Submission to the ISC's privacy and security inquiry

February 2014

Executive summary

1 This submission addresses the implications of the alleged "Tempora" programme in terms of the technical background and its detrimental effect on both privacy and security. The authors find the present legal authority and oversight scheme inadequate and propose reform in several aspects. Firstly, untargeted mass surveillance must end. Secondly, we make some technical suggestions regarding the means by which GCHQ may retain some capability to monitor internet transmissions. Finally, we propose reforms to the intelligence and security committee.

Introduction

2 The authors of this submission are computer scientists working in academic and commercial contexts. Like many of our peers, we had long jokingly suspected that the NSA (or GCHQ) were exploiting modern technology to spy on ordinary people without clear legal authority. When Edward Snowden's revelations came to light, we were both shocked at the scale of the exploitation, which was beyond our imagining, and felt betrayed - as several of us work to make computer systems more secure under the assumption that democratic governments are not our adversaries. The exploitation that has been revealed has an impact on both our private and professional lives.

3 There have been a wide range of revelations of varying technical depth, but this submission will focus on what we regard as the most distressing: the "Tempora" programme, allegedly operated by GCHQ.

Technical background

4 To avoid any confusion in the later parts of this submission, it would appear prudent to introduce some technical information. The statements made by members of the intelligence and security committee (not least this inquiry's call for evidence) lead us to believe that they are not adequately advised on the essential principles of the internet or computer science in general. Without this background information it is not surprising that oversight is weak and legislation is passed without realising its consequences. The committee need not take our word for the technical detail – they can call Sir Tim Berners-Lee, Jimmy Wales or Baroness Lane-Fox as witnesses.

5 We must first understand what information is being transmitted and hence intercepted by GCHQ, and where it is transmitted. *Data* is a generic term for information, while the term *metadata* is commonly used to distinguish some parts of the data that pertain to how the *contents* are to be processed or transmitted. The classical scholars on the committee may note that metadata uses a Greek prefix to denote "data about data". However, the distinction is not exact, and can vary according to the different purposes for which the data is used.

6 In a telephone call, there is quite a clear distinction between metadata (the fact a call took place between two numbers at a certain time, for a certain duration of time, in a certain location) and contents (what is said during the call). This distinction is easy to make as the metadata is "textual" whereas the contents are audio. As soon as both the metadata and contents are digital, the distinction breaks down. Indeed, the contents of SMS (text) messages are transmitted using a metadata channel of the GSM protocol, hence the limitation of 160 characters.

7 The most common uses of the internet can be roughly divided into three categories: the *world-wide web*, *e-mail*, and *instant messaging*. The first category is the broadest and includes Google Search, Google Maps, Facebook, Twitter, Amazon, Wikipedia, online banking, and BBC iPlayer. We will describe what information can be gleaned from interception of web traffic, and briefly describe e-

mail traffic.

8 All computers connected to the internet have an *IP address*. This is used like an ordinary street address to identify where to send information. An IP address is a number, but it can reveal the location of the computer, the *internet service provider* (e.g. BT), and sometimes the organisation to which the computer belongs. The internet service provider will often be able to divulge the person who pays for the internet subscription corresponding to the IP address. This information alone can sometimes be compromising. For instance, transmission of information between IP addresses known to belong to Wikipedia and the Palace of Westminster can give reasonable suspicion that an MP has been sprucing up his or her Wikipedia page.

9 In order to use a *website* (Google Search, Wikipedia, etc.), the user's computer must send an HTTP message to the IP address of the computer which hosts the website, and that computer must send a reply containing a *web page*. If the website is hosted outside the United Kingdom, these messages will certainly leave the UK – possibly via the alleged GCHQ tap in Bude – and even if the website is hosted within the UK the messages may be transmitted out of the UK, only to be returned (this gives the internet its resilience).

10 Below is an example of an HTTP request message to the Bing Search website, and the response message. The parts highlighted in grey can be considered metadata, and the rest as contents, although as we have indicated above there is no precise distinction. This relatively innocuous pair of messages can be highly revealing, even when considering the metadata alone.

Request:

```
Source: 80.42.XXX.XXX:2348 (Tiscali, London UK)
Destination: 204.79.197.200:80 (Bing.com, Chicago USA)

GET /search?q=sadomasochism&form=MOZSBR&pc=MOZI HTTP/1.1
Host: www.bing.com
Mozilla/5.0 (Windows NT 6.1; Win64; x64; rv:25.0) Gecko/20100101 Firefox/25.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: en-gb,en;q=0.5
Accept-Encoding: gzip, deflate
DNT: 1
Connection: keep-alive
```

Response (truncated):

```
Source: 204.79.197.200:80 (Bing.com, Chicago USA)
Destination: 80.42.XXX.XXX:2348 (Tiscali, London UK)

HTTP/1.1 200 OK
Cache-Control: private, max-age=0
Transfer-Encoding: chunked
Content-Type: text/html; charset=utf-8
Content-Encoding: gzip
Expires: Tue, 04 Feb 2014 XX:XX:XX GMT
Server: Microsoft-IIS/8.0
P3P: CP="NON UNI COM NAV STA LOC CURa DEVa PSAa PSDa OUR IND"
Set-Cookie: _FS=NU=1; domain=.bing.com; path=/
Set-Cookie: SRCHD=MS=3206709&D=3206709&AF=MOZSBR; expires=Thu, 04-Feb-2016
XX:XX:XX GMT; domain=.bing.com; path=/
Set-Cookie: SRCHS=PC=MOZI; domain=.bing.com; path=/
Edge-control: no-store
Date: Tue, 04 Feb 2014 XX:XX:XX GMT

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"><html lang="en"
xml:lang="en" xmlns="http://www.w3.org/1999/xhtml"
xmlns:Web="http://schemas.live.com/Web/"><head><meta content="text/html;
charset=utf-8" http-equiv="content-type" /><script type="text/javascript">//<!
[CDATA[si_ST=new Date//]]></script><script type="text/javascript">...
```

11 The information transmitted when using other websites is even more revealing. On Google Maps there are locations, addresses and routes; on Facebook there are private photographs and messages; on Twitter there are private and public messages; with online banking there are account balances and statements; on Amazon there are product orders and credit card details; on other sites there is legal pornography, LGBTQ resources, and medical and legal advice.

12 Some, but not all, of these website interactions are *encrypted*. Encryption is used when the network between the user's computer and the website's computer is not trusted. This provides some protection of the contents of messages but, as described above, the mere fact of two IP addresses interacting can be compromising in itself.

13 E-mail is transmitted in a different manner (although services like GMail involve interactions with a website), but similar issues arise. The "to" and "subject" lines of an e-mail might be considered metadata, and the body as the contents. Obviously, the contents may be of a highly sensitive nature, but even the subject line and the mere fact of contact between two people may be revealing. E-mail is not usually encrypted.

Alleged current situation

14 It has been alleged that GCHQ's Tempora programme exploits relationships with telecommunications companies to gather and store internet and telephone traffic passing in and out of the UK. GCHQ stores all such data for three days, and the part that GCHQ defines as metadata for thirty days. The nature of the internet means this data comprises website traffic and e-mails between British citizens, between British citizens and foreigners, and between foreigners. The data belongs to millions, if not billions, of innocent people who are no threat to the national security of the United Kingdom.

15 Although it has not been made clear by the government, it seems that this programme has its legal basis in the Regulation of Investigatory Powers Act (RIPA). RIPA requires that targeted interception takes place under warrant from the foreign or home secretary. However, the Tempora programme is not targeted at individuals: it is a mass surveillance system running counter to the principles of presumption of innocence and freedom of speech. We believe it constitutes the largest violation of privacy in history, and an unprecedented increase in the power of the state.

16 Some arguments have been made that (i) "the haystack is needed to find the needles" and (ii) "no human being is reading your e-mails". The first of these arguments is a ridiculous simplification of the situation that appeals to those with little technical knowledge. If we were to modify the metaphor to be more accurate, the internet is a continual stream of "straw" passing though a computer. The computer could choose to store anything it considers a "needle" and let the rest pass untouched. Instead what we have today is that GCHQ stores the entire stream of straw, only to identify the needles in the subsequent three days. If the claim is made that only a human can identify needles, then the second argument – that no human is reading e-mails – has been contradicted.

17 The second argument is also tailored for those with no technical knowledge. If no human were reading e-mails, then the data would not need to be stored for three days. Seconds or milliseconds would likely suffice. Moreover, the absence of a human in the process is not necessarily a cause for comfort. Automated classification and profiling systems are inaccurate, as the recent debacle with the "Great Firewall of Cameron" has shown. It is one thing to be inconvenienced by not having access to an innocent website, but it is quite another thing to be accused of terrorism.

18 There has been a further allegation that GCHQ in conjunction with the NSA have deliberately weakened encryption standards and the security of commercial products such as the iPhone and Windows. This risks both the privacy *and the security* of billions of computer and mobile phone users.

19 The inquiry's call for evidence asks how internet surveillance compares to CCTV. While we believe

that the use of CCTV has grown out of all proportion in the UK, it is self-evidently a less intrusive form of untargeted surveillance than the Tempora programme. CCTV in effect records public information – that an anonymous person visited one location or another. There is no CCTV in private homes. CCTV is not, as far as we are aware, nationally or globally centralised, which means that there is little chance of a national agency building up a complete picture of a person's movements. Finally, and most importantly, CCTV does not record the entire transcript of what is said or written in each place visited. It cannot record bank balances, credit card numbers, medical records, or letters to lovers. Tempora seeks to do each and every one of these things.

Legislative reform

20 First and foremost we believe that the Tempora programme must have a clear and explicit legal basis provided for by Parliament. RIPA is woefully inadequate.

21 The new legislation must make precise what powers it is granting to GCHQ, and reform the oversight bodies. In our view, untargeted mass surveillance must be confirmed as immoral and illegal. The deliberate subversion of encryption standards must be made illegal. GCHQ may continue to operate its tap with the requirement that only "needles" are stored. A time limit (a deadline), chosen by independent technical advisors, could apply to the computer systems that determine whether an item of data is a "needle" or "straw". This time limit will be extremely short. In addition, the "needles" may only be stored if they originate from persons who have been individually and temporarily targeted under warrant.

22 If GCHQ continues to make a distinction between metadata and contents, it will be necessary to define the terms legally (and publicly) and to revise their definitions on the basis of technical advice.

23 The intelligence and security committee must be reformed. First, its membership must include a prominent representative of a civil liberties organisation such as Shami Chakrabarti. This representative will have the same access to classified material as the parliamentarians. As indicated above, the committee must have substantial and continual technical advice, as technology continues to evolve. GCHQ must be compelled to inform the committee about new programmes, rather than waiting for the committee to ask the right questions.

24 Finally, some aggregate statistics must be made public for Parliament and voters to determine whether the intrusion has in fact resulted in greater security. These statistics could include the volume and percentage of data classified as "needles" each year, and the number of terrorism convictions arising thereof.

Submitters (this section should remain confidential)

[REDACTED]