Chapter One:

Deep Packet Inspection and Its Predecessors[1]

By Christopher Parsons[2]

February 6, 2012 :: Version 3.5

[2] Christopher Parsons is a PhD candidate in the Department of Political Science at the University of Victoria. His research interests focus on how privacy (particularly informational privacy, expressive privacy and accessibility privacy) is affected by digitally mediated surveillance, and the normative implications that such surveillance has in (and on) contemporary Western political systems. Feedback on this draft is welcomed, and can be sent to parsons@uvic.ca.

Table of Contents

The earliest social choices and administrative decisions guiding the Internet's growth emphasized packet delivery over infrastructural or data security.[3] These early choices have led to an Internet that is fundamentally predicated on trust and radical vulnerability, insofar as individuals must trust that their data will arrive at its destination without interference. The 'default-setting' of Internet communications is to hope that no other agent will take advantage of the fact that most people's communications are transmitted throughout the Internet in easily read plain text. Methods that secure this vulnerable data traffic, such as encryption, obfuscation, and forensic real-time packet analysis, are effectively a series of kludges that are bolted onto an architecture designed primarily to ensure packet delivery. Whereas packet inspection technologies initially functioned for diagnostic purposes, they are now being repositioned to 'secure' the Internet, and society more generally, by taking advantage of the Internet's vulnerabilities to monitor, mediate, and modify data traffic. Such inspection capabilities reorient the potentialities of the digital medium by establishing new modes of impacting communications and data transfers, thus affecting the character of messages on the Internet. Whereas the early Internet could be characterized as one of trusting the messenger, today the routing infrastructure responsible for transferring messages may have secretly inspected, recorded, or modified messages before passing them towards their destination.

This chapter traces the lineage of contemporary packet inspection systems that monitor data traffic flowing across the Internet in real time. After discussing how shallow, medium, and deep packet inspection systems function, I outline the significance of this technology's most recent iteration, deep packet inspection, and how it could be used to fulfill technical, economic, and political goals. Achieving these goals, however, requires that deep packet inspection be regarded as a surveillance practice. Indeed, deep packet

---

[3] S. Landau. (2011). *Surveillance or Security: The Risks Posed by New Wiretapping Technologies*. Cambridge, Mass.: The MIT Press. Pp. 39.

inspection is, at its core, a surveillance-based technology that is used by private actors, such as Internet service providers, to monitor and mediate citizens' communications. Given the importance of Internet-based communications to every facet of Western society, from personal communications, to economic, cultural and political exchanges, deep packet inspection must be evaluated not just in the abstract but with attention towards how society shapes its deployment and how it may shape society.

A Lineage of Data Packet Inspection

Network administrators initially logged some network activity to identify and resolve network irregularities when ARPANET, the predecessor to the public Internet, was under development.[4] Logging let administrators determine if packets were being delivered and whether network nodes were functioning normally. At this point security was an afterthought, at best, given that the few people using the network were relatively savvy users. While the military, which invested in the early funding of ARPANET, moved to systems that were segregated from networks used by researchers and civilians, there were no effective means preventing packets from being sent to, or received from, ARPANET.[5] Compounding these security challenges were the UNIX systems connected to the Internet: these system were generally recognized as insecure because neither they nor ARPANET more generally had been designed with security in mind.[6] Before the first piece of software that intentionally exploited the network was released, ARPANET and its accompanying workstations operated in a kind of "network of Eden."

For ARPANET, the poison apple was the Morris worm. Whereas viruses tend to be attached to files, worms are typically autonomous programs that burrow into computers and simply spread. Their primary function is to be self-replicating, with other functionality, such as viral attack code, often being appended to them. Morris compromised computers connected to ARPANET without damaging core system files, instead slowing down computers until they had to be rebooted to restore their usability.[7] The worm spread to hundreds of computers and led to significant losses in available computing time. In Morris' aftermath the security of the network became a more prominent concern in the minds of researchers and general users, alike.

To mitigate or avoid subsequent disseminations of malware (harmful software intended to impair or act contrary to the computer owners' intentions or expectations), "computer

---

[4] K. Hafner and M. Lyon. (2006). *Where Wizards Stay Up Late: The Origins of the Internet*. New York: Simon & Schuster. Pp. 161-165.
[5] H. Orman. (2003). "The Morris worm: a fifteen-year perspective," *Security and Privacy, IEEE* 1(5). Pp. 36.
[6] H. Orman. (2003). "The Morris worm: a fifteen-year perspective," *Security and Privacy, IEEE* 1(5). Pp. 35-36.
[7] While there are claims that thousands of computers were infected by the worm, no one can be certain of such numbers. Paul Graham has stated that he was present when a 'guestimate' of 6,000 infected computers was arrived at. This estimate was based on the assumption that about 60,000 computers were attached to the network, and roughly 10 percent assumed compromised. P. Graham. (2005). "The Submarine," PaulGraham.com. Published April 2005. Last accessed May 4, 2011. Online: http://www.paulgraham.com/submarine.html#f4n <http://www.paulgraham.com/submarine.html#f4n>

science departments around the world tried to delineate the difference between appropriate and inappropriate computer and network usage, and many tried to define an ethical basis for the distinctions."[8] The diagnosis of the Morris worm also provoked extended discussion about computer ethics by the Internet Engineering Task Force (IETF),[9] the Internet Activities Board,[10] National Science Foundation,[11] Computer Professionals for Social Responsibility,[12] as well as in academic, professional, and popular circles.[13] Further, the Computer Emergency Response Team (CERT), which documents computer problems and vendor solutions, was formed. Computer firewalls also received additional attention. While firewalls, which are designed to permit or deny transmissions of data into networks based on rules established by a network administrator, had been in development before the Morris worm, in the aftermath of the worm and the shift towards a broader public user base led to firewalls being routinely deployed by 1994-5.[14] Firewalls are effectively packet analysis systems, and are configured to "reject, allow, or redirect specific types of traffic addressed to specific services and are (not surprisingly) used to limit access to certain functions and resources for all traffic traveling across a device."[15] They have evolved in three general waves since the mid-90s: shallow packet, medium packet, and deep packet inspection.

While early packet analysis systems merely examined information derived from data packets' headers, they now examine both the header and the payload. The header includes the recipient's Internet Protocol (IP) address, a number that is used to reassemble packets in the correct order when recompiling the messages and to deliver packets to their destination(s). At a more fine-grained level, the information used to route packets is derived from the physical, data link, network, and transport layers of the packet. The payload, or content, of the packet includes information about what application is sending the data, whether the packet's contents are themselves encrypted, and what the precise content of the packet is (e.g. the actual text of an email). More specifically, the payload can be understood as composing the session layer, presentation layer, and application layers of the packet.

---

[8] H. Orman. (2003). "The Morris worm: a fifteen-year perspective," *Security and Privacy, IEEE 1(5)*. Pp. 40.

[9] J. Reynolds. (1989). "RFC 1135: The Helminthiasis of the Internet," IETF Network Working Group. Online: < http://tools.ietf.org/html/rfc1135>.

[10] Internet Activities Board. (1989). "Ethics and the Internet," *Communications of the ACM 32(6)*.

[11] National Science Foundation. (1989). "NSF Poses Code of Networking Ethics," *Communications of the ACM 32(6)*.

[12] Computer Professionals for Social Responsibility. (1989). "CPSR Statement on the Computer Virus," *Communications of the ACM 32(6).*

[13] See Section 9: Bibliography of J. Reynolds. (1989). "RFC 1135: The Helminthiasis of the Internet," IETF Network Working Group. Online: < http://tools.ietf.org/html/rfc1135>.

[14] H. Orman. (2003). "The Morris worm: a fifteen-year perspective," *Security and Privacy, IEEE* 1(5). Pp. 35-43.

[15] M. Zalewski. (2005). *Silence on the Wire: a Field Guide to Passive Reconnaisance and Indirect Attacks*. San Francisco: No Starch Press. Pp. 174.

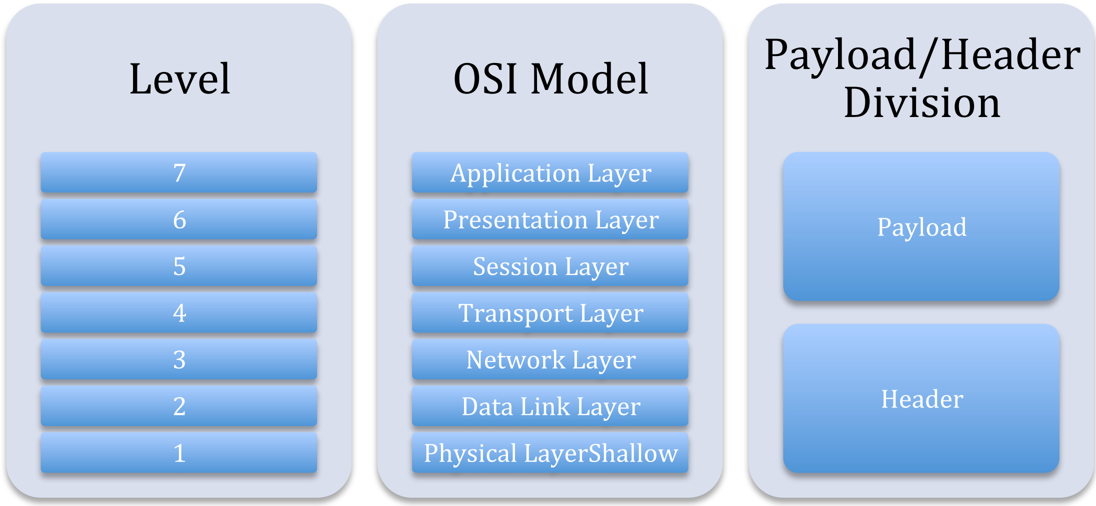

Figure 1: Levels in the OSI Packet Model

These granular divisions of header and payload are derived from the Open Systems Interconnect (OSI) model (Figure 1), which is composed of seven layers. This model was developed by the International Standards Organization (ISO) in 1984 to standardize how networking technologies were generally conceptualized, though it was later abandoned for practical networking activities in favor of the Transmission Control Protocol and Internet Protocol Suite (TCP/IP). OSI's most significant contribution to network development efforts has been to force "protocol designers to be more conscious of how the behavior of each protocol would affect the entire system."[16] OSI stands in contrast to TCP/IP's key contribution, which was to create a fungible system that maximized interoperability by minimizing system interfaces (IP) and checking for packet delivery and network congestion (TCP). TCP/IP's other key contribution was that it ensured that the ends of the network, as opposed to the core, would govern the flow of data packets. In a TCP/IP network, client computers are responsible for controlling the flow of packets and, as such, limit network owners' control over what, why, and how packets course across the Internet.[17]

When sending a packet of data, the Application Layer interacts with the piece of software that is making a data request, such as the email client, web browser, instant messaging software and so on. For example, when you enter a URL into a web browser, the browser makes a HTTP request to access a webpage, which is passed to the lower layers of the stack. When the browser receives a response from the server on the Internet that hosts the requested page, the browser displays the content associated with the URL. The Presentation Layer is concerned with the actual format that the data is presented in, such as the JPEG, MPEG, MOV, and HTML file-types. This layer also encrypts and compresses data. In the case of a webpage, this stage is where the data request is identified as asking for an HTML file. The fifth layer, the Session Layer, creates, manages, and ends communications within a session between the sender(s) and

[16] J. Abbate. (1999). *Inventing the Internet*. Cambridge, Mass.: The MIT Press. Pp. 177.
[17] *Ibid*. Pp. 194.

recipient(s) of data traffic; it effectively operates as a 'traffic cop' by directing data flows. When navigating to a URL, this layer regulates the transmission of data composing the web pages, the text, the images, the audio associated with it, and so on. These three layers broadly compose what is termed the 'payload' of a packet.

The fourth through first layers of a packet compose what is commonly referred to as the 'header'. The Transport Layer segments data from the upper levels, establishes a connection between the packet's point of origin and where it is to be received, and ensures that the packets are reassembled in the correct order. This layer is not concerned with managing or ending sessions, only with the actual connection between the sender(s) and recipient(s) of packets. In terms of a web browser, this layer establishes the connection between the computer requesting data and the server that is hosting it. It also

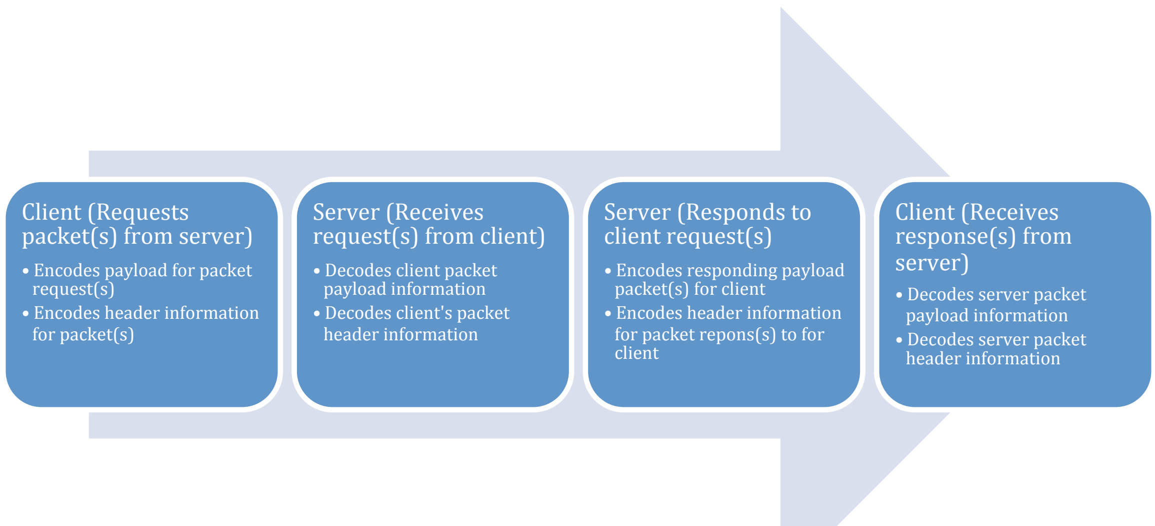

Figure 2: Client-Server data transaction

ensures that packets are properly ordered so that the aggregate data they contain are meaningfully (re)arranged when arriving at their destination. The Network Layer provides the packet's addressing and routing; it handles how the packet will get from one part of the network to another, and it is responsible for configuring the packet to an appropriate transmission standard (e.g. the Internet Protocol). This layer is not concerned with whether packets arrive at their destination error free; the transport layer assumes that role. The Data Link Layer formats the packet so that it can be sent along the medium being used to transmit the packet from its point of origin to its destination. As an example, this layer can prepare packets for the wireless medium when sending an email from a local coffee shop, then re-packaged to be sent along an Ethernet connection as it travels to an ISP and through its wireline networks, and then back to a wireless format when being received by a colleague in their office whose laptop is connected to their local network using wireless technology. The Physical Layer doesn't change the packet's actual data; it defines the actual media and characteristics along which the data are being transmitted.

Packets are typically transmitted from clients to servers. Figure two provides a visual presentation of a basic client-server transaction. These transactions begin with a client computer requesting data from a server by encoding a packet using the OSI layer model (i.e. creating a packet that contains the information from layers 7 to 1). The server receives the request, decodes it, and then encodes a packet response for the client, which subsequently receives and decodes the packet to provide the application with the requested information.

Shallow Packet Inspection

Shallow Packet Inspection (SPI) technologies depend on (relatively) simplistic firewalls. They limit user-specified content from leaving, or being received by, the client computer. When a server sends a packet to a client computer, SPI technologies examine the packet's header information and evaluate it against a blacklist. In some cases these firewalls come with a predefined set of rules that constitute the blacklist against which data are evaluated, whereas in others network administrators are responsible for creating and updating the rule set. Specifically, these firewalls focus on the source and destination IP address that the packet is trying to access and the packet's port address. If the packet's header information – either an IP address, a port number, or a combination of the two[18] – is on the blacklist then the packet is not delivered. When SPI technology refuses to deliver a packet, the technology simply refuses to pass it along without notifying the source that the packet has been rejected.[19] More advanced forms of SPI capture logs of incoming and outgoing source/destination information so that a systems administrator can later review the aggregate header information to adjust, or create, blacklist rule sets.

SPI cannot read beyond the information contained in a header and focuses on the second and third layers in the OSI model; SPI examines the sender's and receiver's IP address, the number of packets that a message is broken into, the number of hops a packet can make before routers stop forwarding it, and the synchronization data that allows for reassembling the packets into a format that the receiving application can understand. This means that SPI cannot read the session, presentation, or applications layers of a packet; it cannot peer into a packet's payload and survey the contents.

Medium Packet Inspection

Medium Packet Inspection (MPI) is typically used to refer to 'application proxies', or devices that stand between end-users' computers and ISP/Internet gateways. These proxies can examine packet header information against their loaded parse-list.[20] Parsing involves structuring data as "a linear representation in accordance with a given grammar."[21] While finite languages can provide infinite numbers of sentences/linear representations, a parse list holds a set of particular representations and, upon identifying them, takes specified action against them. In effect, this means that MPI devices bridge

[18] T. Porter. (2010). "The Perils of Deep Packet Inspection," Symantic Corporation. Available: <http://www.symantec.com/connect/articles/perils-deep-packet-inspection>

[19] The action of rejecting packets without notifying their source is sometimes referred to as 'blackholing' packets. It has the relative advantage of not alerting the sources that are sending viruses, spam messages, and so on that their packets are not reaching their destination.

[20] It should be noted that, in addition to MPI being found in application proxies, some security vendors such as McAfee and Symantec include MPI technology in their 'prosumer' firewalls, letting their customers enjoy the benefits of MPI without paying for a dedicated hardware device.

[21] D. Gune and C. Jacobs. (1990). *Parsing Techniques: A Practical Guide*. West Sussex: Ellis Horwood Limited. Pp. 1.

connections between computers on a network and the Internet at large, and they are configured to look for very particular data traffic and take preordained actions towards it.

More specifically, in the case of MPI devices this entails examining packet headers and a small amount of the payload, which together can assume an infinite number of representations, for particular representations. Importantly, parse-lists are subtler than blacklists. Whereas the latter establishes that something is either permissible or impermissible, a parse-list allows specific packet-types to be allowed or disallowed based on their data format types and associated location on the Internet, rather than on their IP address alone. Further, parse-lists are meant to be easily updated to account for new linear representations that network administrators want to remain aware of, or modify existing representation-sets to mitigate false-positives. As such, MPI constitutes an evolution of packet awareness technologies, insofar as it can more comprehensively 'read' the packet and take a broader range of actions against packets that fall within their parse-lists.

Application proxies intercept data connections and subsequently initiate new connections between the proxy and either the client on the network (receiving data from the Internet) or between the proxy and data's destination on the Internet (when transmitting data to the Internet).[22] These devices are typically placed inline with network routing equipment – all traffic that passes through the network must pass through the proxy device – to ensure that network administrators' rule sets are uniformly applied to all data streaming through the network. Figure three offers a visual example of how this might appear in a network diagram. Placing devices inline has the benefit of separating the source and destination of a packet – the application proxy acts as an intermediary between client computers and the Internet more broadly – and thus provides network administrators with the ability to force client computers to authenticate to the proxy device before they can receive packets from beyond the administrator's network.

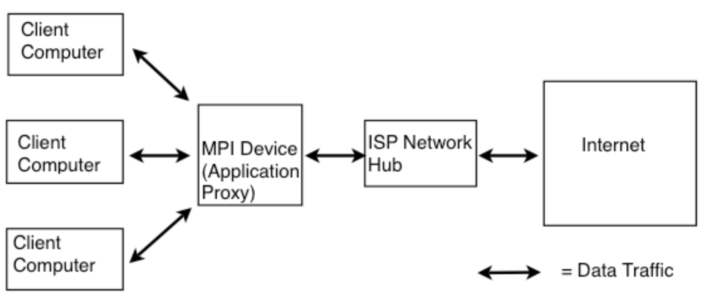

Figure 3: MPI Device Inline with Network Routing Equipment

Using MPI devices, network administrators could prevent client computers from receiving flash files from YouTube, or image files from social networking sites. MPI technologies can prioritize some packets over others by examining the application

[22] M. Zalewski. (2005). *Silence on the Wire: a Field Guide to Passive Reconnaissance and Indirect Attacks*. San Francisco: No Starch Press. Pp. 146.

commands that are located within the application layer[23] and the file formats in the presentation layer.[24] Given their (limited) insight into the application layer of the packet, these devices can also be configured to distinguish between normal representations of a data protocol such as HTTP and abnormal representations, and filter or screen abnormal representations from being passed to a client within the network. They can also dig into the packet and identify the commands that are being associated with an application protocol and permit or deny the data connection based on whether the command/application combination is on the parse-list. Thus, an FTP data request that included the 'mget' command, which copies multiple files from a remote machine to a local machine might be prevented, whereas FTP connections including the 'cd', or change directory command, might be permitted. Given MPI devices' status as application proxies, they also assume characteristics of offering full logging information about packets as opposed to just header information, and when integrated into a trust-chain can decrypt data traffic, examine it, re-encrypt the traffic, and forward it to the traffic's destination.

Unfortunately, MPI devices suffer from poor scalability; each application command or protocol that is examined requires a unique application gateway, and inspecting each packet reduces the speed at which the packets can be delivered to their recipients.[25] Given these weaknesses, MPI devices are challenging to deploy in large networking operations where a large variety of applications must be monitored. This limits their usefulness for Internet Service Providers, where tens of thousands of applications can be transmitting packets at any given moment.

While MPI devices suffer from limitations, they act as a key facet in technological developments towards deep packet inspection. Specifically, their capability to read the presentation layer of the packet's application layer acts as a transition point for reading the entire payload. As a result, this inspection technology constitutes a stepping-stone in the path towards contemporary deep packet inspection technologies.

Deep Packet Inspection

Deep Packet Inspection (DPI) equipment is typically found in expensive routing devices that are installed in major networking hubs. The equipment lets network operators precisely identify the origin and content of each packet of data that passes through these hubs. Arbor/Ellacoya, a vendor of DPI equipment, notes that their e100 devices use DPI "to monitor and classify data directly from your network traffic flow. Inspecting data packets at Layers 3-7 allows the e100 to provide crucial information to your operations and business support systems, without compromising other services."[26] Whereas MPI devices have very limited application awareness, DPI devices can potentially "look inside

---

[23] Application commands are typically limited to Telnet, FTP, and HTTP.
[24] T. Porter, A. Zmolek, J. Kanclirz and A. Rosela. (2006). *Practical VoIP Security: your hands-on guide to Voice over IP (VoIP) security*. Rockland, Mass.: Syngress Publishing, Inc.
[25] C. Tobkin and D. Kligerman (2004). *Check Point Next Generation with Application Intelligence Security Administration*. Rockland, Mass.: Syngress Publishing, Inc.
[26] Arbor Ellacoya (2008). "Arbor Ellacoya e100: Unmatched Scale and Intelligence in a Broadband Optimization Platform (Datasheet)". Last accessed: March 14, 2011. Online: <http://www.arbornetworks.com/index.php?option=com_docman&task=doc_download&gid=355>

all traffic from a specific IP address, pick out the HTTP traffic, then drill even further down to capture traffic headed to and from Gmail, and can then reassemble e-mails as they are typed out by the user."[27] While MPI devices have scaling issues, DPI devices are designed to determine what programs generate packets, in real-time, for hundreds of thousands of transactions each second. They are designed to scale in large networking environments and behave reactively, insofar as actions against certain data packets can be taken when particular pre-set conditions are met.

At its most basic level, DPI equipment examines a particular packet in its totality and examines the packet's characteristics against a predefined rule set. Such examinations entail looking at layers 2-7 to examine packet headers and payloads to search for indications of protocol non-compliance, malicious code, spam, and any predefined data types that the network owner wants to monitor or take action towards. The equipment identifies and classifies packets based on a signature database. Signatures are developed by extracting characteristic elements of packets that are associated with applications of interest. These characteristics are used to develop signatures in port addresses, string matches, and the packets' numerical properties. Port address analysis behaves similar to SPI and MPI techniques: the equipment examines which data port is in use and, where that port is uniquely assigned to a single application or protocol (e.g. port 25 is assigned to SMTP email traffic), then packets that are being transmitted to or from the port may have an action taken against it. String analysis entails examining the packet for unique numeric and alphabetic characteristics, such as the name of the application responsible for transmitting the packet. String analysis enables the operator to 'catch' packets that use a common port, such as port 80, to either avoid detection or take advantage of more relaxed rules. Thus, a peer-to-peer application might transmit data using port 80 but, if it declares its name, a string analysis may identify the application's traffic. When examining numerical properties, the DPI device will examine the specific size of the data packet; where very specific sizes are identified and the packet accords with other characters (e.g. port or string) then action may be taken.[28] No specific analytic technique needs to be used in isolation; taken together these variables constitute signatures.

Upon identifying a packet-of-interest it can be redirected, marked or tagged, blocked or dropped, rate limited, or reported to the network administrator. A redirection could see particular packet signatures forwarded to a specific location within the network; perhaps all STMP (email) traffic is forwarded to a specialized piece of equipment that evaluates whether the traffic is spam or not, and then subsequently sends the email to its destination. Packets can also be marked to assign them a quality of service level; packets that are sensitive to high levels of latency, which is the measure of delay experienced in the packet exchange system, might be given a higher priority to be routed to their destination than packets that are less affected by latency. Packet tagging, in contrast, is predominantly used to assign internal identifiers to packets than can then be acted upon.

[27] N. Anderson. (2007). "Deep Packet Inspection meets 'Net neutrality, CALEA," *Ars Technica*. Published July 25, 2007. Last accessed March 20, 2011. Online: <http://arstechnica.com/articles/culture/Deep-packet-inspection-meets-net-neutrality.ars>

[28] Allot Communications. (2007). "Digging Deeper Into Deep Packet Inspection (DPI)," Last accessed July 28, 2011. Online: <https://www.dpacket.org/articles/digging-deeper-deep-packet-inspection-dpi>

Tagging can often be performed by one device that can modify packets, such as DPI, and then a subsequent element of the network can read the tag and tag action based on the tag. This might include routing the packet through a particular network gateway or only moving it along a particular set of friendly/secure routers. When either blocking or dropping packets, the equipment will refuse to forward the packet to the next hop towards its destination, often without notifying the source of the packet that it is being blocked. Rate limitations establish particular levels of data transmission capacity depending on the application responsible for generating the data traffic. Such limitations are particularly common where certain applications, such as FTP and peer-to-peer, are well known to use large amounts of data capacity (measured in data transferred per second) and data volume (measured in the total amount of data that is being transferred over time).

In some cases DPI equipment cannot immediately identify the application that has produced a packet. When this occurs, network operators can often use 'Deep Packet Capture' (DPC) technologies to collect packets in the device's short or long memory. DPC lets network administrators perform forensic analysis of packets to determine "the real causes of network problems, identify security threats, and ensure that data communications and network usage complies with outlined policies."[29] Packets can be either fully captured, or only have particular characteristics captured, such as IP destination, the port the packet used or application-type. After a DPC process, packet streams can be evaluated against sets of known applications and their corresponding data stream patterns, which lets ISPs evaluate whether their customers are conforming to security or data usage policies. To elucidate, using this technology a new file sharing program's packet stream, which was unfamiliar to the DPI device, could be captured and subsequently analyzed and identified. Following this identification of this new program's packet stream, each packet from that program could have rule sets applied to it that corresponded with the ISP's networking policies.

To properly identify a packet, hundreds or thousands of packets can be stored in the memory of the inspection device until it has enough information to appropriately match the packets against the devices' list of known packet-types.[30] Once the device can match the previously ambiguous packets against its list of known packet contents, it knows what application (or application-type) is generating and sending the packet, and rules can be applied to allow or disallow the application(-type) from continuing to send and receive packets. Rules could, alternately, moderate the rates of data flowing to and from the application – this intentional alteration of data flow rates is often referred to as 'throttling'. While it is theoretically possible for all data to be captured using DPC technologies and subsequently analyzed using DPI functionality, this would substantially slow the transmission of packets and degrade user experiences when they were streaming content. Further, if the network environment has a large number of client devices or users, such as in mid-to-large sized businesses and ISPs generally, then the storage

---

[29] Bivio Networks and Solera Networks. (2008). "White Paper: Complete Network Visibility through Deep Packet Inspection and Deep Packet Capture. Lindon, Utah: Solera Networks." Last accessed March 21, 2011. Online: <www.soleranetworks.com/products/documents/dpi_dpc_bivio_solera.pdf>

[30] Allot Communications Ltd. (2007) "Digging Deeper into Deep Packet Inspection," Published 2007.

requirements will prohibit even short-term full data retention. As a result, DPC is not marketed as a means to persistently capture all of the data that ISPs' customers send and receive, but to enable targeted capturing of packets. Such data captures can be used to improve subsequent network performance and to comply with regulatory demands, such as government wiretap or data retention and preservation requests.

DPC capabilities can also be used to compose the unique hash of files that a user is receiving from, or transmitting to, the Internet. After computing the hash the device can examine it against a hash database and take action against the file. Hash-based approaches fail, however, when the file itself has been modified in any manner, such as when a word processing file has text added or subtracted. Fingerprinting is a more computationally intensive process, which entails generating a unique representation of the file and examining the file itself – not the hash – to see if the representation is present. As a result, in the case of a word processing document the DPI device would identify a modified file by reference to the common fingerprinted data that was shared between the original (unmodified) document and the modified one. Such processing is extremely expensive, however, and thus presently ill-suited for large-scale fast network conditions.[31]

When a DPI device cannot identify the application responsible for sending packets by examining the packets' headers and/or payloads, it examines how packets are being exchanged between the computers that are exchanging packets. The device evaluates the spikes and bursts of traffic that occur as the unknown application sends and receives data to and from the Internet, and it correlates the traffic patterns against known protocols that particular programs use to exchange data. This heuristic evaluation effectively bypasses the challenges that data encryption pose to packet inspection devices; full-packet encryption prevents DPI devices from examining payload data.

To make this latter process a bit clearer, let us turn to an example. Skype hinders packet inspection devices from identifying its packets by masking its legitimate packet header information and encrypting payload data. Given that the packets themselves are fully encrypted and the information contained in the headers is bogus, ISPs must adopt a different method for detecting Skype traffic. As a solution, DPI devices must watch for a particular exchange of data that occurs when Skype users initiate a voice chat. Each time you contact someone using Skype, the seemingly random initial burst of packet exchanges follows a common pattern that can be heuristically identified and correlated with the Skype application.[32] After the application is identified, it is possible to impede or prioritize the packets generated by this application.

---

[31] Ipoque. (2009). "Copyright Protection in the Internet," Germany, Ipoque. Last accessed July 4,2011. Online: <http://www.ipoque.com/sites/default/files/mediafiles/documents/white-paper-copyright-protection-internet.pdf>

[32] Bonfiglio, Dario, Marco Mellia, Michela Meo, Dario Rossi, and Paolo Tofanelli (2007). "Revealing Skype Traffic: When Randomness Plays With You," *Computer Communications Review* 37(4), pp. 37-48. P. Renals and G. A. Jacoby. (2009). "Blocking Skype through Deep Packet Inspection," *42nd Hawaii International Conference on System Sciences*. See also: Allot Communications Ltd. (2007) "Digging Deeper into Deep Packet Inspection," Published 2007.

Effectively, DPI lets network owners inspect, stop, or manipulate unencrypted data exchanges flowing across their network in real time. Where encrypted data is transferred using a known pattern, administrators can intuit what is likely transmitting the data and similarly take action. This level of awareness concerning packet contents lets administrators interact with packets at a granular level, and in an automated fashion, before the packets leave the originating network or arrive at a recipient within that network. This interrogation capacity has implications for how large network providers, such as Internet Service Providers (ISPs), develop their network and also establishes an appealing technical infrastructure that non-ISPs may be interested in influencing. Having discussed the capabilities of deep packet inspection, let us turn to how they might be utilized to fulfill various technical, economic, and political goals.

Technical Capabilities and Their Potentials

Deep packet inspection devices are designed to accomplish a range of goals; they are deployed for security, network management, content identification, and data modification purposes. A range of socio-political enabled potentialities is integrated into the design of the technology and is responsible for driving its characteristics and technical capacities. While detailed investigations into the theory of social-technical relationships, and empirical data on actual uses of deep packet inspection will follow in later chapters, we must first consider the potentialities linked with the technology. To this end, I suggest that there are technical, economic, and political uses to which the technology may be put

The Technical Possibilities of DPI

Network administrators are concerned with the functioning of the network itself: are security incidents logged and kept to a minimum? Do network policies simultaneously ensure the functioning of the network and meet users' expectations and needs? Are the network's nodes appropriately configured to address congestion? Deep packet inspection helps administrators improve network security, implement access requirements, guarantee quality of service, and tailor service for particular applications. Each of these functions is dynamic, insofar as the technology can utilize layered rule sets and is incorporated within a broader networking assemblage to dynamically react to changes in the network. As a result of DPI's penetration into packet transfers, combined with its potentialities, the technology can be helpful in daily and long-term network operations.

DPI was initially meant to offer network providers improved intrusion detection and prevention mechanisms that could recognize and respond to contemporary threats.[33] To respond to emerging threats, DPI appliances are reconfigurable and scale to monitor high volumes of traffic, and also provide logging and anomaly detection. Logging establishes a pattern of known behavior and lets the system (and system administrator, if they examine the logs) examine traffic 'offline'. Offline analysis facilitates a granular analysis of the traffic because it needn't occur in real-time, thus mitigating some of the technical challenges associated with in depth analysis of data packets while maintaining high data transit speeds. As a result of logging traffic, systems and administrators can 'learn' how to sub-classify network traffic within applications. To make this a bit clearer, consider a

[33] I. Sourdis. (2007). *Designs & Algorithms for Packet and Content Inspection*. Delft: TU. Delft.

process of logging unencrypted HTTP, or web browser, traffic. The system could identify HTTP traffic, and then sub-classify traffic associated with social-media websites, further classify traffic to differentiate between downloading and uploading traffic, and go one step further by identifying whether a user is involved in transmitting or receiving images, movies, or other types of content within a social media website.[34]

It is also possible to use logging-based learning to develop expected use-patterns for individual users and applications, and set notifications to administrators if deviations from the norms are detected. Such deviations may indicate that a known client's credentials are being used by a third-party to access the network, based on suspicious or deviant data transmissions and receptions, or that an application has been infected with malware. Because DPI systems afford high levels of control, if a particular detection signature is too 'chatty' – if a signature is being identified regularly but is uninteresting to the network administrator – the DPI system can be set to either ignore or more carefully monitor the signature in question. A more careful monitoring schema might narrow down the parameters of the inspection, such as shifting from monitoring for all encrypted communications across a corporation to monitoring for encrypted communication in specific business units that are not expected to engage in secure communications.[35] Alternately, the system might be set to avoid establishing a 'normal' activity pattern for authenticated 'guest' accounts because the logged in user(s) regularly changes, though the equipment could still watch for anomalous application behavior.

More generally, as a component of an integrated security processes, DPI can examine inbound and outbound data traffic and flag packets that warrant a more sustained analysis of its contents. This might happen when the device cannot positively identify the application responsible for the packet, or when configured to forward some packets to a proxy server prior to delivery. At the intermediary between the DPI appliance and destination an algorithmic analysis may be performed. Such an analysis might examine whether an email attachment contains material that cannot enter or exit the network, or generate an alert requiring a human to evaluate the information, such as when abnormal packets are being received or transmitted[36] or to vet the appropriateness of email

[34] This level of functionality is provided by Q1 Labs' 'QRadar 7.0' product.

[35] This specific attention to encryption from systems and business units that have not been configured to use encryption by IT staff is a reasonably common practice in some businesses in Canada. This kind of activity is monitored because abnormal instances of encrypted data traffic may indicate that either an employee is engaged in espionage or (more commonly) has established an encrypted proxy connection to evade business policies and watch online television or download movies.

[36] A properly configured DPI device may have been helpful in diagnosing a problem with network equipment run by Telekomunikacja Polska, Poland's national telco. They had network equipment that was mangling traffic by stripping TCP headers from the packet payload, which resulted in their network transmitting unusual and suspicious traffic to ports 21536, 18477 and 19535. Had DPI been in place at the outskirts of their network, the telco might have identified the traffic and corrected its implementation of TCP/IP itself, rather than relying on third-party researchers to identify the packets and their source. For more on this, see M. Zalewski. (2008). *Silence on the Wire: a Field Guide to Passive Reconnaisance and Indirect Attack*. Pp. 186-187.

attachments.[37] When directing data traffic beyond the network that the DPI is integrated into, it might add a prefix to a packet's header to indicate the quality of service it should receive, whether the packet is the bottom of a 'stack' or series of related packets, or impose a time-to-live value[38] that overwrites the value set by the client sending the packet. Stacks of tags might nest a series of attributes, such as Quality of Service or where the packet should be forwarded, and may be coded so that egress or ingress networks can act on the attributes.[39] Such prefixes can also be used in establishing virtual private networks when partnered with perimeter edge routers capable of maintaining their own routing tables. Perimeter routers will identify what other routers traffic can be forwarded to, and separates traffic so that users cannot see data outside of their network. Using this approach, encryption is not required because traffic cannot deviate from pre-programmed traffic routes.[40]

Existing policy management tools and servers will often guide the technical management of data traffic. Policy control is "a broad concept" that "is usually based on the use of an automated rules engine to apply simple logical rules which, when concatenated, can enable relatively complex policies to be triggered in response to information received from networks."[41] Network managers can examine which account is authenticated to a particular data stream, call the rules dictating how that user can transmit and receive data, and then examine their entire packet stream and mediate data flows as dictated by the policy governing the user. This may mean that a client for an ISP on an entry-level service package is prevented from transmitting packets that are not related to HTTP (web-based) or SMTP (email-based) traffic, whereas premium users have all of their data traffic prioritized over that of other users of the network. Policy controls permit a vast range of rules, which may prioritize or deprioritize some kinds of traffic either in perpetuity, at certain points in the day, or for certain users, block some content if the user's account does not permit its reception or transmission, or modify some data traffic in real-time. Modifications might include changing HTTP traffic so that users see a banner in their web browser that notes whether users are nearing or exceeding the volume

---

[37] Sonicwall. (2008). "10 Cool Things Your Firewall Should Do," *Sonicwall Slide Deck*. Pp. 11. Last accessed February 3, 2013. Online: < http://www.sosonicwall.com/lib/deciding-what-solution/10_Things_Your_Firewall_Should_Do.pdf >.

[38] Time-to-live (TTL) is a value that identifies the maximum number of 'hops' that a packet can take before the Internet's routing structure will cease to pass it to another router. It is meant to prevent endless loops of packets being sent through the Internet – thus consuming router resources – when something has gone awry with routing tables. Each packet has a number assigned to it by the client application, in tandem with the client computer's implementation of the TCP/IP stack, and that number decreases by one for each 'hop' to a new network component that it travels along.

[39] Y. Rekhter, B. Davie, E. Rosen, G. Swallow, D. Farinacci, and D. Katz. (1997). "Tag Switching Overview," *Proceedings of the IEEE* 85(12).

[40] Cisco. "Introduction to Cisco MPLS VPN Technology," Last accessed June 26, 2011. Online: <http://www.cisco.com/en/US/docs/net_mgmt/vpn_solutions_center/1.1/user/guide/VPN_UG1.html> ; K. DeGeest. (2001). "What is an MPLS VPN Anyway?" *SANS Institute*. Last accessed June 25, 2011. Online: <http://www.sans.org/reading_room/whitepapers/vpns/mpls-vpn-anyway_718>; E. Rosen, Y. Rekhter. (2006). "RFC 4364: BGP/MPLS IP Virtual Private Networks (VPNS)," *IETF*. Last accessed June 25, 2011. Online: <http://tools.ietf.org/html/rfc4364>.

[41] G. Finnie. (2009). "(Report) ISP Traffic Management Technologies: The State of the Art," or the CRTC Public Notice on the Review of the Internet traffic management practices of Internet service providers. Pp. 12.

of data they are provided within a billing cycle[42] or warning them that they are possibly infected with a virus, worm, or other piece of malware. The policies, and their associated servers, work hand-in-hand with DPI devices, often to guide how the devices themselves take action on packets traversing the network.

Digital networks are involved in transmitting more and more data and key points in the network require regular upgrades to keep pace with growth patterns. While growth adheres to a well-known rate,[43] the general patterns of aggregate expanded bandwidth requirements do not necessarily identify the expanded bandwidth requirements placed on particular routers. When routers experience high-levels of usage – when so many data packets are sent to a router that it reaches or exceeds the maximum amount of packets it can forward to the next hop per second – they become congested. Congestion simply means that for a period of time more data is being forwarded to the router than it can pass forward. As a result, some packets are not forwarded to their next hop on the Internet and thus are not delivered to their destination.[44]

Deep packet inspection equipment is meant to limit the inconveniences associated with router congestion. By identifying and prioritizing packets in real-time, DPI appliances can ensure that time-sensitive packets, such as those associated with voice over Internet protocol (e.g. Skype) communications, are moved up in the 'queue' of packets and those that are less sensitive, such as email, are dropped to be resent. Alternately, if the network operator has identified particular applications or application-types that significantly contribute to router congestion then particular rules can be established to limit the amount of the router's bandwidth they can consume. Thus, 20% of a router's aggregate bandwidth might be allocated to the 'problem' application or application-type and the remaining 80% of aggregate bandwidth might be available to all other data traffic. The administrator could forgo assigning bandwidth to any particular application and instead limit the amount of bandwidth that it could consume. This would establish a limit to its data rates and, as a result, lessen 'problematic' applications' contributions to router congestion. These techniques have raised concerns: there is a fear that analyzing packets using DPI to assign packet priority levels may actually worsen congestion by ultimately requiring higher-levels of packet retransmission than would occur without DPI-enhanced analysis[45] and that such analysis may not identify the real cause of congestion, the expansion of router buffers to hold more and more packets for transmission instead of

---

[42] One Internet service provider in Canada, Rogers Communications, currently modifies data traffic to alter customers when they are nearing their permitted monthly data volume allowance.

[43] The Minnesota Internet Traffic Studies research group and Cisco alike publish expected bandwidth growth rates, and both typically project roughly similar rates. For more, see: http://www.dtc.umn.edu/mints/

[44] It is important to note that dropped packets are a common event in digital networks. Each packet as a sequencing number and when a client does not receive a packet that composes a larger aggregate communication it will request that the packet be resent. Resent packets may take an alternate pathway to their destination, thus avoiding the previously congested network link.

[45] M. C. Riley and B. Scott. (2009). "Deep Packet Inspection: The end of the Internet as we know it?" *Freepress*. Last accessed: June 18, 2011. Online: <http://www.freepress.net/files/Deep_Packet_Inspection_The_End_of_the_Internet_As_We_Know_It.pdf>

dropping them more rapidly.[46] Such concerns have not prevented network administrators from installing DPI equipment in their networks, nor from monitoring and acting on data packets.

In the approaches noted above, the network operator has made some kind of decision about the appropriateness of the applications that end-users are employing: either some applications are more important than others, or some are identified as problematic and thus have special rules crafted to mediate their ability to generate congestion. Using DPI a network operator can also shift focus from the application to the user. In this situation an administrator might establish conditions concerning how clients can utilize available bandwidth. As an example, when a client used their maximum allotted bandwidth for a 15-minute interval they might have *all* of their traffic deprioritized or delayed for a period of time following the interval. This has the effect of prioritizing 'bursty' traffic, that which transmits data in short intervals rather than over a long period of time. Accessing webpages generates bursty traffic, whereas long file transfers using either peer-to-peer applications or a file transfer client are non-bursty types of traffic. This user-centric approach can be seen as 'application agnostic', insofar as it does not target specific applications, though the rule set will disproportionately affect some applications, such as peer-to-peer and FTP clients, over others, such as web browsing clients. Taken together, it is apparent that DPI equipment provides network administrators with tools to better secure their networks, implement access requirements, and enhance quality of service for some applications. Whether this is a prominent driver for the actual *adoption* of these technologies, however, will be explored in subsequent chapters.

Economic Potentials of DPI

The ability to examine and act upon data packets in real-time affords new revenue opportunities for ISPs and third-parties alike, as well as offers measures to ways to curtail threats to revenue maximization. Specifically, Internet service providers may be motivated to offer differential service plans that compete based on what applications customers can use to connect to the web, the priority that applications' packets are given at routers, or the speed at which users can access websites. ISPs may also prioritize their own 'value added' services, such as voice over Internet protocol, email, or home security systems, over services offered by their competitors. Parties other than network owners may also be interested in DPI: copyright holders may try to limit the sharing of files that infringe on copyrights, and advertisers may monitor and mine data traffic to identify consumer habits and subsequently modify packets to serve targeted ads.

ISPs have long offered differential service plans since dial-up modem pools were used to connect to the Internet. Today, broadband connections mean that ISPs compete based on the rate that data is exchanged between the client's location and the Internet, the volume of data they are permitted to transfer each month, value added services such as email

[46] The expansion of router buffers to hold more packets is referred to a 'bufferbloat' and causes high levels of latency which may, in turn, worsen Internet connections. Bufferbloat afflicts both client devices, such as home computers, mobile phones, and anything else with a TCP/IP stack, as well as routing devices. For more, see J. Gettys (2011). "Bufferbloat: Dark Buffers in the Internet," *IEEE Internet Computing* 15(3). The project investigating bufferbloat is online at: <http://www.bufferbloat.net/projects/bloat>

accounts, and cost. DPI lets ISPs further distinguish their offerings by selectively letting applications connect to the Internet; a web browser and email client connect might be included in a 'basic' Internet package, whereas video game applications or streaming music applications might be included in a 'premium' package. The fungibility of DPI, and deep integration with policy control servers, affords advantages over prior networking technologies, such as MPI, insofar as the same device is better able to mediate multiple different data forms and formats. Further, whereas some data-types, such as web browsing, or data sources, such as a national online newspaper, might not be counted towards a monthly data quota, other data-types and sources could.[47] Alternately, an ISP could limit or prevent access to the Internet unless customers pay for each connected device; DPI can be used to examine data traffic and ascertain whether 'registered' or 'unregistered' devices are attempting to access the Internet and, in the case of unregistered devices, limit their access until a fee is paid. Figure four gives a theoretical example of what these kinds of pricing formats might look like.

Figure 4: A tiered 'app-based' pricing model for the Internet[48]

[47] For a brief report on these kinds of differentiations of service, see N. Anderson. (2010). "Can ISPs charge more to make gaming less laggy? They already do," *Ars Technica*. Published December 15, 2010. Online: <http://arstechnica.com/tech-policy/news/2010/12/can-isps-charge-more-to-make-gaming-work-better-they-already-do.ars>

[48] Image produced by 'Quink' and first made available October 28, 2009 at http://www.reddit.com/r/pics/comments/9yj1f/heres_a_new_scenario_i_just_created_illustrating/

This limitation by device is part of an 'app-model' for the Internet, where connectivity is bundled with a particular application, such as an online movie watching application, or a particular device, such as a PC or tablet computer. In an app-based model, users may never see how much bandwidth volume or capacity they are afforded and instead only enjoy selective access to the Internet based on the services paid for on a monthly basis.[49] This approach to Internet pricing might be combined with, or supplemented by, a prioritization of an ISP's own services to the detriment of competitors. The ISP's voice over Internet protocol client, or a client belonging to a company that had paid an ISP, might be 'free' with the basic package whereas competitors' VoIP traffic is given a lower priority. This approach could buttress an ISP's complementary products or enhance revenue when competitors pressure those complementary product lines.[50] DPI could be used to identify favored applications and give them preferential treatment by guaranteeing higher levels of priority, making larger volumes of bandwidth available to them, or by not counting the data they generate against users' monthly volume limits. An ISP's exclusion of competing services or rent-seeking is logical from the stance of economics. More specifically, "[a]s long as the exclusion of rival from its Internet-service customers translates into more sales of its complementary product, and the additional profits are larger than the costs of exclusion, exclusion will be a profitable strategy."[51] Given the relative prevalence of viral, malware, and spyware the exclusion of competing applications may be couched simultaneously in the language of service and security, masking core economic drivers behind the mask of technical improvements to the network.

DPI also provides copyright holders with a tool to (try to) limit or monitor the traffic of infringing computer files and data streams that course across the Internet. To date, most analyses of infringing data traffic rely on questionable statistics or shoddy methodologies. In the case of the former, the United States' Government Accountability Office (GAO) has publicly rebuffed the monetary losses that American corporations claim to experience from infringement. The GAO notes that for widely cited statistics there are no studies that support estimated losses, and that efforts to evaluate actual losses suffer from methodological limitations.[52] The introduction of detailed packet analysis equipment begins to resolve some of the methodological problems associated with quantifying infringing data traffic; by monitoring packets and cross-referencing them against their point of origin – are they from 'legitimate' digital retailers – and their contents – are the files copyrighted – it is possible to develop an index of how much data traffic is

---

[49] N. Anderson. (2010). "Imagine a world where every app has its own data plan," *Ars Technica*. Published December 15, 2010. Online: <http://arstechnica.com/tech-policy/news/2010/12/net-neutrality-nightmare-a-world-where-every-app-has-its-own-data-plan.ars>

[50] C. Parsons, A. Ly, S. Anderson, S. Sinnott. (2011). "The Open Internet: Open for Business and Economic Growth," *Casting and Open Net: A Leading-Edge Approach to Canada's Digital Future*. S. Anderson and R. Yeo (eds.). Online: <http://openmedia.ca/files/OpenNetReport_ENG_Web.pdf>. Pp. 107.

[51] B. van Schewick. (2010). *Internet Architecture and Innovation*. Cambridge, Mass.: The MIT Press. Pp. 253.

[52] United States Government Accountability Office. (2010). "Intellectual Property: Observations on Effects to Quantify the Economic Effects of Counterfeit and Pirated Goods." United States Government.

potentially infringing.[53] If the copyright monitoring system isn't intended to prevent the movement of data, but merely log it, then a DPI system could be established to do a quick analysis of packets to identify their likely contents. Where it identifies the packets as potentially holding infringing content they could be passed to their destination, while copies were made and stored in a short-term offline storage system. Once in that system, a computer program could develop a hash value for the files and compare it against a known list of copyrighted files. Where the file was protected under copyright and the source of the transmission was an illegitimate online content provider the storage system could call on the subscriber database, correlate the subscriber's personal information with the inappropriate exchange of infringing material, and notify an ISP administrator or member of council, copyright holders, authorities, or some other designated party.

One problem with using a hash-based analytic system is that minor changes in the file can result in different values being generated.[54] These values would not align with the database of known hashes, and thus the DPI or offline analysis system would not identify the files as potentially infringing. To identify files that have had slight modifications, or elements of files that have been combined to create a 'mash-up' of multiple content sources, file fingerprinting could be employed. Because fingerprinting is a computationally expensive process it is not tenable to fingerprint files in real-time. It is, however, useful of offline search and analysis of files.[55] If DPI were used to 'prescreen' data traffic that might be mobilizing infringing data – perhaps targeting applications and application-types that are believed to be prominently involved in moving infringing material – then an offline analysis, tied to a database with content fingerprints and subscriber database associated personal information with instances of infringement, could be used to monitor and react to the transfer of copyrighted content.

Alternately, if copyright holders have identified a particular application or protocol as principally involved in exchanges of copyrighted material then they might demand that DPI equipment scan packets for that application or protocol. Upon detecting 'suspicious' packets the equipment might block the packets, degrade their priority levels, delay their transmission speeds, or inject 'reset' packets into the data stream. By injecting reset packets a connection between clients is terminated, thus ending the transfer of potentially infringing data between the clients involved in the transaction.[56] Resetting connections

---

[53] C. Parsons. (2009). "Aggregating Information About CView," *Technology, Thoughts, and Trinkets*. Published December 17, 2009. Online: <http://www.christopher-parsons.com/blog/privacy/aggregating-information-about-cview/>

[54] It should be noted that, while small changes can modify a hash value, for most infringing works there are 'only' 3-6 popular variants on the Internet at any time. While further changes might prevent perfect monitoring and enforcement of copyright-related policies, arguably a significant amount of infringing data transfers could theoretically be identified. For more, see: K. Mochalski, H. Schulze, and F. Stummer. (2009). "Copyright Protection in the Internet (Whitepaper)," *ipoque*. Online: <http://www.ipoque.com/sites/default/files/mediafiles/documents/white-paper-copyright-protection-internet.pdf>

[55] K. Mochalski, H. Schulze, and F. Stummer. (2009). "Copyright Protection in the Internet (Whitepaper)," *ipoque*. Online: <http://www.ipoque.com/sites/default/files/mediafiles/documents/white-paper-copyright-protection-internet.pdf>. Pp. 4.

[56] RadiSys. (2010). "DPI: Deep packet inspection motivations, technology, and approached for improving broadband service provider ROI," *RadiSys*. Online:

between applications can also be used to disrupt large-scale data transfers that are believed to contribute to congestion at nodes in the network.[57] While copyright holders may be independently motivated to 'encourage' using DPI to address copyright infringement such motivations may be enhanced where network operators are *also* rights holders. In such a case, limiting copyright infringement might be positioned as ensuring user security – protecting users against malware-ridden files integrated with music files users are interested in – as well as ensuring 'appropriate' uses of the network, all while protecting content-based revenue streams that might be reduced by copyright infringing behaviours.

The injection of foreign code into data transfers can also facilitate enhanced behavioral advertising systems. Behavioural advertising is the "practice of tracking consumers' online activities to target advertising to individual consumers based on their online history, preferences and attributes."[58] When DPI is used to facilitate advertising it can modify data packets that customers request from the Internet and add a tracking code to otherwise legitimate data traffic. To do this the DPI router will conduct a series of packet redirects, as described below.

1. A user tries to request access to a website but, if the requestors machine does not already have a cookie – a small text-based computer file – that is associated with the DPI equipment the request is redirected to the DPI router.
2. At the DPI router the user's machine is assigned an identifier and then routed to a another element of the advertiser's network, where they receive a cookie that mimics those presented by the requested website but contain a unique tracking code.
3. The user is finally presented with the website they had requested, but now possess a modified first-party cookie[59] that is used to track online activities and, based on the activities, insert advertisements that are intended to resonate with the user's online behaviours.[60]

<http://www.radisys.com/Documents/papers/DPI_WP_Final.pdf>, Pp. 3. For a discussion on detecting packet injections, see S. Schoen. (2007). "Detecting Packet Injection: A guide to observing packet spoofing by ISPs," *Electronic Frontier Foundation*. Online: <https://www.eff.org/files/packet_injection.pdf>

[57] M. C. Riley and B. Scott. (2009). "Deep Packet Inspection: The end of the Internet as we know it?" *Freepress*. Last accessed: June 18, 2011. Online: <http://www.freepress.net/files/Deep_Packet_Inspection_The_End_of_the_Internet_As_We_Know_It.pdf>, Pp. 4-5.

[58] J. Lo. (2009). "A "Do Not Track List" for Canada?" *Public Interest Advocacy Clinic*. Published October 2009. Online: <http://www.piac.ca/files/dntl_final_website.pdf>, Pp. 4.

[59] There are generally two kinds of cookies; first-party and third-party. The former are used by websites to maintain session information and are useful in 'remembering' that a user has logged into a website as they navigate through it, or to maintain an online shopping cart. Third-party cookies are used by different servers than the website you are visiting, and are often used for advertising and analytics purposes. To clarify, when you visit CBC.ca and log into the website, a first-party CBC cookie will be placed on your computer so that you remain logged into the website as you navigate between pages. Simply by visiting the website you will also have third-party cookies from Doubleclick – Google's advertising company –placed on your computer to target ads based on past online behaviours.

[60] For a full step-by-step analysis of how this system works, see: R. Clayton. (2008). "The Phorm "Webwise" System." Last revised May 18, 2008. Online: <http://www.cl.cam.ac.uk/~rnc1/080518-phorm.pdf>

Alternately, a DPI-based system could use the methodology below:

1. Tie a subscriber's customer record that is maintained by an ISP to a unique hash code that lets the advertiser uniquely and persistently identify individuals without ever having access to the personal information associated with the ISP's records.
2. The DPI system monitors the subscriber's online web activity, which will include examining web pages that are browsed to, search terms that are entered, and words that appear on web pages. Using this information the advertising network will identify the subscriber's interests according to pre-set categories. The intelligence developed about the subscriber is associated with the unique hash code that was generated.
3. The DPI appliance will ensure that the subscriber's web browser is preloaded with cookies that uniquely identify the subscriber. Where partners of the advertising firm have purchased ad space, the presence of the preloaded cookies lets advertisers display targeted advertisements. Even after deleting cookies on a computer the DPI appliance will reload the cookie, with the unique identifier, as soon as the subscriber opens another web browsing session.[61]

Regardless of whether the first or second approach is taken to track and advertise to consumers, the presence of DPI technology is a mandatory component to this mode of advertising. For these approaches, monitoring users' transactions online demands placing cookies on computers in a way that users cannot prevent. The process of modifying data streams and packet contents to inject tracking code is only possible using technologies that penetrate the payload of a packet, and this is only possible using DPI-based technologies.

Political Potentials of DPI

States have been invested in monitoring and analyzing citizens' telecommunications since the telegraph, to the point of retaining encrypted text and banning certain modes of communications for fear that they would undermine state surveillance. "Most European countries, for example, forbade the use of codes except by governments, and in Prussia there was even a rule that all copies of all messages had to be kept by the telegraph company. There were also various rules about which languages telegrams could be sent in: any unapproved language was regarded as a code."[62] Whereas telegraph operators had to personally examine telegrams for inappropriate means of communication, or forward baskets of messages to state authorities for subsequent evaluation, DPI lets network operators monitor communications remotely and in real-time for content of interest. Given its capacity to monitor the content of communications, DPI can be helpful in supporting 'lawful access' legislation and limiting the transmission of content the state has outlawed.

---

[61] R. M. Topolski. (2008). "NebuAd and Partner ISPs: Wiretapping, Forgery, and Browser Hijacking," *Free Press* and *Public Knowledge*. Published June 18, 2008. Online: <http://www.freepress.net/files/NebuAd_Report.pdf>. Pp. 2-3.

[62] T. Standage. (1998). *The Victorian Internet: The remarkable story of the telegraph and the nineteenth century's on-line pioneers*," New York: Walker and Company. Pp. 111.

Lawful access legislation enhances policing and intelligence powers. There are typically three types of access powers associated with such legislation: search and seizure provisions, interception of private communications powers, and production of subscriber data.[63] Deep packet inspection equipment is most useful in intercepting communications, and can be thought analogously as installing wiretap capabilities into digital networks.[64] By installing DPI routers at key points in ISPs' networks it is theoretically possible to remotely monitor communications of those suspected of engaging in illegal acts by making copies of all data traffic or specifically targeting one type of traffic (e.g. VoIP, web browsing, or peer-to-peer) and not logging or monitoring traffic that falls outside of the specified rule set. It is important to recognize that, while on the one hand using DPI might be seen as the logical technology to facilitate state-based surveillance, this mode of monitoring differs from traditional wiretapping capabilities because of the breadth of communications that occur online. Whereas a traditional wiretap would capture voice communications, DPI-facilitated surveillance can capture and perform front-line analysis on *any* type of digital transaction, be it a voice communication, text-based chat, web browsing session, or any other kind of non-encrypted transmission. As such, DPI-based 'wiretapping' arguably stretches what it meant by wiretapping a considerable degree, and may not constitute 'maintenance' of state surveillance powers but an expansion of it.[65]

As private copyright holders may be motivated to monitor for infringing files coursing across digital networks for civil reasons, the government may be concerned with monitoring and preventing content transmission it has deemed illegal. Using techniques similar to those exercised to monitor for copyright infringement, but with policies designed to take action on data traffic rather than just watching the wire for it, government could try and blacklist files known to contain child pornography, viruses, malware, disapproved encryption protocols, confidential or secret government documents, and so forth. Blocking or monitoring content could take the format of a government requiring certain routing equipment be installed in network providers' infrastructure or demanding that those same providers install and operate the equipment on the government's behalf.

The relative fungability of DPI-based analysis and blocking technologies may be appealing to governments; for ISPs already using DPI for their own business purposes a 'minor modification' that repurposes existing systems may prove an easier political win than forcing entirely new technical systems to combat particular content- and traffic-types on Internet intermediaries. The mixing of legal and illegal/disliked content has led

---

63 CIPPIC. (2007). "What is "lawful access?" Last updated June 2, 2007. Online: <http://www.cippic.ca/en/projects-cases/lawful-access/#LA01>

64 S. Lerman Langlois. (2009). "Net Neutrality and Deep Packet Inspection: Discourse and Practice," *Deep Packet Inspection: A Collection of Essays from Industry Experts*. Ottawa: Office of the Privacy Commissioner of Canada. Pp. 25-26.

65 Policy Engagement Network. (2009). "Briefing on the Interception Modernisation Programme," *London School of Economics and Political Science.* Online: < http://www.lse.ac.uk/collections/informationSystems/research/policyEngagement/IMP_Briefing.pdf>

Western governments to limit what is blocked online;[66] while various governments may *want* to block content, such as pornography generally, an attempt to do so would risk *overblocking* insofar as could also limit access to non-pornographic content. This having been said, these same nations around the world – including Canada, the US, and UK[67] – already block certain (limited) content on the Internet such as child pornography; it is not far-fetched that these nations could force the adoption of next-generation technologies to build upon and enhance existing blocking regimes.

The political capacity to monitor, mine, and censor for certain data traffic will almost certainly depend on framing. Governments have historically used the language of safety, security, and order to justify blocking communications content. This language of "securitization," a process whereby issues, problems, and phenomena are defined in "security" terms and associated with a "protectionist reflex" can be used to legitimize extraordinary means to solve a perceived problem.[68] While state agents could be responsible for ensuring that content is appropriately mediated, it is possible that the same end – blocking content – could be achieved by a shift towards intermediary liability.

Under such a liability approach "the *intermediaries*, or companies transmitting or hosting users' communications or other content, are held *liable* for their users' and customers' behavior."[69] As noted by Morozov, intermediary liability is attractive to government because "[i]t's the companies who incur all the costs, it's the companies who do the dirty work, and it's the companies who eventually get blamed by the users."[70] Companies' awareness of their technical capabilities, combined with their (perceived) protection from individual complaints about violations of freedoms of speech and association, make them the ideal party to which to outsource Internet censorship. Of course, a widespread shift to this liability structure – where ISPs are held accountable for what their subscribers transmit and receive – would constitute a significant transition away from common carrier protections.

Such protections, in theory, immunize ISPs from legal liabilities for what their subscribers transmit so long as the ISPs themselves are not aware of what their networks are carrying. A shift towards ISP liability, however, would effectively mandate awareness of what traffic is being carried. Such a shift might serve to largely formalize already existing practices: today social networking companies, ISPs, journalism sites, and other interactive content communities often censor or block the sharing and posting of content deemed offensive or problematic by the organization in question. Scaling the magnitude

---

[66] J. Goldsmith and T.Wu. (2006). *Who Controls the Internet? Illusions of a Borderless World*. Toronto: Oxford University Press. Pp. 83-4.
[67] See the country summaries for more detailed. R. Deibert, J. Palfry, R. Rohozinsky, and J. Zittrain. (2010). *Access Controlled: The Shaping of Power, Rights, and Rule in Cyberspace*. Cambridge, Mass.: The MIT Press.
[68] U Beck. (1998). *World Risk Society*. Cambridge, UK: Polity.
[69] R. MacKinnon. (2012). *Consent of the Networked: The Worldwide Struggle for Internet Freedom*. New York: Basic Books. Pp. 93.
[70] E. Morozov. (2011). *The Net Delusion: The Dark Side of Internet Freedom*. New York: Public Affairs. Pp. 101.

of what is blocked, or reported to authorities, and formalizing the existence of such policies may constitute a quantitative shift but not necessarily a qualitative one.

When simultaneously considering the technical, economic, and political potentialities of deep packet inspection technologies it's helpful to keep in mind that the potentials uses of the technology may not necessarily be practically *instantiated* in real-world networking situations. Further, we can also see how some of the "pure" technical capabilities are infused with the values of control and awareness of the network, and those advocating that the technology be used to meet technical, economic, or political goals may differentially express such values. It is only as we move into our case studies, however, that we will ascertain both the specific drivers and configurations of technologies *as well as* whether the potentialities of the technology can be, or are being, practically instantiated in the real world.

DPI as a Surveillance Technology

DPI devices are more, however, than just controlling and monitoring technologies: they are surveillance technologies. DPI provides network operators with heightened capacities to survey data flows at broad, network-wide, and particular, user-specific, levels. As a result, these technologies are not just concerned with the capture of personal information but are also involved in broader ordering processes. In what follows in this section I briefly discuss the monitoring of the individual – and why it matters – and then turn to address aggregate ordering and its significance. Together, this means of understanding surveillance processes will let us, in the case studies, ask whether instantiations of DPI reveal surveillance of the individual or broader subscriber base, and whether focuses on either the individual or group affect the language used to frame DPI.

Lyon defines surveillance as "the focused, systematic and routine attention to personal details for purposes of influence, management, protection or detection." It is also "deliberate and depends on certain protocols and techniques."[71] Insofar as surveillance is focused on the individual, the individual may self-consciously reduce the scope of their actions and behaviours. This is corroborated by scholars; Judith Wagner DeCew, an American privacy and legal scholar, argues that the "surveillance of normal, everyday activities can lead one to be distracted and feel inhibited."[72] Further, Julie Cohen warns that "[p]ervasive monitoring of every move or false start will, at the margin, incline choices toward the bland and mainstream." Persistent, individually-targeted, surveillance thus "threatens to chill the expression of eclectic individuality, but also, gradually, to dampen the force of our aspiration to it."[73]

Recent contributions to the surveillance and privacy literatures take pains to recognize contemporary surveillance – and, correspondingly, privacy infringements – as not

---

[71] D. Lyon. (2007). *Surveillance Studies: An Overview*. Cambridge, UK: Polity. Pp. 14.
[72] Wagner DeCew, Judith. (1997). *In Pursuit of Privacy: Law, Ethics, and the Rise of Technology*. Ithica, New York: Cornell University Press.
[73] Cohen, Julie. (2007). "Examined Lives: Informational Privacy and the Subject as Object," 52 *Stanford Law Review* 1373.

equivalent to Orwell's 'Big Brother'. Solove suggests that we adopt the metaphor of Kafka's *The Trial* to understand surveillance and privacy invasion, on the basis that it is not a single actor that watches or acts upon us, but instead a set of often shadowy or hidden actors using direct and indirect means alike to influence individuals *and* the groups they are associated with.[74] In a common vein, Haggerty and Ericson propose that surveillance studies ought to study the 'assemblage', or the groups, parties, technologies, practices, and discourses that, in aggregate, constitute contemporary surveillance. They further suggest that surveillance technologies "do not monitor people *qua* individuals, but instead operate through processes of disassembling and reassembling. People are broken down into a series of discrete informational flows which are stabilized and captured according to pre-established classificatory criteria."[75] Together, these authors' writings are suggestive that no particular, singular, body be seen for surveillance or overall privacy-impacting actions but, instead, nuance and complexity of surveillance behaviours and practices must be sought out and understood. For our purposes, this means that the technologies, practices, policies, and potentialities of network surveillance have to be read in relation to one another to create a complete story, rather than focusing on any one particular element of the network surveillance.

But while a multitude of actors can collaborate to monitor individuals, the same – or different – actors can also surveil networks to derive aggregate insights into the subscriber base. Surveillance of data can extend to 'dataveillance' if there is a "systematic use of personal data systems in the investigation or monitoring of the actions or communications of one or more persons."[76] The integration of personal and "non-personal" data can be linked to processes of social sorting, where only the smallest of facets of individuals are sorted into profiles that correlate with the interests of the actors conducting and overseeing the surveillance. The process of sorting is, at a high level, meant to "plan, predict and prevent by classifying and assessing those profiles and risks."[77] Importantly, the profiles that are established tend to be non-transparent to individuals that are affected by the profiles; knowledge of why a credit score was weakened, or internet connection terminated, or particular consumer ads are shown tend to be mysterious or poorly understood. Ultimately, such sorting behaviours have the effect of ordering a population by segregating it into a set of discrete groups that have predicable, and understood, behavioural patterns.

Beyond a discussion of surveillance of the specific or the population is a question of its breadth: how much of an individual's environmental characteristics are paid attention to, and what is and is not watched for? Actors can search for particular, or specific, information about individuals or, alternately, engage in surveillance to "ensnare a

---

[74] Daniel Solove. (2004). *The Digital Person: Technology and Privacy in the Information Age*. New York: New York University Press.

[75] Kevin D. Haggerty and Richard V. Ericson. (2007). "The New Politics of Surveillance and Vsibility," in Kevin D. Haggerty and Richard V. Ericson (Eds). *The New Politics of Surveillance and Visbility*. Toronto: University of Toronto Press. Pp. 2.

[76] R. Clarke. (1988). "Information Technology and Dataveillance," *ACM* 31(5). Pp. 499.

[77] David Lyon. (2003). "Surveillance as social sorting: Computer codes and mobile bodies," in David Lyon (Ed.). *Surveillance as Social Sorting: Privacy, Risk and Digital Discrimination*. New York: Routledge. Pp. 13.

significant amount of information beyond any originally sought."[78] While broad surveillance may accidentally capture information beyond that sought, 'search' surveillance – the specific targeting of an information-type – may provide the surveying party with a deep field of data that is relatively limited in its scope. There is a link between broad and narrow surveillance, insofar as the former "is concerned with groups of people and involves a generalized suspicion that some (as yet unidentified) members of the group may be of interest. Its purposes are to identify individuals who may be worth subjecting to personal surveillance, and to constrain the group's behavior."[79] The distinction between broad and narrow surveillance processes raises questions of the felt and realized impacts of surveillance, and whether multifaceted responses to different calibers of surveillance are needed when addressing surveillance technologies. Moreover, in both broad and narrow surveillance procedures, questions of who is, or may be, discriminated against must also be raised, as must the possibilities of 'social sorting' that may arise following the deployment of DPI technologies. Such discrimination is linked to conceptions of ordering processes: which groups or profiles might individuals be assigned to, and what are the ramifications for such assignments? Does mass surveillance lead to the targeting of individuals, and vice versa?

When we take up DPI as a surveillance technology that is invested with normative and practical concerns we must consider the context in which the surveillance is conducted; surveillance can be empowering or positive when it ensures entitlement or guards against harm to users and disempowering or negative when acting in an opaque manner to damage users' interests. We might imagine the technology being seen positively when used to protect Internet subscribers from malware and viruses, and less positively when its uses are unclear or negatively disrupt legal protocols and content. As a result, how the technology is both specifically deployed at the ISP-level, and governed at the level of the ISP and government regulatory, is key to understanding the normative values guiding its instantiations in network infrastructures. Regardless of its actual uses, however, DPI "enlarges the surveillance toolkit primarily by allowing many more actors to collect data and use them for their own purposes."[80] The purposes to which this toolkit is configures and used for, however, remain in question and will be responded to in our case studies. Only in our concluding chapters will we consider the broader normative questions of appropriateness of integrating these surveillance technologies within democratic nations' communications networks.

Conclusion

While packet inspection technologies are not new in and of themselves, their newest iteration carries with it advances that may have a broad impact on digital communications networks. Whereas SPI and MPI enable network administrators to prevent content from reaching clients, and varying levels of packet awareness, the advent of DPI restructures

---

[78] D. Solove. (2008). *Understanding Privacy*. Cambridge, Mass.: Harvard University Press. Pp. 109.
[79] Clarke, Dataveillance. At http://www.rogerclarke.com/DV/CACM88.html#PDV
[80] B. Jaap-Koops. (2009). "Deep Packet Inspection and the Transparency of Citizens," in *Deep Packet Inspection: A Collection of Essays from Industry Experts*. Ottawa: Office of the Privacy Commissioner of Canada.

the possible range of surveillance that Internet subscribers may be subject to. Gillespie, writing in the context of copyright, argues that some technologies "come to herald and stand for the cultural and economic shifts that helped produce them, and as such they become flashpoints for the legal dispute that follows."[81] As we will see, DPI has reinvigorated economic, cultural, and political discussions surrounding the monitoring, censoring, and modifying of communications content in real-time. DPI's technical potentials have excited a series of prospective policy actors, which operate with different aims and objectives but are mutually interested in framing DPI's usage so that long-term policy decisions are reflective of each actor's particular ends.

Depending on how it is actualized, DPI may function as a beneficial technology that assuages security worries and permits limited mediation of bandwidth usage until infrastructure is provisioned to relieve congestion. Alternately, it could be used to radically undermine the neutrality of networks, where ends have more power over the communications flow than the 'core' that routes data between the networks' ends, to the detriment of the Internet's generativity.[82] There are concerns that DPI could be used to hold ISP subscribers hostage, using the technology to extract higher rents than otherwise possible from content providers[83] or undermine competition between businesses and stymie innovation.[84] The technology could be used to regulate content dissemination, sidestepping legal judgments born of slow court decisions and instead automatically limiting expression.[85] Law enforcement and intelligence might also abuse the technology, and DPI infrastructures they establish could ultimately create gaping communications security vulnerabilities.[86] The range of potentials accompanying DPI are indicative of the interests that might drive its practical instantiations, and remind us that technical artifacts "affect us not merely by dint of physical or material properties but by properties they acquire as systems and devices embedded in larger material and social networks and webs of meaning."[87]

In light of the *potential* impacts associated with this technology, it is important to evaluate the *actual* uses and anticipated potentials of the technology. Moreover, we must ask whether implementations of the technology are uniform and, if they are not, what particular social, economic and political conditions have mediated the application of the technology. To ascertain how and why DPI is deployed and regulated we will, in the next chapter, establish three theoretical frameworks to evaluate the politics of DPI in Canada, the US, and UK. Specifically, we will consider how path determinacy, international

---

[81] T. Gillespie. (2007). *Wired Shut: Copyright and the Shape of Digital Culture*. Cambridge, Mass.: The MIT Press. Pp. 31.
[82] J. Zittrain. (2008). *The Future of the Internet – And How To Stop It*. New Haven: Yale University Press.
[83] B. van Schewick. (2010). *Internet Architecture and Innovation*. Cambridge, Mass.: The MIT Press.
[84] C. Parsons, A. Ly, S. Anderson, S. Sinnott. (2011). "The Open Internet: Open for Business and Economic Growth," *Casting and Open Net: A Leading-Edge Approach to Canada's Digital Future*. S. Anderson and R. Yeo (eds.). Online: <http://openmedia.ca/files/OpenNetReport_ENG_Web.pdf>
[85] M. Mueller. (2010). *Networks and States: The Global Politics of Internet Governance*. Cambridge, Mass.: The MIT Press. Pp. 188.
[86] S. Landau. (2011). *Surveillance or Security: The Risks Posed by New Wiretapping Technologies*. Cambridge, Mass.: The MIT Press.
[87] H. Nissenbaum. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford: Stanford University Press. Pp. 6.

governance, and domestic policy activities may or may not advance Internet-related policies and practices, and where DPI may figure into each framework. We follow these frameworks with case studies of Canadian, American, and British agendas associated with DPI, studies that will identify which potential uses of the technology have been, or are planned to be, instantiated. It is only after unpacking the reality of DPI deployments that we will be closer to understanding the actual, rather than sensationalized or theoretical, politics that are actually driving DPI technologies.