

Chapter 1

1 Collect it all: everyday lives turned into passive signals intelligence

1.1 Introduction

The NSA and GCHQ have stated their desire to “collect all the signals, all the time”.ⁱ To fulfil this aim, they want the ability to collect all data generated by our daily use of the Internet, phones and other technology. Through the wiretapping of international fibre optic cables, the communications of billions of people are intercepted every day. These include our personal phone calls, emails, text messages and web searches. The information provided by these communications can be broadly split into the actual content of the communication and the communications data, also called metadata, the information on when, where and to whom it was communicated.



GCHQ collects both the content of communications, what is being said, and the metadata of communications. Traditionally metadata collection was perceived as less intrusive than the interception of content – reading the envelope was not as bad as opening and reading the letter. But nowadays it is widely acknowledged that metadata can provide an intimate picture of an individual’s life. We can map social relationships and traveling patterns just from metadata, without anyone looking at the contents of messages.

At present, content and communications data are treated differently under the law but they can both be used to gain intrusive insights into our life.

As well as collecting our personal communications data, GCHQ and the NSA have programmes that can collect data from apps, web cams and social media. The data we

generate will proliferate as more and more household products, from cars to fridges, use digital technology.

Below we provide an outline of the main components of the UK mass surveillance system as documented in leaked documents, with some important caveats.

The focus of this paper is the mass surveillance of global civilian communications. But we should make clear that these signal agencies have and must retain some legitimate capabilities to eavesdrop military signals and adversarial official state communications. The government will also need to be able to access communications for law enforcement purposes, including suspected terrorists, but this function should be strictly separated from intelligence gathering because mixing them up creates intractable regulatory problems, as we can see in this report.

There are some important activities not fully covered in the UK-related leaked documents, such as the authorisation of targets and programmes and how exactly the data collected is incorporated in the disseminated intelligence. In contrast, the NSA has received scrutiny on a much broader spectrum. Journalists specialising on intelligence such as Duncan Campbellⁱⁱ, have provided information on some of those areas, but a comprehensive picture of all the activities of GCHQ is out of scope for this paper.

1.2 Collaboration with companies to tap international optic cables

Among the first documents leaked by Edward Snowden, and released by the Guardian on June 6, 2013,ⁱⁱⁱ were revelations that GCHQ engages in bulk collection and storage of communications content and data. The main source of data comes from wiretapping international fibre optic cables with the collaboration of telecommunications companies, as part of the GCHQ programme Global Telecoms Exploitation (GTE).

Most cables in and out of the UK, and also some at overseas locations, could be tapped at any time. The agreements force these companies to install equipment^{iv} that allows for the information travelling through the cables to be duplicated and sent to GCHQ with minimal interference with the traffic.^v The Interception Capabilities 2000 report for the European Parliament^{vi} provided details of how the NSA and GCHQ were tapping into submarine cables around the world in clandestine operations. This appears to now be largely superseded by arrangements with companies.

This kind of activity has been documented in the US in the past, with AT&T being sued by civil rights groups for transferring bulk internet data to the NSA from their facilities in San

Francisco^{vii}. In August 2013, the German newspaper, Süddeutsche Zeitung, made public the names and codenames of cable companies known to cooperate with the British authorities:^{viii}

- Verizon Business - DACRON
- British Telecommunications - REMEDY
- Vodafone Cable - GERONTIC
- Global Crossing - PINNAGE
- Level 3 - LITTLE
- Viatel - VITREOUS
- Interoute – STREETCAR

Because of the way the Internet routes data around the globe, and the geographical position of the UK, intercepting online traffic in and out of Britain inevitably scoops the communications of billions of people worldwide. The data collected also includes telephone calls, as these are also sent via fibre optic. In 2012, GCHQ were handling 600 million 'telephone events' each day.

A complete 2010 list of cable intercepts has been leaked showing that GCHQ also has access to cables around the world, in places as far apart as Malta, Japan and the Caribbean.^{ix} Of particular importance is a facility in Seeb, Northern Oman, which intercepts cables running through the Persian Gulf^x and has extensive processing capabilities. The image below is part of leaked GCHQ list of cable access opportunities.

Cable	UK?	REMEDY	GERONTIC	DACRON	Partner LITTLE	PINNAGE	STREET CAR	VITREOUS
Apollo	UK	IRU/LC	DCO	IRU/LC	IRU/LC		IRU/LC	
CANTAT 3	UK	DCO	IRU/LC					
Concerto	UK						DCO	
EIG	UK	DCO	DCO	DCO				
Flag Atlantic 1	UK			IRU/LC	IRU/LC			
Flag EA	UK	IRU/LC	IRU/LC	IRU/LC				
Hibernia	UK				IRU/LC			IRU/LC
Solas	UK		DCO					
SMW-3	UK	DCO	IRU/LC	DCO				
Tangerine	UK				DCO			
TAT-14	UK	DCO	DCO	DCO	DCO			
Tata TGN-Atlantic	UK	IRU/LC						
Tata TGN-Western Europe	UK	IRU/LC						
UK-France 3	UK	DCO	DCO					
UK-Germany 6	UK	DCO	DCO					
UK-Ireland (GX)	UK					DCO		
UK-Netherlands 14	UK	DCO	DCO					
Ulysses	UK			DCO				
Yellow/AC-2	UK	IRU/LC			DCO	DCO		
AAG		DCO						
AC-1		IRU/LC		IRU/LC	IRU/LC	DCO		
Americas II			IRU/LC	DCO	DCO	DCO		
APCN-2		DCO	DCO	DCO				
APCN		IRU/LC	IRU/LC	DCO				
ARCOS					IRU/LC	DCO		
Antillas 1				DCO				
Atlantis II				DCO				
Australia-Japan Cable		IRU/LC	IRU/LC	DCO				
Bahamas 2				DCO				
Carac			DCO					
Cayman-Jamaica FS			DCO					
China-US			IRU/LC		DCO			
Circe N								DCO
Circe S								DCO
Columbus III				DCO				
Denmark-Poland 2		IRU/LC						
Denmark-Russia 1		IRU/LC						
Flag Falcon							IRU/LC	
Flag North Asia Loop				IRU/LC				
Gemini Bermuda			DCO					
Globenet			IRU/LC					
GO-1							IRU/LC	
Guam-Philippines				DCO				
Italy-Malta							IRU/LC	
Japan-US		DCO	DCO	DCO	DCO	IRU/LC		

In some cases, more than one company appears to be involved in the tapping. In separate documents, Süddeutsche Zeitung claimed that Vodafone and BT have helped GCHQ access the cable TAT-14 that routes communications to Germany.^{xi} The cable list shows that this cable is potentially accessed via four companies who have “Direct Cable Ownership” (DCO).

In other cases, for example the cable running from Denmark to Poland, this is achieved through the “Indefeasible Rights of Use/Lit Capacity” (IRU/LC) of a company that is not the owner of the cable, in this case BT. The documents show that GCHQ has secret agreements with companies providing at least 114 access points to 63 undersea cables or landing stations. Fibre optic cables carry data through light, and normally have more than one fibre optic strand.

Within each strand data is split to create independent channels of communications that can go in opposite directions and improve overall capacity^{xii}. GCHQ works at the channel level by attaching what they call “processors”. In 2010 GCHQ was able to access 592 channels of 10 Gigabits per second each, and could egress (send home for analysis) around a tenth of that capacity.^{xiii} GCHQ hopes to more than treble that capacity in the future.

1.2.1 CASE STUDY: Cable and Wireless/Vodafone, INCENSER and NIGELLA

Documents released as part of an investigation by Channel 4 News and several German media outlets provide details of one such corporate partnership. Cable and Wireless – owned by Vodafone since 2012 – has been providing GCHQ with access to their cables.^{xiv} The company claims that it is merely complying with lawful warrants, but the documents appear to show that they went above and beyond by actively helping shape the traffic to improve collection. At some point GCHQ seconded a full time employee to the company, which has been paid millions of taxpayers' pounds by UK authorities for providing this access to data.

The documents also suggest that Cable & Wireless/Vodafone assisted GCHQ in harvesting data from other cable companies not served by warrants. The Guardian newspaper had raised concerns about this practice in 2013 but did not provide concrete evidence at the time.^{xv} The operation – codenamed INCENSER – provides GCHQ with access to what appears to be an intersection of two cables connecting the Atlantic with Europe and Asia, which is itself codenamed NIGELLA.

The cables were owned by the Indian company Reliance Globalcom - now called Global Cloud Xchange. According to the documents, GCHQ hacked into the NIGELLA cable system using Cable and Wireless/Vodafone for access as “Landing Partner” in Cornwall. BT and Verizon are also named in documents as providing GCHQ access to the cables involved, but not in relation to this operation. Vodafone is also GCHQ's getaway driver, helping to backhaul the intercepted data - some 7.5% of the total – to the agency's Regional Processing Centre (RPC) in Bude, where the data is processed and shared with close allies under the programme WINDSTOP.^{xvi} The NSA names INCENSER as the fourth largest corporate data operation providing them with about 10% of their total collection. INCENSER is also used for wider hacking purposes as part of the TURMOIL system described elsewhere in this report.

1.3 TEMPORA: making the Internet manageable

Processing the increasing amounts of data collected in the operations described above has long presented a considerable challenge to signals agencies, which have responded by increasing the computing capacity of their systems.

The European Parliament report, *Interception Capabilities 2000*, documents how the NSA and GCHQ have operated their now international Internet-like networks since the 1980s, and were well prepared to snoop on the open web. The NSA was monitoring Internet data at nine sites in 1995. Efforts to capture the small proportion of Internet data that had intelligence value were already in place.^{xvii}

The latest development in bulk collection and processing in the UK is called TEMPORA, and brings Big Data technologies to state surveillance. Since the 1990s most communications, including voice calls, have moved to submarine cables and many to the Internet.

The system was launched in 2011 after several years of trials that started in 2008, and it is a joint effort of several GCHQ programmes, including Mastering the Internet (MTE) and Global Telecoms Exploitations (GTE). The TEMPORA system has been described as a time machine that can “slow down the Internet” to better allow the agency to sift through the information. TEMPORA does this by storing a few days of almost unfiltered Internet data, effectively creating an “Internet buffer”.

According to Snowden, the British TEMPORA is the first “full take” system developed by any intelligence agency^{xviii} that allows wholesale Internet traffic to be collected, instead of forcing a preselection of materials. Not every bulk collection system from GCHQ is fed into TEMPORA, but apparently the majority are and there is a growing trend. As of May 2012, there were TEMPORA capabilities at three processing centres – which appear to correspond to Bude, Oman and Cheltenham.

The exact capacity is difficult to assess and it is clearly evolving in any case. Der Spiegel reported that overall the agency has installed the relevant equipment to access 200 of these channels – but not all at the same time - and there are plans to extend TEMPORA’s potential reach to 1500 processors.^{xix} Other documents indicate that in May 2012 GCHQ had the capacity to feed 46 channels of 10 Gb/s at any time each into their “full take” system^{xx}.

An NSA document about TEMPORA says: “*We’ve all heard about Big Data; now you can get Big Access to Big Data*”. TEMPORA itself comprises different components, including the actual access points to fibre-optic cables described above, and several tools to make the data manageable. The following are the steps performed by TEMPORA in the collection of data, but in practice it appears that much of this process is currently incorporated into the XKEYSCORE subsystem described below and may happen simultaneously.

1.3.1 Access and collection

1.3.1.1 Rules for ingestion into the system

Despite its extensive capabilities, GCHQ has more access to data in cables than the available capacity to process it and send it home for further analysis – called to egress by the agency, so the raw material has to be preselected. According to the documents, the system has tools to create rules that “promote traffic into the Internet buffer capability”

Deep packet capture inspection (and injection)

The bulk collection of Internet data itself involves huge processing. A cable splitter may initially divide the light for collection, but this flow of data is not very useful for the agencies. Raw Internet data is composed of small “packets” of data. Any processing of data requires the ability to look into individual packets. The equipment physically connected to the cables is

also able to inject data back into the pipe. These machines with dual capacity to read and write directly to the backbone of the Internet form the global TURMOIL system, are run by the NSA, but with extensive UK participation.

1.3.1.2 Volume Reduction

The selective ingestion involves discarding traffic that takes up a lot of space but has low intelligence value, such as consumer videos and file-sharing media downloads. They then keep things like email, chats, etc. Some 30% of the total traffic is ingested. Several documents mention a system called “massive volume reduction” (MVR)^{xxi} - and clarify that MVR is not available at every TEMPORA instance. However, it is not completely clear whether there are two separate systems for initial selection and volume reduction or just one.

In any case, the initial processing and rules appear to be a completely internal process that falls in a regulatory gap between the initial warrants and the application of safeguards by operatives.

1.3.2 PROCESSING

1.3.2.1 Improving and cleaning up the data

After getting rid of non-useful kinds of data, bulk surveillance systems try to automatically carry out an initial optimisation and clean up. According to leaked documents GCHQ uses a tool called POKERFACE, but there are no available details of its operation. This is unfortunate, as this initial automated processing of data receives little scrutiny and has few legal protections against abuse.

We understand the optimisation process involves some filtering and preselection on the basis of certain rules, for example grabbing certain VPNs, or isolating all mobile phone location records. This intermediate optimisation isn't part of the MVR process, or the subsequent “selector” process as described below.

1.3.2.2 Reconstruction of communications

As described above, all email or Internet communication is broken down in packets, which are sent separately and regrouped at the other end. GCHQ has to recreate the messages by pulling the packets together. They call this to “sessionise”, and the aim is to reconstruct the full sessions of communications, including individual messages and full conversations.

1.3.2.3 Automated targeting and storage

Any information – both content and metadata - related to known persons of interest is treated differently as a priority. Operatives elsewhere have fed intelligence systems with “selectors” – the term used by intelligence agencies for search items such as phone numbers, usernames and email addresses.

The term for the creation of these target lists is “tasking”. The NSA and GCHQ have separate specialist tools for the tasking of selectors across their digital estate called UTT and UDAQ.

Tasked data is pulled aside for long term storage and analysis in large databases. This would include the full contents of emails, entries on Facebook, history of Internet use, telephone and Skype calls, their location, etc. But the rest is not deleted, it gets further processed as we see below.

Leaked documents show that the NSA uses a tool called SCISSORS that flags data through the use of so-called “defeat” categories. For example, the NSA appear to have a problem with the proliferation of “ownerless address-books” captured in the data streams, so SCISSORS sends them to long term databases for de-duplication and storage, without clogging the precious space required for active processing.^{xxiii} Importantly, a lot of this data is not dropped but simply stored elsewhere. NSA documents puts SCISSORS in these terms: “*Memorialize what you need*” versus “*Order one of everything off the menu and eat what you want*”.

The agencies also use so-called “defeat lists” which contain specific types of data that the agencies are not meant to collect. It is expected that GCHQ's defeat lists would include known UK identifiers to avoid, such as emails from known British companies.

1.3.2.4 Metadata processing

From the leaked documents we understand that the rest of the data - relating to millions of people under no suspicion whatsoever - is processed to separate the content from the “metadata” - the who, when, where and how of the communication- which is then kept for at least a further 30 days.

1.3.2.5 Processing and analytics through the XKEYSCORE front-end

XKEYSCORE indexes and sorts the raw data providing search capabilities, similarly to Google. Operatives have access to the raw data, where they use a special language called GENESIS to sift through the unfiltered *content* of communications. Operatives can also modify the ingestion rules.

The communications of suspects and the metadata of innocent people plus any other useful bits of information, documents, etc. are also available through XKEYSCORE. The system also operates defeat lists for information to drop.

1.3.2.6 Long term data storage

The database systems of the NSA have been thoroughly documented (see section on data sharing with the US) but their British equivalent have not. It is known that metadata and content are kept separate by the NSA, as are phone calls and Internet communications. In addition, there are many specialist databases, for example dedicated to mobile phone location records.


1.4 XKEYSCORE

TEMPORA is run by GCHQ, but it is also described^{xxiii} as a component of the US National Security Agency's global mass surveillance system XKEYSCORE.^{xxiv} This exact relationship is a critical aspect of the UK surveillance regime, but very difficult to establish in detail due to the surrounding secrecy. US staff have direct access to the UK's XKEYSCORE.

This system allows the NSA and partners to monitor “everything a user does on the Internet” in real time, including the content of emails, websites visited and searches, as well as their metadata”.^{xxv} It is impossible to overstate how powerful and privacy intrusive XKEYSCORE is, according to the leaked documents. There are plugins for searching specific types of data, or an operator can choose to replay a whole Internet session from a selected target. PRISM data and older upstream systems are accessible through this unified search system. Some databases such as MARINA are also connected with the aim to provide a “one-stop-shop” for searches and creation of new leads.

TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL

What is XKEYSCORE?



1. DNI Exploitation System/Analytic Framework
2. Performs strong (e.g. email) and soft (content) selection
3. Provides real-time target activity (tipping)
4. “Rolling Buffer” of ~3 days of ALL unfiltered data seen by XKEYSCORE:
 - Stores full-take data at the collection site – indexed by meta-data
 - Provides a series of viewers for common data types

1. Federated Query system – one query scans all sites
 - Performing full-take allows analysts to find targets that were previously unknown by mining the meta-data

TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL

XKEYSCORE has been thoroughly described in leaked documents but can be very confusing, because it integrates quite different functions, which may present a challenge to regulators.

As we saw above, XKEYSCORE is a central component of the UK's bulk Internet surveillance system TEMPORA. But it is also a common technology platform initially created by the NSA and shared with the UK. The platform involves the back end storage and processing of collected data, but also the front end search and query capabilities, including asking for new collection in real time.

XKEYSCORE was created to solve the limitations of the older bulk data collection systems of the NSA.

Under the umbrella FAIRVIEW,^{xxvi} some of these have been active since 1978, based on partner companies such as AT&T and Verizon installing equipment that copies, analyses and sends data to the NSA. Former leaks identified US company Narus as the main provider of this equipment. Their “semantic analysis” tools — which are also used for security purposes by telecommunications companies worldwide – market the following capabilities:^{xxvii}

Intercept and surveillance application for real-time precision targeting of any type of IP traffic

- Provides real-time, surgically precise targeting, allowing full IP session reconstruction and visibility for targeted traffic such as webmail, e-mail, IM, chat, Voip and other Internet Protocol-based communications
- Enables the capture of packet-level, flow-level and application-level usage information for forensic analysis, surveillance and regulatory compliance

The commercial machines from Narus however had limited computing capacity to process the initial ingest rules for selection and processing of packets in real time.^{xxviii} This drove the overall process towards an increase in ingestion to avoid missing potentially useful data. But this increase in collection creates a new problem with storage and processing as more data is kept.

This is the often mentioned 'haystack' required to find the 'needles'^{xxix}.

XKEYSCORE, operational since at least 2007,^{xxx} solves these problems through a decentralised architecture of nodes that work independently but talk to each other. These nodes ingest bulk data from around 150 global packet capture collection points. Many of these are TURMOIL network of probes at the core of the Internet, including the UK's TEMPORA programme. But they also include data collected from embassies, satellites, drones and other sources.

The latest versions of XKEYSCORE - called “deep dive” - doesn't even try to carry out an initial filtering of the traffic, but syphons as much as possible for processing to local clusters of commodity Linux computers. GCHQ has 1000 such machines dedicated to processing 40 billion pieces of content a day.

A major innovation from XKEYSCORE is the massive increase in storage, keeping a cache of three days of raw data and thirty days of metadata. Previous systems operated locally, but tried to send larger amounts of data for processing directly to the NSA's headquarters. Moving the storage closer to the initial collection point, either at the same place or through

the use of regional processing centres reduces cable bottlenecks in the hauling back of data. The UK version will obviously operate slightly differently to the US.

The other major innovation of XKEYSCORE is providing a single interface to analysts, who are able to query the temporary cache of data both on metadata and content, link local temporary records with remote intelligence databases and create new ingest rules for the collection. The latter is termed Signal Development (SIGDEV).

1.5 Other sources of data

In addition to the bulk collection of cable data through TEMPORA, US and UK spy agencies target other communication technologies in their quest to collect “all the signals, all the time”.

95% of the world’s communications are routed via submarine cables but satellites still cover some strategically important niches, such as satellite phones, industrial data links and all communications to ships, airplanes and oil rigs.

The US and the UK have a global network of sites^{xxxii} collecting communications traffic from satellites, called FORNSAT in spying terminology. GCHQ operates stations in places like Cyprus,^{xxxiii} Kenya^{xxxiii}, Oman^{xxxiv} and also in Bude, Cornwall. The latter has been partially reconverted for use in the TEMPORA system, given its proximity to the landing point of many cables.

GCHQ "at least intermittently, kept tabs on entire country-to-country satellite communication links, like Germany-Georgia and Germany-Turkey, for example, of certain providers".^{xxxv}

Satellite data is also processed as part of the global XKEYSCORE^{xxxvi} system of acquisition and analytics.

Historically GCHQ has also monitored a broad range of technologies, such as high frequency radio - used for diplomatic communications - and microwave radio relays connecting inter city networks and also used to connect with Ireland.^{xxxvii} The bulk monitoring of communications with Ireland was successfully challenged by civil rights groups at the European Courts of Human Rights in a high profile case.^{xxxviii}

1.6 What is being collected

The agencies are always on the lookout for new data streams and ways to make sense of and enrich what is already available. The documents leaked by Edward Snowden contain many examples of data collection activities. We also know that GCHQ carries out experiments on data and many forms of analytics.

Most programmes revealed in leaked documents, such as the monitoring of webcams, or internal business communications, rely on the bulk collection technologies described above to provide the underlying data.

We would assume that the single data repository from XKEYSCORE would be used as the basis for most extra collection activities, with ancillary medium term databases separate from their long term databases of known intelligence related data. But there is little information on how these other processes relate to the core bulk data collection system.

Here we summarise some of the documented cases that have caused most controversy for their intrusiveness and potential lack of clear legal basis.

1.6.1 Internal data cables of Internet companies

The NSA and GCHQ have been intercepting the private cables that connect the data centres of some Internet companies, such as Google and Yahoo. The joint MUSCULAR^{xxxix} programme is based in Britain and mainly run by GCHQ. In a period of 30 days, MUSCULAR collected 181 million records^{xl}. With “full take” access to the internal data links of the companies, the agencies can intercept communications in real time and take “a retrospective look at target activity” (ibid.).

MUSCULAR is fed into the global NSA system as part of the WINDSTOP programme of second party collaboration with close allies such as the UK.

1.6.2 Monitoring social networks

The newer kinds of 'communications data' that the intelligence agencies can access paint an intimate picture of our lives. Researchers have confirmed that simple endorsements in social media reveal likely political opinions, sexual preferences, lifestyle preferences, social circles, personal habits and patterns of behaviour. This is just based on clicking activities, without the target providing any personal details.^{xli}

The SQUEAKY DOLPHIN programme^{xlii} gives GCHQ the ability to monitor Facebook, Twitter and YouTube in real-time, collect addresses from videos watched and other user information, on a daily basis.

According to the leaked presentation, this programme currently appears to be an experiment in mass psychology and socio-cultural research. The objective seems to be understanding trends, rather than building individual profiles. But there are ethical problems with any non-consensual research of this kind.

In addition, GCHQ is expanding this programme into more targeted forms of intervention. They are using social media for the targeted enrichment of profiles, for example by, adding geolocation. The agency, is also using these tools to identify specific individuals, such as top Twitter users in an area.

GCHQ has several programmes dedicated to monitoring other social media, such as Google +,^{xliii} LinkedIn,^{xliv} bulletin boards and online fora.

1.6.3 Apps data

The agencies monitor the data sent by mobile phone apps, including Angry Birds, about their users.^{xlv} According to leaked NSA documents, mobile apps data provide an intimate picture of people's lives, depending on the information they have personally supplied. This information can include home country, current location (through geolocation), age, gender, postcode, marital status (the options included single, married, divorced, swinger and more), income, ethnicity, sexual orientation, education level and number of children.

1.6.4 Private webcams

There are documents showing that GCHQ has tapped into the private webcam communications of innocent Yahoo subscribers without clear legal authorisation. The agency collected millions of pictures, including substantial amounts of explicitly sexual materials.^{xlvi} The programme, called OPTIC NERVE, apparently unknown to Yahoo, targeted 1.8 million unwitting users in a six month period without any form of minimisation or filtering. The images were apparently used to improve facial recognition software. Metadata and images were also fed into the NSA database and search engine XKEYSCORE.

US senators have launched an investigation^{xlvii} into these activities, accusing GCHQ of 'breathtaking lack of respect for privacy and civil liberties'. GCHQ has simply provided a boilerplate response about compliance with UK laws, but this programme seems very hard to justify under current legislation. We believe this particular episode deserves full investigation as one of the most egregious privacy intrusions documented in the leaked documents.

1.7 The future of data collection

Intelligence agencies and security services very soon will have routine access to a wealth of data from cars and home appliances, such as thermostats and fridges. The so-called Internet of Things will eventually see most electronic gear connected to the Internet in order to exchange data with users, manufacturers and third parties. A particularly concerning development is the emergence of wearable technologies and health sensors which can track not just minute movements but also a broad range of physiological information.

Given the current strategy of GCHQ and other signals intelligence agencies, it is to be expected that the agency will attempt to integrate any available data in their collection systems. But the umbrella group of EU data protection authorities, Article 29 Working Party, has raised concerns about the potential inferences derived from such data:

“Apparently insignificant data originally collected through a device (e.g. the accelerometer and the gyroscope of a smartphone) can then be used to infer other information with a totally different meaning (e.g. the individual’s driving habits). This possibility to derive inferences from such “raw” information must be combined with the classical risks analysed in relation to sensor fusion, a phenomenon which is well-known in computer science.”^{xlviii}

The amalgamation - or "triangulation" - of databases is a well known problem for privacy, and can provide a much richer picture of an individual's life, or even allow for the re-identification of previously anonymous data.^{xlix} A reformed regulation of surveillance should take this into account.

A review of interception will also need to address not just the handling of such data by agencies, but the threat of sabotage of domestic appliances. There is widespread evidence that GCHQ engages in actively hacking computer systems, including those belonging to innocent third parties. The former chief of the CIA, David Petraeus, told a conference in 2012 that the Internet of Things presented an opportunity to spies.¹

1.8 Conclusion

It is clear that GCHQ possesses a huge breadth of collection capabilities and techniques. They have an unprecedented global access to data. From analysing the documents and the wealth of programmes and methods for data collection, we can witness a vast breadth of ambition.

GCHQ demonstrates a technological sophistication that is rare in government. They have solved many problems that people would have thought nearly impossible, such as reconstruction of content.

The implication of these revelations is that in the future people will expect the worst from our agencies. This is something which perhaps companies who worked with them should have expected, but did not.

There is a more or less full integration of the UK’s collection regime with USA’s NSA., including the sharing of the XKEYSCORE technology. This pattern continues as we will see in the next section.

-
- i One internal document quotes the head of the NSA, Lieutenant General Keith Alexander, on a visit to Menwith Hill in June 2008, asking: "Why can't we collect all the signals all the time? Sounds like a good summer project for Menwith." <http://www.theguardian.com/uk/2013/jun/21/gchq-cables-secret-world-communications-nsa>
- ii <http://www.duncancampbell.org>
- iii <http://www.theguardian.com/uk/2013/jun/21/gchq-cables-secret-world-communications-nsa>
<http://www.theguardian.com/uk/2013/jun/21/gchq-mastering-the-Internet>
- iv <http://www.pcpro.co.uk/news/security/382666/how-spies-could-tap-fibre-cables>
- v <http://www.theguardian.com/uk-news/2013/oct/25/leaked-memos-gchq-mass-surveillance-secret-snowden>
- vi http://www.duncancampbell.org/menu/surveillance/echelon/IC2000_Report%20.pdf
- vii <https://www.eff.org/nsa/hepting>
- viii <http://www.sueddeutsche.de/digital/internet-ueberwachung-snowden-enthuellt-namen-der-spaehenden-telekomfirmen-1.1736791>
- ix https://edwardsnowden.com/wp-content/uploads/2014/11/partner_cables.pdf
- x http://www.theregister.co.uk/2014/06/03/revealed_beyond_top_secret_british_intelligence_middleeast_internet_spy_base/?page=2
- xi <http://www.sueddeutsche.de/politik/nachrichtendienst-gchq-briten-schoepfen-deutsches-internet-ab-1.1704670>
- xii Cables divide the signals using what is termed wavelength-division multiplexing (WDM). This technique uses different infrared light colours (wavelengths), in some cases dozens of them, to separate the channels.
- xiii https://edwardsnowden.com/wp-content/uploads/2014/11/WHERE_WE_ARE.pdf
- xiv <http://www.channel4.com/news/spy-cable-revealed-how-telecoms-firm-worked-with-gchq>
<http://international.sueddeutsche.de/post/103543418200/snowden-leaks-how-vodafone-subsiadiary-cable>
- xv <http://www.theguardian.com/business/2013/aug/02/telecoms-bt-vodafone-cables-gchq>
- xvi <http://electrospace.blogspot.co.uk/2014/11/incenser-or-how-nsa-and-gchq-are.html>
- xvii http://www.duncancampbell.org/menu/surveillance/echelon/IC2000_Report%20.pdf
- xviii <http://www.spiegel.de/netzwelt/netzpolitik/Internetueberwachung-so-maechtig-sind-xkeyscore-tempora-und-prism-a-914300.html>
- xix <http://www.spiegel.de/netzwelt/netzpolitik/Internetueberwachung-tempora-ist-schlimmer-als-prism-a-907337.html>
- xx https://www.eff.org/files/2014/06/23/gchq_report_on_the_technical_abilities_of_tempora.pdf
<http://www.spiegel.de/media/media-34103.pdf>
- xxi <http://www.wired.co.uk/news/archive/2013-06/24/gchq-tempora-101>
- xxii <http://cryptome.org/2013/10/nsa-sso-overview.pdf>
- xxiii <http://www.spiegel.de/media/media-34103.pdf>
- xxiv <https://en.wikipedia.org/wiki/XKeyscore>
- xxv Ibid.
- xxvi [https://en.wikipedia.org/wiki/Fairview_\(surveillance_program\)](https://en.wikipedia.org/wiki/Fairview_(surveillance_program))
- xxvii <http://blogs.law.harvard.edu/surveillance/2008/11/11/narus-security-through-surveillance/>
- xxviii <http://arstechnica.com/information-technology/2013/08/building-a-panopticon-the-evolution-of-the-nsas-xkeyscore/>
- xxix <http://www.theguardian.com/world/2013/nov/07/heads-of-gchq-mi5-and-mi6-appear-before-intelligence-committee-live>
- xxx <http://www.theguardian.com/world/2013/jul/31/nsa-top-secret-program-online-data>
- xxxi http://upload.wikimedia.org/wikipedia/commons/7/74/NSA_Primary_FORNSAT_Collections.jpg
- xxxii <http://www.telegraph.co.uk/news/worldnews/europe/cyprus/10427890/British-military-base-in-Cyprus-used-to-spy-on-Middle-East.html>
- xxxiii <http://www.spiegel.de/international/world/snowden-documents-show-gchq-targeted-european-and-german-politicians-a-940135-2.html>

-
- xxxiv https://www.privacyinternational.org/sites/privacyinternational.org/files/file-downloads/eyes_wide_open_v1.pdf
- xxxv <http://electrospace.blogspot.co.uk/2015/01/how-gchq-prepares-for-interception-of.html>
- xxxvi <https://upload.wikimedia.org/wikipedia/commons/d/d9/Xkeyscore-hierarchy.jpg>
- xxxvii http://www.duncancampbell.org/menu/surveillance/echelon/IC2000_Report%20.pdf
- xxxviii [http://hudoc.echr.coe.int/sites/eng/pages/search.aspx?i=001-87207#{"itemid":\["001-87207"\]}](http://hudoc.echr.coe.int/sites/eng/pages/search.aspx?i=001-87207#{)
- xxxix <http://apps.washingtonpost.com/g/page/world/how-the-nsas-muscular-program-collects-too-much-data-from-yahoo-and-google/543/>
- xl http://www.washingtonpost.com/world/national-security/nsa-infiltrates-links-to-yahoo-google-data-centers-worldwide-snowden-documents-say/2013/10/30/e51d661e-4166-11e3-8b74-d89d714ca4dd_story.html
- xli <http://www.pnas.org/content/110/15/5802.full.pdf+html>
- xlii http://investigations.nbcnews.com/_news/2014/01/27/22469304-snowden-docs-reveal-british-spies-snooped-on-youtube-and-facebook?lite
- xliii BUGSY
- xliv FATYAK
- xlv <http://www.theguardian.com/world/2014/jan/27/nsa-gchq-smartphone-app-angry-birds-personal-data>
- xlvi <http://www.theguardian.com/world/2014/feb/27/gchq-nsa-webcam-images-internet-yahoo>
- xlvii <http://www.theguardian.com/world/2014/feb/28/nsa-gchq-webcam-spy-program-senate-investigation>
- xlviii http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf
- xlx http://ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation
- l <http://www.wired.com/2012/03/petraeus-tv-remote/>