*Wikimedia Foundation v. NSA*
No. 15-cv-0062-TSE (D. Md.)

# Plaintiff's Exhibit 1

**IN THE UNITED STATES DISTRICT COURT**
**FOR THE DISTRICT OF MARYLAND**

| | |
|---|---|
| WIKIMEDIA FOUNDATION,<br><br>*Plaintiff*,<br><br>v.<br><br>NATIONAL SECURITY AGENCY / CENTRAL SECURITY SERVICE, *et al.*,<br><br>*Defendants*. | No. 1:15-cv-00662-TSE |

**DECLARATION OF SCOTT BRADNER**

**I. INTRODUCTION ..... 1**

**II. SUMMARY OF MY CONCLUSIONS ..... 2**

**III. QUALIFICATIONS ..... 4**

A. Employment ..... 5
B. Publications ..... 5
C. The Internet Engineering Task Force ..... 6
D. Involvement in Data Network Design and Operation ..... 6
E. Compensation ..... 7

**IV. BACKGROUND OF THE TECHNOLOGY IN THIS CASE ..... 7**

A. History of the Internet ..... 8
1. Pre-1960s ..... 8
2. Advanced Research Projects Agency (ARPA) ..... 9
3. The origin of Packet Data Networks ..... 10
4. Packets ..... 10
5. The ARPANET ..... 14
B. Definitions ..... 17
1. A Communication ..... 17
2. Layers, links and nodes ..... 18
3. Flow ..... 21
4. Transaction ..... 22
5. Network ..... 23
6. Network Node ..... 24
7. Circuit ..... 24
8. Packet ..... 25
9. Switch ..... 29
10. Router ..... 29
11. Mirroring ..... 29
12. Routing ..... 30
13. Internet Protocol ..... 30
14. Internet Service Provider (ISP) ..... 30
15. Proxy ..... 31
16. Tunnel ..... 31
17. Metadata ..... 31
C. The Key Internet Protocols ..... 32
1. The Internet Protocol Suite ..... 32
a. The Internet Protocol (IP) ..... 32
i. IP addresses ..... 33
ii. Viewing IP header information ..... 34
iii. Sizes of IP packets ..... 34
iv. Multiple packets in a communication ..... 35
b. Transport Protocols ..... 35
i. The User Datagram Protocol (UDP) ..... 36
ii. The Transmission Control Protocol (TCP) ..... 37
2. Application Protocols ..... 39

a. The Hypertext Transfer Protocol (HTTP)....39
i. HTTP commands....39
ii. Encrypted HTTP (HTTPS)....39
iii. HTTPS Handshake....41
iv. IP addresses in HTTP packets....42
b. Email....42
i. Email Header Information....43
ii. Email Servers....44
iii. Simple Mail Transfer Protocol (SMTP)....45
(1) SMTP Metadata....47
(2) IP addresses in email packets....47
iv. Internet Message Access Protocol (IMAP)....47
c. Telephone Calls....48
3. Plain Text in Application Protocol Headers....48
4. Number of Packets in a Communication....49
D. Other Features of the Internet and its Architecture Relevant to this Case....50
1. Internet Architecture....50
2. Internet Backbone....51
3. Internet Service Providers (ISPs)....53
a. Address assignments for ISPs....55
4. ISP Interconnection....56
5. Customer Networks....57
a. Address assignments for customer networks....58
6. Customer Network Interconnection....58
7. Network Address Translators (NATs)....59
E. Routing in the Internet....59
1. Autonomous System (AS)....61
2. Routing an IP Packet....61
3. Volatility of Routing Information....64
4. Asymmetric Data Paths....66
F. International Connections....67
1. Details of Undersea Fiber-Optic Cables....70
2. Details of Terrestrial Fiber-Optic Cables....73
3. Public Internet Communications on International Fiber-Optic Cables....73
4. Undersea Fiber-Optic Cable Landing Locations....75
5. Terrestrial Fiber-Optic Cable Terminations....77
G. Places to Monitor International Public Internet Communications....78
H. Locating Network Nodes Using IP Addresses....81

**V. NSA'S SECTION 702 COLLECTING OF COMMUNICATIONS....82**
A. Selectors....83

**VI. PRISM COLLECTION PROGRAM....88**

**VII. OPINIONS A, B & C: THE NSA'S UPSTREAM COLLECTION PROGRAM INVOLVES COPYING, REASSEMBLING AND REVIEWING INTERNET TRANSACTIONS....88**
A. Upstream Collection Program....90
1. A Description of NSA's Upstream Collection Program....91
2. Upstream Collection Process....93
a. Stage 1: Copying the Packets....94
i. Copy-Then-Filter....96
(1) Fiber-optic splitter....96
(2) Link-Layer Copying....97
(3) Filtering the packets....97
ii. In-Line Filter....98
iii. Implementation....98
b. Stage 2: Filtering....101
c. Stage 3: Reassembling Transactions....106
d. Stage 4: Reviewing Transactions....109
i. "multiple communications transaction (MCT)" collection....111
ii. "about" collection....112
iii. Collection of Encrypted Internet Transactions....114
e. Stage 5: Ingesting Transactions....116
3. Upstream Collection Monitor Placement....116

**VIII. OPINION D: WIKIMEDIA COMMUNICATIONS ARE TRANSPORTED ON ALL INTERNATIONAL CIRCUITS ORIGINATING OR TERMINATING IN THE UNITED STATES.....119**
A. Wikimedia....120
1. Wikimedia Websites....120
2. Wikimedia International Communications....121
3. Protocol Support on Wikimedia Websites....124

**IX. OPINION E: THE NSA HAS COPIED, REASSEMBLED AND REVIEWED WIKIMEDIA COMMUNICATIONS....124**

**X. DR. SCHULZRINNE'S DECLARATION....126**
A. Surveillance Configurations....127
B. Selectively Filtering Internet Traffic....129
C. Selectively Filtering Wikimedia IP addresses....132
D. U.K. Surveillance Disclosures and Court Proceedings....134

## I. INTRODUCTION

1. My name is Scott Bradner. I have been asked by the plaintiff's counsel in *Wikimedia Foundation v. National Security Agency*, No. 1:15-cv-006622-TSE (D. Md.), to provide an expert report addressing the following questions:

   a. What is the basic structure of the Internet and how do communications traverse it?

   b. How does upstream collection work, based on official government acknowledgments and my expertise in network design and operation?

   c. What is the likelihood that the government has copied and reviewed the plaintiff's international text-based Internet communications in the course of upstream collection?

2. In this declaration I will address these questions based on my own technical expertise and experience, and on relevant technical principles. The information I used to understand and explain the Section 702 collection of Internet transactions came from my review of documents provided to me by plaintiff's counsel. Counsel has informed me that all of the U.S. government documents they provided have been officially released by the government.

3. A list of the documents provided to me by plaintiff's counsel is attached as Appendix B.

4. After explaining my conclusions, I address the declaration of Dr. Henning Schulzrinne, filed in support of the government's motion for summary judgment.

5. In this declaration, I refer to actions such as copying, reassembling and reviewing as if they were wholly performed by the NSA. But in doing so, I am specifically including the possibility that some or all of those actions are performed by

others at the direction of the NSA. In addition, when I refer to "copying," as in "copying a packet," I am including any process which results in one or more duplicate copies of the original packet, for example, splitting a beam of light and reconstructing packets from each portion of the split light beam, as well as using an electronic device to produce a copy of a packet it received on a network.

## II. SUMMARY OF MY CONCLUSIONS

6. After reviewing the materials available to me in this case I have concluded the following. These conclusions apply both before and after the "about" collection was stopped:

a. It is my opinion that, to conduct upstream collection of international Internet communications traversing any particular circuit, as this operation has been described by the government, the NSA must be copying at an absolute minimum the packets constituting the transactions it wishes to review for the presence of selectors. Based on other practical necessities I describe below, it is also my opinion that the NSA is almost certainly either (1) copying all packets traversing that circuit or (2) copying all of the packets that an IP address filter test determines are not part of a wholly domestic transaction.

b. It is my opinion that, in order to review Internet transactions to determine if a selector tasked for collection is present, the NSA must be reassembling the packets of the transactions it intends to review.

c. It is my opinion that the NSA must review the reassembled Internet transactions in order to identify those that include a tasked selector and thus are subject to collection under the upstream collection program.

d. It is my opinion that it is virtually certain that Wikimedia's international communications traverse every circuit carrying public Internet traffic on every international cable connecting the U.S. to other countries.

e. It is my opinion that it is virtually certain that the NSA has, in the course of the upstream collection program, copied, reassembled and reviewed at least some of Wikimedia's communications.

7. I have carefully reviewed the declaration of Dr. Schulzrinne, and nothing in it alters the above conclusions. I will address parts of his declaration at various places in my declaration and more fully at the end of my declaration. In summary, I conclude as follows:

a. Dr. Schulzrinne does not directly address the likelihood that the NSA has, in the course of the upstream collection program, copied, reassembled and reviewed at least some of Wikimedia's communications. Accordingly, he does not deny that it is virtually certain that the NSA has, in fact, done so.

b. Dr. Schulzrinne speculates that the NSA could, in theory, have designed its upstream collection program to have avoided the copying, reassembly and review of *any* of Wikimedia's communications, but as I explain in detail below, his speculation is technically inaccurate and it is, as a practical matter, simply implausible that the NSA designed and operated its upstream collection program as Dr. Schulzrinne speculates it could have to avoid such copying, reassembly and review. For example, he speculates that the NSA could have been and is "blacklisting" Wikimedia's IP addresses or could have been and is filtering out all web traffic from upstream collection. Blacklisting

Wikimedia's IP addresses would not in fact avoid the copying, reassembly or review of Wikimedia's communications, as I explain below. Moreover, there is no reason to believe that the NSA has been or is currently attempting to filter out Wikimedia's traffic, and there are compelling reasons to believe that it isn't. Finally, it strains credulity to suggest that the NSA is, in the course of an Internet surveillance program, deliberately filtering out all *web activity*, one of the most common modes of communication on the Internet. (The NSA has in any event confirmed that it monitors web activity under upstream collection, as I note below.)

c. For these reasons, Dr. Schulzrinne's speculation about technically possible but exceedingly unlikely measures the NSA might have been taking or might currently be taking to avoid Wikimedia's communications do not alter my conclusion that it is a virtual certainty that the NSA has, in the course of the upstream collection program, copied, reassembled and reviewed at least some of Wikimedia's communications.

8. I rely on my own knowledge as well as public technical publications for the background of the technology section and on the documents supplied to me by Wikimedia's counsel to understand the NSA's upstream collection program and to support these opinions.

## III. QUALIFICATIONS

9. My background and expertise that qualify me as an expert in the technical issues in this case are as follows:

### A. Employment

10. I worked at Harvard University ("Harvard") in a number of information technology roles, from 1966 to 2016, at which time I retired. My last role at Harvard was as a Senior Technology Consultant in the office of the Harvard University Chief Technology Officer (CTO) where I worked on identity management projects. Before joining the Harvard CTO's office I was the Harvard University Technology Security Officer (UTSO) for 8 years. I currently teach courses on *Technology, Security, Privacy, and the Realities of the Cyber World* at the Harvard University Extension School and have supervised masters and Ph.D. theses for students in Harvard University itself and in the Harvard University Extension School. In the past I have taught classes for undergraduate and graduate students at Harvard University and multi-day tutorials in the 1990s to thousands of students at the largest U.S. Internet-related trade show as well as at a number of major technology companies including IBM, Oracle and Nortel. I have also consulted for many technology companies, a number of universities and for multiple departments within the U.S. government.

### B. Publications

11. I have authored or co-authored 4 books and over 90 articles or other publications in peer-reviewed journals, conference proceedings, popular publications, monographs and standards organizations. These publications span a range of topics including analyzing network hardware, Internet technology, technology policy and standards processes. In addition, between 1992 and 2013 I wrote a regular column in the technical journal *Network World*, which was read around the world.

### C. The Internet Engineering Task Force

12. The Internet Engineering Task Force (IETF) is a primary standards creation and maintenance body for the Internet. The work of the IETF is conducted in Working Groups and IETF Working Groups are organized into Areas. Each of the technical areas in the IETF is managed by one to three Area Directors. At various times I served as the Director or co-Director of the IETF's Operational Requirements, Operations and Management, IP Next Generation, Transport and Sub-IP areas. As an Area Director, I served as one of the members of the Internet Engineering Steering Group (IESG), the IETF's standards approval and general management committee from 1993 to 2003. As a member of the IESG, I reviewed and evaluated hundreds of IETF working documents that were proposed by IETF working groups or IETF participants to be approved as IETF standards. The documents I was involved in approving covered all areas of IETF technology and included all aspects of Internet design, operation and evolution. I will note in passing that I worked often with Dr. Schulzrinne in the IETF.

### D. Involvement in Data Network Design and Operation

13. I was involved in the design, operation and use of data networks at Harvard University since the early 1970s, and was involved in the design, implementation and operation of the original Harvard data networks, the Longwood Medical Area network (LMAnet) and the New England Academic and Research Network (NEARnet).

14. Additionally, I was the founding chair of the technical committees of LMAnet, NEARnet and the Corporation for Research and Enterprise Network (CoREN). I was involved in the day-to-day operation of these networks as well as their evolution.

15. I have also served as a consultant on network design, management and security to educational institutions, federal agencies, international telecommunications enterprises and commercial organizations ranging from Fortune 500 companies to small businesses, from 1989 to the present. I have served as an expert witness in the Communications Decency Act challenge (*Reno v. ACLU*, 521 U.S. 844 (1997)) in U.S. federal court and in a number of patent cases.

16. In addition, I have also served on the technical advisory boards of about two-dozen companies in various technology fields, mostly relating to the Internet and other data networks, and I have been a frequent speaker at technical conferences.

17. My CV and list of previous cases is attached to this declaration as Appendix A.

### E. Compensation

18. I am not being compensated for my work in this case other than for travel expenses, if any.

## IV. BACKGROUND OF THE TECHNOLOGY IN THIS CASE

19. I agree in general with the background information Dr. Schulzrinne provides in ¶¶ 16-44 of his declaration. I note below where we disagree. The following involves more detail and sometimes a different focus from Dr. Schulzrinne's background section.

20. This case involves communications over the Internet. The Internet is the world-wide collection of interconnected networks that operate following the standards that define the Internet Protocol. The different networks that make up the Internet are operated independently. There is no overall manager of the Internet, nor is there any general form of governance of the Internet. The Internet operates by mutual agreement

among the companies that produce the computers that connect to the Internet and the companies that operate the independent networks that make up the Internet to implement the same set of technical standards in the software of the computers and to operate the networks in ways that are consistent with generally ad-hoc operational standards. See below for a fuller description of the Internet.

21. To put the relevant technologies and concepts in context, I will provide a brief history of the Internet, define some of the terms I will be using, explain the key protocols in use on the Internet today, and describe other key features of the Internet and its architecture relevant to this case.

### A. History of the Internet

22. I will now provide a short history of the Internet as a way to introduce the technology of the Internet that is relevant to this case.

#### *1. Pre-1960s*

23. The wiring of the world started with the Samuel Morse patent for the telegraph in 1847 and accelerated with the Alexander Graham Bell telephone patent in 1876. Until the late 1960s the networks that supported the telegraph and telephone services only supported those services—that is, they were specific-purpose not general-use networks. In describing the environment that led to the Internet, I will focus on the telephone network.

24. By the beginning of the 1960s, telephone networks had evolved into a general hierarchical hub-and-spoke architecture. The telephones in an area, for example a town, were connected to a telephone switch in a local central telephone office in that town with dedicated pairs of wires. As many as tens of thousands of telephones could be connected to each of these local central telephone switches. These local central office

telephone switches were connected to a more central telephone switch, which, in turn, was connected to an even more central switch.

25. To make a telephone call, a caller would dial a telephone number. The telephone number was sent, digit-by-digit, to the local central office telephone switch over the dedicated pair of wires. Using this telephone number, the local central office switch would then cause a dedicated path to be set up between itself and the local central office telephone switch connected to the telephone assigned the telephone number the caller had dialed. The path might traverse a number of telephone switches. The dialed telephone would then ring and, if someone picked up the dialed phone, a conversation could be held over the dedicated path. When the caller or called person hung up, the dedicated path established to support the call would be "torn down"—that is, the individual wires that had been used to make up the path would be released to be used for future telephone calls.

26. Two significant limitations of this telephone system architecture included:

a. That the wire between the telephone and the local central office telephone switch could only be used for one thing, a single telephone call, at a time.

b. That the failure of a telephone switch or of a connection between pairs of telephone switches would terminate all telephone calls whose paths went through the switch or link that failed.

### *2. Advanced Research Projects Agency (ARPA)*

27. Parallel developments in the Cold War between the U.S. and the Soviet Union set the stage for the development of the modern Internet. In the 1950s and 1960s, in that context, the launch of the Sputnik spacecraft by the Soviet Union on October 4, 1957 was a profound shock to the U.S. scientific and political establishment. In direct

response to the launch of the Sputnik, President Dwight David Eisenhower established the Advanced Research Projects Agency (ARPA) in the U.S. Department of Defense within three months of the launch. ARPA was established with a very broad mandate to undertake advanced research in any area that might be helpful to the U.S. military and, hopefully, to minimize the chance of another Sputnik-like surprise. ARPA came to play an important role in the development of the Internet.

***3. The origin of Packet Data Networks***

28. ARPA was not alone in supporting advanced research within the U.S. Department of Defense. Relevant to this history, the U.S. Air Force supported research efforts at RAND Corporation. One of the researchers at RAND was Paul Baran. Mr. Baran was very worried about the survivability of the telephone system that the U.S. military would need to use for communication in the aftermath of a nuclear attack on the U.S. Mr. Baran developed an alternative architecture that would have a much better chance of surviving mass destruction. That alternative architecture became the basis of today's Internet.

***4. Packets***

29. As noted above in ¶ 26, one of the issues with the architecture of the telephone system in the 1960s was that the failure or destruction of a single one of the large telephone switches or links between switches would terminate any call currently running though that switch or link. Mr. Baran developed the idea of using a large number of small switching nodes interconnected by links as a redundant array. The switching systems are represented by the dots and the links by the interconnecting lines in the sample distributed network shown in the figure below from Mr. Baran's 1962 paper *On Distributed Communications Networks*:

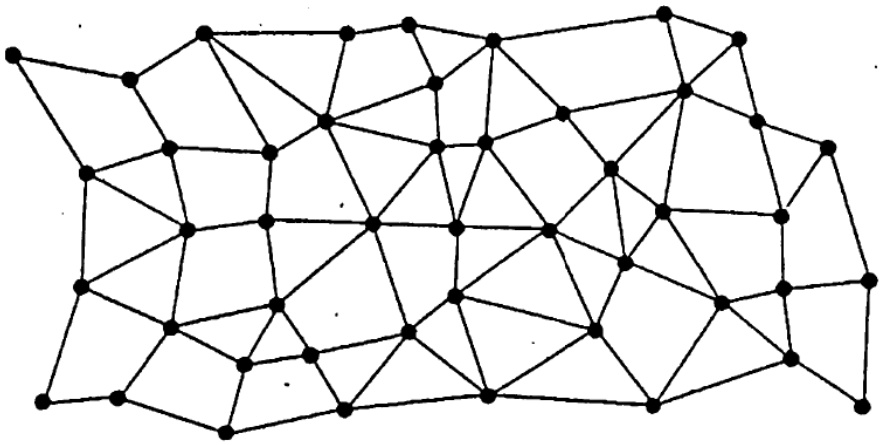


**Figure 1**[1]

30. With this type of redundant architecture, connections can get rerouted in case of a failure of a link or of a switching node. The following figure shows a sample path (green line) that could be used through a network. The path traverses a number of switching nodes and links:

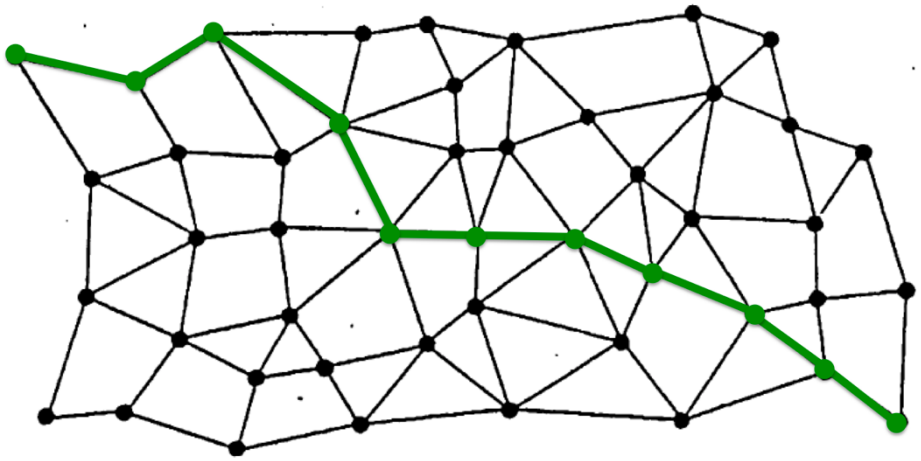


**Figure 2**

[1] Appendix G at 5 (Paul Baran, RAND Corp., *On Distributed Communications Networks* at 4 (Sept. 1962)).

31. The following is the same network showing a sample path after the failure of a switching node in the network (marked by the red X).

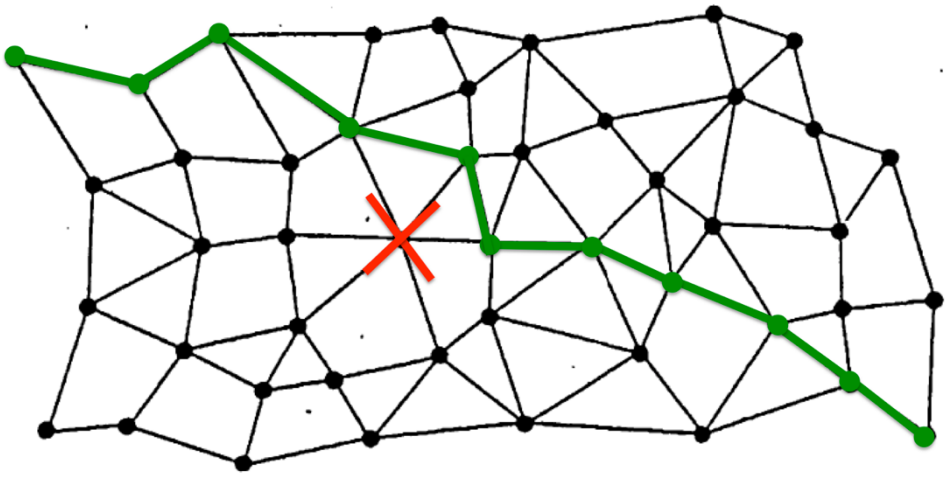


**Figure 3**

32. But redundancy, by itself, is not sufficient. A communication, such as a voice call, would be disrupted during any reroute of the communication path. So Mr. Baran developed the concept of breaking each communication up into multiple autonomous chunks, which he called message blocks but which are now known as "packets", the term which I will use in this report. Mr. Baran's diagram of a packet is shown in the following figure from his 1962 paper:

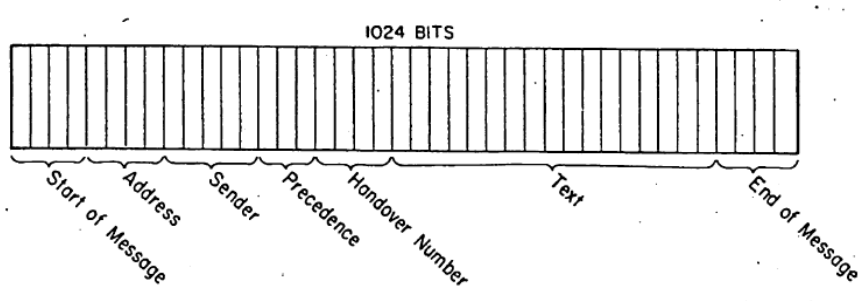


**Figure 4**[2]

33. Each of the packets includes a field at the start of the packet that tells the network to which network node this packet is to be delivered (the "Address" field), a field that says what network node sent the packet (the "Sender" field), some other control information (the "Precedence" and "Handover Number" fields) and a payload field which contains the chunk of information being transported in the packet (the "Text" field).

34. These same types of fields are present in the packets that traverse today's Internet. For example, the following figure shows the format of an Internet Protocol version 4 (IPv4) packet:

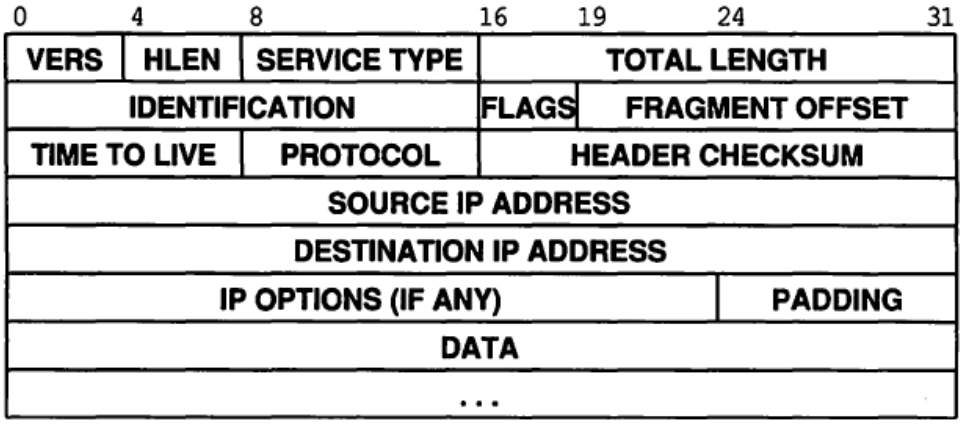



**Figure 5** — IP packet format[3]

35. IPv4 is the version of the Internet Protocol (IP) that was deployed in 1983 and is still the predominant version in use today. A revised version of IP, known as IPv6, is being deployed but is not yet in general use. The IPv4 and IPv6 headers differ but not in ways that alter this discussion.

[2] Appendix G at 27 (Baran, *supra* note 1, at 26).

[3] Douglas E. Comer, *Internetworking with TCP/IP: Principles, Protocols, and Architecture* (2nd ed. 1991).

36. The figure above shows a "source IP address" field (Baran's "sender" field), a "destination IP address" field (Baran's "address" field), a number of fields that correspond to Baran's control information (e.g., "service type", "protocol", etc.) and a "data" field (Baran's "text" field). (See ¶¶ 95-104 for a fuller discussion of the Internet Protocol.)

37. Breaking the communication into packets means that only a small part, if any, of the communication will get lost, and perhaps have to be retransmitted, if the path is disrupted by some failure rather than having the whole communication be terminated.

38. Another big advantage of using packets to carry communications is that multiple communications can be run over the same link at the same time by intermingling packets from different communications. Many, even hundreds or thousands, of separate communications can be running over a single link at the same time, and if the link is in the center of a network, such as the network shown in the Baran figure, these communications can be to and from many different sending and receiving nodes.

### *5. The ARPANET*

39. Meanwhile, back at ARPA, there was an interest in sharing big research computers among multiple researchers located around the country or even outside of the country. At that time computers that were needed for large-scale computation were physically very large and very expensive—much too expensive for the government to be able to provide a computer for each research institution. Thus ARPA had an interest in making it possible for researchers at different locations to be able to share the use of the large computers.

40. The approach ARPA decided to take was to build a nation-wide network to interconnect the big computers and the institutions where the researchers were located. ARPA also decided to use the basic concepts that Mr. Baran had developed, even though ARPA at that time was more interested in sharing computing resources than surviving nuclear attacks. The same technology, packet-based data networking, would support both types of needs.

41. The initial parts of the resultant network, known as the ARPANET, were installed in four locations on the U.S. west coast in late 1969. Within a few years the network had been extended to the U.S. east coast and to dozens of nodes. A few years later there were a few hundred ARPANET nodes including a few in the U.K. and Europe.

42. The original ARPANET design had a significant limitation. The ARPANET operated using the Network Control Protocol (NCP). NCP was designed to interconnect network nodes, generally a single node at a location such as a university but occasionally two or three. Bob Kahn realized that, in order to be able to grow, the design had to be changed such that the ARPANET would interconnect networks rather than nodes. Each location, such as Harvard, could have its own network with as many nodes as it wanted to have. The nodes on the networks at multiple sites could then communicate with nodes at other sites with an almost unlimited ability to grow the number of nodes.

43. Dr. Kahn enlisted the help of Dr. Vint Cerf, and together they developed the Internet Protocol. IP defines a way to interconnect networks (thus “inter-net”) such that a node on one network can communicate with another node on the same network or with a node on a different network. The Internet Protocol specifications define the

format of Internet Protocol packets, and, in a general way, how packets are constructed, transported and processed.

44. The ARPANET transitioned from NCP to the Internet Protocol starting on January 1, 1983. This was the start of the Internet, as the concept is understood today.

45. ARPA operated the ARPANET as a backbone network—i.e., a network that interconnected other networks—until 1990. Note that the ARPANET did not have a single link that was its backbone carrying all of its traffic. Instead, as shown in Figure 6 from October 1980, the ARPANET, like Internet Service Providers these days, had a mesh-like set of links that provided for redundancy and shared the traffic load. Traffic would only traverse as much of the ARPANET links as it needed to in order to reach its destination.

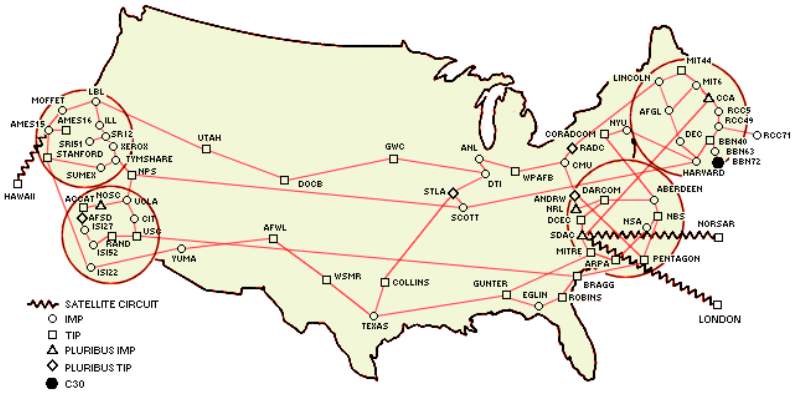



**Figure 6** — The ARPANET in October 1980[4]

46. By the time ARPA shut down the ARPANET, the U.S. National Science Foundation (NSF) was operating its own backbone network (NSFNET) to interconnect networks at NSF-sponsored universities and research centers. The NSF replaced the ARPANET until the NSFNET was closed down in 1995.

[4] *Internet Technology*, Technology UK, http://www.technologyuk.net/telecommunications/internet/internet-technology.shtml.

47. Starting in the late 1980s, commercial Internet service providers operating in parallel to the NSFNET began to appear. By the mid-1990s there were thousands of small local ISPs and a growing number of nation-wide ISPs. By the end of the 1990s, a few of the U.S. ISPs had expanded internationally.

### B. Definitions

48. The Internet today remains a packet data network, following Baran's original concept of redundant network connections and autonomously routed chunks of data called packets. Before explaining the key protocols and architecture of the Internet today, I will first specify what I mean by the terms that I will be using in this report. Unless otherwise noted, these definitions are widely accepted and consistent with the use of these terms by experts in the field of Internet communications and architectures.

#### *1. A Communication*

49. The term ***communication*** does not have a single precise definition in the field of Internet communications, but in the context of this report I will generally use the term "communication" to mean data exchanged between a pair of nodes on a network. Communications include phone calls, email messages, data files, requests for web pages and web pages. Communications are broken up into chunks, called ***packets***, for transmission over the network. Communications are bidirectional with packets flowing in both directions even when a user is viewing a web page.

### *2. Layers, links and nodes*

50. Networks are organized into ***layers*** to simplify design and operation. Each layer provides services to the layer above it and shields the layer above it from the complexities of providing that service. The Internet follows the 4-layer model shown in the following figure:

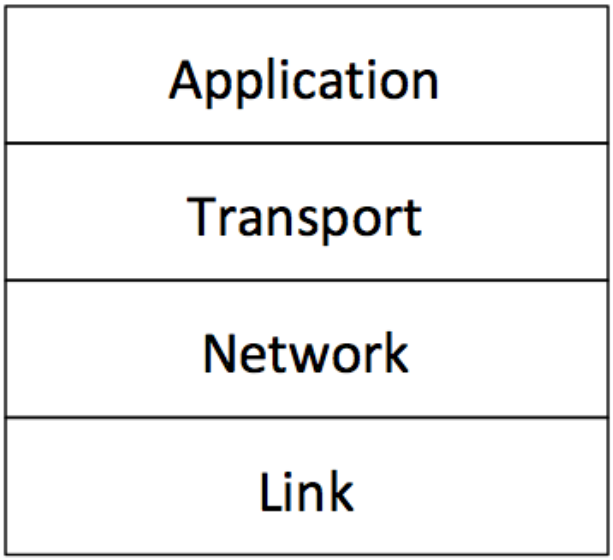



**Figure 7** — Internet 4-layer model

51. The above figure is consistent with the description in ¶ 27 of Dr. Schulzrinne's report except that he includes the physical network that the packets ride on as a layer below the data-link layer.

52. The ***Link Layer***, also known as the ***Data-Link Layer***, is responsible for delivering data in the form of packets over a physical or virtual network link between network devices. A ***physical network link*** is a direct connection between two network devices using a physical medium such as copper wire or fiber optic cable. Another type of network link I will mention later is a virtual network link. A ***virtual network link*** appears to the two nodes communicating over the virtual link to be a physical network

link, but the virtual network link is not restricted to being a physical connection just between the two nodes. Instead it can be implemented as a continuous communication over a network consisting of multiple physical network links. See, for example, the discussion of ***tunnel*** below.

53. One example of a physical network link is Ethernet, the most common type of physical network link used in enterprise data networks. Another example is WiFi, a radio-based equivalent of Ethernet used with portable network devices such as laptops and smartphones. A third example is fiber-optic cable. Short fiber-optic cables are used between buildings in a campus network, longer ones are used between cities and very long fiber-optic cables are used to interconnect continents. A fiber-optic cable contains multiple individual optical fibers. Each individual fiber in a fiber-optic cable can be used as a network link, or individual fibers can be divided up into many different colors of light, known as ***lambdas***. An individual lambda can be used as a network link or multiple lambdas can be combined into a network link.

54. Those network links, such as Ethernet and WiFi, which can interconnect more than two network devices, make use of ***link-layer addresses*** to specify the source and destination of the packets making up communications running over the link-layer network. A link-layer address is a numerical value that uniquely identifies a node on a particular network. The network links that only interconnect two devices, such as lambdas in an optical fiber, generally do not need such addresses since there is only one possible source and one possible destination on any particular link.

55. Sets of interconnected network links are often referred to as ***Local Area Networks*** (***LANs***). If a LAN consists of more than a single network link, the individual

network links in the LAN are interconnected with switches. See below for a description of switches.

56. The ***Network Layer*** is responsible for delivering data between network devices on different LANs. The Internet Protocol defines the network layer in the Internet. See ¶¶ 94-104 for more information about IP. The network layer uses network addresses, rather than link-layer addresses to specify the source and destination of the packets running over a network layer network. A ***network address*** is a numerical value that uniquely identifies a node on a particular network. If the network is the Internet, the network address must be unique across the Internet. The network addresses used in the Internet are ***Internet Protocol (IP) addresses***. See below in ¶¶ 97-98 for a discussion of IP addresses.

57. Devices on the Internet normally have both a link-layer and network address. The link-layer address is used to deliver the packet to the correct device on a particular LAN, and the network address is used to get the packet to the correct LAN. I will describe this further below.

58. The Internet is composed of LANs interconnected with routers. See below in ¶¶ 84, 86 for a description of routers.

59. The ***Transport Layer*** is responsible for managing the flow of packets in each direction that make up a communication between two network devices. As part of this function the transport layer is responsible for splitting the data into packets for transmission and reassembling them into continuous data when they are received. The Transmission Control Protocol (TCP) and the User Datagram Protocol (UDP) are the two

most common transport layer protocols used in the Internet. See the discussions on TCP (¶¶ 110-115), flows (¶ 62) and packets (¶¶ 74-82) below.

60. The ***Application Layer*** is responsible for handling an Internet data flow in a way defined for the specific application the flow is a part of. Applications are the way that people use the Internet. Internet applications most relevant to this case include electronic mail (email) and the world wide web. See below for discussions of these Internet applications.

61. Portions of every packet transferred across the Internet provide support for each of the above layers. See the description of a packet below in ¶¶ 74-82.

***3. Flow***

62. A ***flow*** is a set of packets that are part of a single communication and that are transported from one network node to another network node. While communications are generally bidirectional, flows are unidirectional. The packets that make up a flow are distinguished from other packets when the following five fields in a packet are identical between the packets: the source and destination IP addresses, the protocol field and the source and destination port numbers. This information is often called a ***five tuple*** (or ***5-tuple***). See below for a discussion of packets that includes a discussion of these fields.

#### *4. Transaction*

63. The government's response to the Foreign Intelligence Surveillance Court's Briefing Order of May 9, 2011 defines ***transaction*** as follows:

> *a complement of 'packets' traversing the Internet that together may be understood by a device on the Internet and, where applicable, rendered in an intelligible form to the user of that device.*[5]

64. The government's use of the term "transaction" is not a common way that the term is understood in Internet communications. Merriam-Webster's definition of "transaction" relating to communications is the more common understanding:

> *a communicative action or activity involving two parties or things that reciprocally affect or influence each other.*[6]

65. But I will adopt the government's definition for the term "transaction" for this report where the term is used in regards to upstream collection. In practice, a "transaction", as defined by the government, appears synonymous with a "flow" as I define the term above in ¶ 62.

66. The NSA also talks about ***multi-communication transactions*** (***MCTs***), which contain more than one individual communication, such as more than one email message, not all of which would be proper candidates for collection on their own:

> *NSA Defendants respond that to their understanding (i) the term "single communication transaction," when used in reference to Upstream Internet collection, meant in unclassified terms an Internet transaction that contained only a single, discrete communication, and (ii) the term "multi-*

[5] Appendix C at 1 (FISC Submission (June 1, 2011)).

[6] *Transaction*, Merriam-Webster Dictionary, https://www.merriam-webster.com/dictionary/transaction.

> *communication transaction" meant, in unclassified terms, an Internet transaction that contained multiple discrete communications.*[7]

67. The NSA says that an MCT might consist of, for example, multiple email messages.[8]

68. The NSA says that it is not technically feasible to only collect the individual transactions in an MCT that qualify for collection under the upstream collection program:

> *The NSA's acquisition of MCTs is a function of the collection devices it has designed. Based on government representations, the FISC has stated that the "NSA's upstream Internet collection devices are generally incapable of distinguishing between transactions containing only a single discrete communication to, from, or about a tasked selector and transactions containing multiple discrete communications, not all of which are to, from, or about a tasked selector."*[9]

69. Also see below at ¶¶ 316-320.

***5. Network***

70. A ***network*** consists of a set of computers and the network links and routers and switches that permit the computers to exchange communications. The Internet is a network of networks.

[7] Appendix D at 13 (NSA Response to Plaintiff's Interrogatory No. 8 (Dec. 22, 2017)).

[8] Appendix E at 15-16 n.17 (FISC Opinion (Apr. 26, 2017)).

[9] Appendix F at 45 (Privacy and Civil Liberties Oversight Board, *Report on the Surveillance Program Operated Pursuant to Section 702 of FISA* at 40 (July 2, 2014) ("PCLOB Report")).

### *6. Network Node*

71. A ***network node*** is a computer connected to a network. Network nodes include the ***end-systems*** between which communications are exchanged and the network nodes (such as switches and routers) that forward the packets that make up a communication between the end-systems. Such end-systems include user desktop or laptop computers and smartphones as well as computers that provide services to the users such as web servers—for example www.cnn.com and www.wikipedia.org.

### *7. Circuit*

72. In its response to one of Plaintiff's interrogatories, the NSA described a ***circuit*** as follows:

> *NSA Defendants respond that to their understanding a "circuit," within the context of Internet communications, traditionally consists of two stations, each capable of transmitting and receiving analog or digital information, and a medium of signal transmission connecting the two stations. The medium of signal transmission can be electrical wire or cable, optical fiber, electromagnetic fields (e.g., radio transmission), or light. Individual circuits may be subdivided further to create multiple "virtual circuits" through application of various technologies including but not limited to multiplexing techniques.*[10]

73. This description is consistent with the definition for "network link" I provided above in ¶¶ 52-55, with the addition of the nodes at each end of the link. I will adopt the government's definition of circuit for the purpose of this report.

[10] Appendix D at 6 (NSA Response to Plaintiff's Interrogatory No. 2 (Dec. 22, 2017)).

***8. Packet***

74. A ***packet*** is a chunk of a communication. Packets in the Internet can vary in size and are autonomous, meaning that they can be processed independently by devices within the network (explained below). An example Internet packet is shown in the following figure. As explained below, each layer in this figure depicts the corresponding layer within the four-layer model of the Internet, described above in ¶¶ 50-61:

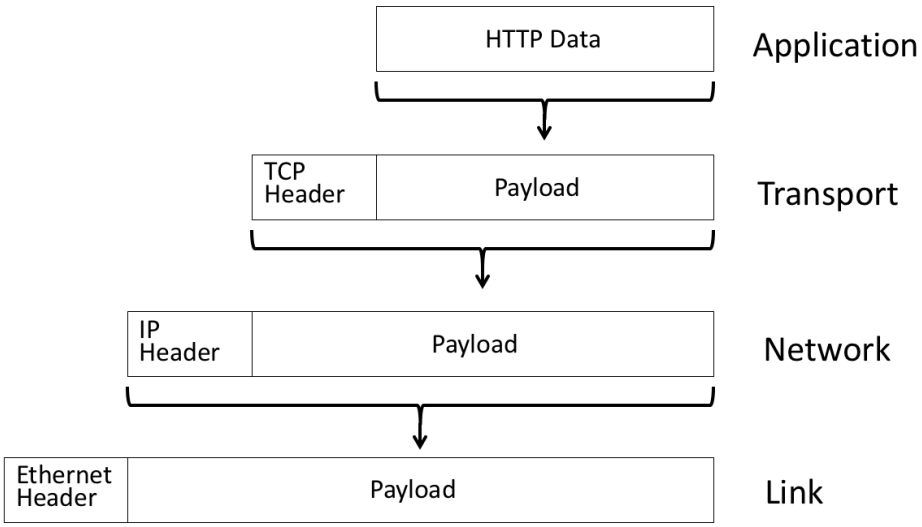



**Figure 8** — Packet format showing web data over Ethernet

75. Figure 7 shows an Internet packet as I described above in the definition of layer. In this case, the figure shows an Ethernet packet that is transporting world wide web (HTTP) data.

76. The lowest pair of boxes represents an Ethernet packet (also known as a ***frame***). The left box is the Ethernet header and the right box is the Ethernet payload, which is the entire IP packet. An Ethernet header is shown in the following figure:




**Figure 9** — Ethernet header

77. The information is transmitted onto the Ethernet starting with the left edge of the figure. The first information transmitted is the link-layer destination address, followed by the link-layer source address, then finally the type field. The link-layer destination address specifies the specific network device on the LAN to which this packet is to be delivered. The link-layer source address contains the link-layer address of the network device that is sending the packet. Finally, the value of x800 in the type field identifies the payload in this Ethernet packet as an IP packet.

78. The IP part of the packet is shown in the two connected boxes above the Ethernet packet in Figure 7. The format of an Internet Protocol (IP) version 4 packet is shown in the following figure (which is the same as Figure 5, above):

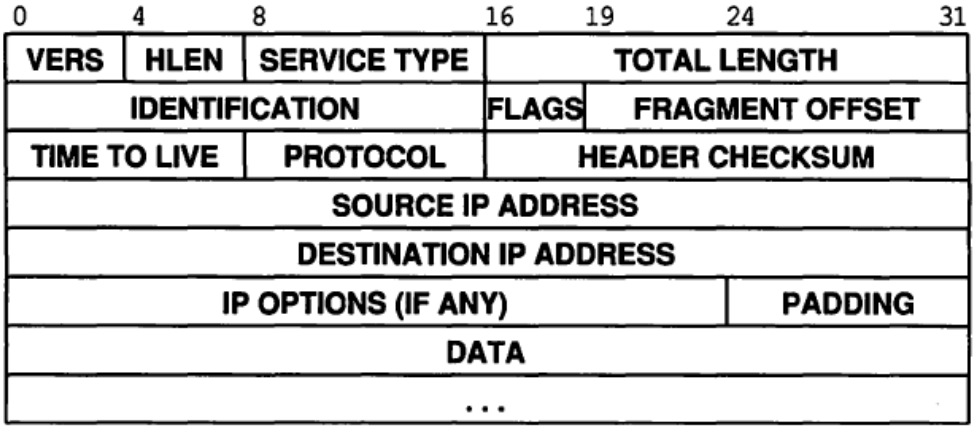



**Figure 10** — IP packet format[11]

79. Information is sent on the Ethernet starting in the upper left box of the figure and continuing, row by row, to the lower right. Figure 10 shows the source and destination IP addresses. These are the addresses described above in the definition of layer as network addresses. See below for a fuller description of IP addressing. In this example case, the protocol field will be set to a value of 6 to indicate that the payload of the IP packet (labeled as "data" in the figure) is a TCP packet.

[11] Comer, *supra* note 3.

80. The TCP part of the packet is shown in the two connected boxes above the IP part of the packet in Figure 7. The format of a TCP packet is shown in the following figure:

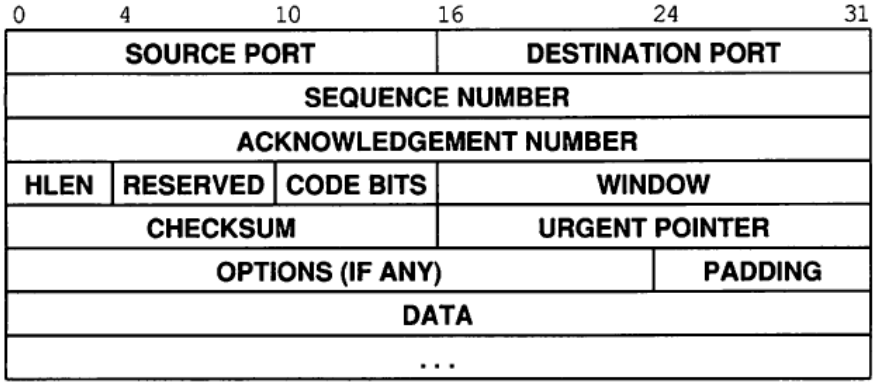



**Figure 11** — TCP packet format[12]

81. Information in a TCP packet is sent following the same pattern as with the IP packet. The only field important for this section of this report is the Destination Port field. Since this example packet is carrying world wide web data, the destination port field will be set to a value of 80 or 443. (See below in ¶¶ 110-115 for a fuller description of TCP.) The value 80 in the destination port field indicates that the payload portion (labeled "data" in the figure) is HTTP (world wide web) information and the value 443 indicates that the payload portion (labeled "data" in the figure) is HTTPS, the encrypted version of HTTP.

82. The HTTP part of the packet is shown as the box above the TCP part of the packet in Figure 8. See ¶¶ 117-124 for a fuller description of HTTP.

[12] Comer, *supra* note 3.

***9. Switch***

83. A ***switch*** is a network node that is connected to two or more network links. A switch receives packets on these network links and forwards each of the packets it receives onto one or more of the other network links based on the destination link-layer address in the link layer of a packet received by the switch. Thus switches are used to forward packets *within* a LAN. Typically there would be an Ethernet switch in some central location on a floor of an office building. Ethernet links would then connect individual desktop computers to the switch.

***10. Router***

84. A ***router*** is a network node that, like a switch, is connected to two or more network links. A router receives packets on these network links and forwards each of the packets it receives onto one or more of the other network links based on the destination Internet address in the network layer of the packet received by the router. Thus a router is used to forward packets *between* LANs.

***11. Mirroring***

85. Some switches and some routers have the ability to make copies of some or all of the traffic sent or received on one network link and send that traffic out of a second network link. This is the copying function Dr. Schulzrinne describes in ¶ 58 of his report.

### *12. Routing*

86. ***Routing*** is the process by which a router in a network decides onto which network link the router should forward a packet it has received in order to get the packet closer to the packet's destination, where the destination is represented by the destination Internet address in the received packet. Routers decide where to forward the packets they receive in one of three ways:

a. Routers can be manually configured to determine a forwarding decision.

b. Routers can exchange information with other routers to build a dynamic database of information on which to make forwarding decisions.

c. Routers can be configured to use a combination of the two.

87. See ¶¶ 175-199 for additional discussion on routing in the Internet.

### *13. Internet Protocol*

88. The ***Internet Protocol*** is defined by a set of standards that specify the format of packets in the Internet and how the packets are to be generated by the sender of the packet and processed by the receiver of the packet to enable the transfer of communications between nodes in the Internet. See below for a fuller description of the Internet Protocol.

### *14. Internet Service Provider (ISP)*

89. An ***Internet service provider*** (***ISP***) is a company that provides connectivity between a set of customers and the rest of the Internet. The customers could be individuals using smartphones or computers in their own homes or in enterprises that run their own Internet Protocol-compatible enterprise networks. ISPs range from ones

that service a small part of a small town to ISPs that service customers around the globe. See below for a fuller description of ISPs.

***15. Proxy***

90. A ***proxy*** is a network node that serves as a forwarding agent for communications between other Internet nodes. In most cases a proxy rewrites the IP packet header information in the communication such that the proxy appears to be the origin or destination of the communication rather than the network node the proxy is serving.

***16. Tunnel***

91. A ***tunnel*** is a type of virtual network link used to establish what appears to be a direct network link between network nodes by transporting packets flowing between the two nodes within other packets. The transporting packets may traverse multiple network nodes, both switches and routers, on a path between the two tunnel nodes. In many cases the packets being transported over a tunnel are encrypted. An example of an encrypted tunnel is a ***virtual private network*** (***VPN***) that a traveler uses to connect his or her laptop computer in a coffee shop back to his or her employer's enterprise network. Such VPNs are used to protect communications between the laptop and an enterprise network from eavesdropping and to protect communications between enterprise networks.

***17. Metadata***

92. ***Metadata*** is information about a communication that is not within the communication itself. Examples of metadata include the source and destination IP addresses for a communication, and the time the communication starts and ends.

### C. The Key Internet Protocols

93. In the following section, I will describe the key protocols that are used in the Internet today (i.e., the Internet Protocol Suite) and several of the most common ***application protocols*** used on the Internet (i.e., HTTP/HTTPS for web access and IMAP/SMTP for email).

#### *1. The Internet Protocol Suite*

94. Kahn and Cerf defined more than just the format of IP packets and how IP packets were created and processed; they defined a suite of protocols. The suite includes the Internet Protocol itself as well as a few "higher-level" protocols that use IP packets for transport and that define ways to support specific types of communication between network nodes. I will describe the Internet Protocol more fully and then mention two of those higher-level protocols below.

##### a. The Internet Protocol (IP)

95. As mentioned above, there is a defined format for IPv4 packets, which is shown in the following figure which I repeat from above to make it convenient for the reader:

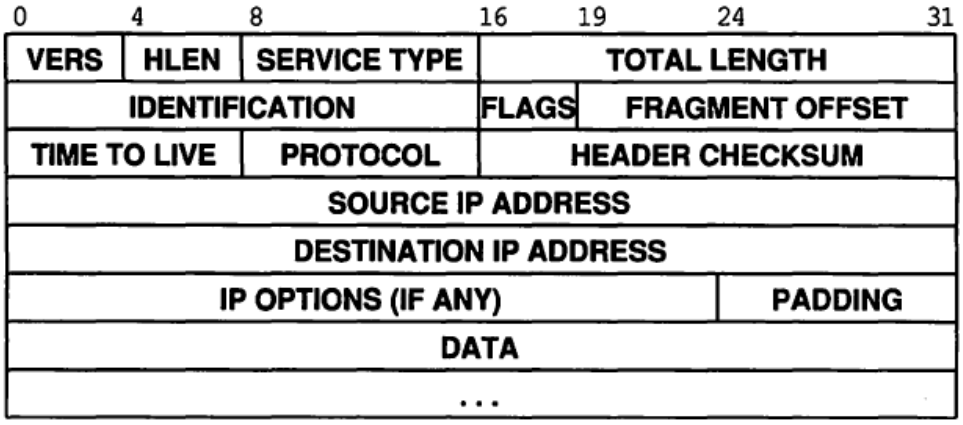


**Figure 12** — An IP packet[13]

96. The first part of an IP packet is known as the ***IP header***. The IP header comprises the fields shown in Figure 12 through the optional "padding" field (that is, the first six rows). In addition to the source and destination IP address fields in the IP header that I have already described, there is one other field in the IP header that is relevant to this case. The "protocol" field is used to indicate what higher-level protocol is using the IP packet for transport. When an IP packet is created and sent by a network node, for example by a user's personal computer, the node will put its own IP address into the Source IP Address field and the IP address of the node that it wants to send the packet to into the Destination IP Address field. The computer will also put a value in the protocol field so that the receiving node will know what to do with the packet when it is received.

i. IP addresses

97. An ***IP address*** is a number that is used to identify a particular network device on a network that is using the Internet Protocol for communication. IPv4 addresses are 32-bits long and can identify about 4 billion individual network devices. IPv6 addresses are 128-bits long and can identify trillions of trillions of individual network devices. I will focus on IPv4 in this report, but when I use the term "IP address" it should be taken to mean the type of IP address used in the version of IP in use in the particular situation.

[13] Comer, *supra* note 3.

98. An IPv4 address is represented as a set of 4 numbers separated by periods. For example, the IP address for the web server I run in my house is 173.166.5.74 and, as of this writing, one of the IP addresses of the University of Oxford's website, www.ox.ac.uk, was 129.67.242.155.

ii. Viewing IP header information

99. The IP header information is visible throughout the path a packet takes through the Internet. Except in the cases where the IP addresses are modified in transit, (I will mention some cases of this below), the actual source and destination of each packet in the Internet can be determined by just looking into its IP header.

100. The IP header information must be unencrypted even when the information being transported is encrypted. To transport an email message, for example, the IP header information for the packets that make up the email must be unencrypted so that the routers forwarding the packets know where to send them and so that the receiving node knows what to do with them.

101. Information beyond the IP addresses and protocol can be observed in IP packets by looking further into the packet to get the port numbers and application-specific information. The function of looking into packets to better understand the application-level communications they transport is often referred to as "deep packet inspection (DPI)." I will discuss DPI further below.

iii. Sizes of IP packets

102. IP packets in the Internet are variable in length. They range from a minimum size of 68 bytes long to 1,500 bytes long. The 1,500 byte limit derives from the maximum packet size that is supported on Ethernet, the most common type of local

physical network. A 1,500 byte packet is big enough to transport the body of an email message of up to a thousand characters—about 200 four-letter words (including spaces between each word).

iv. Multiple packets in a communication

103. A particular communication will be broken up into multiple packets by the sending node if the communication cannot fit in a single large (1,500 byte) packet. The packets are reassembled into the communication by the destination node in order to recover the originally transmitted message.

104. The reassembly must be done by the destination node because the Internet does not guarantee that all of the packets that make up a particular communication will be present at any other place along the path from sender to receiver. Two features of the Internet cause this to be the case:

a. The paths that packets take through the Internet can change at any time, even between successive packets in a single communication.

b. The paths packets take are asymmetric, in that packets in a two-way communication traveling in one direction will generally not follow the same path as packets traveling in the opposite direction.

**b. Transport Protocols**

105. As described above in ¶¶ 50-61, ***transport protocols*** are used to break communications into packets and to provide the desired level of reliability. The two transport protocols I will describe here are the User Datagram Protocol and the Transmission Control Protocol. These are the dominant transport protocols currently in general use on the Internet.

i. The User Datagram Protocol (UDP)

106. The ***User Datagram Protocol*** (***UDP***) provides a way to send packets from one network node to another network node over IP packets. Many applications use UDP for transport, including certain voice and video streaming applications, tunneling protocols, and domain name lookups (see ¶ 184 for a description of how to do domain name lookups).

107. UDP information is carried in the "data" portion of those IP packets that make up a communication using UDP as its transport. UDP has its own header as shown in the figure below:

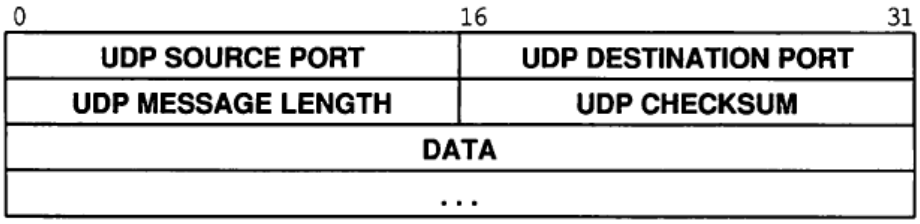



**Figure 13** — The UDP header[14]

108. The destination UDP port field is used to specify the application that is running over UDP. Hundreds of applications have been defined to date, many in "open" standards but quite a few in non-public and proprietary ones.

109. UDP port numbers can range from 1 to 65,535. UDP port numbers 1 to 49,151 are "registered" for use by particular applications. Port numbers between 49,152 and 65,535 are "unassigned" and open for use by any application, although the node that receives a UDP packet using an unassigned port number must have been preconfigured to know what to do with a packet with that unassigned destination port number. Note that

[14] Comer, *supra* note 3.

port number assignments are, in a way, advisory. As long as the two ends of a communication agree on which port numbers to use, any port numbers will work, even port numbers that have already been assigned to specific applications. Thus, by changing the port numbers in use, someone can change the apparent application being used. For example, quite a few applications use ports 80 or 443, the ports nominally assigned to the world wide web, because these ports are often passed by firewalls that would block unassigned ports.

ii. The Transmission Control Protocol (TCP)

110. Whereas UDP is used to just deliver packets from one network node to another without worrying about the rate of transmission or even if the packet in fact makes it to the destination network node, ***Transmission Control Protocol*** (***TCP***) is used to provide a *reliable data stream* between network nodes. TCP is used by most major Internet applications including email, the world wide web, file transfer and the control channel of Internet calling protocols such as Skype. TCP has its own header that is present in all packets in a communication making use of TCP:

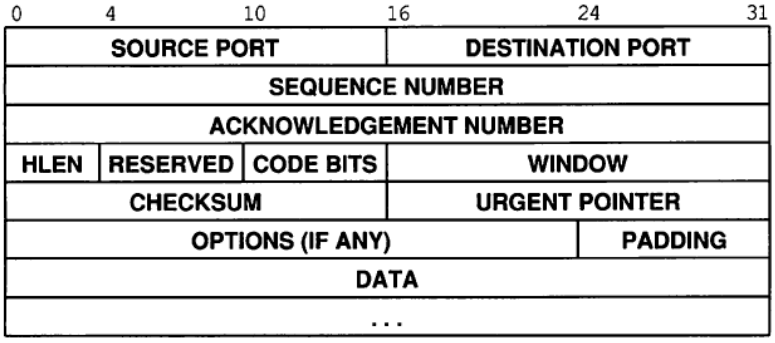



**Figure 14** — The TCP header[15]

[15] Comer, *supra* note 3.

111. Two network nodes can use TCP to create and maintain a two-way communications session, to control the rate of packet transmission as appropriate, and to ensure that all of the information in the session will be reliably delivered.

112. TCP can be used to transport a discrete piece of information such as an email message. It can also be used to support continuous streams of information such as a telephone call, although UDP can also be used to transport information streams, including phone calls.

113. Ports in the TCP header are assigned and used in the same way as ports are used in UDP, except that source ports are required. The set of information in the (1) Source and (2) Destination IP addresses fields and the (3) protocol field in the IP header, along with the information in the (4) source and (5) destination port fields in the TCP header, uniquely identifies packets that are part of a particular TCP communication between two network nodes. As explained in ¶ 62, this information is often called a five tuple (or 5-tuple).

114. The sequence number field in the TCP header is used to ensure that all of the packets comprising a communication have been received and that they are in the correct order. This is important because IP networks do not guarantee that packets will not be lost, duplicated or reordered during their travel though the Internet.

115. The Internet protocol suite includes the Internet Protocol itself plus the transport protocols TCP and UDP as well as other signaling protocols and is frequently referred to as "TCP/IP."

### *2. Application Protocols*

116. UDP and TCP are used to transport packets that implement Internet applications. I will discuss a few of the hundreds of applications that have been defined for the Internet.

#### a. The Hypertext Transfer Protocol (HTTP)

117. The ***Hypertext Transfer Protocol*** (***HTTP***) is used to transport web page content between web servers and web browser software on user computers.

##### i. HTTP commands

118. HTTP consists of a number of plain text commands sent by a web browser to a web server. The basic HTTP commands are shown in the following figure:

| *Command* | *Description* |
|---|---|
| GET | Return the contents of the indicated document. |
| HEAD | Return the header information for the indicated document. |
| POST | Treat the document as a script and send some data to it. |
| PUT | Replace the contents of the document with some data. |
| DELETE | Delete the indicated document. |

**Figure 15** — HTTP commands[16]

119. The HTTP GET command is used to request that the HTTP server return a file to the user's web browser. The GET command includes the name of the requested file. The POST command is used to upload a file to a web server.

##### ii. Encrypted HTTP (HTTPS)

120. An encrypted version of HTTP, referred to as ***HTTPS*** (for "HTTP Secure") was introduced in 1994 by Netscape Communications to support electronic commerce over the Internet. The entire HTTP application layer communication is

[16] Lincoln D. Stein, *How to Set Up and Maintain a World Wide Web Site: The Guide for Information Providers* 49 (1995).

encrypted when using HTTPS. The IP packet and TCP header that HTTPS rides on top of are not encrypted, so an observer can determine that an HTTPS session is running between two nodes identified by the IP addresses in the IP header.

121. It is worth noting that not all encryption used on the Internet is "unbreakable." *See* Schulzrinne Decl. ¶ 42. When properly implemented, modern public standards-based encryption itself is generally considered to be unbreakable. But encryption standards are not enough. The software implementing the encryption standard has to be well designed and bug-free, the systems that make use of the encryption must also be well designed and well implemented, and these systems must be properly and carefully operated for the communications to actually be protected.

122. Not all implementations of HTTPS in use on the Internet today are "unbreakable", and the computers making use of HTTPS are all too frequently compromised because of software bugs or user errors. Once a computer is compromised, it is generally easy to compromise any communications as they are being sent or received by that computer. In addition, some developers decide to create their own encryption protocols and algorithms and most of them turn out to be far from unbreakable.[17] In the cases where the NSA determines that the type of encryption protocol or algorithm being used is weak, it would make sense for the NSA to collect encrypted communications from targeted individuals knowing that, with enough effort, for example, with large amounts of computing power the encryption could be broken. The NSA could also be collecting encrypted communications to subject them to quantum cryptanalysis in the

[17] Joseph Cox, *Why You Don't Roll Your Own Crypto*, VICE: Motherboard (Dec. 10, 2015), https://motherboard.vice.com/en_us/article/wnx8nq/why-you-dont-roll-your-own-crypto.

future. Quantum cryptanalysis, which relies on quantum computers, may make it significantly easier to break certain types of encryption in wide use today. It is not publicly known whether the NSA or any other intelligence agency currently has the capacity to conduct quantum cryptanalysis, but encryption standards bodies have been preparing for a number of years for the possibility that intelligence agencies or malicious actors will. The above factors may help to explain the permissive rules (as discussed in ¶¶ 325-327) for the NSA's collection of encrypted communications under Section 702.

iii. HTTPS Handshake

123. Not all of the HTTP information is hidden when using HTTPS. A single physical web server can be used to support many websites. The web server that I run in my house, for example, supports www.sobco.com, www.sobco.org, www.scottbradner.com, and www.kaybradner.com. Because a single web server may be supporting multiple different websites, a web browser must send the domain name of the website to the web server during the setup phase of an HTTPS session so that the web server knows which website the user wants to access and so that the proper security association can be setup. Since the security association has not yet been set up, the domain name must be sent unencrypted. Thus, HTTPS does not protect the confidentiality of the domain name of the website that is being accessed. For example, an observer would be able to determine that a user had requested a web page from https://en.wikipedia.org, but they would not be able to determine from the HTTPS request that the user had requested the specific web page https://en.wikipedia.org/wiki/Addiction.

iv. IP addresses in HTTP packets

124. There are some cases where the IP addresses in HTTP packets do not accurately identify the original sender of a HTTP packet or its ultimate destination. For example, HTTP proxies are sometimes used in enterprise networks, including hotels, and in some Internet service providers to improve the performance of user's web browsers and to control access to improper websites. HTTP packets sent from all web browsers used by everyone behind an HTTP proxy will have the IP address of the HTTP proxy as the IP Source Address in the header. Likewise, the Destination IP Address in all HTTP packets destined to web browsers that are behind an HTTP proxy will have the IP address of the proxy as their Destination IP Address. There are also cases where there are no proxies or NATs (see below in ¶¶ 173-174) where the IP addresses in the packets identify the sender and receiver of a packet.

**b. Email**

125. As a formal matter, ***electronic mail*** or ***email*** refers to "*a system for sending messages from one individual to another via telecommunications links between computers or terminals using dedicated software*".[18] Email is the third oldest Internet application, behind remote access and file transfer.

[18] *Email*, Merriam-Webster Dictionary, https://www.merriam-webster.com/dictionary/email.

i. Email Header Information

126. Individual email messages have a format defined in specifications from the IETF[19]. The start of an email message consists of a series of plain text "headers" that include the names and email addresses of the sender and intended receiver(s) of the message, the date the message was sent, a subject for the message, some information about the path the message took through the Internet which generally includes the IP address of the email server that sent the message, and some information about the format of the body of the message, i.e. the part of the message following the header lines. Very often, an email message will not fit in a single packet. In such cases the header lines will start in the first packet of the communication, but sometimes the header lines will extend into the second packet.

127. An example of some of the entries in an email header are shown in the following figure:

```
From: "Scott O. Bradner" <sob@sobco.com>
Content-Type: text/plain;
        charset=us-ascii
Content-Transfer-Encoding: 7bit
Mime-Version: 1.0 (Mac OS X Mail 11.3 \(3445.6.18\))
Subject: need to reschedule Taveras appt
X-Universally-Unique-Identifier: 78BC10A7-0A77-4C96-96D0-9C394F696E8E
Message-Id: <40D15B68-D405-46CC-A511-24F809988261@sobco.com>
Date: Thu, 5 Apr 2018 09:52:37 -0400
To: "Cheryl F. Chapman" <cfc@sobco.com>
```

**Figure 16** — Sample email header

128. The above is the header portion of an email message from me to my wife. The "From:" header line provides my name and email address as the sender of the email

[19] See, e.g., *Internet Message Format*, Qualcomm Inc., Network Working Group (October 2008), https://www.ietf.org/rfc/rfc5322.txt.

message. The "To:" line shows my wife's name and email address as the destination of the message. The "Subject:" line shows what I said was the subject of the message. The "Date:" line shows the time I sent the message. Finally, the "Message-Id:" line is a unique identifier for this particular message. The body of the message that follows the header lines could be plain text, one or more photos, one or more pieces of video or music, a spreadsheet, a Microsoft Word document, a pdf, or any one of dozens of other things. In addition, the body of an email message may or may not be encrypted.

ii. Email Servers

129. As a general rule, email messages do not go directly from a sender to a receiver. Instead, there could be an email server at the sending end, and there is almost always an email server on the receiving end. Email servers maintain databases of sent and received email messages for each of their users.

130. Email users access their email servers by using a web browser or by using a piece of software called a "mail user agent" on their own computer. With the web browser or mail user agent, an email user can create and send email messages and also read any email he or she might have received.

131. Many large commercial email services, such as Hotmail and Gmail, are accessed via web browsers. Some large commercial email services, for example Microsoft Exchange, are accessible via web browsers but are also accessible via their own special mail user agents. In addition, some computers come with their own generalized mail user agents that can connect to multiple commercial email services. One example of the latter is the Mail program that comes with Apple computers. This is the mail user agent that I use. I use the Apple Mail application to connect to Harvard's Microsoft Exchange server, Google Gmail and to the email server that I run in my house.

132. Mail user agents generally download all new email messages to the user's computer whenever the mail user agent is started. Thus, when an email user turns on their laptop after a few days "off line" a burst of email messages can be transferred to the laptop. Such bursts will often be done over a single communications session between the email server and the mail user agent, resulting in multiple individual email messages in the same communication. Some web mail implementations do the same type of burst fetch of unread email. This behavior may be an example of what the NSA has called a multi-communication transaction (MCT) since the NSA says that an MCT can consist of multiple email messages.[20] (See above at ¶¶ 66-68 and below at ¶¶ 316-320.)

133. There are a number of IETF protocols that define the communications between email servers and between email servers and mail user agents. In addition, there are some proprietary protocols. I will discuss the two most common, standards-based protocols:

a. ***Simple Mail Transfer Protocol*** (***SMTP***): used between email servers and between email servers and some mail user agents

b. ***Internet Message Access Protocol*** (***IMAP***): used between most mail user agents and email servers.

iii. Simple Mail Transfer Protocol (SMTP)

134. The Simple Mail Transfer Protocol (SMTP) is used to transport email messages between email servers and, less frequently, between mail user agents and email servers. The SMTP protocol defines a ***handshake*** that is used to start up a session to transfer an email message. A sample of an SMTP handshake used when a user is sending

[20] Appendix E at 15-16 n.17 (FISC Opinion (Apr. 26, 2017)).

an email message is shown in the following figure, where "S" identifies text sent by the email server and "C" identifies text sent by the email client:

```
S: 220 Beta.GOV Simple Mail Transfer Service Ready
C: HELO Alpha.EDU
S: 250 Beta.GOV

C: MAIL FROM:<Smith@Alpha.EDU>
S: 250 OK

C: RCPT TO:<Jones@Beta.GOV>
S: 250 OK

C: RCPT TO:<Green@Beta.GOV>
S: 550 No such user here

C: RCPT TO:<Brown@Beta.GOV>
S: 250 OK

C: DATA
S: 354 Start mail input; end with <CR><LF>.<CR><LF>
C: ...sends body of mail message...
C: ...continues for as many lines as message contains
C: <CR><LF>.<CR><LF>
S: 250 OK

C: QUIT
S: 221 Beta.GOV Service closing transmission channel
```

**Figure 17** — SMTP startup handshake[21]

135. The SMTP handshake includes the message sender's email address (Smith@Alpha.EDU) and the email address of the intended recipients of the message (Jones@Beta.GOV, Green@Beta.GOV and Brown@Beta.GOV). Even if parts of the body of an email message are encrypted, the SMTP handshake is not, although the entire SMTP exchange could take place within an encrypted connection, in which case the

[21] Comer, *supra* note 3.

SMTP handshake would be encrypted, and any encrypted parts of the email message would be doubly encrypted.

(1) SMTP Metadata

136. The sender's and receiver's email addresses as well as the date and time that the mail was sent and the IP addresses of email servers would all be considered email metadata. This metadata is included in the SMTP startup handshake as well as in the email headers.

(2) IP addresses in email packets

137. The IP addresses in the packets exchanged between email servers identify the email servers but often have no relationship to the actual sender or receiver of an email message. Some mail user agents are configured to use SMTP to send email messages directly to the email server associated with the intended recipient. In such cases the source IP addresses in packets sent to the email server will identify the computer that is running the mail user agent.

iv. Internet Message Access Protocol (IMAP)

138. IMAP defines the formats and meanings of the messages exchanged between a mail user agent and an email server. In general, these messages are used to maintain a copy of the email user's portion of the email server on the user's own computer.

139. As mentioned above in ¶ 132, when a mail user agent connects to an email server using IMAP, all new messages will be downloaded to the user's computer in a batch.

140. In my own case, the mail user agents on my laptops, desktops and smartphone are configured to use IMAP to connect to the email server I run in my house

when sending email. The IP addresses in the packets my email server sends will be the IP address of that server, no matter where in the world I might be. Similarly, packets comprising an email message sent by a Gmail user will include the IP address of the Gmail server in their source address field no matter where the Gmail user is actually located.

**c. Telephone Calls**

141. While it's not part of this case, a number of NSA documents say that the NSA collects telephone calls and that telephone numbers are one type of selector that is used to target Internet transactions, under the upstream collection program. Since almost all international telephone calls are currently transported over the Internet using the IETF-developed Session Initiation Protocol (SIP), it is easy to include them in the upstream program. SIP has HTTP-like headers that are used to specify the source and destination telephone numbers and the IP addresses between which the audio portion of the phone call will flow.

***3. Plain Text in Application Protocol Headers***

142. Many Internet applications, including the applications mentioned above, include "plain text" (i.e., not encrypted and not otherwise encoded) fields in their headers. Such fields can be searched for specific strings such as a name or email address or other string that might indicate that a packet is part of a communication that is of interest, even if portions of the underlying communication are encrypted. These text fields will sometimes be entirely in the first packet of a flow of packets that makes up a communication but often do extend into successive packets.

### *4. Number of Packets in a Communication*

143. As described above in ¶ 49, a particular communication between nodes over the Internet is broken up into packets for transmission. The number of packets in any one communication varies greatly. The sample email message between me and my wife shown above was short enough to be contained within a single packet (although the SMTP handshake that would've preceded the email when sent between email servers would have required an exchange of multiple separate packets), but I sent an email message to a colleague recently that contained two image files. The message was 2.7 million bytes (MB) long so it took at least 1,860 packets to transport that message. I frequently send email messages that are 10 MB or more. It takes thousands of packets to transport each of those messages.

144. I ran a command on the router that interfaces the network in my house to my Internet service provider that asked for statistics on the number of packets in a flow. The results of that command are shown in the following figure:

| Protocol | Total Flows | Flows /Sec | Packets /Flow | Bytes /Pkt | Packets /Sec | Active(Sec) /Flow | Idle(Sec) /Flow |
|---|---|---|---|---|---|---|---|
| TCP-Telnet | 14222087 | 3.3 | 1 | 43 | 4.1 | 1.0 | 15.5 |
| TCP-FTP | 113213 | 0.0 | 2 | 60 | 0.0 | 1.9 | 14.4 |
| TCP-FTPD | 13567 | 0.0 | 1 | 40 | 0.0 | 0.0 | 15.4 |
| TCP-WWW | 2967948 | 0.6 | 81 | 846 | 56.1 | 24.5 | 8.2 |
| TCP-SMTP | 3454819 | 0.8 | 7 | 261 | 5.9 | 2.0 | 5.7 |
| TCP-X | 328461 | 0.0 | 1 | 40 | 0.0 | 0.0 | 15.4 |
| TCP-BGP | 14779 | 0.0 | 1 | 40 | 0.0 | 0.0 | 15.4 |
| TCP-NNTP | 8667 | 0.0 | 1 | 40 | 0.0 | 0.0 | 15.4 |
| TCP-Frag | 174 | 0.0 | 1 | 460 | 0.0 | 0.1 | 15.4 |
| TCP-other | 50802851 | 11.8 | 20 | 432 | 245.9 | 4.5 | 9.1 |
| UDP-DNS | 324539 | 0.0 | 2 | 63 | 0.1 | 0.7 | 15.4 |
| UDP-NTP | 597821 | 0.1 | 1 | 76 | 0.1 | 0.0 | 15.4 |
| UDP-TFTP | 28463 | 0.0 | 1 | 42 | 0.0 | 0.0 | 15.4 |
| UDP-Frag | 36396 | 0.0 | 1 | 540 | 0.0 | 0.0 | 15.5 |
| UDP-other | 68219772 | 15.8 | 1 | 224 | 17.6 | 0.3 | 15.4 |
| ICMP | 1701074 | 0.3 | 5 | 63 | 2.2 | 9.4 | 15.4 |
| IPv6INIP | 14 | 0.0 | 1 | 80 | 0.0 | 0.0 | 15.4 |
| GRE | 57855 | 0.0 | 2111 | 209 | 28.4 | 29.1 | 15.3 |
| IP-other | 451 | 0.0 | 1 | 53 | 0.0 | 0.0 | 15.5 |
| Total: | 142892951 | 33.2 | 10 | 459 | 360.9 | 2.5 | 12.8 |

**Figure 18** — Average flow lengths in my home router

145. The printout shows the statistics since the router was last rebooted a few years ago. The results show that the average length of the email messages that I sent or received over the past few years was 7 packets (TCP-SMTP) and the average length of my web sessions over the same time period was 81 packets (TCP-WWW).

146. I do not think that these statistics are necessarily representative of general Internet traffic, but they do show that much Internet traffic consists of communications comprising multiple packets.

### D. Other Features of the Internet and its Architecture Relevant to this Case

147. In the following section, I will describe other features of the Internet and its architecture that are relevant to this case, including the general structure of the Internet, the role of Internet Service Providers, the way in which networks comprising the Internet connect to one another, the meaning of the "Internet backbone," the undersea fiber optic cables that connect the U.S. to the rest of the world, and the way that packets are routed on the Internet.

#### *1. Internet Architecture*

148. There is no fixed architecture to the Internet. Each customer and service provider is free to design and operate their network or networks in any way they want as long as they are able to transport IP packets along a path from the packet source to the packet destination. Each network operator is also free to interconnect their networks with networks run by other network operators in any way that the two operators agree to, as long as they can properly transport IP packets between the networks.

149. The result is that the Internet structure appears almost random as shown in the following figure from the Opte Internet mapping project:

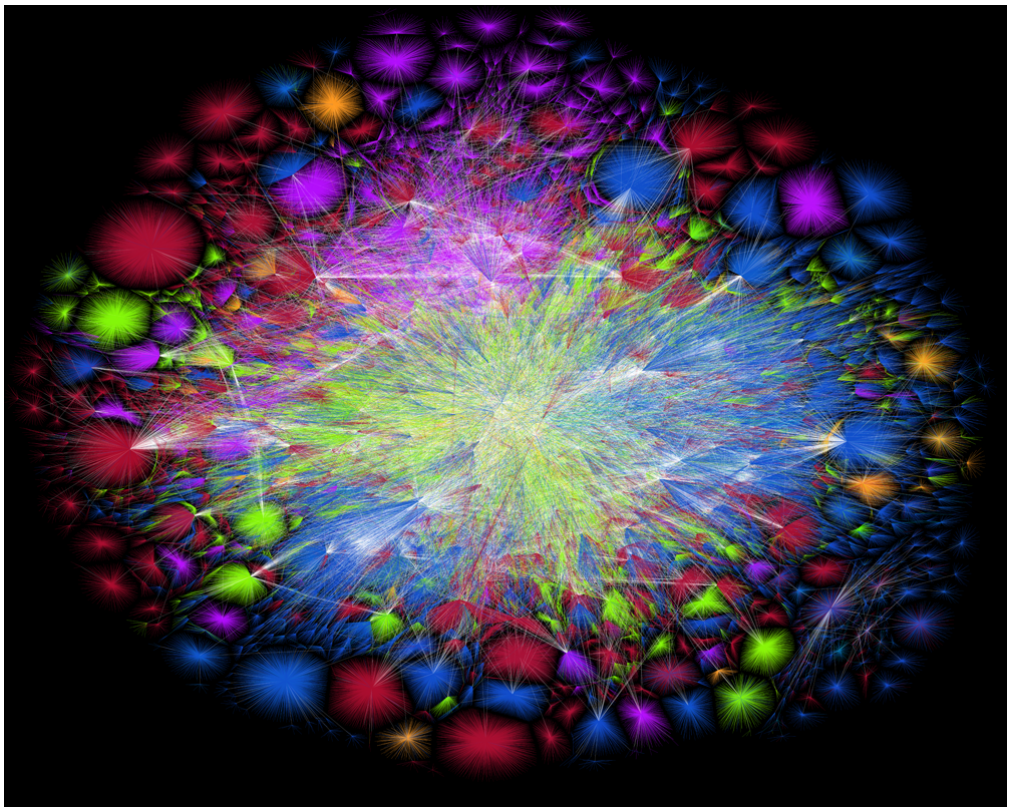


**Figure 19** — The Internet[22]

### *2. Internet Backbone*

150. One of the terms used in this case is "the Internet backbone." Once upon a time, between 1983 and about 1990, it was easy to define the Internet backbone in the U.S. In 1983 it was the ARPANET. The ARPANET was the only nation-wide network that was being used to interconnect other networks, so it was "the Internet backbone." By 1990, the ARPANET had been joined by the NSFNet and the first few commercial ISPs. But there were very few of these ISPs that were nation-wide, so it was reasonable

[22] The Internet 2015, The Opte Project (July 11, 2015), http://www.opte.org/the-internet.

to say that the Internet backbone consisted of the long distance connections in the ARPANET, NSFNet and those ISPs that provided nation-wide service.

151. Since then the growth of ISPs of all sizes and the end of the ARPANET and NSFNet have painted an increasingly more complex picture, to the point that today it is not possible to isolate a single backbone for the U.S. Internet, much less the global Internet. The term "Internet backbone" is one that shows up in the popular press from time to time, but my experience is that experts in the field tend not to use that term. Occasionally, I have seen reference to the "Internet backbones" (plural), referring to the largest ISPs, but more often I've seen references to "ISP backbones", not to an Internet backbone. In an ISP, the backbone is the set of high-speed lines that interconnect routers in different parts of the ISP's geographic footprint.

152. The NSA has provided one interrogatory response and two admissions in regard to their use of the term "Internet backbone":

a. *NSA Defendants respond that to their understanding the Internet backbone is no longer well defined due to the growth of direct peering arrangements, but may be understood as the principal high-speed, ultra-high bandwidth data-transmission lines between the large, strategically interconnected computer networks and core routers that exchange Internet traffic domestically with smaller regional networks, and internationally via terrestrial or undersea circuits.*[23]

b. *NSA Defendants respond that yes, the Internet backbone includes but is not limited to international submarine telecommunications cables that carry Internet communications.*[24]

[23] Appendix D at 18 (NSA Response to Plaintiff's Interrogatory No. 12 (Dec. 22, 2017)).

[24] Appendix H at 6 (NSA Response to Plaintiff's Request for Admission No. 3 (Jan. 8, 2018)).

c. *NSA Defendants respond that yes, the Internet backbone includes but is not limited to high-capacity terrestrial telecommunications cables that carry Internet communications within the United States.*[25]

153. In summary, the government's definition of the Internet backbone includes (1) the high-speed circuits (network links) and routers that are used to interconnect ISPs, (2) the circuits in the undersea cables that connect the U.S. with other countries, and (3) the high speed terrestrial network links (circuits) within the U.S and between the U.S. and other countries. The latter two may be network links between ISPs or within an ISP. I will adopt the government's definition for this report.

154. As stated above in ¶ 70, the Internet is a network of networks. Some of these networks are very small, like the one in my house, and some are very large such as AT&T's IP network, which spans the globe. These networks include customer networks and service provider networks. Each of these millions of networks is under its own management—there is no central manager for the Internet.

### *3. Internet Service Providers (ISPs)*

155. The purpose of service provider networks, known as Internet service providers (ISPs), is to provide "the Internet" to the customer networks that purchase Internet connectivity from the ISP. Each ISP itself consists of multiple interconnected networks. ISPs connect to their customer networks through a link between an IP router in the ISP network and a switch or router in the customer network.

[25] Id. (NSA Response to Plaintiff's Request for Admission No. 4 (Jan. 8, 2018)).

156. According to broadbandnow.com, an Internet site providing information to people looking for ISPs in their area, there are over 2,600 ISPs in the U.S.[26] The ISPs range in size from the big carriers (such as AT&T Wireless and Verizon Wireless, which offer services in all 50 states plus some territories), to the large cable TV companies (which offer ISP service in as many as 40 states), to very small ISPs (such as Surge Communications, which offers Internet services in two just zip codes).

157. For example, Comcast offers its Xfinity Internet service in parts of 40 states. The Xfinity coverage is shown in the following figure:

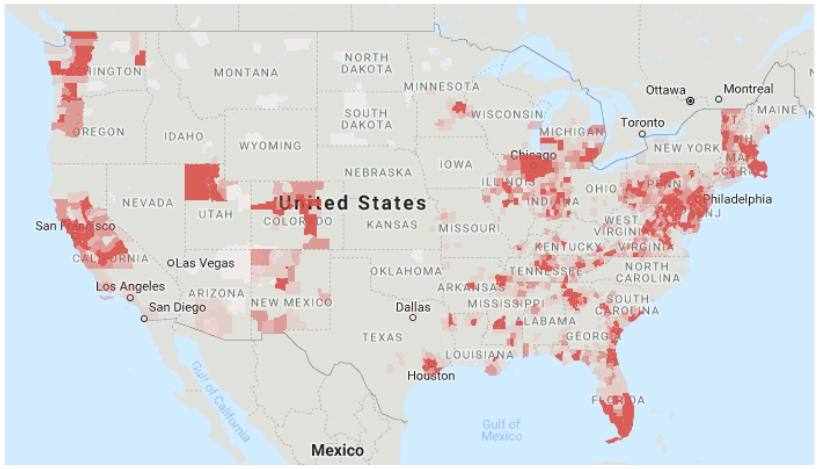



**Figure 20** — Xfinity coverage[27]

[26] *Internet Providers in the U.S.*, Broadband Now, https://broadbandnow.com/All-Providers.

[27] *Xfinity From Comcast Availability Map*, Broadband Now, https://broadbandnow.com/XFINITY (last updated Dec. 1, 2018).

158. A small ISP such as Orca Communications has a still smaller service area, in this case a small part of the southwest coast of Oregon:




**Figure 21** — Orca Communications service area[28]

### a. Address assignments for ISPs

159. Larger ISPs are assigned ranges of IP addresses by one of five Regional Internet Registries (RIRs), each of which is responsible for a part of the globe. The ISPs use the assigned addresses for their own networks, and they subassign some of the addresses to their customers for use in the customer's own networks.

160. Over the last few years a commercial market has developed for the right to use blocks of IP addresses.[29] Individual ISPs or companies can purchase the right to use a block of addresses from someone who currently has that right and then register the

[28] *Orca Communications Availability Map*, Broadband Now, https://broadbandnow.com/ORCA-Communications (last updated Dec. 11, 2018).

[29] Paul McNamara, *MIT Selling 8 Million Coveted IPv4 Addresses; Amazon a Buyer*, Network World (Apr. 21, 2017), https://www.networkworld.com/article/3191503/internet/mit-selling-8-million-coveted-ipv4-addresses-amazon-a-buyer.html.

block with one of the RIRs. The addresses do not have to be used in the same geographic area as they were being used before they were purchased.

#### *4. ISP Interconnection*

161. Because no one ISP connects to all of the customer networks that make up the customer network part of the Internet, ISPs must interconnect with other ISPs to get connectivity to the customer networks they do not directly serve. Each ISP decides on its own how to interconnect with other ISPs to get full Internet connectivity.

162. As a general rule, similarly sized ISPs interconnect with each other with little or no fees exchanged for the interconnection. This type of interconnection is known as ***peering***. Small ISPs must become the customers of larger ISPs in order to be able to interconnect with the larger ISP. The smaller ISP must pay for the interconnection, as any customer must. In general, the interconnections any one ISP maintains are considered proprietary information.

163. ISPs interconnect with other ISPs, either as customers or as peers, through private interconnections and through Internet exchange points.

164. Private interconnections are direct links from a node in one ISP's network to a node in another ISP's network. When large ISPs peer with other large ISPs, they do so at multiple geographically dispersed locations to ensure that traffic between the ISPs can be as distributed as the traffic sources or destinations are, and to ensure reliability through redundancy. For example, AT&T's peering policy requires a minimum of 6 peering points.[30] Large ISPs that peer with multiple other large ISPs are sometimes referred to as ***Tier 1 ISPs***. The ISPs generally considered to be Tier 1 ISPs in the U.S.

[30] *AT&T Global IP Network Peering Policy*, AT&T Business, https://www.corp.att.com/peering.

include AT&T, Verizon, Sprint, Century Link and Level 3. Tier 1 ISPs in Europe include FranceTelecom, Telefonica and Deutsche Telecom.[31]

165. An ***Internet exchange point*** (known as an ***IX*** or an ***IXP***) is a node, usually an Ethernet switch, which has links to nodes in multiple ISPs. Each ISP connected to the exchange point can use the exchange point to interconnect with any other ISP connected to the same exchange point subject to bilateral agreements between the ISPs. The operator of the exchange point need not be a party to any agreement between ISPs to exchange traffic.

166. A single ISP, particularly the large ones, can be connected to multiple Internet exchange points, sometimes in multiple countries or even continents[32]

***5. Customer Networks***

167. Customer networks in the Internet include the small ones such as the one in my house, as well as much larger networks such as the Harvard University's network, Google's internal network and the network at the U.S. Department of Agriculture. Most customer networks themselves consist of many interconnected individual networks.

168. The individual networks that make up a customer network might consist of one or more links, such as physical Ethernet links, interconnected with one or more switches or it might just consist of a single WiFi (wireless) network. The different individual networks that make up a customer network are interconnected with IP routers. For example, I have a physical Ethernet network with multiple Ethernet switches and two

[31] *Who Are the Tier 1 ISPS?*, Dr. Peering International, http://drpeering.net/FAQ/Who-are-the-Tier-1-ISPs.php.

[32] For example, see the list of the exchange points the Australian ISP Telstra peers at: Telstra (International), PeeringDB, https://www.peeringdb.com/net/1459.

WiFi networks in my house. These networks are connected together through an IP router, which I manage.

169. Harvard's network consists of a few hundred separate physical Ethernet networks, each consisting of Ethernet links to individual computers and Ethernet switches to interconnect the Ethernet links. The Harvard network also includes a few dozen WiFi networks. The individual Ethernet networks and the individual WiFi networks are interconnected with many IP routers. Google's internal network spans the globe and consists of an unknown (to me) number of individual networks interconnected through routers.

170. Each individual network in a customer network is assigned its own range of IP addresses to be used by the nodes, such as users' computers attached to that network. Generally, the overall customer network is assigned one or more larger blocks of IP addresses and the individual networks are assigned sub parts of the larger blocks.

**a. Address assignments for customer networks**

171. Most residential or small enterprise customer networks do not have fixed IP addresses on the Internet. Instead they use one or more IP addresses assigned by their ISP that may change from time to time. Larger enterprises can obtain fixed address assignments directly or, for an extra fee, from their ISPs. With some exceptions, networks that are not assigned fixed IP addresses cannot support Internet services such as email servers or web servers.

***6. Customer Network Interconnection***

172. As a general rule with some exceptions, customer networks do not interconnect directly with other customer networks. Instead customer networks connect to ISP networks to get Internet connectivity, including connectivity to other customer

networks. Customers expect to get access to the whole Internet when they purchase Internet service from an ISP.

#### *7. Network Address Translators (NATs)*

173. ***Network address translators*** (***NATs***) are network nodes that sit on the edge of an individual network, a group of networks or even a whole customer network. Their purpose is to translate the IP addresses in the header of an IP packet and the port numbers in the TCP or UDP header such that all of the network nodes on the network appear to have the same IP address. By sharing IP addresses in this way, NATs reduce the demand for the somewhat limited number of IPv4 addresses, and they can hide the internal structure of a network from observers outside of the network, which is seen as a security advantage.

174. But an effect of NATs is that individual computers whose packets pass through a NAT do not have separate IP addresses; they all have the same IP address that was assigned to the NAT, so the communications cannot be distinguished merely by looking at the IP addresses in the packets that make up the conversation.

### E. Routing in the Internet

175. Networks comprise one or more network links interconnected with switches. Networks are connected to other networks through routers.

176. As described above in ¶¶ 71, 84, the network nodes that are used to connect one network to another in the Internet are called routers. This is the case within a customer network, within an ISP network, between a customer network and an ISP network, and between ISPs.

177. For routers to know where to forward packets, they must understand the topology of a relevant part of the network. They gain this understanding by exchanging information with other routers within the name network. The same is true for the routers used to interconnect ISPs—they exchange information so that they can understand the Internet topology well enough to know where to forward packets they receive.

178. Routing protocols define the mechanisms the IP routers use to exchange this topology information. IP routers within a customer network or within an ISP network use a type of routing protocol designed to be used where all the IP routers are run by the same organization such that information from them can be trusted. Such a routing protocol is called an ***Interior Gateway Protocol*** (***IGP***). The two most common IGPs are Open Shortest Path First Routing Protocol (OSPF) and Intermediate System to Intermediate System Routing Protocol (IS-IS).

179. The routing protocol used between ISP networks and other ISP networks or between ISP networks and some of their larger customers is called an ***Exterior Gateway Protocol*** (***EGP***). The only EGP in current use in the Internet is Border Gateway Protocol version 4 (BGP4). ISPs do not generally run a routing protocol between themselves and their customer networks unless the customer has connected their network to multiple ISPs. In such cases, BGP4 is used.

180. Unlike with IGP routing protocols, EGPs operate in an environment where the different routers are operated by different organizations, and an ISP needs to be able to define the level of trust it wants to have in particular information from particular other ISPs or from their customers. Thus, BGP4 has an extensive set of mechanisms to let the operators of routers configure just what information they want to accept from other

routers and what information they want to provide to other routers. The configuration of these mechanisms in a router is done by the router operator. There are no general rules as to what the configuration should be.

### *1. Autonomous System (AS)*

181. A set of routers under common administrative control, such as the routers within a customer network or within an ISP, are assigned an ***Autonomous System number*** for identification. For example, many of the routers at Harvard are assigned AS 11. AS numbers are used by routing protocols as a way to refer to a part of a network or to a whole network such as an ISP.

### *2. Routing an IP Packet*

182. I will now walk through the process by which an IP packet is transported across the Internet, taking as an example my connecting to a web server.

183. In the first step, I type a URL which specifies a particular resource, such as a picture, on a specific website into the window at the top of my web browser, or I click on a link that specifies the same resource. I will use the website for the University of Oxford in England (www.ox.ac.uk) as an example website.

184. For my computer to be able to send a packet containing an HTTP request to www.ox.ac.uk, the computer needs to find out what IP address has been assigned to www.ox.ac.uk. This address is needed so it can be put in the destination IP address field of the packets my computer wants to send to www.ox.ac.uk. Computers use the ***Domain Name System*** (***DNS***) to convert the domain name in the URL into an IP address. At the time of this writing, one of the IP addresses for www.ox.ac.uk was 129.67.242.154.

185. My computer then creates a packet containing the HTTP command my browser wants to execute, likely a GET command, and puts the IP address of www.ox.ac.uk in the destination IP address field in the packet. My computer also puts its own IP address into the source IP address field in the packet. Then, using link-layer addressing, my computer sends the packet to my local router.

186. My local router then looks up the destination IP address in the router's ***routing database*** (also called a ***routing table***). This is the database maintained by the routing protocol. Using the information in the routing database, my local router determines which router the packet needs to go to next on its way toward the web server.

187. In general, my local router's routing table will not have an entry for the specific range of IP addresses that includes the IP address for www.ox.ac.uk. This is because there are many millions of such address ranges and my local router does not have the memory space or processing power to keep track of them all. Instead my local router, after determining that it does not have an appropriate entry in its routing table, uses a ***default route*** configured into the router to identify the ***next-hop router***. Using link-layer addressing, my local router then forwards the packet to that "next-hop router".

188. As a general rule, unless specifically configured otherwise, a router will try to find the "best" next-hop router where the determination of "best" is based on the "cost" of sending a packet through that next-hop router to the destination.

189. In an IGP, cost is generally determined by the number of routers the packet will need to traverse within a customer or ISP network in combination with the speed of the links between the routers.

190. In an EGP, cost is generally based on the number of ISPs (identified by their AS numbers) that the packet will need to traverse across the Internet to reach a destination. I say "generally" because the operator of the router can modify the router's configuration so as to determine the criteria. ISP operators configure the routers they use to connect to other ISPs to filter the routing information they accept from the other ISPs and the routing information they send to those ISPs. ISP operators do this to reject known bad routing information, to prefer next-hop routers in ISPs they have peering contracts with, to prefer some next-hop routers for load balancing reasons, and for a number of other operational reasons. (Dr. Schulzrinne's declaration states that a router may route packets to avoid congested connections. No IGP or EGP routing protocols currently in use on the Internet take "congestion" into account in routing packets. *See* Schulzrinne Decl. ¶ 40. That said, some ISPs do manually reroute traffic to avoid overloaded links.)

191. The next-hop router performs the same type of address lookup process to determine the router that is the next-hop from its point of view.

192. This process continues, hop by hop, until a router recognizes that the address is one on a link directly connected to that router. When a router recognizes this, it uses link-layer addressing to forward the packet to the web server.

193. The decision as to the next-hop router can change at any time based on the most up-to-date information in the routing table in the router, so the next packet in my message to www.ox.ac.uk could be sent to a different next-hop router. I will discuss routing table volatility in the next section.

### *3. Volatility of Routing Information*

194. The Internet today consists of millions of network links and millions of nodes, including switches and IP routers. Changes in state may not occur all that often in each one of these routers and links, but with millions of routers and links, each of which are subject to failures, the overall rate of state change can be significant. Each of these state changes can result in a routing update propagated throughout the Internet. Each of the routers receiving the update updates its own routing, which may produce a change in the next hop a particular packet may be forwarded to and, thus, the links a packet will traverse. Changes in router or link state can result from many things, including local power outages, equipment failures, management induced changes (e.g., turning off a link for debugging or, as mentioned above, rerouting traffic to avoid overloaded links) and physical damage to wires.

195. The following figure shows the rate of changes seen at a particular exchange point in September 2013:

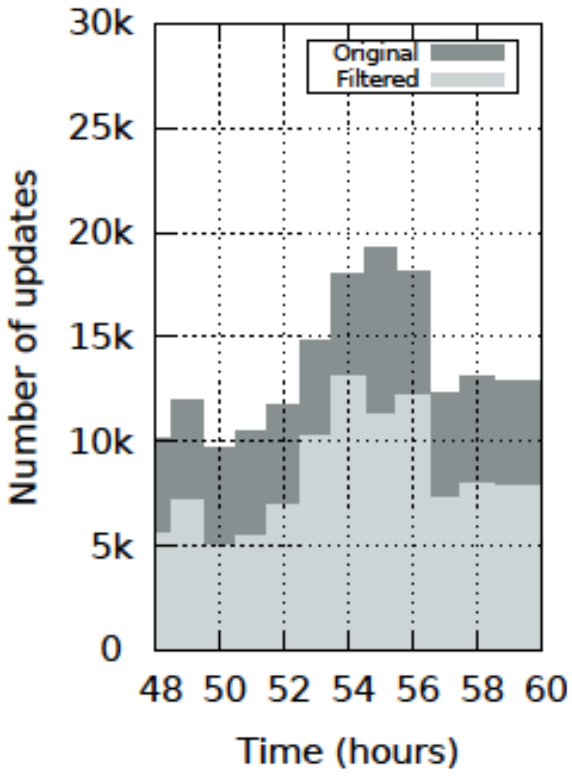



**Figure 22** — BGP updates at an Equinix exchange point in September 2013[33]

196. The figure shows the number of updates seen per hour over a 12-hour period starting September 19, 2013 at midnight GMT. The lighter grey area shows the number of unique updates per hour. For example, during the hour between 1 AM and 2 AM there were about 7,000 updates—a rate of almost two updates per second. Since BGP routing updates are propagated throughout the Internet, the same rate of updates will be seen by BGP routers all over the Internet.

[33] Appendix I at 4 (David Hauweele et al., *What Do Parrots and BGP Routers Have in Common?*, Computer Comm. Rev. (July 2016), https://ccronline.sigcomm.org/wp-content/uploads/2016/07/sigcomm-ccr-paper26.pdf)).

### *4. Asymmetric Data Paths*

197. The packets the web server sends back to my web browser in response to my hypothetical request follow the same process. Each router along the path makes its own determination of the next-hop router. Because of this there is no guarantee that the return packets will follow the same path that the request packets took.

198. I mentioned above in ¶ 164 that when large ISPs interconnect with other large ISPs, they generally do so at multiple geographically distinct places. As a general rule, ISPs configure their routers with special rules for the forwarding of packets that are destined to pass through another ISP. The ISPs generally configure the routers to send such packets to the other ISP through the closest interconnect even if that would not otherwise be the "best" path. Since both ISPs do the same, the paths packets take going in one direction can be very different than the paths packets take coming back. This configuration results in asymmetric paths for packets going in opposite directions between two network nodes. This type of routing is known as "nearest exit routing" or "hot-potato routing"—i.e., the ISP passes the packets off to another ISP as fast as it can.

199. Dr. Schulzrinne's description of routing in ¶¶ 41, 89 is incomplete in his failure to mention the asymmetric routing of communications. He states in ¶ 41, for example, that "*packets traveling between two points on the Internet generally follow the same path for long distances*". This is generally true for packets traveling in a particular direction, unless the ISP decides to change the path as I mention above in ¶ 190. But packets going in one direction between two points commonly take a very different path than packets going in the other direction between those same two points, due to asymmetric routing. You can think of the ISP forwarding rules that result in asymmetric routing as similar to one-way streets, causing the route you take from home to the

restaurant, for example, to be different from the route you take from the restaurant back home.

### F. International Connections

200. The heavily redundant connections between ISPs are reduced somewhat when it comes to intercontinental connections due to the relatively few undersea physical connections. Note that I'm referring to all of the cables connecting the U.S. to other countries as ***undersea*** even though one of them runs under Lake Ontario and would be more properly called an underlake cable. In addition to these undersea fiber cables, the U.S. is interconnected with Canada and Mexico with many ***terrestrial*** fiber cables. There are also some satellite-based interconnections, far fewer than there used to be before so many fiber cables were installed. Satellite-based connections are of far lower capacity than fiber-based ones and, because of the extra distance the signal has to travel up to the satellite and back, have added delays. Thus, satellite-based international communications are generally limited to islands that have not yet been connected with fiber cables, places far away from civilization and expensive satellite telephones. Since the vast majority of international Internet communications is transported over fiber, I will concentrate on that transport mode.

201. There are over 50 undersea fiber optic cables that connect the U.S. to other countries.[34] In addition, there are a number of fiber optic cables connecting the U.S. to Canada and to Mexico. The following figures are from TeleGeography, a well-regarded source of information about the telecommunications industry including, in particular,

[34] Appendix J (Report on International Submarine Cables Landing in the US, based on information compiled from Telegeography (Jan. 2018)).

maps of undersea cables. The first figure shows the undersea fiber cables connecting the U.S. to other countries as of early 2018:




**Figure 23** — Undersea fiber cables[35]

202. The following figure shows the trans-Atlantic cables:

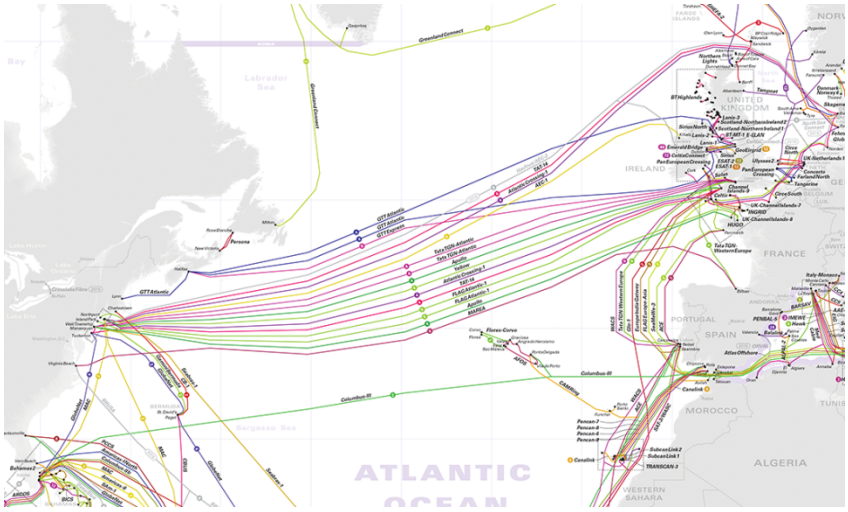



**Figure 24** — Trans-Atlantic undersea fiber cables[36]

[35] *Submarine Cable Map 2018*, TeleGeography, https://www.submarinecablemap.com/#/submarine-cable/tat-14.

[36] Id.

203. The following figure shows the trans-Pacific cables:

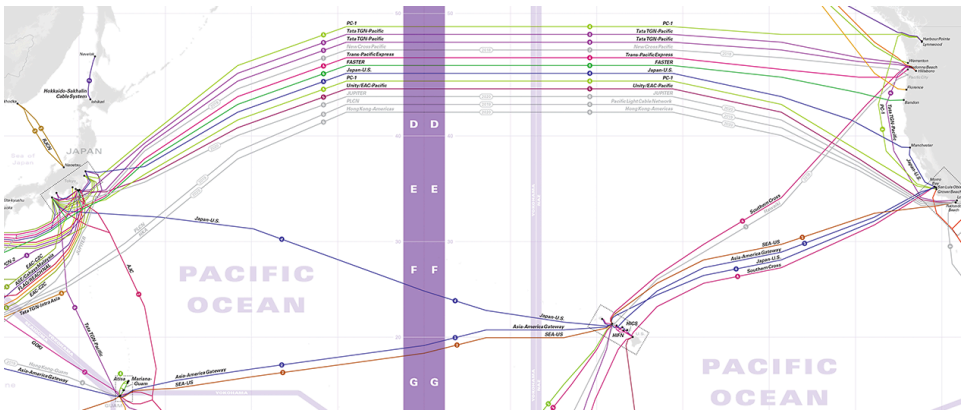
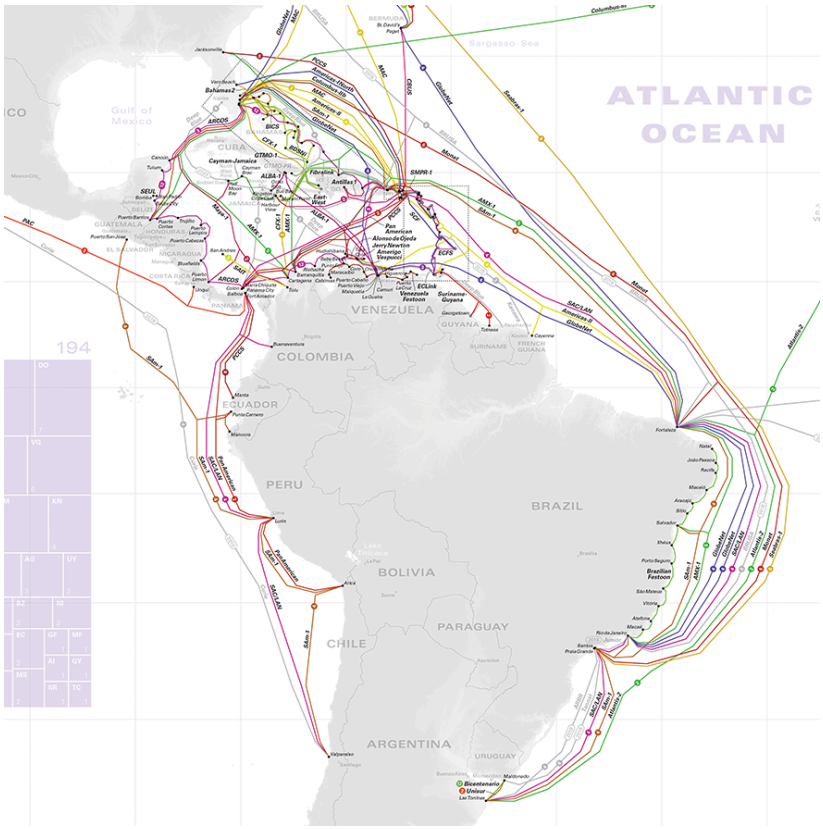



**Figure 25** — Trans-Pacific undersea fiber cables[37]

204. The following figure shows the undersea cables servicing South America and the Caribbean:




[37] Id.

**Figure 26** — Undersea cables between South America, the Caribbean and Europe[38]

205. The following figure shows terrestrial cables between the U.S. and Canada and between the U.S. and Mexico.



**Figure 27** — Terrestrial cables between the U.S. and Canada and between the U.S. and Mexico[39]

### *1. Details of Undersea Fiber-Optic Cables*

206. Each of the undersea cables contains multiple fiber pairs. One fiber in each pair is used to send traffic in one direction, and the second fiber in a pair is used to send traffic in the other direction. Each fiber can support multiple different simultaneous circuits, one on each of a number of colors of light, referred to as ***lambdas***. For example, one of the older transatlantic cables, the TAT-14 cable, has 4 pairs of fibers, each fiber of which supports 40 lambdas, for a total of 160 lambdas in each direction.[40] Each lambda

[38] Id.

[39] *ITU Interactive Transmission Map*, Int'l Tele-Comms Union, https://www.itu.int/itu-d/tnd-map-public (last updated Nov. 2018).

[40] *About the TAT-14 Cable Network*, TAT-14 Cable System, https://www.tat-14.com/tat14.

can support up to 40 gigabits per second (Gbps).[41] The full TAT-14 cables currently support 3.15 terabits per second (Tbps) each. A map of the TAT-14 cables is shown in the following figure:

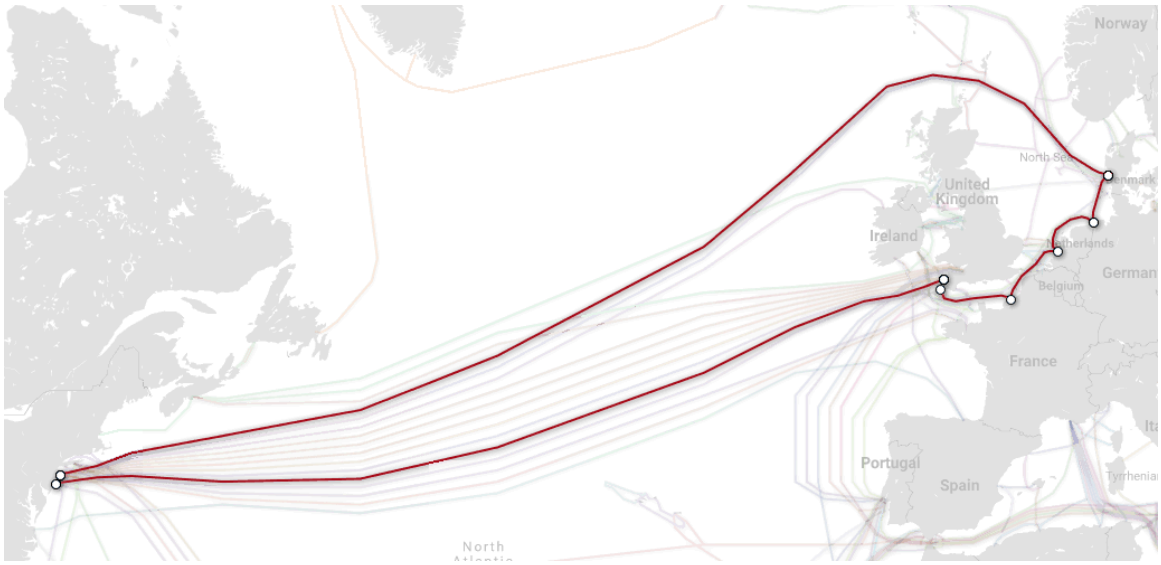



**Figure 28** — TAT-14 cable[42]

207. MAREA, a newer cable, installed by Microsoft, Facebook and Telxius (a global telecommunications infrastructure company) that connects the U.S. to Spain, contains 8 pairs of fibers and can support up to 160 Tbps.[43] A map of the MAREA cable is shown in the figure below:

[41] Gigabits per second (Gbps) is a measure of the speed of data transmission. A gigabit is a billion bits of information, and a bit is the smallest unit of digital information, represented by a one or zero. A terabit is 1,000 gigabits. For comparison, 8 bits make up a *byte*, a single text character is represented by a pattern of bits in a byte. A gigabit is enough data to carry about 30 million 4-character words or about 50 copies of Tolstoy's *War and Peace.*

[42] *Submarine Cable Map: TAT-14*, TeleGeography, https://www.submarinecablemap.com/#/submarine-cable/tat-14 (last updated Dec. 6, 2018).

[43] Deborah Bach, *Microsoft, Facebook and Telxius Complete the Highest-Capacity Subsea Cable to Cross the Atlantic*, Microsoft (Sept. 21, 2017), https://news.microsoft.com/features/microsoft-facebook-telxius-complete-highest-capacity-subsea-cable-cross-atlantic.

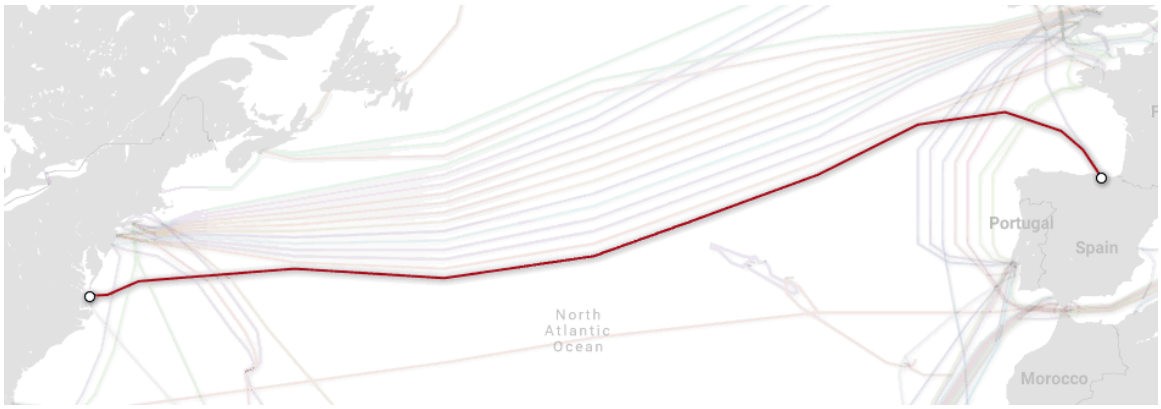



**Figure 29** — MAREA[44]

208. The number of fiber pairs in the undersea cables terminating in the U.S. ranges from 4 to 12 (with BICS being the only cable with 12 pairs, and only 4 cables having 8 pairs).[45]

209. Attached as Appendix J is a list of the international undersea cables that terminate in the U.S.[46] The list was compiled from the information on the TeleGeography website.

210. The above description is consistent with the government's description of the submarine cables:

> *The NSA Defendants respond further that, according to data available from Telegeography, international submarine cables typically contain 2-8 pairs of fiber-optic cables. Each fiber-optic pair is typically capable of carrying between approximately 15 and 120 individual communications circuits on different light wavelengths, depending on age and technology used. As a result, an individual submarine cable may carry between approximately 30 and 960 communications circuits. (Individual circuits*

[44] *Submarine Cable Map: MAREA*, TeleGeography, https://www.submarinecablemap.com/#/submarine-cable/marea (last updated Dec. 6, 2018).

[45] *Submarine Cable Map 2018*, TeleGeography, https://www.submarinecablemap.com.

[46] Appendix J (Report on International Submarine Cables Landing in the US, TeleGeography (Jan. 2018)).

> *may be subdivided further to create multiple "virtual circuits" through application of various technologies.) Each wavelength carried on a fiber-optic pair is typically capable of transporting between 10 and 100 gigabits of data per second (10-100 Gbps), meaning that a typical submarine cable can carry between approximately 300 and 96,000 Gbps of data.*[47]

211. Devices at the transmitting end of a fiber use electronics to convert packets into modulated beams of light at specific frequencies (a.k.a., lambdas), and they then use optics to combine multiple lambdas into a single beam of light to send onto the fiber. Devices at the receiving end of a fiber use optics to split the beam of light from the fiber into the individual lambdas, and they then use electronics to reconstitute streams of packets from each of the lambdas.

#### *2. Details of Terrestrial Fiber-Optic Cables*

212. Terrestrial fiber-optic cables, ones that cross borders or ones that are a part of an ISP's infrastructure, are much shorter than undersea cables and tend to have far more fibers but, otherwise, operate in the same way that undersea cables do.

#### *3. Public Internet Communications on International Fiber-Optic Cables*

213. An individual company can own or lease a whole cable, pairs of fibers within a cable or pairs of lambdas within fibers. In some cases, the cable, fibers or lambdas are owned or leased by ISPs and used as part of the ISP's internal network, as circuits for peering with another ISP or as circuits to Internet exchange points. In the cases where the circuits are connecting to another ISP or to an exchange point, all communications on the circuit would be what I will call in this report ***public Internet***

[47] Appendix H at 4-5 (NSA Response to Plaintiff's Request for Admission No. 1 (Jan. 8, 2018)).

***communications*** or ***public Internet traffic***. That is, communications between Internet users. In the case where the circuit is used as part of the ISP's own network, some of the communications will be to support the ISP operations—to manage their routers for example. These communications would not be considered public Internet communications, while the rest of the traffic on such an internal communications link would be public Internet communications since it would be between Internet users.

214. Not all of the fibers in these cables are used for public Internet communications. Some of the undersea cables, fibers or lambdas are owned or leased by companies for use as part of their own internal networks or for corporate telephone and video communications. Communications on these cables would not be considered public Internet communications. In addition, many cables were built with more fiber than were initially required to allow for future expansion and have not yet been made active, or "lit."

215. Thus, public Internet communications are transported on a subset of the lambdas operating as circuits in a subset of the fibers that these undersea cables are capable of supporting. Since many ISPs consider their internal architecture and the number and location of the other ISPs they peer with to be proprietary, the ISPs and cable operators often do not publicly disclose the specific circuits that are used to transport public Internet communications.

216. Internet sites such as TeleGeography have done a very good job of cataloging the undersea cables that tie together countries around the world, but these sites do not break down which circuits on which fibers on which cables are used for public

Internet traffic and which are used for other purposes such as video distribution or internal corporate networks.

217. Viewing the available information, it is reasonable to infer that the distribution of circuits transporting public Internet communications roughly matches the overall distribution of undersea international cables and terrestrial international cables because the cables, in general, connect population centers where large numbers of Internet users live and work.

#### *4. Undersea Fiber-Optic Cable Landing Locations*

218. There are 47 sites where the international undersea cables that were identified from the TeleGeography information come ashore in the U.S.[48] Some of the cables come ashore in more than one U.S. location. TAT-14, which has branches that come ashore in two towns about 40 miles apart on the New Jersey shore, is an example of such a cable. When an undersea cable comes ashore, it is run to an enclosure where the individual fibers are broken out of the cable. The fibers can terminate in network devices (such as routers) in such an enclosure as shown in ¶ 23 of Dr. Schulzrinne's declaration, or they could be patched through to another cable that connects that enclosure to a location, such as a data center, where the network devices are located. The second option is shown in the following figure from a Virginia Beach planning presentation for the MAREA cable termination. The figure shows a conduit path from an enclosure at the beach where the cable comes ashore to a data center where the network devices are:

[48] See Appendix J for a list of the termination sites.




**Figure 30** — Conduit path for MAREA cable[49]

219. International Internet links can terminate at a variety of different types of physical facilities within the U.S. For example, some terminate at patching stations such as the one in Virginia Beach shown in Figure 31 below, cable landing stations such as the one shown in ¶ 23 of Dr. Schulzrinne's declaration, Internet exchange points, peering points, or ISP points of presence.

[49] Appendix CC (City of Virginia Beach Dep't of Info. Tech., *Next Generation Network and Transoceanic Subsea Cable Updates* (Oct. 4, 2017), https://www.vbgov.com/government/departments/communications-info-tech/Documents/NGN-and-Transoceanic-Subsea-Cables.pdf).



**Figure 31** — Manhole for fiber patching in Virginia Beach[50]

220. As discussed above in ¶¶ 200-201, the vast majority of the U.S. international Internet communications—i.e., communications that start or end in the U.S. where the other end is outside the U.S.—go through the undersea or terrestrial fiber cables shown in the figures above in ¶¶ 201-204.

### *5. Terrestrial Fiber-Optic Cable Terminations*

221. International terrestrial fiber-optic cables do not require as distinct terminations as do undersea cables. Many of them are simple ISP interconnects or connections to Internet exchanges and are indistinguishable from any other terrestrial fiber-optic cables.

[50] Id.

### G. Places to Monitor International Public Internet Communications

222. As can be seen in the figures and discussion above, the U.S. termination points of the circuits carried on international undersea cables (see ¶¶ 218-220), as well as the U.S. ends of the international terrestrial cables (see ¶¶ 200, 216, 221) are prime locations to monitor communications between Internet users in the U.S. and Internet users in other countries, because essentially all of the public Internet communications between the U.S. and other countries flow over these circuits.

223. U.S. ends of the circuits carried on the trans-Atlantic and trans-Pacific cables are also attractive places to monitor public Internet communications between some non-U.S. and non-U.S. sites (other than Mexico and Canada). As can be seen from Figure 26, there is only one 2-pair fiber cable connecting South America to Europe and there are no cables connecting South America or the Caribbean with the Far East. Thus, almost all public Internet communications passing between South America, the Caribbean and the rest of the world will pass through the U.S. The same is true, but to a lesser extent, for public Internet communications in circuits in undersea cables between the Far East (China, Japan, Taiwan, and South Korea) and Europe. This means that the U.S. ends of the circuits carried on the trans-Atlantic and trans-Pacific cables are prime locations for monitoring public Internet communications between many non-U.S. locations. Monitoring at those locations also means that any monitoring equipment need only be in U.S. territory. Such monitoring locations would generally not capture communications entirely within a region such as communications between Europeans or such as communications between residents of the Far East.

224. As can be seen in the figures above, the total number of international undersea and terrestrial cables is relatively small, and there are even fewer physical locations where the cables terminate because multiple cables terminate at some of the locations. It is certainly not out of the question that the NSA would have been able to deploy upstream collection devices at all of these sites.

225. As I discuss below in ¶ 291, the FISC has confirmed that the NSA does in fact monitor at least some "*international Internet link[s]*",[51] which are the circuits connecting a network node in the U.S. to a network node in a foreign country. This of course makes sense, given that public Internet traffic on international Internet links will consist almost entirely of communications being sent or received (or both) by a node outside the U.S., which is the traffic that the NSA is authorized to monitor under its Section 702 procedures. It is not relevant to my report or to the conclusions I come to what type of facilities or physical locations at which the NSA is monitoring international Internet links; the relevant point is that the NSA is monitoring at least some international Internet links.

226. NSA representative Rebecca J. Richards, during her deposition, did not specifically say that the NSA monitors at the U.S. ends of the circuits carried on the trans-Atlantic and trans-Pacific cables, but she did say that the NSA did monitor at least one "*Internet backbone circuit*",[52] and she agreed that the international undersea cables can be part of the "*Internet backbone*".[53]

[51] Appendix P at 45 (FISC Opinion (Oct. 3, 2011)).

[52] Appendix K at 122:20-123:5 (Transcript of Deposition of Rebecca J. Richards (Apr. 16, 2018)).

[53] Id. at 79:15-20.

227. In several of its officially disclosed documents, the government has confirmed that it conducts upstream collection on multiple circuits. For example, the PCLOB Report states that upstream collection occurs with the compelled assistance "*of the providers*"—plural—"*that control the telecommunications backbone*".[54] The report also states that the providers facilitating upstream collection must "*assist the government in acquiring communications across these circuits*"—again, plural.[55] That said, it seems very obvious, as the PCLOB Report confirms, that the NSA must be monitoring more than one circuit carried on the international undersea cables. The NSA's thousands of surveillance targets are, presumably, in many parts of the world, and so if the NSA monitored only a single circuit in a single international undersea cable, it could not capture many or most of the communications of those geographically dispersed targets. Moreover, asymmetric routing (as discussed above in ¶¶ 197-199) means that monitoring only a single link could only ever capture those packets in a communication going in one direction, and monitoring only a single link could easily miss all of a target's packets if the routing changed as described above in ¶¶ 194-196.

228. Based on the NSA's description of the capability of undersea fiber cables, cited above at ¶ 210, the international undersea and terrestrial cables that terminate in the U.S. are capable of supporting thousands of individual communications circuits. Some fraction of these circuits are used to transport public Internet communications. It may be that the NSA has deployed enough upstream capture systems to provide full coverage of the international circuits that are used to transport public Internet communications, or the

[54] Appendix F at 40 (PCLOB Report at 35); see also id. at 12 (PCLOB Report at 7).

[55] Id. at 36-37.

NSA may have not done so yet. In any case, I find it hard to believe that the NSA has left many such circuits unmonitored considering the high number of surveillance targets, the variety of circuits that targets' Internet communications may travel into and out of the U.S., the variable routing of Internet communications, the importance the government attributes to the upstream collection program, and the NSA's stated desire to be comprehensive in its collection.[56]

### H. Locating Network Nodes Using IP Addresses

229. The use of regional assignment of IP addresses coupled with companies which have developed databases of the geographic locations of specific IP address ranges mean that determining where on the globe a network node using a particular IP is located has become quite reliable. One example of a use of such lists is a system that needs to restrict access to copyrighted material for licensing reasons. For example, Apple iTunes is only usable in specific countries. One commercial database of U.S. IP address ranges includes more than 66,000 individual entries.[57]

230. Locating where a network node is in the real world using the IP address in packets sent to or from a network node is generally but not always accurate. A NAT (see ¶¶ 173-174) will make a whole network's worth of network nodes appear to be in a single location even if the network nodes were actually located anywhere on a nation-wide or world-wide enterprise network. In addition, network nodes using VPNs or tunnels (see ¶ 91) will appear to be where the VPN or tunnel ends rather than where the node actually is. Thus, an IP address filter which uses a list of "U.S. IP addresses" to include or

[56] Id. at 10, 123, 143.

[57] *Create Country ACL*, Country IP Blocks, https://www.countryipblocks.net/country_selection.php.

exclude communications to be reviewed will likely exclude some communications that should be included or include some communications that should be excluded from or to U.S. Internet nodes because of the use of VPNs and NATs.

## V. NSA'S SECTION 702 COLLECTING OF COMMUNICATIONS

231. The NSA collects copies of communications involving non-U.S. persons under the authority of Section 702 of the Foreign Intelligence Surveillance Act, as amended. As the government has acknowledged, some of the communications also involve U.S. persons.[58] Two of the NSA's collection programs fall under the authorization of Section 702: ***PRISM*** and ***upstream collection***.[59] I will describe both of these programs below.

232. Under these programs the NSA collects, at least, recordings of phone calls and copies of Internet communications, which the NSA refers to as "*transactions*" (see ¶¶ 63-65), as well as metadata about the communications.

233. The NSA stores these copies in multiple NSA systems and data repositories:

> *Communications provided to NSA under Section 702 are processed and retained in multiple NSA systems and data repositories. One data repository, for example, might hold the contents of communications such as the texts of emails and recordings of conversations, while another, may only include metadata, i.e., basic information about the communication,*

[58] Appendix F at 7, 11 (PCLOB Report at 2, 6).

[59] Id. at 12 (PCLOB Report at 7).

> *such as the time and duration of a telephone call, or sending and receiving email addresses.*[60]

234. These NSA systems and data repositories are also referred to collectively as "*Section 702 databases*".[61]

235. NSA analysts use search tools to identify copies of communications that are stored in the Section 702 databases and which may be relevant to a particular investigation.

**A. Selectors**

236. Both PRISM collection and upstream collection programs make use of ***selectors*** to identify the communications that are to be collected.

237. The following excerpt describes how selectors are determined:

> *Once the NSA analyst has identified a person of foreign intelligence interest who is an appropriate target under one of the FISC-approved Section 702 certifications, that person is considered the target. The NSA analyst attempts to determine how, when, with whom, and where the target communicates. Then the analyst identifies specific communications modes used by the target and obtains a unique identifier associated with the target - for example, a telephone number or an email address. This unique identifier is referred to as a selector. The selector is not a "keyword" or particular term (e.g., "nuclear" or "bomb"), but must be a specific communications identifier (e.g., e-mail address).*[62]

[60] Appendix L at 7 (NSA Director of Civil Liberties & Privacy Office, *NSA's Implementation of Foreign Intelligence Surveillance Act Section 702* at 6 (Apr. 16, 2014) ("DCLOP Report")).

[61] Appendix F at 132 (PCLOB Report at 128).

[62] Appendix L at 5 (DCLOP Report at 4).

238. The May 2, 2011 letter from a Department of Justice official to Judge Bates of the FISC describes the selectors used in the upstream collection program as including "*electronic communication accounts*", "*electronic communication addresses*" and "*electronic communications identifiers*":

> *As previously described to the Court, in conducting upstream collection using electronic communication accounts/addresses/identifiers (hereinafter "selectors") pursuant to Section 702, NSA acquires Internet communications that are to or from a tasked selector, or which contain a reference to a tasked selector.*[63]

239. The December 8, 2011 DoJ, NSA & DNI joint statement notes that accounts can be tasked:

> *Thus although upstream collection only targets Internet communications that are not between individuals located in the United States and are to, from, or about a tasked account, there is some inevitable incidental collection of wholly domestic communications or communications not to, from, or about a tasked account that could contain U.S. person information.*[64]

240. The Privacy and Civil Liberties Oversight Board (PCLOB) July 2, 2014 report provides additional details on what can be a selector and what cannot:

> *The Section 702 certifications permit non-U.S. persons to be targeted only through the "tasking" of what are called "selectors." A selector must be a specific communications facility that is assessed to be used by the target,*

[63] Appendix M at 1 (FISC Submission (May 2, 2011)); *see also, e.g.*, Appendix N at 4-5 (FISC Submission (Aug. 16, 2011)).

[64] Appendix O at 8 (Joint Statement at 7, *FISA Amendments Act Reauthorization: Hearing Before the H. Permanent Select Comm. on Intelligence* (Dec. 8, 2011)).

> *such as the target's email address or telephone number. Thus, in the terminology of Section 702, people (non-U.S. persons reasonably believed to be located outside the United States) are targeted; selectors (e.g., email addresses, telephone numbers) are tasked. The users of any tasked selector are considered targets—and therefore only selectors used by non-U.S. persons reasonably believed to be located abroad may be tasked. The targeting procedures govern both the targeting and tasking process. Because such terms would not identify specific communications facilities, selectors may not be key words (such as "bomb" or "attack"), or the names of targeted individuals ("Osama Bin Laden"). Under the NSA targeting procedures, if a U.S. person or a person located in the United States is determined to be a user of a selector, that selector may not be tasked to Section 702 acquisition or must be promptly detasked if the selector has already been tasked.*[65]

241. Note that the selector **must be a specific communications facility** such as a telephone number for a telephone facility or an email address for an email facility and cannot be some generic word (e.g., "bomb") or someone's name, since neither of these would be an identifier that was specific to a particular communications facility.

242. Most of the documentation the NSA has publicly released only lists telephone numbers and email addresses as examples of selectors. But some of these documents describe selectors as "*electronic communication accounts/addresses/identifiers*".[66]

[65] Appendix F at 37-38 (PCLOB Report at 32-33).

[66] Appendix M at 1 (FISC Submission (May 2, 2011)).

243. Examples of "*electronic communications accounts*" or "*electronic communications identifiers*" could include Twitter handles, Skype, Snapchat, Snow (a Chinese Snapchat), WhatsApp or Instagram IDs, Wikimedia usernames and similar application-specific identifiers or account names. URLs of target websites or services would also meet the description of "*electronic communications addresses*".

244. In theory, IP addresses could be selectors because they are unique identifiers that qualify as "*electronic communication addresses*". It is worth noting however, that there are many circumstances in which IP addresses do not uniquely identify individual Internet users, which might present difficulties for the NSA in using them as selectors, depending on the circumstances. As the FISC summarized the NSA's explanation:

> *Internet communications are "nearly always transmitted from a sender to a recipient through multiple legs before reaching their final destination." June 1 Submission at 6. For example, an e-mail message sent from the user of [redacted] to the user of [redacted] will at the very least travel from the [redacted] user's own computer, to [redacted], to [redacted] and then to the computer of the [redacted] user. Id. Because the communication's route is made up of multiple legs, the transaction used to transmit the communication across and particular leg of the route need only identify the IP address at either end of that leg in order to properly route the communication. Id. at 7. As a result, for each leg of the route, the transaction header will only contain the IP addresses at either end of that particular leg. Id.*[67]

[67] Appendix P at 34-35 n.33 (FISC Opinion (Oct. 3, 2011)).

245. In other words, packets making up the communication on each of these legs would have the IP addresses of the ends of the individual leg in their source and destination IP address fields. Thus, the IP addresses in the packets of the communications could change multiple times between the source and destination.

246. In addition, the IP addresses in the packets that make up email messages sent or received by a mail server on behalf of any of its users will have the same IP address—the IP address of the server—as their source or destination address, and all packets sent to or from the network nodes behind a NAT or VPN will have the NAT's IP address in the packet's source or destination address fields. (See ¶¶ 173-174.)

247. For these reasons, IP addresses will frequently not be effective selectors for identifying the communications of targets. This, in turn, means that it is more likely that the NSA is reassembling communications in order to determine if they contain selectors.

248. The above sorts of identifiers and others would be uniquely identifying in the way that selectors must be, and so could very well be the type of selectors the NSA uses in conducting upstream collection. The NSA has not publicly disclosed whether it uses them, however, and at least with respect to URLs, the NSA refused during its deposition to say whether it uses them as selectors.[68]

[68] Appendix K at 207:6-208-11 (Richards Depo.)

## VI. PRISM COLLECTION PROGRAM

249. Although this case is about upstream collection, understanding how PRISM collection works may be useful in understanding the distinguishing features of upstream collection. (Note that the NSA now refers to PRISM collection as "*downstream collection*".) The Privacy and Civil Liberties Oversight Board (PCLOB) described the PRISM process as follows:

> *In PRISM collection, the government sends a selector, such as an email address, to a United States-based electronic communications service provider, such as an Internet service provider ("ISP"), and the provider is compelled to give the communications sent to or from that selector to the government. PRISM collection does not include the acquisition of telephone calls. The National Security Agency ("NSA") receives all data collected through PRISM. In addition, the Central Intelligence Agency ("CIA") and the Federal Bureau of Investigation ("FBI") each receive a select portion of PRISM collection.*[69]

## VII. OPINIONS A, B & C: THE NSA'S UPSTREAM COLLECTION PROGRAM INVOLVES COPYING, REASSEMBLING AND REVIEWING INTERNET TRANSACTIONS

250. In the subsections that follow, I explain how the NSA's upstream collection program must work at a technical level, in the monitoring of any particular circuit. As discussed below in ¶¶ 265-329, I conclude that the NSA's upstream collection process must, as a technical matter, involve copying at an absolute minimum the packets constituting the transactions it wishes to review for the presence of selectors. I also conclude that, as a matter of practical necessity, upstream collection involves either:

[69] Appendix F at 12 (PCLOB Report at 7).

a. copying all of the packets flowing on the circuit, so that the packets can be sent to an IP filter to eliminate those that are part of a wholly domestic transaction, if necessary; or

b. copying all of the packets that an IP address filter test determines are not part of a wholly domestic transaction.

251. In either case, at least the packets that are not part of a wholly domestic transaction are copied.

252. **Opinion A:** Thus, it is my opinion that, to conduct upstream collection of international public Internet communications traversing any particular circuit, as this operation has been described by the government, the NSA must be copying at an absolute minimum the packets constituting the transactions it wishes to review for the presence of selectors. Based on other practical necessities I describe below, it is also my opinion that the NSA is almost certainly either (1) copying all packets traversing that circuit or (2) copying all of the packets that an IP address filter test determines are not part of a wholly domestic transaction.

253. As discussed below in ¶¶ 301-309, I also conclude that to determine whether an Internet transaction that passes the NSA's filter contains a selector, the NSA must first reassemble captured packets into transactions.

254. **Opinion B:** Thus, it is my opinion that, in order to review Internet transactions to determine if a selector tasked for collection is present, the NSA must be reassembling the packets of the transactions it intends to review.

255. As discussed below in ¶¶ 310-327, I also conclude that to determine whether an Internet transaction that passes the NSA's filter contains a selector, the NSA must review all of the reassembled copies of Internet transactions by scanning them to determine if the reassembled Internet transactions contain one of more selectors.

256. **Opinion C:** Thus, it is my opinion that the NSA must review the reassembled Internet transactions in order to identify those that include a tasked selector and thus are subject to collection under the upstream collection program.

### A. Upstream Collection Program

257. This case concerns the NSA's upstream collection program, also referred to as ***upstream surveillance***.

258. In the following section I will describe the NSA's upstream collection program as it existed in 2015, when Wikimedia filed its amended complaint. Between 2015 and now, the NSA suspended one part of the program—the part referred to as ***about collection***. As I explain further below, ***about collection*** involved the collection of communications that included a selector in the body of the communication and were therefore "about" a target.

259. In summary, the NSA uses the upstream collection program to collect Internet transactions that contain selectors (see ¶¶ 236-248) and that are from or to a non-U.S. person outside the U.S. The actual collection is done by devices that execute a type of what is known as ***deep packet inspection*** (***DPI***). DPI is a well-known and widely used tool used in enterprise and ISP networks to scan network communications for various purposes, including the detection of security threats. Billions of dollars of DPI equipment are sold annually around the world by many different equipment

manufacturers.[70] For example, since 2008, the Department of Homeland Security has been using successive generations of a DPI system—known as EINSTEIN 2 and EINSTEIN 3 Accelerated—to help protect a number of federal agency networks.[71]

### *1. A Description of NSA's Upstream Collection Program*

260. The government has made a number of statements describing the upstream collection program.

a. The PCLOB described upstream collection as follows:

*upstream collection . . . occurs with the compelled assistance of providers that control the telecommunications 'backbone' over which telephone and Internet communications transit, rather than with the compelled assistance of ISPs or similar companies.*[72]

b. The PCLOB also said that the term "upstream" refers to the fact that the surveillance

*does not occur at the local telephone company or email provider with whom the targeted person interacts . . . but instead occurs 'upstream' in the flow of communications between communication service providers.*[73]

c. In the March 19, 2014 PCLOB hearing, Rajesh De, General Counsel of the NSA, stated

*upstream collection refers to collection from the, for lack of a better phrase, Internet backbone rather than Internet service providers."*[74] *In a*

[70] See, for example: *Deep Packet Inspection (DPI) Market Research Report, Analysis, Trends, Market Size Estimations and Forecast to 2022*, Reuters (Sept. 12, 2017), https://www.reuters.com/brandfeatures/venture-capital/article?id=16008.

[71] U.S. Dep't of Homeland Security, *EINSTEIN*, https://www.dhs.gov/einstein (last updated May 17, 2018).

[72] Appendix F at 12 (PCLOB Report at 7).

[73] Id. at 40 (PCLOB Report at 35).

> *declaration, Miriam P. stated "Upstream collection, in contrast, involves the compelled assistance (through a Section 702 directive) of certain providers that control the telecommunications backbone over which telephone and Internet-based communications transit. Unlike PRISM, Upstream collection generally involves the acquisition of certain communications as they traverse the telecommunications backbone.*[75]

261. All of these statements differentiate upstream collection from PRISM collection based on where the surveillance takes place and the manner in which the surveillance is conducted. Whereas PRISM collection involves compelling electronic communications service providers to turn over communications of their users, upstream collection involves compelling telecommunications providers to turn over communications that transit their networks. And whereas PRISM collection involves the collection of communications to or from the government's targets, upstream collection involves the collection of communications to, from, or (until April 2017) "about" the government's targets.

262. The government's public statements concerning the locations at which upstream collection is conducted are somewhat inconsistent. The second PCLOB statement above, ¶ 260.b, describes upstream collection as taking place in the "*flow of communications between communication service providers.*" The other statements, ¶ 260.a & c, refer to upstream collection as occurring on the "*Internet backbone*," which, as discussed above in ¶¶ 150-153, the government defines more broadly as including (1) the high-speed circuits (network links) and routers that are used to interconnect ISPs, (2)

---

[74] Appendix Q at 26:6-8 (PCLOB, Transcript of Public Hearing Regarding the Surveillance Program Operated Pursuant to Section 702 of the Foreign Intelligence Surveillance Act (Mar. 19, 2014)).

[75] NSA Decl. ¶ 7, *Jewel v. NSA*, No. 4:08-cv-04373 (N.D. Cal. Nov. 7, 2014) (ECF No. 300).

the circuits carried on the undersea cables that connect the U.S. with other countries, and (3) the high speed terrestrial network links (circuits) within the U.S. and between the U.S. and other countries, whether the undersea or terrestrial network links are **between ISPs or within an ISP**.

263. In her deposition, the NSA's representative Rebecca J. Richards agreed that the Internet backbone included connections between ISPs and within ISPs.[76]

264. I will assume for this report that upstream collection may take place on circuits either between ISPs or within an ISP.

### *2. Upstream Collection Process*

265. The process followed for upstream collection was described in the PCLOB report as follows:

> *Once tasked, selectors used for the acquisition of upstream Internet transactions are sent to a United States electronic communication service provider to acquire communications that are transiting through circuits that are used to facilitate Internet communications, what is referred to as the "Internet backbone." The provider is compelled to assist the government in acquiring communications across these circuits. To identify and acquire Internet transactions associated with the Section 702–tasked selectors on the Internet backbone, Internet transactions are first filtered to eliminate potential domestic transactions, and then are screened to capture only transactions containing a tasked selector. Unless transactions pass both these screens, they are not ingested into government databases. As of 2011, the NSA acquired approximately 26.5 million Internet transactions a year as a result of upstream collection.*[77]

[76] Appendix K at 47:18-22, 52:16- 53:12, 54:20- 55:7 (Richards Depo.).

[77] Appendix F at 41-42 (PCLOB Report at 36-37).

266. The "*government databases*" mentioned in the above extract are the same ones referred to as the "*Section 702 databases.*" (See ¶¶ 233-234.)

267. Upstream collection program-related documents refer to both screening, as the above extract does, and "*scanning*."[78] I will use the term ***reviewing*** in this report for this function.

268. The extract at ¶ 265 describes a 3-stage upstream collection process, but given the manner in which upstream collection must be conducted (as I explain below), it is clearer to describe upstream collection conceptually as having 5 stages.

### a. Stage 1: Copying the Packets

269. As described in ¶ 38, multiple communications are simultaneously run over each Internet circuit by intermingling packets from different communications on the circuit. This is shown in the following figure:



**Figure 32** — Packets on a circuit

270. The small rectangles in the above figure represent packets flowing from left to right over a circuit. The different colors represent packets from different communications. For this explanation, the circuit is one of the ones that the NSA refers to as an Internet backbone circuit and is operated by an electronic communication service provider, which I will refer to as an ***ISP***.

271. I refer to a ***monitoring system*** in the section below. By that I mean, one or more devices that perform the processing required to implement upstream collection.

[78] See, e.g., Appendix R at 3, 24 (FISC Submission (June 28, 2011)); Appendix S at 6 (NSA Section 702 Minimization Procedures (2014)); Appendix F at 124 (PCLOB Report at 119).

Some of these devices may be ones designed by the NSA specifically for the upstream collection program. The government has acknowledged using "*NSA-designed upstream Internet collection devices*" in the upstream collection process.[79] Some of the devices may be off-the-shelf networking devices. I do not mean to imply any particular arrangement of such devices by using the term "system."

272. As a technical matter, there are only two possible configurations the NSA could be using to accomplish the copying of transactions necessary for upstream collection:

a. Copying all the traffic on a circuit so that the traffic can be passed on to one or more devices that then isolate the Internet transactions of interest. I will refer to this configuration as the ***copy-then-filter*** configuration.

*or*

b. Copying a subset of the traffic on the circuit, for example only the packets that are a part of Internet transactions that are not wholly domestic, and then passing the copied traffic on to one or more devices that then isolate the Internet transactions of interest. I will refer to this configuration as the ***in-line filter*** configuration.

[79] Appendix F at 44 (PCLOB Report at 39).

i. Copy-Then-Filter

273. The copy-then-filter configuration is shown in the following figure:

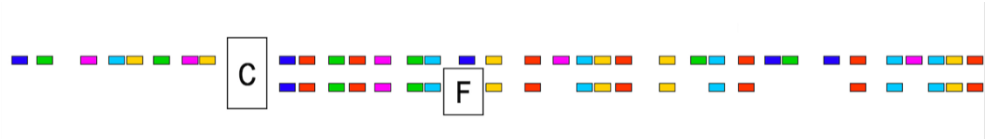



**Figure 33** — Copy-then-filter

274. The box marked "C" in Figure 33 represents a device that copies the traffic. The copying in the copy-then-filter configuration could be done in one of two ways; both ways use devices that are placed into a fiber or a circuit:

a. at the physical layer using a fiber-optic splitter;

*or*

b. at the link layer using a device that makes a copy of all the packets on a circuit.

(1) Fiber-optic splitter

275. A fiber-optic splitter splits the light on a fiber into two parts, each of which is put on its own fiber. Such a splitter could be placed on a fiber carrying traffic from an ISP's terrestrial network into an international cable (Figure 34) or a fiber carrying traffic from an international cable into an ISP's terrestrial network (Figure 35).

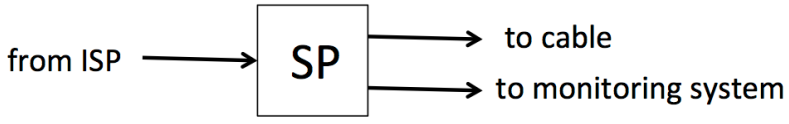



**Figure 34** — Fiber-optic splitter on fiber **into** an international cable

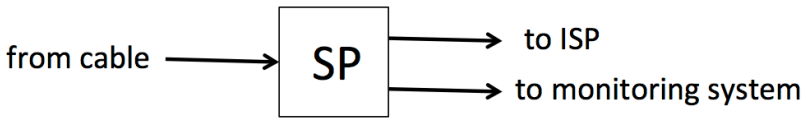



**Figure 35** — Fiber-optic splitter on fiber **from** an international cable

276. For the configuration shown in Figure 34, in which the ISP is sending traffic onto the international cable, one fiber from the splitter goes to the international cable and the other fiber goes to the monitoring system. In the other configuration (Figure 35), where the ISP is receiving traffic from the international cable, one fiber from the splitter carries traffic to the ISP, and the other one goes to the monitoring system.

277. In both cases, as discussed above in ¶ 211, the monitoring system must optically split out the lambdas of interest then reconstitute streams of packets from those lambdas. This process results in two copies of the packets: one copy to the ISP or the international cable and one copy to the monitoring system. Dr. Schulzrinne discusses the use of a fiber-optic splitter in ¶ 55 of his declaration.

(2) Link-Layer Copying

278. A link-layer copying of packets can be done by a separate in-line device or by the ISP's router, using for example the router's mirroring function. Dr. Schulzrinne discusses using a router's mirroring function to copy packets in ¶ 58 of his declaration. The use of either a separate copying device or the mirror function in the ISP's router results in all the packets on the circuit being copied and forwarded to the monitoring system.

(3) Filtering the packets

279. The box marked "F" in Figure 33 represents a filtering function in the monitoring system. This filter function can be used to implement the IP address filter

that accepts only Internet transactions that are not wholly domestic, as described in the above PCLOB extract (see ¶ 265) and described below under Stage 2 (see ¶¶ 290-300). The filter function could also be used to implement more extensive filtering.

ii. In-Line Filter

280. In the in-line filter configuration, all the packets on a circuit being monitored by the NSA are sent through an in-line device configured to copy only those packets that meet a set of criteria. This configuration, which is described in ¶ 57 of Dr. Schulzrinne's declaration, is shown in the following figure:




**Figure 36** — In-line filter

281. The box marked "F" in Figure 34 represents the filter that (a) copies the subset of the packets on the circuit that meet the filter criteria and (b) sends them on for further processing. Dr. Schulzrinne notes in ¶ 60 of his declaration that the mirroring function in many ISP routers can be configured to perform this filtering function by selectively copying packets based, for example, on access control lists that are configured to use the IP addresses or port numbers in packets.

iii. Implementation

282. For a number of reasons explained below, I consider it most likely that the NSA is using the copy-then-filter configuration implemented using fiber-optic splitters or using link-layer copying. I consider it less likely that the NSA is using an in-line filter and very unlikely that the NSA would be using an in-line filter with sensitive or complex filtering criteria such as those described as possibilities by Dr. Schulzrinne.

283. The copy-then-filter configuration is the easiest configuration for both the NSA and the ISP to implement and operate. This configuration requires no or minimal support from the ISP or its personnel and leaves the NSA in full control of the upstream collection process. All the ISP has to do is to hand the NSA copies of all of the packets on a circuit, which is very easy to do using the router mirroring function, or a portion of the light on a fiber, which is very easy to do with a fiber-optic splitter. Thus the ISP is not a party to any proprietary information other than the basic fact that monitoring is being done at a particular location.

284. In contrast, the in-line filter configuration would require either that the ISP agree to place an NSA-operated device into the heart of its network—unlikely because of the potential impact on the ISP's network in the event of an equipment failure or misconfiguration—or that the ISP's personnel have enough knowledge of the filter criteria to configure the ISP's router.

285. Under Section 702, the NSA can compel an ISP to provide assistance to the NSA as part of upstream collection.[80] Thus, the NSA could compel an ISP to configure its routers to provide the in-line filter functionality. But, compelling an ISP to conduct complex in-line filtering on the ISP's routers would require that ISP personnel know what the NSA's filter criteria were. This would not be a real issue if the filter criteria were not sensitive—for example, if the criteria merely excluded packets with U.S. source and destination IP addresses. But if the filter criteria were more selective, as

[80] 50 U.S.C. § 1881a(i)(1)(A) (Attorney General and Director of National Intelligence may direct that providers "*immediately provide the Government with all information, facilities, or assistance necessary to accomplish the acquisition in a manner that will protect the secrecy of the acquisition and produce a minimum of interference with the services that such electronic communication service provider is providing to the target of the acquisition*").

postulated by Dr. Schulzrinne in ¶ 60 of his declaration, then the ISP personnel would have access to sensitive or classified filtering criteria the NSA uses in its upstream collection process, which I believe the NSA would want to minimize to the greatest degree possible.

286. I do know from personal experience that some parts of the U.S. government consider network device configuration details to be secret. When I was involved in the U.S. government Trusted Internet Connections (TIC) Program as a consultant, I was told that the configurations for the EINSTEIN filtering devices were considered secret because they could disclose what the government knew about cyber attackers.

287. Because of the sensitivity of the filter criteria, I consider it most likely that, if the NSA relies on in-line filters operated by an ISP, the filter criteria would not include blacklisting or whitelisting of individual IP addresses or rejection of individual ports such as 443, because if that information were to ever get out it would provide a roadmap for people who wanted to avoid NSA upstream collection. Note that complex filtering could easily be done using the copy-then-filter configuration, which would not require ISP personnel to have access to the NSA's filtering criteria, because the filter itself would be operated by the NSA.

288. Dr. Schulzrinne suggests that the in-line filter configuration is "*desirable from the perspective of reducing the volume of communications that must be processed (electronically reviewed) to identify the communications of interest*," see Schulzrinne Decl. ¶ 57, but he overstates that benefit. Modern deep packet inspection devices individually or operating in parallel, can process or review Internet communications at

the same rate that those communications traverse high-bandwidth Internet links. In addition, adding even a "simple" IP address-based filter to an ISP's router in order to exclude wholly domestic transactions would require adding tens of thousands of lines to the router's configuration and would place potentially significant additional demands on the router's processing power which could affect the performance of the router and create a risk of overloading the router, thereby interfering with the ISP's ability to support its customers' traffic.

289. In my opinion, the copy-then-filter configuration gives the NSA the greatest operational control and confidentiality in carrying out upstream collection with the least risk of interference with the ISP's ordinary network operations. For these reasons, I consider it more likely that a copy-then-filter implementation is used rather than the in-line filtering that Dr. Schulzrinne hypothesizes. But if an in-line filter is used, in my opinion the filter is almost certainly a simple one as discussed in the next section. (See ¶ 298.) In either case, packets are copied, whether before the filter or by the filter.

**b. Stage 2: Filtering**

290. The publicly released documents show that the NSA uses IP address filters to eliminate wholly domestic transactions prior to scanning for selectors, though, as explained below, the documents indicate that the NSA may not filter packets by IP address on certain international Internet circuits it is monitoring. The PCLOB extract in ¶ 265 notes that the "*Internet transactions are first filtered to eliminate potential domestic transactions.*"

291. The publicly released NSA documents reveal, however, that not all Internet transactions are filtered to eliminate wholly domestic communications before

being reviewed for the presence of selectors. For example, the NSA's 2014 targeting procedures says:

> *In addition, in those cases where NSA seeks to acquire communications about the target that are not to or from the target, NSA will either employ an Internet Protocol filter to ensure that the person from whom it seeks to obtain foreign intelligence information is located overseas, or [redacted] In either event, NSA will direct surveillance at a party to the communication reasonably believed to be outside the United States.*[81]

292. The above passage may explain why a wholly domestic "about" transaction could be acquired if the transaction was routed through an international link. Such routing of wholly domestic communications over international circuits does occasionally happen.[82] This situation is described in the following passage from the Foreign Intelligence Surveillance Court's October 3, 2011 opinion describing the operation of the NSA's upstream collection program:

> *the government readily concedes that NSA will acquire a wholly domestic 'about' communication* ***if the transaction containing the communication is routed through an international Internet link being monitored by the NSA*** *or is routed through a foreign server.*[83]

293. This passage indicates that the NSA does not use IP filtering at least on some of the international circuits it is monitoring. This is unsurprising because, by definition, the packets on international circuits are destined for or come from (or both)

[81] Appendix T at 2 (NSA Section 702 Targeting Procedure (2014), at 2).

[82] See, e.g., Shaun Waterman, *Internet Traffic Was Routed Via Chinese Servers*, Wash. Times (Nov. 15, 2010), https://www.washingtontimes.com/news/2010/nov/15/internet-traffic-was-routed-via-chinese-servers.

[83] Appendix P at 45 (FISC Opinion (Oct. 3, 2011)) (emphasis added) (citing the government's June 1, 2011 FISC Submission at 29).

non-U.S. locations and thus cannot have U.S. IP addresses as both source and destination addresses except in the case of a routing abnormality.

294. If the NSA were passing all transactions through an IP filter to eliminate wholly domestic transactions before copying, reassembly and review for selectors, then the NSA would never collect a transaction between U.S. IP addresses. That is because the NSA cannot review transactions for selectors, and therefore potentially collect them, without copying the packets and reassembling them into transactions first. (See ¶¶ 301-0.) Since the NSA admits to collecting wholly domestic "about" transactions from international links, it must not be applying an IP address filter in at least those cases. In addition, since the NSA admits it "*will acquire*" wholly domestic transactions from at least some international links, the NSA must be copying, reassembling and reviewing *all* the transactions on those links—otherwise the NSA would not see the selectors in the wholly domestic transactions and would not be collecting them. This is true for upstream collection of communications "to" and "from" the NSA's targets, not just collection of communications "about" its targets.

295. Where the NSA uses an actual IP address filter, it has further described the filtering mechanism as follows:

> *NSA Defendants respond that to their understanding the term "filtering mechanism," as used in the above-referenced brief when filed, meant, in unclassified terms, the devices utilized in the upstream Internet collection process that were designed to eliminate wholly domestic Internet transactions, and transactions that did not contain at least one tasked selector, before they could be ingested into Government databases. Today the term "filtering mechanism" would mean, in unclassified terms, the devices utilized in the Upstream Internet collection process that are*

> *designed to eliminate wholly domestic Internet transactions, and to identify for acquisition Internet transactions to or from persons targeted in accordance with the current NSA targeting procedures.*[84]

296. Most references to the filter function in NSA documents refer to an ***IP filter*** or ***Internet protocol address filter***. An IP filter is a device that can filter Internet packets based on information available in the IP header. The IP header includes a variety of data, but most importantly, it contains the source and destination IP addresses of the packet. (See ¶¶ 96-101.) There are a few places where the NSA refers to its upstream collection filter function as an ***IP address filter***.[85] I believe that the "IP filter" referred to in the other documents is an IP address filter because of these citations and also because the only way that an IP filter could be used to eliminate potential domestic transactions would be to filter based on IP addresses. As discussed above in ¶¶ 229-230, as a general rule, ranges of IP addresses are assigned to ISPs or, through ISPs to their customers in such a way that an individual IP address can be geographically located to a reasonable degree of accuracy. The accuracy is not perfect since blocks of IP addresses are reassigned to different networks in different locations, including in different countries, from time to time. The frequency of these changes has increased significantly in the last few years because of the commercial market for the right to use IPv4 addresses, which I discuss above in ¶ 160. This may be what the NSA is referring to when it says "*[b]ecause NSA's filters will be looking at the best available information*."[86]

---

[84] Appendix D at 7-8 (NSA Response to Plaintiff's Interrogatory No. 3 (Dec. 22, 2017)).

[85] Appendix U at 24 (FISC Hearing Transcript, *In Re: DNI/AG 702(g) Certification* [*Redacted*] (2008)); Appendix C at 32, 37 (FISC Submission (June 1, 2011)).

[86] See Appendix C at 11 (FISC Submission (June 1, 2011)).

297. Thus, the source and destination IP addresses in each individual packet can be checked to see that both of them are from ranges of IP addresses assigned to a network inside the U.S. Using an IP **address** filter provides the function the NSA described for the IP filter:

> *NSA is required to use other technical means, such as Internet protocol ("IP") filters, to help ensure that at least one end of an acquired Internet transaction is located outside the United States.*[87]

298. Note that even a "simple" filter configured to just reject wholly domestic transactions by using an IP address-based filter is no easy task. There are over 66,000 entries in one of the lists of U.S. address blocks. (See ¶ 229.) Adding and maintaining that many entries to a production router's configuration is a significant task and would have a significant chance of adversely impacting the router's performance.

299. As Dr. Schulzrinne points out in ¶¶ 60-64 of his declaration: In general, such a filter could also be configured to perform other checks such as rejecting any packets transporting protocols that an entity is not interested in, or the reverse, accepting any packets transporting protocols the entity is interested in. The filter could also be configured to reject packets destined to or from particular network addresses an entity might not want to monitor. It should be noted that the more complex the filtering configuration, the more effort is required to keep the filter configurations up to date. As discussed above in ¶¶ 285-289 doing any filtering other than simple U.S. vs. non-U.S. addresses would likely have to be managed by NSA personnel on an NSA operated device or by ISP personnel with security clearances.

[87] Appendix F at 43 (PCLOB Report at 38).

300. To state the obvious, filtering out packets at this stage would eliminate the NSA's ability to collect the Internet communications to which those packets belong, and would thus foreclose its ability under this program to collect and analyze any foreign intelligence information those communications contain.

**c. Stage 3: Reassembling Transactions**

301. The next step is to reassemble the packets that make up individual communications so that they can be reviewed using DPI for the presence of selectors. As computer researchers Shuhui Chen and Yong Tang put it, "*Stream Reassembly is an indispensable function of Deep Packet Inspection*."[88] What Chen and Tang call a "stream" is another name for what the NSA calls "transactions." (See ¶¶ 63-65.)

302. Transaction reassembly is required before the DPI device can review for selectors because: (1) the packets that make up a particular transaction are intermingled with packets from other transactions (see ¶ 38) and must be isolated from the other packets by selecting the packets with the same source and destination address and ports and the same protocol value (the 5-tuple) and adding them to an assembly buffer[89] (see ¶ 113), (2) the packets may also have to be reordered to be in the right sequence (see ¶ 114), and finally, (3) the selectors that the NSA's reviewing devices look for may be split between the packets that make up the transaction.

[88] Appendix V (Shuihui Chen & Yong Tang, *A Stream Reassembly Mechanism Based on DPI*, Inst. of Electrical & Electronics Engineers (2012)).

[89] By ***assembly buffer***, I mean a temporary storage place in the collection device's memory

303. There are DPI designs that can review for keys such as the NSA's selectors without reassembling the streams (transactions),[90] but since the NSA does need the reassembled transactions to be able to store any with selectors in its databases, transaction reassembly is required even if the reviewing process itself does not need to work on reassembled transactions.

304. The reassembly process is shown in the following figure:



**Figure 37** — Reassembling transactions

305. The figure above shows the packets that were passed by the filter being reassembled into Internet transactions. Each transaction comprises all of the packets related to a particular communication (i.e., that have the same 5-tuple) that pass by the monitoring point.

306. The assembly needs to continue until there is an indication that the Internet transaction is complete or there has been some period during which no new packets with a matching 5-tuple have been received.

307. Since the Internet does not guarantee that the order of packets will be maintained during their journey through the network, packets in the buffer may have to be swapped around so that the packets making up the transaction are in the right order. This is required so that any selector that extends across a packet boundary will be made whole for the reviewing process (see below) and be properly recognized.

[90] See, e.g., Appendix W (U.S. Patent No. 8,813,221).

308. Because the paths taken by successive packets as they travel through the network may occasionally change, there is no guarantee that all the packets that make up an Internet transaction will pass by any particular monitoring point. (See ¶ 194.) This will result in some incomplete Internet transactions being assembled. An incomplete Internet transaction might not have a complete selector and thus be missed in the collection process. Conversely, even incomplete Internet transactions may contain complete selectors, and those transactions would thus be collected.

309. Also, because of asymmetric routing paths, the packets that make up the Internet transaction in each direction of the bidirectional exchange of packets that make up most Internet communications (see ¶ 111) will generally not pass through the same monitoring point. (See ¶¶ 197-198.) In those cases where a selector appears in both directions of a communication and where the packets in each direction pass through NSA monitoring points, the upstream collection process will result in both Internet transactions being collected, and they can later be associated during the analysis process. But, it would not be common for some types of selectors, such as a source email address, to be present in both directions of an Internet transaction; normally it would only appear in one direction. The effect of asymmetric routing is one more reason that it is likely the NSA has multiple monitoring points.

### d. Stage 4: Reviewing Transactions

310. The Internet transactions that have been reassembled from packets that passed the NSA's IP address filter then need to be reviewed for the presence of selectors. This stage is shown in the following figure:

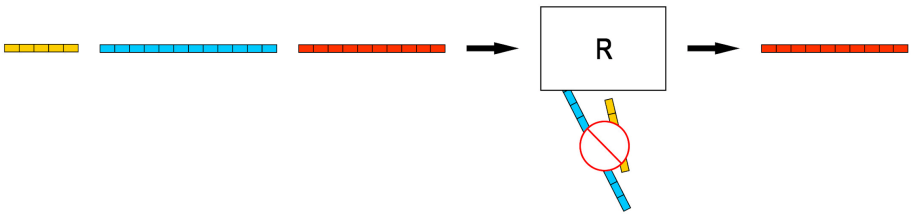



**Figure 38** — Reviewing transactions for selectors

311. The above figure shows a series of reassembled transactions being sent to a reviewing device (the box marked "R") and only the transactions containing selectors exiting the device. The remaining transactions are discarded.

312. As discussed above in ¶¶ 236-240, the selectors the NSA uses in the upstream collection program include "*electronic communication accounts/addresses/identifiers*".[91] The examples provided in the NSA documents are account identifiers, telephone numbers and email addresses. These are carried in the application layer in Internet communications. (See ¶ 60.)

313. For example, email addresses are carried in the SMTP handshake (see ¶¶ 134-137), in email headers (see ¶¶ 126-128), in IMAP (see ¶¶ 138-140), and in HTTP/S (see ¶¶ 117-123) when HTTP is the user's interface to an email server (see ¶¶ 129-133). In all of these cases, the email addresses are carried in the application layer of an Internet communication. Email addresses are also sometimes carried in the body of

[91] Appendix M at 1 (FISC Submission (May 2, 2011)).

Internet communications, which is also in the application layer, such as the body of email messages or in the contents of web pages. Telephone numbers in voice over IP are carried in the SIP headers (see ¶ 141), which are also in the application layer. Thus, the NSA must be reviewing the application layer of Internet communications if it is looking for these types of selectors within the communications.

314. In her deposition, Rebecca J. Richards acknowledged that the NSA, at least in 2015, did review the application layer of Internet communications.[92] Following an order from her counsel, she refused to answer the same question about upstream collection today because she said that the answer would be classified.[93] It is strange that the NSA considers classified the answer to the question of whether upstream collection today involves reviewing the application layer of communications. There is no question that it must involve that sort of reviewing, because the email addresses and other user identifiers in Internet communications are transported in the application layer. The NSA has acknowledged using "*NSA-designed upstream Internet collection devices*" in the collection process.[94] The NSA has also acknowledged reviewing web traffic:

> *Results were reviewed for three randomly selected days in April, averaged to produce an estimated figure of collection of [redacted] for the month of April. This figure was then compared to the total take of Section 702 upstream collection of web activity for the month. From this sample NSA estimates that approximately 9% of the monthly Section 702 upstream collection of [redacted].*[95]

[92] Appendix K at 263:11-18 (Richards Depo.).

[93] Appendix K at 266:4-13 (Richards Depo.).

[94] Appendix F at 44 (PCLOB Report at 39).

[95] Appendix C at 30 (FISC Submission (June 1, 2011)).

315. ***Web communications*** are the communications carried by HTTP or HTTPS. (See ¶¶ 117-123.) Thus, since the NSA was comparing the amount of collection of a particular redacted type of communication against the amount of collection of "web activity" to get a percentage, they must have been comparing the amount of web (HTTP/S) collection.

i. "multiple communications transaction (MCT)" collection

316. The PCLOB Report described MCT collection as follows:

> *An MCT is an Internet "transaction" that contains more than one discrete communication within it. If one of the communications within an MCT is to, from, or "about" a tasked selector, and if one end of the transaction is foreign, the NSA will acquire the entire MCT through upstream collection, including other discrete communications within the MCT that do not contain the selector.*[96]

317. An example of this type of MCT is the burst of email messages downloaded to a mail user agent when a user reconnects to a mail server after being disconnected for a while. (See ¶ 132.) Under the upstream collection program, the NSA would collect an MCT comprised of multiple email messages if any of the email messages in the burst is from a target outside the U.S. to someone inside the U.S. It might be that only one of the email messages is from the target and ten more are from sources within the U.S., but the entire MCT would be collected.

318. In order to discover that an MCT includes an email message that is from a target, the NSA must be reviewing the entire transaction. This is because each email

[96] Appendix F at 12 (PCLOB Report at 7).

within a burst of email messages has its own header information (e.g., "To:" and "From:" addresses). (See ¶¶ 126-128.)

319. MCT collection is controversial because it can involve the capture of wholly domestic communications, which is generally not authorized under upstream collection. It can also involve the capture of international communications that are not to, from, or about a targeted selector, which again is not generally authorized under upstream collection. But the NSA says that it does not have the technology to separate out the collectable from the non-collectable communications in MCTs.[97]

320. In its April 2017 Order, the FISA Court restricted the NSA to collecting MCTs only "*when the target is a party to the entire MCT*." For example, when the target identified by the selector is in the "To" field of each of the email messages in the MCT.[98]

ii. "about" collection

321. Until April 2017, the upstream collection program collected transactions where selectors appeared **anywhere** in a transaction, not just in the sender or receiver fields of the transaction. For example, upstream collection would collect an email if it contained a selector inside the email message's "body" text. The NSA's "about" collection shows that the NSA was scanning the entirety of each of the reassembled transactions for selectors, likely with the same DPI device that was used to review for other selectors, not just the application headers. (See the discussion above about MCTs.) Prior to April 2017, this scanning led to the ingestion of Internet transactions that were

[97] Appendix F at 45 (PCLOB Report at 40).

[98] Appendix E at 26 (FISC Opinion (Apr. 26, 2017)).

"about" a target in addition to transactions sent by or addressed to a target. The PCLOB Report described "about" collection as follows:

> *An "about" communication is one in which the selector of a targeted person (such as that person's email address) is contained within the communication but the targeted person is not necessarily a participant in the communication. Rather than being "to" or "from" the selector that has been tasked, the communication may contain the selector in the body of the communication, and thus be "about" the selector.*[99]

322. This procedure was controversial because it involved the warrantless reviewing of the contents of Americans' communications and because it involved the collection of many wholly domestic communications where both the sender and receiver of the message were within the U.S. After an extensive review, apparently prompted by the findings of the Foreign Intelligence Surveillance Court that the NSA had not complied with certain procedures related to the upstream collection program, the NSA decided to stop the "about" collection and destroy most of the transactions that had been collected under the "about" collection process.[100]

323. The NSA has not said that it stopped reviewing the entire contents of transactions when it stopped the "about" collection. As mentioned above in ¶¶ 258-259, about collection likely used the same DPI devices that were used to look for communications *to* or *from* a selector, which the NSA still needs to do. "About" collection merely involved retaining transactions with selectors located in parts of a transaction other than in the application headers.

[99] Appendix F at 12 (PCLOB Report at 7).

[100] Appendix X (NSA Press Releases (Apr. 28, 2017)).

324. The recent extension of Section 702 permits the NSA to resume "about" collection under the upstream collection program if it gives proper notice before doing so.[101]

iii. Collection of Encrypted Internet Transactions

325. Under Section 702, the NSA is authorized to collect encrypted Internet transactions and to retain them for an extended period so they can attempt to decrypt them[102]. An HTTPS transaction is an example of an encrypted Internet transaction. In theory, the NSA could configure its IP filters to reject HTTPS traffic by rejecting packets with a destination or source TCP port of 443 but, during her deposition, Rebecca J. Richards followed her lawyer's order to not say if the NSA had done so.[103]

326. In fact, there are obvious reasons that the NSA would seek to collect traffic on port 443 even though it is encrypted.

a. The NSA may, currently or in the future, be able to decrypt important encrypted messages. It is this possibility that justifies the NSA's retention of encrypted communications longer than it is permitted to keep unencrypted communications.[104]

b. For example, the NSA could have compromised the end systems generating or receiving the HTTPS traffic and thus have obtained the keys to permit the transaction to be decrypted. (See ¶ 121.)

[101] FISA Amendments Act Reauthorization Act of 2017, Pub. L. No. 115-118, § 103(b).

[102] *See e.g.*, Appendix F at 65, 68 (PCLOB Report at 60, 63); Appendix S at 10 (NSA Section 702 Minimization Procedures (2014)).

[103] Appendix K at 280:13-281:11 (Richards Depo.).

[104] See e.g., Appendix F at 65, 68 (PCLOB Report at 60, 63); Appendix S at 10 (NSA Section 702 Minimization Procedures (2014)).

c. Even if the NSA is not able to decrypt all HTTPS traffic, there is nonetheless useful information that can be obtained from HTTPS transactions including the IP addresses of the Internet user and of the web server. In addition, as discussed in ¶ 123, the domain name of the web server (e.g., www.government.ru) is disclosed in the setup phase of an HTTPS session. In short, even if encrypted, HTTPS communications can reveal who a target is communicating with or which Internet domains he or she is visiting.

d. In addition, as noted in ¶ 109, port numbers are not always a perfect indicator of what application protocol is being used because port numbers can be changed as long as both ends of a communication agree on what port numbers to use. Because of this it is not uncommon for applications to use port 443, the port number assigned for HTTPS, for other uses just to bypass security filters blocking packets using unknown or unwanted ports. Ignoring HTTPS traffic would thus entail ignoring many other types of communications that also use port 443.

e. Finally, HTTPS is one of the most common application-layer protocols used to transmit Internet communications around the world today. Ignoring HTTPS traffic would create a large and needless blind spot.

327. For at least the above reasons, it is very likely that the NSA is reviewing HTTPS transactions whose constituent packets meet the origin or destination criteria for review under upstream collection.

328. Many of the above reasons for collection of HTTPS communications also apply to collecting other forms of encrypted communications, such as communications in

VPNs. For at least these reasons, the NSA would have an incentive to collect encrypted communications of all types.

**e. Stage 5: Ingesting Transactions**

329. The Internet transactions that pass the reviewing stage are then ingested into the NSA's Section 702 databases. (See ¶¶ 231-235.) The following figure shows this stage:

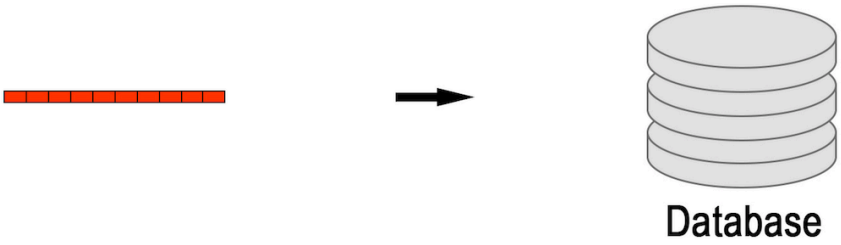



**Figure 39** — Ingesting transactions that contain one or more selectors

330. The above figure shows an Internet transaction, in which the reviewing mechanism found one or more selectors, being ingested into the NSA's Section 702 database.

***3. Upstream Collection Monitor Placement***

331. The NSA has admitted that the upstream collection program collects information from the Internet backbone, and that the Internet backbone consists of high-speed network links between and within ISPs, including terrestrial and undersea fiber cables. (See ¶¶ 150-153.)

332. Since the upstream collection program is limited to collecting Internet transactions where at least one end is outside of the U.S., the logical places to locate upstream collection systems would be at the U.S. end of circuits carried on the undersea and other fiber cables that go between the U.S. and other countries. The FISC, citing the

government's submissions, has confirmed that at least some of the upstream collection program occurs at these points. It has described upstream collection of transactions routed through "*an international Internet link being monitored by the NSA*."[105] As discussed above in ¶¶ 222-223, these locations are also logical places for the NSA to collect communications where both ends are outside the U.S. This non-U.S. collection is feasible because so much of the world's Internet traffic flows through the U.S. (See ¶¶ 222-228.)

333. The NSA has not provided any public information on what percent of the total international public Internet capacity is covered by the upstream collection program, but the government has repeatedly stated that the intention of the upstream collection program is "*to comprehensively acquire communications that are sent to or from its targets*"[106] as long as the communications are not wholly domestic.[107] The NSA refers to the Internet communications it acquires through upstream collection as "*transactions*." (See ¶¶ 63-65.) In order to comprehensively acquire its targets' transactions, the NSA must be comprehensively reviewing Internet transactions to see if they are transactions to or from NSA targets, since the NSA cannot know in advance which of the many transactions on the Internet could be to or from one of the NSA's targets. In order to comprehensively review Internet transactions, the NSA must be comprehensively monitoring the places on the Internet where the non-wholly domestic transactions to or from its targets will transit. If the NSA is not comprehensive in where it does

[105] Appendix P at 45 (FISC Opinion (Oct. 3, 2011)).

[106] Appendix F at 15, 128 (PCLOB Report at 10, 123); see also id. at 148 (PCLOB Report at 143).

[107] Appendix F at 148 (PCLOB Report at 143) ("*[T]he NSA takes additional measures, including the use of IP filters, to try to avoid collecting wholly domestic communications.*").

monitoring, then it cannot be comprehensive in its collection of the transactions to or from its targets. The places where non-wholly domestic transactions to or from its targets will transit include the U.S. ends of the Internet backbone circuits transporting transactions between the U.S. and other countries. (See ¶¶ 200-211.)

334. The NSA has disclosed that it has over 120,000 Section 702 targets, all of them located abroad.[108] The paths that transactions will take between those targets and correspondents in the U.S. are controlled by Internet routing protocols. (See ¶¶ 175-180.) Because of this, in general, the packets that make up these Internet transactions will take the shortest path between the sender and receiver. Using the shortest path will mean that the packets sent by a target located outside the U.S. to a site within the U.S. will generally traverse the topologically closest international link that supports public Internet traffic between the sender's location and the U.S. With thousands of targets in different places around the globe, a wide distribution of international circuits will be used by Internet transactions sent and received by the NSA's targets. In addition, people, including the NSA's targets, move around from time to time, and such movement may change which international circuits their communications use. Thus, the number, distribution and movement of the NSA's targets means that the NSA needs to monitor communications carried by most, if not all, such circuits carried on international cables if it wants to ensure that it captures the communications of those targets.

335. Moreover, regardless of which circuits it monitors, the NSA must also be comprehensive in its monitoring of each circuit. That is, if the NSA's goal is to comprehensively obtain its targets' communications, then it must comprehensively copy,

[108] Appendix Y at 14 (ODNI Statistical Transparency Report for 2017 (Apr. 2018)).

reassemble and review all transactions that could conceivably be to or from a target that transit the circuits being monitored. Since all transactions transiting the monitoring points other than the ones that are wholly domestic could be to or from a target, the NSA must be copying, reassembling and reviewing all, or essentially all, international transactions that transit the circuits being monitored.

## VIII. OPINION D: WIKIMEDIA COMMUNICATIONS ARE TRANSPORTED ON ALL INTERNATIONAL CIRCUITS ORIGINATING OR TERMINATING IN THE UNITED STATES.

336. Wikimedia operates servers in multiple countries to optimize the user experience in different regions of the world. This case concerns the international traffic to and from Wikimedia's U.S.-based servers or users, including the communications between Wikimedia's users outside the U.S. and Wikimedia's U.S.-based servers, the traffic between Wikimedia's non-U.S. servers and its U.S-based users, and the international communications of Wikimedia's staff originating in or terminating in the U.S.

337. Comparing the geographic distribution of international undersea and terrestrial cables, which are used to carry public Internet traffic (which I discussed above in ¶¶ 200-204), with the geographic distribution of countries from which users access Wikimedia's U.S.-based servers (which I discuss below in ¶¶ 341-350) makes it clear that communications to and from Wikimedia's U.S.-based servers are carried on all of the circuits transporting public Internet traffic in the cables connecting the U.S. to other countries.

338. **Opinion D:** Thus, it is my opinion that it is virtually certain that Wikimedia's international communications traverse every circuit carrying public Internet traffic on every international cable connecting the U.S. to other countries.

### A. Wikimedia

339. Wikimedia Foundation is a non-profit organization based in San Francisco, California, that operates twelve free-knowledge projects on the internet, including Wikipedia, Wiktionary, Wikinews, Wikibooks, and Wikisource. Wikipedia is one of the top ten most-visited websites in the world.[109] Wikimedia describes its mission as to empower people around the world to collect and develop free educational content. Wikimedia does this by developing and maintaining "wiki"-based projects, and by providing the full contents of those projects to individuals around the world free of charge.

340. This case involves Wikimedia's international Internet communications, described more fully below.

#### *1. Wikimedia Websites*

341. People all over the world make use of Wikimedia websites. Most users access the websites in order to get information about some topic. For example, Wikipedia is an online free encyclopedia, Wiktionary is an online dictionary, Wikinews is an online news site, Wikibooks is an online repository with open-content textbooks, and Wikisource is an online free library. All of these sites, and seven more, are capable of supporting people around the world in their native languages. For example, as of January 2018, Wikimedia projects supported web pages in 288 languages.[110]

342. In addition, many people around the world volunteer as content producers and editors for Wikimedia services.

[109] *The Top 500 Sites on the Web*, Alexa, https://www.alexa.com/topsites.
[110] Appendix Z at 29 (Wikimedia Responses to Defendants' Interrogatories (Jan. 11, 2018)).

### *2. Wikimedia International Communications*

343. Wikimedia operates servers in multiple countries to optimize the user experience in different regions of the world.

344. For purposes of my analysis below, I focus on Wikimedia's web activity, but my conclusions apply to Wikimedia's communications in total. This case concerns three categories of Wikimedia's international communications:

a. Wikimedia's international communications with its community members, which consist principally of the traffic between Wikimedia's users outside the U.S. and its U.S.-based servers, as well as traffic between Wikimedia's U.S.-based users and its Amsterdam-based servers;

b. communications log information sent from Wikimedia's Amsterdam-based servers to its U.S.-based servers[111]; and

c. international communications of Wikimedia's staff that originate in or terminate in the U.S.

345. Wikimedia has maintained servers in the U.S. in the following locations: Ashburn, Virginia; Carrollton, Texas; Chicago, Illinois; Dallas, Texas; San Francisco, California; and Tampa, Florida.[112]

346. For the six-month period between August 1, 2017 and January 31, 2018, Wikimedia engaged in approximately 760 billion international communications.[113] To put the volume of Wikimedia's Internet traffic in comparative perspective, it operates one

[111] According to Wikimedia's discovery responses, "*Every time Wikimedia receives an HTTP/S request from a person accessing a Wikimedia Project webpage, it creates a corresponding log entry*." Appendix AA at 19 (Wikimedia's Second Amended Responses to Defendants' Interrogatories (Apr. 17, 2018)).

[112] Appendix Z at 26-27 (Wikimedia Responses to Defendants' Interrogatories (Jan. 11, 2018)).

[113] Appendix BB Ex. 1 (Wikimedia Response to ODNI Interrogatory No. 19 (Apr. 6, 2018)).

of the top ten most-visited websites in the world, alongside Google.com, Youtube.com, Facebook.com, and Baidu.com.[114]

347. Not only is the volume of Wikimedia's communications immense, but its millions of users are widely dispersed around the globe. For example, Internet users in every country accessed Wikimedia's U.S.-based servers between August 1, 2017 and January 31, 2018. During that time period, Internet users outside the U.S. made over 380 billion web requests to Wikimedia's servers inside the U.S., and Wikimedia's servers sent over 380 billion responses to those requests. See Appendix BB[115] and the map below:

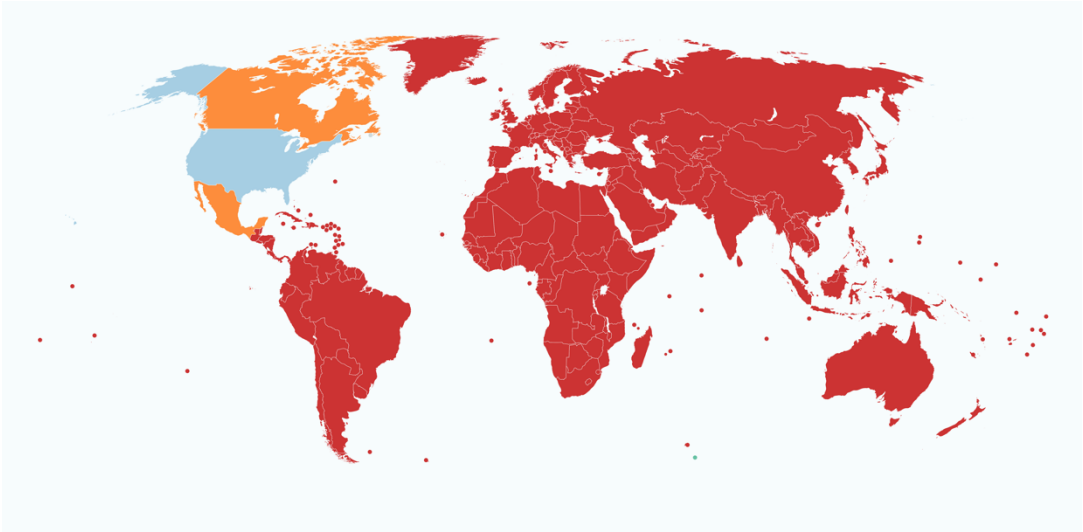


**Figure 40** — Countries from which Wikimedia's U.S. servers received web requests

348. Wikimedia's U.S.-based servers receive web requests via circuits in undersea cables from all the countries colored in red. The websites also receive web requests via terrestrial as well as undersea and under-lake circuits from Canada and Mexico, shown in orange. In summary, Wikimedia U.S.-based servers receive web

[114] *The Top 500 Sites on the Web*, Alexa, https://www.alexa.com/topsites.

[115] Appendix BB (Wikimedia Response to ODNI Interrogatory No. 19 (Apr. 6, 2018)).

requests from all of the world's inhabited continents and islands. Thus, Wikimedia users are very widespread. To provide some context, Wikimedia's U.S.-based servers receive hundreds of billions of requests annually and provide at least as many responses. Even with a large number of international circuits, there are very many Wikimedia communications on each circuit. For example, even if there are thousands of international circuits, there would still be hundreds of millions of Wikimedia communications on the average circuit.

349. The paths that Internet communications take between Wikimedia users outside the U.S. and Wikimedia servers in the U.S. are controlled by Internet routing protocols. (See ¶¶ 175-180.) Because of this, in general, the packets that make up these Internet communications will take the shortest path between the sender and receiver. Using the shortest path will mean that the packets sent by a user located outside the U.S. to a server within the U.S. will generally traverse the topologically closest international circuit that supports public Internet traffic between the user's location and the U.S. With Wikimedia users located in all of the world's inhabited continents and islands, the widest possible distribution of international circuits will be used by Internet communications sent and received by the Wikimedia users.

350. Thus, it is my opinion that it is virtually certain that Wikimedia's international communications traverse every circuit carrying public Internet traffic on every international cable connecting the U.S. to other countries, including the "international Internet links" monitored by the NSA.

### *3. Protocol Support on Wikimedia Websites*

351. Wikimedia websites support both HTTP and HTTPS. Within Wikimedia's foreign-to-U.S. HTTP and HTTPS communications, the percentage of communications that use HTTPS had been growing and is now about 97.7% overall.[116] But there are a number of countries where the percentage is much lower. For example, 38% of Iranian communications with Wikimedia's U.S.-based servers use HTTP, as do 28% of Irish communications, 24% of Chinese communications, 19% of Dutch communications, and 16% of Finnish communications (all with Wikimedia's U.S.-based servers).[117] To provide context, Wikimedia's U.S.-based servers received over 8 billion HTTP requests from foreign users in the six months between August 1, 2017 and January 31, 2018.[118]

352. As discussed above in ¶¶ 122, 326, even encrypted Internet transactions can still reveal important information or can be saved for later attempts at decryption. In other words, just because a communication is encrypted does not mean that the NSA will not copy, scan or collect it.

## IX. OPINION E: THE NSA HAS COPIED, REASSEMBLED AND REVIEWED WIKIMEDIA COMMUNICATIONS

353. Based on my conclusions above in Opinions A–D, as well as the other features of upstream surveillance I've discussed, I conclude that: Even if the NSA were monitoring only a single circuit under upstream collection, it would be copying and

[116] Appendix AA (Wikimedia's Second Amended Responses to Defendants' Interrogatories (Apr. 17, 2018)).

[117] Id.

[118] Appendix BB, Exhibit 1

reviewing at least some of Wikimedia's communications. Moreover, while it is unnecessary to my conclusion here, the government's officially released documents indicate that the NSA is monitoring multiple circuits, which only increases my confidence that the NSA is copying and reviewing Wikimedia's communications. In fact, for the reasons discussed above in ¶¶ 332-333, the NSA is very likely to be monitoring a large number of international circuits, given that it would need to monitor **most, if not all,** such circuits to accomplish its stated (and unsurprising) goal of reliably and comprehensively collecting the communications of its targets.[119]

354. Moreover, the NSA's need to monitor most, if not all, communications carried by international circuits in order to comprehensively acquire its targets' communications makes it highly likely that the NSA is copying and reviewing some of Wikimedia's communications in each of its categories of international communications. (See ¶ 343.)

355. The NSA's monitoring of many circuits would only increase the volume of Wikimedia communications that the government is intercepting, copying and reviewing in the course of its upstream collection program.

356. **Opinion E:** Thus, it is my opinion that it is virtually certain that the NSA has, in the course of the upstream collection program, copied, reassembled and reviewed at least some of Wikimedia's communications.

[119] Appendix F at 15, 128, 148 (PCLOB Report at 10, 123, 143)

## X. DR. SCHULZRINNE'S DECLARATION

357. The government submitted a declaration by Dr. Henning Schulzrinne in support of its motion for summary judgment. Dr. Schulzrinne is a computer scientist at Columbia University. I have known him for many years having first met at the IETF. The government appears to have asked Dr. Schulzrinne to address a different question than Wikimedia's counsel asked me to address. Wikimedia's counsel asked me to address the likelihood that the NSA has, in the course of upstream collection, copied, reassembled or reviewed at least some of Wikimedia's communications. Dr. Schulzrinne's declaration does not address that question. He does not state any opinion about the likelihood that the NSA has copied, reassembled or reviewed Wikimedia's communications.

358. Nor does Dr. Schulzrinne mention many of the critical features of upstream collection on which I base my conclusion that it is a virtual certainty that the NSA has copied, reassembled or reviewed at least some of Wikimedia's communications.

359. For example, he does not address the number of targets of Section 702 surveillance that the government has acknowledged (over 120,000 as of April 2018); he does not acknowledge the NSA's stated goal of "*comprehensively acquir[ing] communications that are sent to or from its targets*";[120] he does not discuss the asymmetric routing of communications on the Internet; he does not mention the special permission the NSA has under Section 702 to collect and analyze encrypted communications; he does not acknowledge that useful information can be obtained from

[120] Appendix F at 15 (PCLOB Report at 10).

the scanning of encrypted Internet communications even if their content cannot be decrypted; and he does not acknowledge that the NSA has publicly conceded that it monitors "*web activity*."

360. In his declaration, Dr. Schulzrinne makes one point about how surveillance can be performed on the Internet, and one about how the NSA could avoid Wikimedia traffic.

361. In regards to the mechanisms of surveillance, Dr. Schulzrinne describes (as I also describe) that there are two configurations of equipment with which the NSA could be obtaining copies of the Internet communications it will review for selectors. Dr. Schulzrinne states that the second configuration (what I call an ***in-line filter***, see ¶¶ 279-281) "*would be desirable*."[121] I disagree with his conclusion. (See ¶¶ 288, 363-365.)

362. Second, Dr. Schulzrinne speculates that the NSA could, as a technical matter, filter out some types of communications so that its surveillance equipment would not copy, reassemble or review any of Wikimedia's communications. Dr. Schulzrinne's explanation is not entirely accurate as a technical matter, and it is simply implausible as a practical matter given everything that is known about upstream collection. (See ¶ 367.)

### A. Surveillance Configurations

363. Dr. Schulzrinne describes the same two surveillance configurations as I do. I referred to them as the ***copy-then-filter*** and the ***in-line filter*** configurations. (See ¶¶ 269-289.) Dr. Schulzrinne says that the in-line filter configuration would be desirable as compared to the copy-then-filter configuration because it would reduce the volume of communications that would need to be scanned. As I mentioned above in ¶ 288, I do not

[121] Schulzrinne Decl. ¶ 57.

think that reducing the volume of communications is all that important because modern DPI equipment can, either singularly or in parallel, keep up with the traffic in the type of channels the NSA is dealing with. To the extent that such an in-line filter would permit cheaper DPI equipment to be used, it might be desirable, but there are other important countervailing factors, as described below.

364. Dr. Schulzrinne describes the filtering being done using the mirror function in the ISP's existing routers. If that were the case, it would avoid the need for extra network equipment (the fiber-optic splitter) that would be required in the copy-then-filter configuration. But, as I discuss above in ¶ 287, if the filter function is implemented using the mirror function in the ISP's router, the filter functions would likely have to be limited to some non-secret set of filters such as the list of IP address ranges that are located in the U.S. Otherwise the ISP technician who configures the router, the router itself and the backup systems used to manage the router would be dealing with secret information (the filter criteria), which, if it were to be compromised, would give a roadmap on how to avoid NSA collection. The copy-then-filter configuration has the advantage that the filter device could be entirely under the control of the NSA and thus avoid the risk of the ISP personnel having access to potentially secret filter configurations.

365. In the copy-then-filter configuration, all Wikimedia traffic that transits a channel that the NSA is monitoring will be copied. In the in-line filter case, unless the filter was set to filter with a higher degree of selectiveness than checking to see if the IP addresses are in the U.S. or not then all international Wikimedia traffic that transits a

channel that the NSA is monitoring will be copied. In both cases all international Wikimedia traffic would be copied.

**B. Selectively Filtering Internet Traffic**

366. Dr. Schulzrinne spends considerable time discussing the possibility that the NSA could use selective filtering to avoid Wikimedia traffic. He describes using the traffic mirror function present in some ISP routers to blacklist or whitelist IP addresses or protocols.[122] While such filtering is technically possible, there are a number of reasons to conclude that Dr. Schulzrinne's hypotheticals are implausible and, accordingly, that it is implausible that the NSA is engaging in such filtering.

a. As discussed above in ¶¶ 285-289, having the mirror function in the ISP router do advanced selective filtering would mean that the configuration of the mirror function would include secret information, complicating the protection of such information.

b. Adding any protocol specific blocklist, for example not including any packets with port 443 (HTTPS) or protocol 50 (IP Sec) in reassembly and review, would create a blind spot that would provide a path by which an NSA target could communicate without the communications being detected. Sophisticated targets could easily probe to find any such blind spots and exploit them.

c. As discussed in ¶ 288, there is no particular reason to think that selective filtering is needed to reduce the load on the DPI devices. In any case, while the total number of Wikimedia's mostly text-based communications is immense, the total amount of those communications in bytes is minuscule as

[122] Id. ¶¶ 63-71.

compared to YouTube's video-based traffic. If filtering traffic for performance reasons were desirable, the NSA would get much more result from filtering YouTube than from filtering Wikimedia.

d. Dr. Schulzrinne mentions using ***whitelists*** (lists of addresses the NSA is interested in) rather than ***blacklists*** (lists of addresses the NSA wants to ignore).[123] As a practical matter, whitelists are almost useless for the type of collection program the NSA is running. Whitelisting requires knowing in advance all of the IP addresses that might be used by each of the NSA's targets as well as assuming that those targets are not moving around and thereby changing their IP addresses. This is not remotely possible. (See ¶¶ 137, 140, 173-174, 229-230, 244-247, 334.)

e. Dr. Schulzrinne suggests selectively filtering applications, for example by using the port number in the transport header.[124] As I discuss in ¶ 109, the use of a particular port number does not mean that a particular application is being used. Port numbers are only advisory in that pairs of Internet devices can decide what application they want to run on a port—for example, running email using port 80 to avoid firewalls. If the NSA were blacklisting traffic using specific ports, it would provide another path that NSA targets could use to avoid collection.

[123] Id. ¶¶ 65-66.

[124] Id. ¶¶ 70-71.

f. One example application Dr. Schulzrinne suggests could be blacklisted is the world wide web (ports 80 and 443).[125] Doing so would leave a very large hole in the NSA's collection ability. The hole would include web email, web chat, web-based editors which have been used to send hidden messages, ISIS videos and the like. In addition, the NSA acknowledges collecting web traffic.[126] (See ¶¶ 314-315.)

g. Dr. Schulzrinne specifically suggests blacklisting HTTPS (port 443). As mentioned just above, the fact that a communication uses port 443 does not mean that the communication is actually HTTPS or even that the communication is encrypted. In addition, as I discuss above in ¶ 326, there are many obvious reasons to believe the NSA is acquiring HTTPS communications, including the fact that the NSA is expressly authorized to collect encrypted Internet communications, and that one can learn a lot from an encrypted HTTPS session, including the IP addresses of the user and server and the domain name of the server.

h. Even if the NSA were blacklisting HTTPS, it would still be virtually certain that the NSA would still be copying, reassembling and reviewing Wikimedia HTTP communications considering the number and distribution of those communications. (See ¶ 351.)

[125] Id. ¶ 79.

[126] Appendix C at 30 (FISC Submission (June 1, 2011)).

### C. Selectively Filtering Wikimedia IP addresses

367. Dr. Schulzrinne posits that the NSA could "blacklist" Wikimedia's IP addresses and suggests that if the NSA did so, "*NSA would receive no access to Wikimedia HTTP or HTTPS communications (or, for that matter, Wikimedia communications of any kind)*" (Schulzrinne Decl. ¶ 81). As I show below, that claim is technologically inaccurate and entirely implausible. Dr. Schulzrinne concedes that he has no evidence to support the possibility that the NSA made such a decision, and he does not offer his view on the **likelihood** that the NSA would make such a decision; he merely claims that it is technically possible.[127]

a) In my opinion it is basically inconceivable that the NSA would have decided to blacklist Wikimedia IP addresses. Given that there are millions of websites on the public Internet, the idea that the NSA would have gone through them to decide which to monitor and which not to, in addition to being an incredibly resource-intensive task, is just totally unbelievable. Any such blacklist would purposefully create blind spots in the upstream collection program that could be exploited by NSA targets to bypass surveillance. Including Wikimedia IP addresses in any such blacklist would deliberately limit the possible collection of information on the use of Wikimedia resources by NSA targets, a potentially valuable source of information about the online research and reading of its targets. Viewed in total, taking into account the total lack of any evidence supporting the possibility that the NSA took such action, the idea that the NSA made a deliberate decision to avoid Wikimedia communications seems entirely implausible.

[127] Schulzrinne Decl. ¶ 77

b) It is also technologically incorrect that blocking Wikimedia's IP addresses would block all Wikimedia traffic. Even if NSA blacklisted Wikimedia's IP addresses, Wikimedia's communications would still be copied, reassembled and reviewed by the NSA in at least several circumstances:

(1) MCTs that contain Wikimedia communications, where the enclosing communication is not to or from Wikimedia, but one or more of the embedded communications are to or from Wikimedia.

(2) In the case where a person located outside the U.S. is using an email service located inside the U.S. to send email to Wikimedia. The first "leg" of the journey the email takes from the user's mail agent to the email server would be subject to copying, reassembly and review because the transaction carrying the email message is not wholly domestic. The same is true in reverse: email from Wikimedia to such a person outside the U.S. would not be seen as wholly domestic in the leg between the email service and the user's mail agent. In both of the above cases, the email transaction transiting the international circuit would not have any Wikimedia IP addresses in the IP headers of the packets such that they could be discarded by an IP address-based blacklist.

(3) The traffic between a VPN service in the U.S. and a user located outside the U.S. would not have Wikimedia IP addresses in the traffic even if the user were accessing a Wikimedia site.

c) In any case, the NSA's descriptions of its IP address filtering all state that the goal is to filter out "*wholly domestic communications*,"[128] and these descriptions do not contain any mention of any other goals for the filtering.

**D. U.K. Surveillance Disclosures and Court Proceedings**

368. The U.K.'s signals intelligence agency, Government Communications Headquarters (GCHQ), is charged with performing the functional equivalent of upstream collection in the U.K.[129] GCHQ's public disclosures reinforce my conclusion that, for various technical and practical reasons, the NSA copies the entire stream of communications on a circuit it is monitoring. The GCHQ has explained in court filings that, for "*technical reasons*" and "*as a matter of practical necessity*," it needs to intercept the entire stream of communications on a circuit (which GCHQ refers to as a "*bearer*") when engaging in its equivalent of upstream collection:

> *As explained in detail in the Observations, the s.8(4) Regime operates in this way as a matter of practical necessity. For technical reasons, it is necessary to intercept the entire contents of a bearer, in order to extract even a single specific communication for examination from the bearer: Observations, §§1.31-1.34.*[130]
>
> *Subjects of interest are very likely to use a variety of different means of communication, and to change those means frequently. Moreover, electronic communications do not traverse the internet by routes that can*

[128] Appendix D at 7-8 (NSA Response to Plaintiff's Interrogatory No. 3 (Dec. 22, 2017)); Appendix F at 46, 125, 148 (PCLOB Report at 41, 120, 143); Appendix H at 7-8 (NSA Response to Plaintiff's Request for Admission No. 6 (Jan. 8, 2018)).

[129] Appendix DD ¶ 12 (*Case of Big Brother Watch & Others v. United Kingdom*, Eur. Ct. H.R., ¶ 12 (2018), http://hudoc.echr.coe.int/eng?i=001-186048).

[130] Appendix EE ¶¶ 7-8 (Further Observations of the Government of the United Kingdom ¶¶ 7-8, *10 Human Rights Organizations v. United Kingdom*, Eur. Ct. H.R. (Dec. 16, 2016), https://privacyinternational.org/sites/default/files/2018-02/2016.12.16%20Government%27s%20further%20obs.pdf).

> *necessarily be predicted. Communications will not take the geographically shortest route between sender and recipient, but the route that is most efficient, as determined by factors such as the cost of transmission, and the volume of traffic passing over particular parts of the internet at particular times of day. So in order to obtain even a small proportion of the communications of known targets overseas, it is necessary for the Services to intercept a selection of bearers, and to scan the contents of all those bearers for the wanted communications.*[131]

369. In its ruling, the European Court of Human Rights repeated this description of how the U.K.'s Internet surveillance program operates.[132] In spite of the fact that the GCHQ may not be operating under the same requirement to exclude wholly domestic U.K. traffic from its collection program, GCHQ's practice—and the reasons it has publicly described—reinforce my conclusions that the NSA relies on the copy-then-filter configuration to conduct the upstream collection program and that it does not selectively filter traffic prior to copying it as Dr. Schulzrinne hypothesizes it could.

370. But even if Dr. Schulzrinne's hypothesis that the NSA is filtering certain traffic before copying the remainder were to be true, for the reasons I set forth above, it is virtually certain that the NSA has, in the course of the upstream collection program, copied, reassembled and reviewed at least some of Wikimedia's communications. This, also for the reasons I set forth above, is also true in the highly improbable scenario that

[131] Appendix FF ¶¶ 1.29-1.31, 4.5-4.6 (Observations of the Government of the United Kingdom, ¶¶ 1.29-1.31, 4.5-4.6, *10 Human Rights Organizations v. United Kingdom*, Eur. Ct. H.R. (Apr. 16, 2016), https://privacyinternational.org/sites/default/files/2018-02/United%20Kingdom%E2%80%99s%20Observations%20on%20the%20Merits.pdf).

[132] Appendix DD ¶ 284 (*Case of Big Brother Watch & Others v. United Kingdom*, Eur. Ct. H.R., ¶ 284 (2018)).

Dr. Schulzrinne hypothesizes that the NSA has been purposefully blacklisting Wikimedia IP addresses from the upstream collection program.

I declare under penalty of perjury that the foregoing is true and correct to the best of my knowledge and belief.

Date: 12/18/18

Scott Bradner

**IN THE UNITED STATES DISTRICT COURT**
**FOR THE DISTRICT OF MARYLAND**

| | |
|---|---|
| WIKIMEDIA FOUNDATION,<br><br>*Plaintiff*,<br><br>v.<br><br>NATIONAL SECURITY AGENCY / CENTRAL SECURITY SERVICE, *et al.*,<br><br>*Defendants*. | No. 1:15-cv-00662-TSE |

**DECLARATION OF SCOTT BRADNER**
**APPENDIX LIST**


Curriculum vitae .....A

List of documents provided by Plaintiff's counsel .....B

FISC Submission (June 1, 2011) .....C

NSA Responses to Plaintiff's Interrogatories (Dec. 22, 2017).....D

FISC Opinion (Apr. 26, 2017).....E

Privacy & Civil Liberties Oversight Board, *Report on the Surveillance Program Operated Pursuant to Section 702 of FISA* (July 2, 2014).....F

Paul Baran, RAND Corp., *On Distributed Communications Networks* (Sept. 1962).....G

NSA Responses to Plaintiff's Requests for Admission (Jan. 8, 2018) .....H

David Hauweele et al., *What Do Parrots and BGP Routers Have in Common?*, Computer Comm. Rev. (July 2016) .....I

Report on International Submarine Cables Landing in the US, TeleGeography (Jan. 2018).....J

Transcript of Deposition of Rebecca J. Richards (Apr. 16, 2018).....K

NSA Director of Civil Liberties & Privacy Office, *NSA's Implementation of Foreign Intelligence Surveillance Act Section 702* (Apr. 16, 2014).....L

FISC Submission (May 2, 2011) .....M

FISC Submission (Aug. 16, 2011).....N

Joint Statement, *FISA Amendments Act Reauthorization: Hearing Before the H. Permanent Select Comm. on Intelligence* (Dec. 8, 2011) .....O

FISC Opinion (Oct. 3, 2011).....P

Privacy & Civil Liberties Oversight Board, Transcript of Public Hearing Regarding the Surveillance Program Operated Pursuant to Section 702 of the Foreign Intelligence Surveillance Act (Mar. 19, 2014).....Q

FISC Submission (June 28, 2011) .....R

NSA Section 702 Minimization Procedures (2014) .....S

NSA Section 702 Targeting Procedures (2014).....T

FISC Hearing Transcript, *In Re: DNI/AG 702(g) Certification* [*Redacted*] (2008).....U

Shuihui Chen & Yong Tang, *A Stream Reassembly Mechanism Based on DPI*, Inst. of Electrical & Electronics Engineers (2012) .....V

U.S. Patent No. 8,813,221.....W

NSA Press Releases (Apr. 28, 2017) .....X

ODNI Statistical Transparency Report for 2017 (Apr. 2018).....Y

Wikimedia Responses to Defendants' Interrogatories (Jan. 11, 2018).....Z

Wikimedia Second Amended Responses to Defendants' Interrogatories (Apr. 17, 2018).....AA

Wikimedia Amended Response to ODNI Interrogatory No. 19 (Apr. 6, 2018).....BB

City of Virginia Beach Dep't of Info. Tech., *Next Generation Network and Transoceanic Subsea Cable Updates* (Oct. 4, 2017) .....CC

*Case of Big Brother Watch & Others v. United Kingdom*, Eur. Ct. H.R. (2018) .....DD

Further Observations of the Government of the United Kingdom, *10 Human Rights Organizations v. United Kingdom*, Eur. Ct. H.R. (Dec. 16, 2016) .....EE

Observations of the Government of the United Kingdom, *10 Human Rights Organizations v. United Kingdom*, Eur. Ct. H.R. (Apr. 16, 2016) .....FF