Duncan Campbell—supplementary written evidence (IPB0124)

This is a supplementary note of documentary material and questions relevant to the committee's review, supplementary to my evidence note of 19 December 2015. My experience and involvement with the matters before the Committee is recited there.

"Internet Connection Records" (ICR) are already created and held

A major issue which may arouse concern with the form of the draft legislation is that the "Internet Connection Records" (ICR) which the Bill proposes should in future be created and retained by Service Providers are already created directly by government agencies and are held, systematically and on massive scale. As of 2012, provision had been made for the storage of 24 trillion (24 thousand billion) such records.[1]

These records include metadata and extensive further metadata derived from analysis of content concerning all types of internet connection, in relation to the totality of UK internet users, all of which is available for any form of analysis and extraction without warrant by UK agencies and by foreign partners. The basis of and sources for these factual statements is described following.

The Committee has asked, *inter alia*:
• Is accessing Internet Connection Records essential for the purposes of IP resolution and identifying of persons of interest?
• Are the requirements placed on service providers necessary and feasible?
• Are the powers sought necessary?
•Has the case been made, both for the new powers and for the restated and clarified existing powers?
• Are there alternative mechanisms?

Given the published facts, the answer to the committee's final question appears to be "yes". ICR (as defined) are currently obtained and generated at the rate of many billions per hour from bulk communications data processed and analysed by GCHQ using a small number of warrants issued under Section 8(4) of RIPA, and have been so obtained since at least 2008.

These Internet Connection Records are derived from a network of probes connected to submarine optical fibre communications cables as they enter and leave the United Kingdom through shore terminal stations, and which existing service providers have been compelled to install by virtue of technical orders made under RIPA and the Telecommunications Act 1984.

The Internet communications data obtained is refined by a process known as "sessionisation". Sessionisation re-assembles the data packets making up individual communications. The technology for sessionisation for Internet optical fibre communications was first developed by an NSA team led by Mr William Binney, whom the Committee have invited to give oral evidence.

I have worked with Mr Binney to examine the UK material employing this type technology, and to consider its relevance to the question of whether Internet Connection Records are in fact necessary given existing deployments. It appears from the UK documents that they cannot be necessary, in that (on the evidence published and cited here) they are already available now (and

[1] https://theintercept.com/document/2015/09/25/gchq-analytic-cloud-challenges Page 6

are filtered) in a far more powerful form than any UK service provider would be able to achieve in the future.

It would follows that the requirements proposed in the draft Bill to be placed on service providers cannot be necessary, whether or not they are in fact feasible to be carried out at the ISP level, or are judged proportional.

There is now abundant evidence that Internet Connection Records of the type proposed to be created and held for Law Enforcement and other purposes already exist and are collected on a massive scale by GCHQ, and that this activity has been taking place since at least 2008. The largest part of this evidence is a corpus of 28 GCHQ documents published by the U.S. online magazine, The Intercept, on 25 September 2015.[2] The documents accompanying the article were, according to the magazine, published in so as to highlight the scope of existing investigation systems installed within the UK Internet, and in anticipation of the expected new legislation.

I would respectfully suggest that the 28 GCHQ documents as a group merit at least the same attention as the Home Office publications accompanying the Bill, for the reason that the GCHQ documents extensively and helpfully explain and define technical and legal practices in the areas to be legislated, as they exist now and as they have evolved over the past 15-30 years.

One GCHQ document in particular, entitled "Operational Legalities", runs to 156 pages and is one of several providing extensive guidance as to current legal practice.[3] One matter of particular concern as to proportionality is current guidance indicating that the all forms of metadata concerning communications between persons in the UK (such as e-mail addresses, e-mail headings, messages, etc, and also including locations and passwords) and taken into GCHQ repositories may currently be examined and analysed without restriction, and without the need for a targetted warrant.

As of 2012, according to a report on "GCHQ Analytic Cloud Challenges" [4], Internet Events records were then being recorded at the rate of 50 Billion Events Per Day, with a capacity then to rise to double that amount. These records included all Internet activity with one or both terminals in the UK, as well as Internet communications events passing through the UK. In 2012, this is said to have included 15 Billion web visit record per day. Each record is an Internet Connection Record, in that it includes all available metadata information about users, their locations, their identifiers and addresses, as well as times and dates and services used, and the user identifiers within their services.

ICR Records and filters in BLACK HOLE data and applications

According to the published documents, the Internet records are accumulated and stored in two depositories in Bude and Cheltenham, named "BLACK HOLE ". The records are then accessed and processed by filters, resulting in the creation of multiple datasets or databases directly

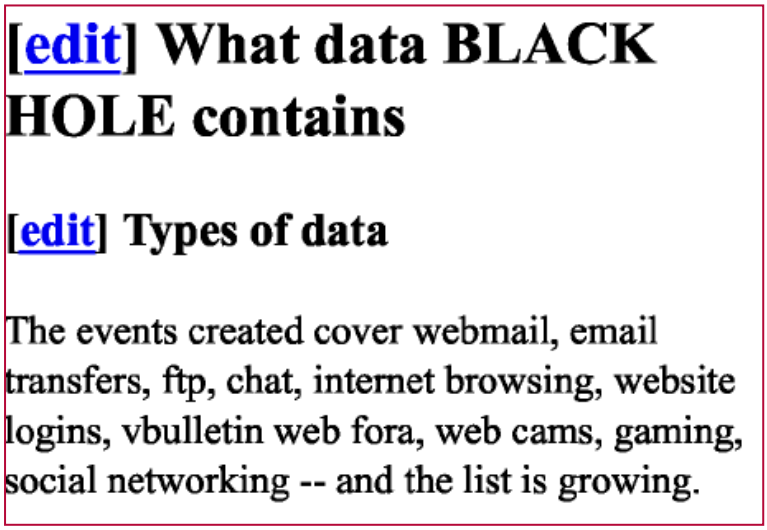
---

[2] https://theintercept.com/2015/09/25/gchq-radio-porn-spies-track-web-users-online-identities
[Declaration of interest: I have written a report for The Intercept.]

[3] https://theintercept.com/document/2015/06/22/operational-legalities-gchq-powerpoint-presentation;
https://theintercept.com/document/2015/09/25/pull-steering-group-minutes;
https://theintercept.com/document/2015/09/25/content-metadata-matrix;
https://theintercept.com/document/2015/09/24/legalities

[4] https://theintercept.com/document/2015/09/25/gchq-analytic-cloud-challenges

capable of answering all the matters set out in the ICR Operational Requirements statement for the Investigatory Powers Bill.[5]



[edit] What data BLACK HOLE contains

[edit] Types of data

The events created cover webmail, email transfers, ftp, chat, internet browsing, website logins, vbulletin web fora, web cams, gaming, social networking -- and the list is growing.

Published GCHQ description of BLACK HOLE Internet Connection Records. [6]

In particular, as shown above, the ICR type of records already contain the "who, when, what, how" type of information that Parliament has been told is currently a "gap" in capability. It follows from this evidence that it may waste public funds, and place an unneeded burden on service providers, to require forced duplication of existing and inferior capabilities.

The sample of requirements for ICR, set out on page 25 of the draft Bill, lists three matters, each of which are shown by the 28 GCHQ documents to already exist in a comprehensive way, providing information far beyond that which service providers do hold or could reasonably be expected to create and retain in future.

The sample suggested requirements were:

> (1) To establish what services a known suspect or victim has used to communicate online, allowing investigators to request more specific communications data;
> (2) To establish whether a known suspect has been involved in online criminality, for example sharing indecent images of children, accessing terrorist material or fraud;
> (3) To identify services a suspect has accessed which could help in an investigation including, for example, mapping services;

The table below appears in the published GCHQ Analytic Cloud Challenges report (foot notes 1 and 4, *supra*), page 5. It demonstrates that all of the questions raised in the ICR are currently answered by the BLACK HOLE system of data and queries.

---

[5] GCHQ's documents sometimes use different names to the Home Office. Internet Connection metadata records held in BLACK HOLE are called "Single Line Records". The Filter or Filters are generally described as "Query Focussed Datasets". These are databases created when filters are applied to BLACK HOLE raw data.

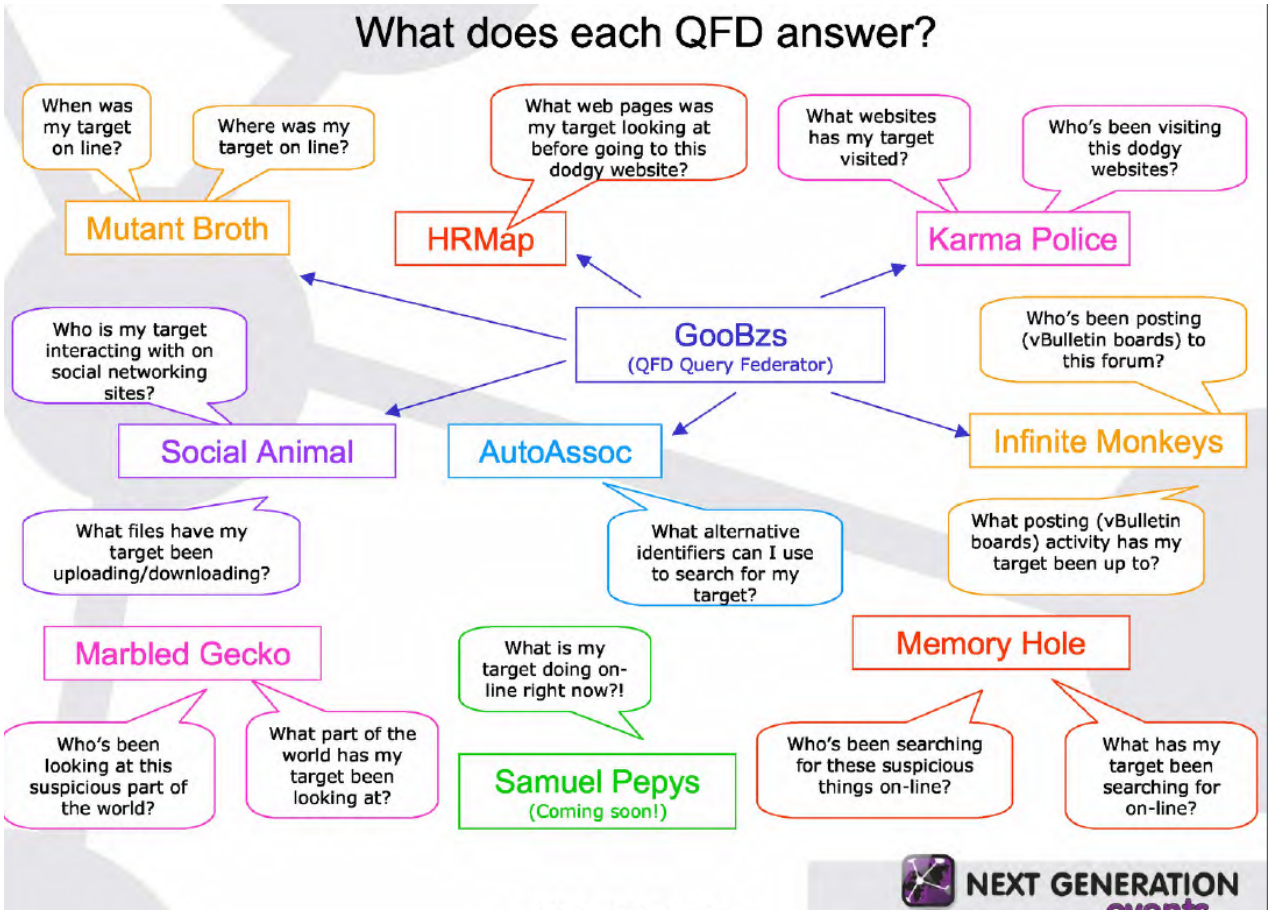
[6] https://theintercept.com/document/2015/09/25/data-stored-black-hole

Name			
AUTOASSOC	Bulk unselected TDI-TDI correlations with confidence scores.	What other TDIs belong to your target ? What technologies your target is using ?	2+1 instances, each 50-70TB storage
Evolved Mutant Broth	Identify when certain TDIs appear in traffic which indicate target usage and their location. Telephony and C2C data provide a converged view.	Where has my target been? What kind of communications devices has my target been using?	10+5 instances, each 70TB storage
Hard Assoc	Provide strongly correlated selectors for both C2C and Telephony traffic taken from TDIs appearing in the same packet	Are there any alternative C2C or Telephony selectors for my target?	3+2 instances, each 70TB storage
HRMap	Host-referrer relationships - information about how people get to websites, including links followed and direct accesses.	How do people get to my website of interest and where do they go to next? What websites have been visited from a given IP?	5+3 instances, each 70TB storage
KARMA POLICE	Which TDIs have been seen at approximately the same time, and from the same computer, as visits to websites.	Which websites your target visits, and when/where those visits occurred. Who visits suspicious websites, and when/where those visits occurred. Which other websites are visited by people who visit a suspicious website. Which IP address and web browser were being used by your target when they visited a website.	11+7 instances, each 70TB storage, 3+1 correlator instances
SOCIAL ANTHROPOID	Converged comms events allowing you to see who your targets have communicated with via phone, over the internet, or using converged channels (e.g. sending emails from a phone or making voice calls over the internet).	What communications your target is engaged in. Who has your target been communicating with. What communications have occurred using a particular locator (IP address, cell tower, etc).	6+3 instances, each 70TB storage

From GCHQ Analytic Cloud Challenges report , page 5

Specified and comprehensible examples of how this type of information directly provided answers to the concerns raised are shown in a further table overleaf identifying the filters, or "Query Focuses'" which extract the relevant data from BLACK HOLE. [7]

[7] https://theintercept.com/document/2015/09/25/demystifying-nge-rock-ridge page 4



From "Demystifying NGE Rock Ridge" page 4

For example, the question "What web pages was my target looking at before going to this dodgy website?" is answered by the filter (or "QFD") HRMAP. The question "What websites has my target visited?" is answered by the filter KARMA POLICE. These would include identifying users who had visited sites offering indecent images of children, or sites offering terrorist materials.

An inquiry to identify services a suspect has accessed which could help in an investigation including, for example, mapping services would be answered by the MARBLED GECKO filter, which records data answering questions such as "Who's been looking at this suspicious part of the world?" or "Find out who has been looking at what on Google Earth".

According the documents the GCHQ KARMA POLICE filter or QFD "aims to correlate every user visible to passive SIGINT with every website they visit, hence providing either (a) a web browsing profile for every visible user on the internet or (b) a user profile for every visible website on the internet." It appears from the reports to hold precisely the material about "what services a known suspect or victim has used to communicate online" that is claimed to be unavailable, and to have done so for at least five years.

Other filtered data derived from BLACK HOLE hold bulk data concerning bulletin board use [INFINITE MONKEYS], Social Networking Site activity [SOCIAL ANIMAL], and search engine requests [MEMORY HOLE].

The information which can be used to identify and access the filtered records includes, according to the documents "web service authentication data", "ID card number or passport number", "driving licence number", "car registration number", and/or "bank card/credit card account numbers".

The existing BLACK HOLE system is on this evidence already more capable than the ICR records system proposed in the Bill. For example, as shown above, a filter or "QFD" called SAMUEL PEPYS will answer the question "What is my target doing on- line right now?".

To my knowledge or in my understanding, all of the internet connection records systems creating the UK's BLACK HOLE repository are built on the Internet fibre cable sessionising systems which Mr Binney's US team devised and which he has explained to the Committee.

Despite the recent disclosures about and avowal of bulk data collection from the Internet, there has been a marked by the government to disclose that the requirement for Internet Connection Records has already been achieved for some time, but that the data recovered has not been made available to law enforcement.

I will be glad to further assist the Committee on any of these matters.

Duncan Campbell

22 December 2015