The Tor Project—written evidence (IPB0122)

Introduction

Background to The Tor Project and the Tor software

1 The Tor Project is a 501(c)(3) non-profit based in the United States, but with employees, contractors, and volunteers worldwide (including the United Kingdom). The Tor Project conducts research, training, and software development to improve Internet privacy and safety, and to promote human rights, free speech, free expression and civic engagement.

2 The Tor Project is predominantly funded by Non-Governmental Organisations (NGOs) and governments, as well as individual and corporate donations. Recent funders include the Swedish International Development Agency (Sweden), the Broadcasting Board of Governors (US), the National Science Foundation (US), the NLnet Foundation (Netherlands) and the Ford Foundation (US).

3 The core software product developed by The Tor Project, "Tor" was originally designed and implemented as a research project by the United States Naval Research Laboratory. The Tor software improves its users' safety while using the Internet by redirecting communications via the Tor network – approximately 7,000 computers ("nodes") operated by volunteers worldwide. The nodes chosen for a particular communication are selected randomly by the Tor software running on the user's computer.

4 Communications sent via Tor typically will pass through three nodes before being sent to the ultimate destination. Each of these Tor nodes will know the source immediately before it, and will know the next destination for the communication, but any one node will not know both the original source and ultimate destination for the communication. Communication between nodes, and between the user's computer and the Tor network are encrypted to protect against eavesdropping and tampering.

5 Through this approach, Tor protects users against someone observing their computer's Internet connection from discovering which websites they are accessing, and with whom they are communicating. This could be of importance, for example, to a journalist collecting information about human rights abuses from sources whose personal safety could be put at risk if the government discovered they were talking to journalists.

6 Tor also prevents websites from discovering the identity of visitors. This could be of importance, for example, to a law enforcement agency collecting intelligence from a website suspected to be involved in criminal activity. Equally, normal Internet users may desire privacy and want to protect their identity from websites who they are concerned might profile their behaviour and use it inappropriately or sell it.

7 A rapidly growing use of Tor is to allow users to circumvent national censorship schemes. Such censorship may be long term, such as the "Great Firewall of China", or can be responsive to events, such as the blocking of Facebook and YouTube by the Tunisian regime in the run-up to the late 2010/early 2011 revolution.

8 Other uses of Tor include victims of crime talking to fellow survivors anonymously, children protecting their personally identifiable information while using the Internet, military personnel working undercover, operators of anonymous tip-lines reducing the risk of their sources being compromised, whistleblowers reporting on corruption, and financial institutions conducting due-diligence.

9 Further information about The Tor Project can be found on our website: https://www.torproject.org/

Use of the Internet by Human Rights Activists

10 This submission is not only based on how the Draft Investigatory Powers Bill would affect The Tor Project and users of its software, but also how the draft bill would affect more general use of the Internet by human rights activists. Information included in this submission is based on experience by Tor Project members in training human rights activists on how to effectively and safely use computers and the Internet.

11 Internet usage by Human Rights Activists can be broadly split into two categories.

12 Firstly there is the use of general-purpose Internet services, such as Facebook, YouTube, Twitter, Flickr, and webmail providers. These are popular amongst human rights activists for organizing their supporters because they are familiar, easy to use, and capable of withstanding bursts in demand that might swamp smaller services. They are also widely used outside of human-rights circles and so may draw less attention by the regime being defended against, and make it easier to get information out of the country to promote their case abroad.

13 Secondly, there are special-purpose tools designed with human rights activists as a significant (although perhaps not exclusive) target user group. Tools in this category include Tor and Martus (a software package developed by Benetech[1] for securely collecting data of human rights abuses). Such tools are developed because there is a lack of security or functionality in general-purpose Internet services and software packages.

14 Both categories of usage are important, although performing a quantitative comparison is difficult. Use of general-purpose Internet services for human rights is likely to be more predominant, but while uses of special-purpose Internet services may be fewer in number they may be greater in their importance.

Comments on the Draft Investigatory Powers Bill

Security of stored communications and communications data

Addressing questions on secure retention of intercepted material, and requirements placed on service providers

15 The draft bill states that communications and communications data that is collected or processed as a result of powers granted by the bill should be protected from unauthorised access. Examples of such requirements include Clause 53, covering filtering arrangements, and Clause 74 covering retained communications data. However, evidence shows that the current state of the art in computer security is not sufficient to adequately protect communications or communications data, or to restrict access to facilities built to collect or process this material. Although there are techniques to protect computer systems from large-scale attacks, there are no effective measures for protecting computer systems from targeted attack by a capable adversary, especially when an adversary with state backing is a possible threat (as is the case with communications and communications data concerning human rights activists).

16 This can be seen from the numerous breaches of security of communications service providers, even those who by far exceed industry standard levels of protection. It is likely that there are other cases of breaches that have not been disclosed due to commercial sensitivity.

17 One such example is the breach of Google's webmail service in December 2009[2]. This attack was specifically targeted against Chinese human rights activists. The breach of

[1] https://www.martus.org/

[2] http://googleblog.blogspot.co.uk/2010/01/new-approach-to-china.html

Google was part of a co-ordinated and sophisticated attack that also included Adobe and other companies that chose not to be publicly disclosed[3]. The attack made use of custom-made malware that was designed to, and succeeded at, avoiding detection by anti-virus software. It also exploited a vulnerability in Microsoft Internet Explorer which was, at the time of the attack, not known publicly. The identity of the attackers remains unknown and was disguised by transmitting their communications through hijacked computers in the US and Taiwan.

18 Another notable incident is the compromise of the Vodafone telephone exchange in Greece[4], allowing attackers to bug the mobile telephone of over 100 high-ranking dignitaries, including the prime minister of Greece. In a highly sophisticated attack, custom-designed software activated the lawful-intercept functionality of the telephone exchange even though Vodafone had not purchased it. The attackers also successfully circumvented the audit logging, thereby hiding their unauthorised access. Eventually, the tampering was discovered, but only after almost a year of being active (the exact date the attack was perpetrated remains unknown).

19 As a final example, a hacker supportive of the Iranian government but who stated that he was not affiliated to the government, compromised the certification authorities DigitNotar and Comodo (and claims to have compromised others), and obtained digital certificates which were used to impersonate Google's website, potentially collecting sensitive information such as passwords, communications data, and content[5]. The same attacker also targeted The Tor Project website, so it is reasonable to suspect that human rights activists were among the targets.

Sensitivity of Communications Data

20 The draft bill requests that communications data, not content, may be collected through a retention notice. The Home Secretary argued that communications data is less sensitive than content ("the modern equivalent of an itemised phone bill"), and thus does not deserve the same safeguards, restrictions on collection, or level of authorisation to access.

21 However, in many cases communications data can be as sensitive as content, and in some cases may be more sensitive than content.

22 For example, Internet Connection Records revealing that someone accessed a website which is collecting evidence on human rights violations could put that person or their family in severe danger. Internet Connection Records would also reveal whether someone had visited a site for people with cancer or alcoholism.

23 Even disclosing that someone was using the Internet at a particular time can be sensitive when it is correlated with, for example, the posting of videos of human rights abuses on YouTube. While the timing of a single instance of a video is unlikely to uniquely identify a person, repeating this exercise, with knowledge of the "usual suspects" for such activity, could single out an individual for repercussions.

24 Experiments have shown that 23.3% of Wikipedia users could be uniquely identified from Internet Connection Records alone, had they been using Tor to protect their privacy[6]. This proportion goes to 95.7% when only Wikipedia users who have posted 50 or more items on Wikipedia are considered.

[3] http://www.wired.com/threatlevel/2010/01/operation-aurora/

[4] http://spectrum.ieee.org/telecom/security/the-athens-affair/

[5] http://arstechnica.com/security/2011/09/comodo-hacker-i-hacked-diginotar-too-other-cas-breached/

25 As another example, communications data showing that a phone call made by a journalist from a particular location could put that journalist at risk. It has been reported that the Syrian government was using Internet communications data analysis to target journalists. This technique has been implicated in the death of Sunday Times war correspondent Marie Colvin[7].

26 Even "entity data", while typically less sensitive than Internet connection records, can be of critical importance. The disclosure of the identity of a person pseudonymously blogging about sexuality, political or religious beliefs could put someone's employment at risk, even within liberal democracies.

27 The reason that communications data can be more sensitive than content is that it is more amenable to automated analysis, particularly when collected in bulk (as proposed by the draft bill). Content is designed for humans to read, and it is a challenging problem for computers to accurately interpret content. In contrast, communications data is designed for computers to interpret and so is far easier for computers to analyse. Communications data allows a more accurate and detailed profile of individuals to be built than is possible with current technology to interpret content.

28 The examples above show that the discussion of the draft bill should not centre on the false tradeoff between civil liberty and security. While it is undoubtedly not the intention of the Home Office, this draft bill will significantly harm the safety of human rights activists. The discussion of the draft bill thus can be framed as a tradeoff between giving additional powers to law enforcement in exchange for taking away the ability of human rights activists and human rights organisations to protect themselves.

29 In making this tradeoff it is also important to note that while a single breach of security is sufficient to compromise the safety of a human rights activist, the inability of law enforcement to obtain communications data relevant to a crime does not mean that the investigation will not succeed. There are frequently alternative sources of information that will result in a successful outcome of the case.

Safeguards

Addressing safeguards on accessing communications, communications data and undertaking equipment interference activities

30 The draft bill proposes safeguards for access to communications and communications data, such as requiring approval by a senior officer before an application can be made, and requiring that service providers retain data securely.

31 As discussed above, it is unlikely that mechanisms to prevent unauthorised access to data, or interception facilities, will work as needed. Audit mechanisms, to detect authorised access, are for the same reasons likely to be possible to bypass.

32 Furthermore, law enforcement agencies and intelligence agencies will likely require that the queries processed under filtering arrangements (Section 51 of the draft bill) be themselves confidential (as the compromise of this data could interfere with investigations). Therefore it will likely not be possible for the service provider to properly audit access, and it will be challenging to safely store logs for any subsequent audit by the Investigatory Powers Commissioner.

[6] http://www-users.cs.umn.edu/~hopper/surf_and_serve.pdf

[7] http://www.telegraph.co.uk/news/worldnews/middleeast/syria/9098511/Marie-Colvin-Britain-summons-Syria-ambassador-over-killing.html

33 Even ignoring the significant possibility of unauthorised access to stored communications or communications data, and ignoring the significant possibility of unauthorised enabling of interception functionality, the mere possibility that the powers in this draft bill will be exercised introduces harm.

34 This is because the cost and risk of adding new functionality to a computer system grows dramatically the later in the development process that the change is introduced. While it may be comparatively cheap to add new functionality while a system is on the drawing board, it will be much more expensive to add the same functionality once the system is deployed in the field.

35 Therefore, the fact that the powers in the draft bill might be exercised will lead to service providers and their equipment suppliers to put in place functionality to intercept and store communications data, even before any powers are exercised. Providers may adopt designs for their systems which facilitate interception, such as through greater centralisation, but which leave the systems more.

36 As a consequence, the risk of interception capability being activated without authorisation will increase. Furthermore, the same equipment will likely be sold to other countries which may use the same interception capability to spy on human rights activists.

37 It is also likely that other countries will use the fact that the UK is proposing such legislation as a justification for their own surveillance proposals. This pattern was recently seen when the Chinese state news agency capitalised on the Prime Minister's statement to the House of Commons contemplating the censorship of social networks during the 2011 riots[8].

Responses from industry

38 The response of service providers to the risks to human rights activists that the proposed bill presents will depend on how important human rights activists, and others who depend in Internet security for their safety, are to the company’s priorities.

39 For general-purpose Internet services, human rights activists are a relatively small proportion of their usage base, and while some providers have been proactive in protecting human rights activists from attack (such as Google[9]), other commercial considerations will likely take priority, and these are better left stated by the companies themselves.

40 In contrast, Internet services designed for human rights activists will likely take a more proactive response in protecting users from harm and so are more likely to avoid being put in the position of having to compromise user safety by avoiding having a UK presence.

41 In the particular example of Tor, recall that it is the user's computer that chooses the path through the network, so if there is sufficient fear that UK nodes are unsafe, users are free to avoid UK nodes without any intervention of The Tor Project.

42 Projects, such as Tor, may also consider that carrying out software development in the UK is too high a risk, because the draft bill may allow developers to be compelled to assist in the implementation of an equipment interference warrant (Clause 101).

[8] http://opennet.net/blog/2011/08/amidst-riots-uk-calls-censor-social-media

[9] http://www.guardian.co.uk/technology/2012/jun/06/google-state-sponsored-hacking

43 The creation of vulnerabilities in software through targeted equipment interference warrants or technical capability notices (Clause 189) not only puts the users of the system at risk, but also the developers because it creates the possibility that someone could intimidate the service provider staff into disclosing communications, private information or equipment data.

Circumvention

Addressing necessity of requirements

44 As can be seen with the attacks on Vodafone in Greece, Google and Adobe in the UK, and DigiNotar in Denmark (in all of these the identity of the attackers is unknown), it is well within the capabilities of sophisticated attackers to hide their traces by hijacking computers and using these as stepping stones. Hijacked computers are effectively being used as a telecommunications service provider, but will not fall under the control of this law because the owner of the hijacked computer will not know that it is being used as a telecommunications service provider.

45 There are well-known techniques[10], and software available, for defeating tracing communications based on communications data. Specifically, messages are delayed, and extra “dummy” messages are added, at each point that communications are relayed. Such techniques incur a high overhead but an attacker who has hijacked a computer to act as a stepping stones will not be paying for the network resources and therefore will have no need to be concerned at the cost.

21 December 2015

[10] http://mixminion.net/