considerable privacy concerns aside).

30. If ICR data retention is adopted, the specific retention requirements should be negotiated with the ISP industry. The ISPs have the technical expertise as to what is technically feasible, and only the ISPs can make realistic projections about the costs. In Denmark, the final session logging requirements were quite different from the original “wish list” of the Ministry of Justice.

31. Even if it is possible to build an ideal ICR retention system, from the viewpoint of law enforcement, it will be really trivial for individuals to circumvent this type of data retention by using VPN connections or the Tor network.