Andrews & Arnold Ltd
Enterprise Court
Downmill Road
BRACKNELL
RG12 1QS

10th Dec 2015

Internet Connection Records

This document is submitted as additional written evidence following oral evidence given on 9th Dec. The purpose it to try and clarify the meaning of “Internet Connection Records” and provide an easy to understand technical background on the challenges facing any communications provider in creating and retaining such records. I appreciate the members time in reading this document.

Adrian Kennard

---

1. History

Once upon a time telephone companies were the only real providers of any sort of electronic communications, and the “telephone call” was the basic building block of that service. The telephone companies did not originally have any sort of logs of telephone calls made, but as telephone exchange equipment become more sophisticated they were able to create itemised telephone bills by recording the details of each call made. These logs are called CDRs (Call Data Records).

The concept of a telephone call is very simple, and the idea of a CDR is simple too. There are some possible complications with diverted calls and three way calls, but even so, the basic log of what number made a call to what number is easy to understand. Logs can also include calls that are being received at a “line”, and can even include calls that were not actually answered.

Obviously police access to such records was invaluable in helping criminal investigations. Eventually this became part of RIPA and the Data Retention Directive and then DRIPA.

It is worth bearing in mind that this started to happen before much was considered on Data Protection or privacy, and if such logging was being introduced now it would no doubt be a major concern for privacy groups.

With the advent of GSM and digital mobile phones, the logging was extended to include text messages.

With the advert of Internet email, the logging was extended to include emails. It is worth noting that email logs are not normally necessary for commercial reasons as there is usually no per-email charge, so this is the point at which the logging became more of a specific service to assist law enforcement rather than simply having access to what data was already there for commercial or operational reasons.

Whilst emails are not quite as simple as telephone calls, they are a relatively simple concept in terms of logging - with an email having a sender, and one or more recipients which can be logged.

Both emails and telephone calls are pretty tangible as a single “communication”, with a start and an end, and a content and addressing for that communication identifying the parties involved. Text messages are, however, an example where this breaks down a little - a logical “communication” may be an ongoing exchange of many messages making for a conversation over a long period.

2. Over the top services

The Internet has become popular with some key services, indeed, some people talk of “the Web” and “the Internet” interchangeably because “web pages” were seen very much as the only thing that the Internet does (apart from, perhaps, email). Many of the more innovative features of modern communications can seem to boil down to “web pages” in that one can access Facebook, and Twitter, and email via “web pages”.

In light of this, the notion of simply logging web page accesses seems a relatively simple concept, and it is easy to see how this is seen as a logical extension of the call, text and email logging of the previous data retention regimes.

Accessing a web page is also seen as a pretty clear cut “communication”, again with a time, and a person involved and an address of a web site where content is fetched or viewed.

However, the problem is that this is not actually how the Internet works.

Even logging of emails is only sensibly done at a point where an “email service” is handled. There are bits of equipment that provide “email” and these bits of equipment make use of the underlying “Internet” connectivity to do so. It is crucial that previous regulations referred to “generated or processed” in terms of logging data, as email is only processed at an “email server” and not in the interconnecting Internet Service Providers. It used to be common for an Internet Service Provider to also provide the email services, but that is much less so these days.

Email is what is called an “over the top” service. It means that email is a service that exists on top of other means of communications, like Internet access. You can log an “over the top” service where there is some service provider who has some processing function such as an “email server”.

As an analogy, in the telephone world, an “over the top service” might be something like “pizza ordering”. You would not expect the phone company to log what pizzas people order (by listening in to they calls) even if that is technically possible, but you might expect every pizza company to log orders that are placed if that had some benefit to law enforcement. The location of the logging relates to the service you are logging.

3. Building blocks

Looking back at telephone service, there is a building block to that service which is the “telephone call” itself. Telephone calls are logged for commercial and operational reasons.

However, the Internet does not work at that level. Even the idea of a “connection” of some sort, such as a “connection to a web site” is an “over the top service” created by the equipment at each end. The underlying “Internet service” uses something called “packets”. Each packet has a destination address (called an Internet Protocol, or IP, address) which works much like a telephone number to identify where the packet is to go.

However, each packet is not really a “communication” in a meaningful sense - it is some small fraction of a communication. The Internet service providers do not work in large chunks like a “telephone call” they work in these small “packets”. It is even possible for some packets to go via one Internet provider and some go via another in the same logical “communication”.

Also, unlike phone calls, there are a lot of these packets - seriously a lot. Even as a small ISP we may pass on literally billions of these packets every minute, and larger ISPs move colossal amounts of data. There is no built in logging of these for commercial reasons - there is no charge based on what the packets are and where they are going. At best, some totals for overall volume of data to/from each customer is recorded. The equipment to make the packets move towards their

destination (called a “router”) is carefully engineered to just look at the destination address of each packet and move that packet one step closer to its destination. This is often done in very fast, expensive, and optimised computer hardware that is designed to do that one job very fast. The packets are not even “looked at” by a “computer program” as such as that would take too long.

Some equipment does have some built in ability to collect some basic statistics, and using such equipment it may be possible to get some logs of some “logical connections” or “flows” that are made - where lots of packets with the same IP addresses are being sent. However not all equipment has this capability, and equipment that does may not be able to record everything in detail.

---

4. Logging web pages

The only logical place to log web page accesses is either at the web browser (the browser history), or at the web server (web access logs). The place that does not make any sense to log web pages is in the Internet Service Provider. This is because, like any “over the top” service, the browser and computer breaks down what it is doing in to packets of data, and sends these over the Internet. The final web server reassembles all of the pieces and accesses the web page in question.

The same is true for logging emails - the sending machine (PC), the email server in the middle, and the receiving machine all see an intact email, and could log it. The ISP sees just lots of small packets in-between. This is why emails are logged at an “email server”.

It is a bit like saying that the postal service have to log letters sent, but they are thwarted by the fact that every sender puts the letter through a shredder first and each shredded bit of each letter is being delivered, mixed in with every other letter, to a destination where it is glued back together.

I appreciate that this sounds crazy - but really, that is how the Internet actually works. **If you want to log anything, you really need to log it where the communication is intact**.

---

5. Beyond “the web”

However, having explained a bit about web pages and email, even if you can log at web servers and email servers, the Internet is changing massively.

Smart phones are the key here, and are used by everyone. Unlike conventional PCs which may only have a web browser and an email client, smart phones have “apps” (application programmes). These talk to services over the Internet.

When a web browser communications with a web site, or an email client communications with an email server, it follows a well documented standard. If you picked up all of the shredded paper (the packets) and reassembled it, you could make sense of what was going on, technically, with a lot of work (and cost).

However, when a smart phone “app” communications with a server, it does not have to follow any such standard. It simply has to be something understood by both ends. So there is no way to know what is going on. Each app can be, and is, different.

They also do not communicate in small busts like “sending an email” or “accessing a web site”, but instead they keep a connection (or many connections) open all of the time - especially social media and messaging apps. That one, on-going connection, can logically be involved in lots of different “communications” with lots of people, none of which is “seen” by the ISP. Much like logging email at a mail server, the only sensible place to log “social media” is at the “social media company”, not the ISP.

Even when you ignore mobile phones you have to consider “games consoles” which again do not follow standards and just need both ends to understand what is communicated. That is a massive area where people can “communicate” in-game. Again, meaningful logging at the ISP is mostly impossible.

Unfortunately, with any “over the top” service, the provider may not be in the UK and subject to UK law, making logging even harder.

But it gets worse - we now see “the Internet of Things” becoming more and more of a reality with the rise of smart phones and intelligent devices, smart thermostats, smart fridges, all sorts of things in people’s homes. This means that more and more of the communication that you see is not a matter of “a person accessing a service”. It means that there is a hell of a lot more “chatter” going on from devices, all of the time.

---

6. Encryption

There is one more complication. I have likened the way the Internet works to shredding the letter you are sending, and sending all of the bits of paper from the shredder separately. This is quite a good analogy as you can see that, with a lot of work, you could put the bits of paper back together and see what is going on. After all, the far end does so. It is very much like the bits of paper are each addressed and numbered to make that a bit easier.

However, encryption is essential to maintaining privacy and security, and this means you cannot see what is on the bits of paper any more. Yes, the addressing is there, but nothing else.

This means that any attempt to create any sort of logs of what is going on with an “over the top” service is thwarted. You cannot see in to the messages being passed to understand what is happening. At best you can see the sender and recipient of those packets of data.

Even the final addressing can be misleading as there are many services that re-route traffic (VPN and Tor and others) to hide the real source and destination of the packets. Even where you can see the destination it can simply be some common “over the top” service like iMessage which gives no clue to the real “communication” that is going on using that service, and the service provider will not see “inside” the messages if they are doing it right.

To make matters even worse a lot of services make use of “content delivery networks”. These are separate service providers that specialise in delivery of data all around the world. If you see the addresses of packets going to/from one of these you have no real clue what is being communicated as the same content delivery network can be hosting data for NASA or BBC or Facebook, or even a terrorist organisation (though CDNs are unlikely to do so knowingly).

On the matter of encryption, and I cannot stress this enough, **the battle against encryption is a lost cause**. You cannot ban encryption or force encryption to have a back door, side door, golden key, escrow key, or weak link. Encryption exists - it is not a secret! It is possible to encrypt a message with no more than pen and paper and dice such that it can never be decoded by a third party without the keys, no matter how much time or computing power they have. It is possible send encrypted communications that are hidden in other data (like images and video) so that there is no way to tell there is a secret message, so even making encryption illegal does not help. Any attempt to reduce the effectiveness of encryption will ultimately have no impact on criminals, even if you make it illegal (they are criminals, remember), but will have an impact on the legitimate use of encryption by normal citizens and businesses. You can never have a back door that is only available with a court order - **mathematics does not understand court orders**, and any sort of back door makes the communications vulnerable to attack by criminals. Please, give up on all attempts to impede encryption. **Embrace encryption** as a crucial tool for security, privacy and the

economy. Encourage encryption, and digital identities, and value the benefits that this brings to society. Find other ways to understand what criminals are saying (getting data at the end points or infiltrating the criminal communities and getting inside their networks).

---

7. Self service

I have mentioned logging phone calls and emails, and that you log those at the point the service is provided as an “over the top” service. Even phone calls over the Internet, whilst almost impossible to log when looking at the packets of data, can be logged at the “telephone service provider”. The same is true for emails, even where the links to email servers are routinely encrypted, the addressing of the email can be logged at the “email servers”.

There is, however, an increasing trend for applications and services to exist which do not rely on a “service provider”. It is, for example, possible to call me using a “number” which, if you have suitable phone, connects directly to my equipment under my control, and there is no “telephone service provider” to see the call, or log that it happened. The same can easily be done with emails where end users can operate their own email servers.

Whilst running your own email or telephone server is more rare, at the moment, applications on phones are more and more working directly, end to end, by themselves without relying on an intermediate service provider. The use of an intermediate service provider is seen as a weakness and point for criminals to attack. They also use encryption end to end. This means that the only logging that could be done is of the packets of data with no visibility in to that data at all and very likely no idea of what application is being used, even.

---

8. What does the Draft Investigatory Powers Bill say?

"Internet Connection Record" is not a defined thing - in the bill or in industry!

In the oral evidence session David Hanson MP seemed adamant that an "Internet Connection Record" was "defined in the bill". He referred to page 25 and asked us to work out costs based on that definition. Page 25 is in the "explanatory notes" and not the bill, and itself is massively unclear. It basically says "*It is a record of the services that they have connected to*".

I fully understand that to someone not technical, saying "*It is a record of the services that they have connected to*" seems reasonably clear. Sadly it really is not, and if you look at the actual wording of the bill, and not just the explanatory notes, it is less clear still. Remember, all an ISP sees is “packets” - those shredded bits of communications passing through the network. I hope much of the above explanation makes that more obvious.

Unlike a telephone call, or even just sending an email, even the definition of the term "connected" is complicated, as is defining the term "service". Actually what happens is packets of data are sent between devices, and as an ISP we send those packets on towards their destination. We don't "see" any sort of "connection" or "service", all was see is “packets”. The idea of “connection” is abstract and defined by the end devices.

Ideally what this means is that web sites log any access, and email servers log any emails, and telephone servers log any telephone calls. Each is an “over the top service” and not something the ISP tries to “log” or “retain”. This is where such logging makes sense and is comparatively simple - though the concerns over storing such data securely still exist. The problem here is that many of these services are not in the UK. If you expect a foreign web site to log web accesses for you and provide data to the UK, they would expect to also provide to any other government too, such as US, or France, or China, or North Korea or Syria. I think most providers are less than keen to do that.

One possible meaning could be that we log the destination IP address of each packet. Sadly this is neither easy nor cheap as there are literally billions of such packets whizzing through our network every minute, and we are a small ISP. I do not see that being useful to law enforcement in any way. Remember may IP addresses are not the real final destination or may be some content delivery network shared by many services.

There is a protocol for a type of "connection" used in the Internet, called TCP. This is only one of many types of connection that can be made but is the most common and is used by email and web pages. It is a standard, which helps a little. So the meaning could be to log each such logical TCP connection. This would mean making something of a jigsaw puzzle of the meta data (the destination and source addresses) in each of those billions of packets as they pass and tracking millions of simultaneous logical "connections" that are happening at any one time, then logging these. Again, this is neither easy nor cheap, and even more work than above. There are also many types of "connection" - an "Internet phone call" using a protocol calls SIP does not normally even use TCP but a "connectionless" protocol called UDP, so somehow that would need to be tracked and logged too. There is no rule that applications have to use these common protocols such as TCP and UDP either, they can make stuff up and use what they like as long as both ends understand it.

Of course, it could be that what we must log is more a matter of logging each "web page" accessed with the name of the web site, and similarly for other “services" that are not actually “web pages”. Indeed, some comments made by the Secretary of State suggested this may be what was meant. This means not only the jigsaw puzzle to construct those TCP connections, but actually looking in to the data that passes on those connections, connecting the data from many packets together, and looking for a part of the information sent called a `Host:` header. This is yet more complexity and work and cost. Again, web pages are just one type of communication that uses a "connection". There are many other types of "connection" that could be made, and new types will come along every day or even every few hours as new applications are developed and new innovations made. Each of these is not published - we know how "web pages" work because they follow a published standard, but mobile phone apps do not have to follow any such standard, they do not even have to use TCP to communicate. So we'd have to constantly research each and every new application and protocol that people invent anywhere in the world, work out what part of that data counts as "Relevant Communications Data" and record it in some format that the police know to ask for and understand. We would not have the help of the developers in this. **Indeed, we'd have to buy and test every app ever published and reverse engineer it to work out what to log.** That would be a huge on-going undertaking at huge cost, made massively worse by the fact that each ISP is on their own not allowed to tell anyone else what they are doing with data retention.

As worded the bill does not define what is to be logged, and nothing stops an order to log and retain “all relevant communications data” with no details being imposed on all ISPs, schools, offices, or even home networks.

So the meaning of recording "what services you connect to" is really very very unclear, and the cost involved in making such logs is not something one can sensibly estimate without actual details.

---

9. Future

The ways that these things work is constantly changing, with new trends in technology, changes in usage by real people and devices, and innovation. This can only mean less useful information and more noise and useless data over time.

Whilst some information is still likely to be obtainable, it is obtainable only in certain places - such as email addresses logged at the mail servers. Trying to extract information from packets of data as they pass through an ISP is pretty futile now, and will become more so over time.

It makes sense for service providers to try and keep some logs and try and help law enforcement where it is proportionate to do so considering costs and privacy. Indeed, one would hope that the likes of Facebook would be keen to help with any serious criminal investigation. Over the top service provides is an obvious target for logging and retention, up to the point that they can - but any sensible provider has end to end encryption and no logging for good security reasons.

ISPs will have some operational data, and will more than likely be able to trace an IP address to a customer for a short period of time - this is often needed in some way for operational reasons, and for some ISPs with “fixed IP addresses” it will be easy to do so. The new protocol - IP version 6 - will help with this, but still not track an address to an individual device at a premises.

Sadly, even normal phone calls and text messaging and emailing, for which logging is comparatively simple, are disappearing and making way for social media and new ways to communicate. Trying to log these new services in the ISP is increasingly pointless - they need logging at the service providers, where that is possible, if the (non UK) service provider co-operates.

Whilst this is a shame for law enforcement, and forces more reliance on “traditional police enquiries”, the increasing trends in use of social media and freely sharing information with friends should help those traditional methods find leads - especially when considering examples like a missing child as often touted as a reason for needing data retention at all.

Indeed, simple cases like a missing child - if the phone is on - it is way simpler for the parent to use an app like “Find my iPhone” on Apple with family sharing to locate the child’s phone within meters than for police to make a RIPA request to a mobile operator (with much less accuracy and taking much longer). People are more and more sharing personal data in smaller family and friend groups (as well as publicly) and this hopefully makes life easier for law enforcement not harder!

---

10. Helping define the data types

I repeat my offer to assist in defining clear data types if that would help clarify the bill. I feel it is crucial to clearly define what is to be logged and by which parties and in what context.

I would also be happy to try and provide more technical training on how the Internet works to members if that would be of use, but I recognise the extremely limited time available to the committee to consider this bill.

I hope this submission has been useful, and welcome any questions or requests for clarification.

—
Adrian Kennard