It's only the beginning:
Metadata Retention laws and the Internet of Things

Clinton Fernandes
University of NSW

Vijay Sivaraman
University of NSW

Abstract:

This article examines the implications of selected aspects of the *Telecommunications (Interception and Access) Amendment (Data Retention) Act 2015*, which was passed by the Australian Parliament in March 2015. It shows how the new law has strengthened protections for privacy. However, focusing on the investigatory implications, it shows how the law provides a tactical advantage to investigators who pursue whistleblowers and investigative journalists. The article exposes an apparent discrepancy in the way 'journalist' is defined across different pieces of legislation. It argues that although legislators' interest has been overwhelmingly focused on communications data, the explosion of data generated by the so-called Internet-of-Things (IoT) is as important or more. It shows how the sensors in selected IoT devices lead to a loss of user control and will enable non-stop, involuntary and ubiquitous monitoring of individuals. It suggests that the law will need to be amended further once legislators and investigators' knowledge of the potential of IoT increases.

Introduction

On 26 March 2015, the Australian Parliament passed data retention laws after five years of community uneasiness, rumours, discussion and debate about the extent to which such laws may violate people's privacy. Known as the *Telecommunications (Interception and Access) Amendment (Data Retention) Act 2015*, it was the Abbott Government's third tranche of national security legislation (after the *National Security Legislation Amendment Act (No. 1)* 2014 and the *Counter-Terrorism Legislation Amendment (Foreign Fighters) Act 2014*). It compels telecommunications providers to retain a defined set of data for a period of two years, whilst reducing the number of agencies able to access such data. The law received Royal Assent in April 2015 and comes into force on 13 October 2015.

The confusion over metadata

When the Attorney-General and the Minister for Communications announced the successful passage of the Bill, they said, 'Metadata is the basic building block in nearly every counter-

http://doi.org/10.18080/ajtde.v3n3.21

terrorism, counter-espionage and organised crime investigation. It is also essential for child abuse and child pornography offences that are frequently carried out online' (Brandis & Turnbull 2015). Earlier that same day, however, the Attorney-General had said in Parliament that 'the word "metadata" ... is a jargon phrase of no clear and specific meaning and no technical standing' (Brandis 2015). Indeed, few aspects of the Bill caused as much controversy and confusion as the term 'metadata.'

The confusion had started as early as 2010, when rumours about the Labor Government's plans for a data retention scheme surfaced in the technology press (Grubb 2010). The Senate Standing Committees on Environment and Communication began an inquiry soon afterwards, and issued a report listing a range of criticisms about the data retention proposals such as its inconsistency with the Privacy Act, its unreasonably differential treatment of online and offline information, and claims that the scheme would in any case be 'unlikely to be effective or useful to law enforcement' (SSCEC 2011).

The Government subsequently announced plans to review the proposals via public consultation, and in July 2012 released, via the Attorney-General's Department, a Discussion Paper on the proposed national security reforms. A central part of the argument for a data retention scheme was that technological change and changes to industry structure, practices and consumer behaviour were rapidly eroding agencies' ability to intercept telecommunications and communications data, which were 'unique and fundamental tools that cannot be replaced by other investigative techniques... cost effective, timely, low risk and extremely successful tools in obtaining intelligence and evidence' (AGD 2012). Without such reforms, the argument went, there would be a serious decline in agencies' capabilities.

As the push for a data retention scheme gained renewed momentum after the September 2013 elections, the new Government pointed to advice from the Australian Federal Police (AFP) that 'between July and September of 2014 telecommunications data was used in 92 per cent of counterterrorism investigations, 100 per cent of cybercrime investigations, 87 per cent of child protection investigations and 79 per cent of serious organised crime investigations' (Brandis & Turnbull 2015).

Confusion over the meaning of 'metadata' intensified after the new Attorney-General George Brandis was unable to explain what the term meant in an interview with Sky News:

> Brandis: "The web address, um, is part of the metadata."
>
> Journalist: "The website?"
>
> Brandis: "The well, the web address, the electronic address of the website. What the security agencies want to know, to be retained is the, is the electronic address of the website that the web user is . . . "

http://doi.org/10.18080/ajtde.v3n3.21

Journalist: "So it does tell you the website?"

Brandis: "Well, it, it tells you the address of the website."

It later transpired that Senator Brandis had been trying to say that investigators wanted telecommunications companies to retain the IP addresses of users' devices rather than 'web address', because URLs and browsing histories were specifically excluded from the data retention proposals.

There is, of course, no formal definition of 'metadata' in Australian telecommunications law. The predecessor to the 2015 law, known as the *Telecommunications (Interception and Access) Act 1979*, was silent on the types of data that telecommunications providers had to retain for law enforcement and national security purposes, and also for how long they had to retain it. Different providers therefore retained different types of information based on their individual billing, marketing and taxation requirements. The result was a considerable variation across the industry. The word 'metadata' did not appear anywhere in the new law either; instead, s. 187A(1) lists certain categories of information such as the characteristics of the subscriber, account, device and service, the source, destination date, time, duration and type of communication, and the location of the equipment or line used in connection with a communication. These categories of information allow investigators to establish primary facts and the likely existence of a (criminal) network in order to determine whether to seek a warrant for a listening device.[1]

The new privacy protections

In listing specific categories of information to be retained, the new law adopts the approach of defining 'metadata' specifically and narrowly, and 'content' non-specifically and therefore broadly. It thus avoids the conceptual error of trying to list content in any detail, because such an approach would have the unwelcome consequence of including within the reach of the law, by omission, content that the Parliament actually wanted to prevent authorities from accessing. The law also contains an explicit prohibition against warrantless access of information that falls outside the listed categories. It specifically prohibits warrantless access of web-browsing histories.

Whereas previously metadata could be accessed by such bodies as the Victorian Taxi Directorate or the Wyndham City Council, the new law ensures that only specified agencies – in the main, crime or corruption investigative agencies – can access metadata. It reduces the number of agencies empowered to access metadata from 85 to 21. However, more agencies can be added to this list. The newly-created Australian Border Force was added a few months after the passage of the Act.

http://doi.org/10.18080/ajtde.v3n3.21

In Parliament, the Attorney-General gave an assurance that the new regime would not be used to prosecute citizens who downloaded content in breach of copyright. Since the metadata access regime is concerned with enforcement of the criminal law, and breach of copyright is a civil wrong, there was, he said, no capacity for anyone other than the 21 crime or corruption investigative agencies to access metadata. Although there are criminal provisions in the *Copyright Act*, they do not relate to breach of copyright but to other actions such as internet piracy. The law also provides (section 281(2)) that a party to a civil action is unable to obtain a subpoena or a notice of disclosure in respect of a breach of copyright claim or any other form of civil claim against a telecommunications service provider that retains metadata solely for the purpose of being compliant with the provisions of the Act.

The problem of whistleblowers and investigative journalists

Despite the new privacy protections, and notwithstanding the ability of technically knowledgeable users to conceal their tracks (Huston 2015), the Act may have a negative effect on journalism in Australia by the all-encompassing and retrospective nature of the data retention regime. By requiring the storage, for a period of at least two years, of everyone's e-mail addresses, phone and VoIP numbers, time, date, type and duration of all electronic communications, location information such as cell tower data, and the names and addresses of the parties, the Act provides the authorities with a digital picture of users' movements, contacts, interests, hobbies, associations, etc over a period of two years. Such a picture may not be comprehensive, but it is nonetheless very revealing. By way of example, one might be able to make inferences about a young woman who makes a long phone call to an abortion clinic and then her boyfriend but not her mother. Police investigators hunting down a whistleblower might similarly be alerted by a public servant who phoned a newspaper's switchboard and whose phone's geo-location then matched that of the journalist who wrote about a leaked document.

The specific problem for journalists (and members of parliaments) and their sources stems from the fact that in Australia, whistleblower investigations are treated as leak investigations. Such investigations are carried out under Section 70 and section 79 of the *Crimes Act*. Section 70 forbids disclosure of certain information by Commonwealth officers ('any fact or document' in their knowledge or possession 'by virtue of having been a Commonwealth officer'). It attracts a penalty of up to two' years in prison. Section 79 of the *Crimes Act* prohibits the disclosure of 'official secrets', which include 'a prescribed sketch, plan, photograph, model, cipher, note, document or article, or prescribed information'. It attracts a penalty of up to seven years in prison if the offence is carried out with the 'intention of prejudicing the security or defence of the Commonwealth'. There are no public

http://doi.org/10.18080/ajtde.v3n3.21

interest defences or exemptions for journalists. Section 79 has two important 'secondary disclosure' provisions (sub-sections (5) and (6)) that criminalise the receipt of leaked documents by journalists if they know or are reckless to the fact that the disclosure was made without authorisation. A journalist faces two years in prison, and up to seven years if the offence is carried out with the 'intention of prejudicing the security or defence of the Commonwealth.' Once again, there are no public interest defences.

Before the 2015 data retention law, police and prosecutors had faced the obstacle of not being able to prove the mental element – that the journalist in question knew or was reckless to the fact that the disclosure was made without authorisation. All journalists had to do was shut up and not incriminate themselves, and the investigators would find it very difficult to disprove that the journalist believed when publishing the material that it was just another one of many government-authorised disclosures.

However, the two-year data retention provisions mean that investigators will be able to look back over suspected whistleblowers' connections, interests, contacts, hobbies and other patterns of life over a period of two years in retrospect – long before the journalist printed the story in question. They may have the ability to see when and how a source contacted a journalist, how the journalist managed the process of interacting with the source, whether there were covert tactics involved, and other evidence that allows the construction of a case to demonstrate that the journalist knew that the disclosure was unauthorised.

As such, the data retention law tips the balance against the media by making it much easier for prosecutors to prove the mental element of the offence. Given their low level of technical proficiency, most journalists will find it almost impossible to guarantee confidentiality to a source. Even if a few journalists are technically proficient enough to encrypt their communications – and most are not – the 'first contact problem' (the fact that a whistleblower has contacted a journalist) remains.

A narrower definition of 'journalist'

Curiously, the new law refers to 'a person who is working in a professional capacity as a journalist' (Division 4C) – a definition apparently borrowed from the *Criminal Code* (Division 119). This is a narrower definition of journalist than that contained in the shield law designed to protect the confidentiality of journalists' sources. This law, known as the *Evidence Amendment (Journalists' Privilege) Act 2011*, says (Division 1A) that a journalist is 'a person who is engaged and active in the publication of news and who may be given information by an informant in the expectation that the information may be published in a news medium'. A news medium means 'any medium for the dissemination to the public or a section of the public of news and observations on news'. If the aim were to protect

http://doi.org/10.18080/ajtde.v3n3.21

journalists, then the broader definition is to be preferred. Reformist politicians might wish to amend the new law's definition of 'journalist' to bring it in line with the *Evidence Act*. Consistency of terms across legislations is a worthy goal.

Journalist Information Warrants

In response to concerns about the chilling impact of the new law on journalism, Division 4C creates a category of 'journalist information warrant', which is a requirement that police obtain a warrant before they obtain a journalist's metadata for the purpose of identifying a source or sources. The law also creates a position of Public Interest Advocate, appointed by the Prime Minister, to make submissions to the warrant issuing authority that a journalist information warrant should not be issued because the public interest would be harmed.

There is a danger that such a Public Interest Advocate may well prove to be only a cosmetic measure, if the past track record of approvals for telecommunications interception warrants is any guide. The 2013-2014 Annual report shows that federal judges and Administrative Appeals Tribunal members approved 4,007 interception warrants out of 4,025 warrant applications. Only 18 were refused or withdrawn (AGD 2015). That means that warrant applications are successful 99.6 per cent of the time – one of the rare areas of public administration where such a high success rate is enjoyed. One commentator has described the warrant process as 'a rubber stamp mechanism' (Dorling 2015).

The Internet of Things

Legislators' interest has been overwhelmingly focused on communications data as an investigative tool. However, of arguably as much importance (or more) is the explosion of data generated by the so-called "Internet-of-Things" (IoT). The term IoT (Gubbi et al 2013) refers to the network of physical objects or "things" embedded with electronics, software, sensors and connectivity to enable objects to exchange data with the manufacturer, operator, and/or other connected devices. Such devices are finding increasing use in a wide range of applications: smart-homes and buildings (smoke-alarms, motion sensors, security cameras), health monitoring systems (wearable medical and fitness appliances), automobiles, industrial automation environments (fridges, climate-control, machine telemetry), and agriculture (soil monitoring, irrigation control). Forecasts by Gartner and ABI estimate the number of IoT devices will reach 26 to 30 billion by 2020. Interestingly, these devices will communicate among themselves and with the 'cloud' automatically, i.e. objects will exchange data without the knowledge of the user, revealing information on individuals: location, interests, health, habits and so on. Studies by Notra, Sivaraman et al (2014) establish some markers on the kinds of information captured by IoT. These studies shed light on the kinds

http://doi.org/10.18080/ajtde.v3n3.21

of information malicious attackers may be able to obtain, and also serve as a pointer to the way in which IoT data may serve as a legitimate investigative tool if it were added to the s. 187A(1) list of information that must be stored for two years.

IoT devices may be categorised into three groups based on the way they communicate –

- those that are directly controlled by the user smartphone (Direct Access);
- those controlled by an external server (External Access);
- those that use the smartphone as a gateway (Transit Access) (Sivaraman et al 2015).

This simple categorisation based on the communication model makes it easier to identify vulnerable points, as briefly summarised next. A detailed study of each of the devices discussed below can be found in (Sivaraman et al 2015).

In the Direct Access model, the user's mobile app communicates directly with the IoT device, typically over the home WiFi network. In turn, the device then updates the cloud-based server on its current status. One example of such a device is the Philips Hue light bulb – the user can change bulb colour and intensity via the app while at home, which is communicated to the bulb via the home network and takes effect immediately; subsequently the bulb updates the external server on its current settings, so that the user can also check and change status of the bulb via a web-portal (e.g. when away from home). The Belkin WeMo motion/switch is another example of a direct access device, and works in a similar way to the bulb. The direct access model is most suitable for devices that the user needs to be able to control from within the house even when Internet connection is lost. Many household appliances such as light bulbs, power switches, and door locks fall into this category. Typically, these devices use WiFi, which makes them susceptible to be snooped upon by eavesdroppers who can steal user credentials to gain unauthorised access.

In the External Access model, the user does not interact directly with the IoT device. Instead, the device communicates over the Internet with the external cloud-based sever. The user's app retrieves relevant data (such as current status) and issues commands (to change status) by communicating with the external server (over the Internet). The advantages of such as approach are that

> (a) the server can process the raw data from the IoT device before serving it to the user for consumption via the app, and
>
> (b) the communication can be better secured at the server via end-to-end encryption and authentication, rather than relying on the mobile app.

The drawbacks of this model are that

http://doi.org/10.18080/ajtde.v3n3.21

(a) the user may lose the ability to monitor or control their device even from within their home if their Internet connection is down, and

(b) the user has no visibility into the data that is being exchanged between the IoT device and the server, which can be a concern that private data is being uploaded without the user's knowledge.

Example devices in this category include the Nest Protect smoke-alarm (now owned by Google) and the Withings baby-monitor.

The Transit Access model is typically used by smaller IoT devices that have only a short-range wired or radio connection (example short-range radio technologies include Bluetooth, BLE, Zigbee, RFID and NFC), and do not run the TCP/IP protocol stack. They tether with the smartphone, which is used as a relay to communicate with the cloud-based server over the Internet. The smartphone app may or may not be able to control the device directly or see the data in transit to the server. Examples of this model include the Fitbit fitness tracker and the Withings weighing scales. The transit model is suitable for wearable devices that need to keep form-factor small and energy use low; in terms of security, they may be vulnerable both on the short-range radio link, and on the Internet link from the smartphone.

Security and Privacy risks

In Sivaraman et al (2015), two representative IoT devices from each category were examined. Their security and privacy risks were profiled in experiments that ranged from passive snooping of personal information to active man-in-the-middle attacks. The findings were as follows:

Direct Access devices are the most susceptible to direct security attacks. Transit access devices are protected by the fact that they use very short-range radio (Bluetooth or Zigbee) that are difficult to snoop on, due to constraints on proximity and sniffer hardware, and moreover they do not run full TCP/IP stacks and are hence not always online. External Access IoT devices are relatively easier to be secured by manufacturers, since they have full control over both ends of the communication (namely the IoT device and the cloud server), and can hence manage their encryption/authentication keys and firmware versions more easily without needing user involvement. Direct Access devices, by contrast, allow users to directly access the device, which not only requires the device to respond to discovery mechanisms (such as the Simple Service Discovery Protocol or SSDP), but also take user input for setting up the necessary security keys. These allow intruders to snoop on the home WiFi network (for which a whole gamut of sniffing and cracking tools are available) to conduct reconnaissance (to discover what devices exist), eavesdropping (to deduce keys/passwords), and active attacks (ARP spoofing to masquerade as a legitimate device by

http://doi.org/10.18080/ajtde.v3n3.21

assuming its address, DNS hijacking to redirect a web-page address to an incorrect server) to compromise these devices. A manufacturer therefore has to weigh up the benefits of direct access (namely providing low-latency in-home access to the device, even if the broadband link goes down) against the extra effort needed to implement foolproof authentication mechanisms for legitimate users in the home. This leads us to believe that direct access IoT devices are likely to be the most insecure in the years to come as they struggle to develop appropriate access control mechanisms.

External Access devices carry the most potential to take private user information without the user's knowledge. The Nest smoke-alarm sends encrypted data daily to its servers, which has the potential to contain information about the location of the user within their house at all times and their bedtime habits (i.e. turning lights off to sleep). This does raise questions in the consumer's mind on what data is actually being sent (since there is no visibility into the raw data) and how it is being used. Though wearable devices (predominantly communicating using the transit access model) also send personal user data to cloud servers (through the phone), their sensing capabilities (e.g. heartrate or steps) are generally more clearly understood by the user who chooses to wear them, unlike household appliances whose sensing capabilities (motion, light, etc.) can be more surreptitious.

The biggest threat for devices that use Transit Access (typically wearables and healthcare devices) lies in the pairing of the device to the user's smartphone. Though the short radio-range gives it some inherent protection, and Bluetooth/Zigbee sniffers are more expensive and less common, there is a non-negligible possibility of illegitimate device access in public places, such as in cafes or on public transport. Short of a Faraday cage in which the pairing is conducted, the most fool-proof way seems to be to give a unique code to the legitimate user (such as the 4-digit one-time code displayed on the Fitbit Charge heart-rate monitor) during the pairing process – this however is only feasible if the device has a display, which may not be cost-effective for very small form-factor devices.

Future directions – the continuing need for techno-policy

Security is clearly only one concern among many that manufacturers of IoT devices are dealing with. The surge in demand for IoT is leading many to rush to market with their product, and increasing user appeal to gain market traction can become more paramount than ensuring foolproof security. Further, the application domain (e.g. healthcare versus home automation) as well as business model (e.g. revenue from device versus from the data) can dictate many aspects such as the form factor, onboard resource availability, and communication patterns of the device, which in turn have a direct bearing on the complexity of the security schemes that can be incorporated. Nevertheless, the classification of IoT

http://doi.org/10.18080/ajtde.v3n3.21

devices based on their communication model gives important insights into the likely security and privacy challenges that will be faced by devices in each of these categories.

The increasing uptake of consumer IoT devices poses security and privacy concerns at an unprecedented level. Although law-makers have mostly focused on communications data as an investigative tool, the data generated by IoT devices may serve as a legitimate investigative tool if it were added to the s. 187A(1) list of information that must be stored for two years. We expect that further research into the technical features of devices as well as their investigatory implications may lead to pressures for further law reform once legislators' and investigators' knowledge of the potential of IoT increases. There is no reason why the definition of 'journalist' cannot be amended right away to bring consistency across different pieces of Commonwealth legislation.

Bibliography

AGD. 2012. *Equipping Australia against emerging and evolving threats*. Canberra, Australia: Attorney-General's Department.

AGD. 2015. Telecommunications (Interception and Access) Act 1979 Annual Report 2013-2014. Canberra, Australia: Attorney-General's Department.

Brandis, G. 2015. *Commonwealth, Parliamentary Debates – Senate, 26 March 2015: 2519* (George Brandis, Senator).

Brandis, G; Turnbull, M. 2015. Data Retention Bill passed by Parliament. *Joint Media Release by the Attorney-General and the Minister for Communications*. (March 26, 2015).

Dorling, P. 2015. Security laws bring us closer to the day when journalists will be jailed for reporting. *Sydney Morning Herald*. (March 17, 2015.)

Grubb, B. 2010. Inside Australia's data retention proposal. *ZDNet*. (June 16, 2010). Retrieved from http://www.zdnet.com.au/inside-australia-s-data-retention-proposal-339303862.htm

Gubbi, J; Buyya, R; Marusic, S; Palaniswami, M. 2013. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, September 2013. 29(7):1645-1660.

Huston, G. 2015. Metadata Retention and the Internet. *Australian Journal of Telecommunications and the Digital Economy, Vol 3, No 1.*

Notra, S; Siddiqi, M; Habibi Gharakheili, H; Sivaraman, V; Boreli, R. 2014. An Experimental Study of Security and Privacy Risks with Emerging Household Appliances, *First*

International Workshop on Security and Privacy in Machine-to-Machine Communications (M2MSec), San Francisco, USA.

Sivaraman, V; Notra, S; Velushomaz, V; Boreli, R. 2015. Security and Privacy Evaluation of six consumer IoT devices. Available at Networks Research Group, NICTA, Sydney, Australia. http://www2.ee.unsw.edu.au/~vijay/publications.html

SSCEC. 2011. The adequacy of protections for the privacy of Australians online. Canberra, Australia: Senate Standing Committee on Environment and Communications.

---

[1] For the sake of brevity this article uses the word 'metadata' to refer to these categories.