



Understanding Deep Packet Inspection (DPI) Technology

George Ou - Policy Director

Digital Society

Abstract

In recent debates on the issues of privacy and the Internet, Deep Packet Inspection (DPI) has been grossly oversimplified as a technology that is primarily harmful to consumers. Much of the commentary has been based on a poor understanding of what DPI is and how DPI works. All too often the debate over DPI is focused on unrelated issues such as free speech and censorship, which are largely political and not technological issues, and all too often is DPI usage described as a violation of consumer privacy when the reality is far more complex and nuanced. What has been missing is a broader discussion on the full nature and application of DPI technology; DPI enables a wide range of applications most of which are not only positive but also essential to the survival of the Internet. This paper will explain how DPI technology works and explore the potential dangers and benefits of DPI technology to consumers and the Internet.

Introduction to Deep Packet Inspection (DPI)	3
Box 1: What are networks and packets?	3
What is DPI?	3
How and where is DPI implemented?	4
Applications of DPI	5
DPI as the immune system of the Internet	5
Network-based DPI can identify Denial of Service attacks and Worms.....	5
Figure 1 – Observed Denial of Service (DoS) attacks in Q1 2009	6
Figure 2 – Mitigating Denial of Service attacks	6
Figure 3 – Attacks from Conficker and other worms on port 445.....	7
DPI in law enforcement and national security	7
DPI and targeted advertising	7
Debunking Myths About DPI	7
Myth: DPI is the same as a postal worker peaking inside letters while en route	7
Myth: DPI violates inviolate Internet protocols	8
Myth: DPI violates privacy	8
Myth: DPI will start a costly encryption arms race.....	8
Conclusion: DPI is Simply a Tool	9

Introduction to Deep Packet Inspection (DPI)

What are networks and packets?

Before we can begin to define what DPI is, we must first understand what networks and packets are. The term “network” in the context of this paper refers to Internet Protocol (IP) networks that comprise the Internet and the private networks that connect to the Internet. Data transmitted on IP networks have to be broken down into smaller units of information called “packets”—the “P” in DPI. The packet is the smallest unit of data that gets transmitted and routed across an IP network, and technology that inspects those packets on a deeper level is called Deep Packet Inspection (DPI).

From an IP network’s perspective, a packet is broken down into three basic elements.

{IP header [Protocol header (Content)]}

IP header: The outermost portion of the packet where information such as the source and destination IP address resides (analogous to the address information on a postcard). It also contains other useful information about the packet, such as the priority level on delivery.

Protocol header: The header that describes the type of protocol used to transmit the packet. The header has traditionally been used as a basic firewall filtering method essential for basic Internet security. Virtually all home, enterprise, and government networks employ firewalls on their Internet gateways today. Protocol headers are also routinely used in private and public IP networks to help classify network traffic so that each traffic type can be given the performance characteristics it needs.

Content: The payload portion of the packet containing the actual content or data being transmitted.

Packets are analogous to a postcard where all the content associated with the card or message is splayed out in the open.

What is DPI?

Deep Packet Inspection (DPI) is a relatively new technology that inspects the content portion of traffic flowing through the network in real-time. More specifically, DPI technology isn’t just about the inspection of individual packets because it often takes more than one packet to form meaningful content. In fact, it might require analyzing hundreds or thousands of packets before malicious code or activity passing through a DPI device can be detected. This ability to analyze a deep string of packets in real-time and perform high level content analysis is where the “deep” in DPI comes from. Deep does not refer to how deep inside an individual packet a network device inspects, nor does it refer to how extensively or completely is the content is examined. Rather, it refers to how many packets deep the DPI technology analyzes, with the purpose of understanding the flow in order to increase performance both for the particular communication, if appropriate, as well as for all users who are sharing the medium. This is a key distinction that needs to be made in order to correct common misconceptions about DPI technology.

Some critics of DPI technology argue that any device that inspects the content of packets or even just the protocol headers of packets is a violation of someone’s privacy, but this definition is very flawed. For one thing, it conflicts with traditional definitions where protocol header analysis falls under the category of plain old packet filtering, or Stateful Packet Inspection (SPI), which has been used in router¹ and firewall² technology since the

early days of the Internet long before the arrival of DPI. Routers and firewalls even have to look at the content portion of the packet to support traditional protocols like File Transfer Protocol (FTP). This is because computers using FTP initiate a call to each other on one port number³ and then negotiate a secondary dynamic (changing) port number to communicate. Firewalls and routers handling FTP traffic have to inspect communications happening on the first channel so that they can open up the correct port for the secondary port.

The other problem is that critics of DPI technology assume that there is something wrong with content inspection. But if we oppose any sort of content inspection, then we would have to oppose Anti Virus (AV) systems, anti spam systems, or Intrusion Detection Systems (IDS).

How and where is DPI implemented?

DPI is a relatively recent technology because it is only in this decade where widespread need for such capability has arisen. Significant research and development had to go into creating computing hardware that could even begin to analyze millions of packets each second passing through a network. As the technology follows the price curve of Moore's Law, DPI capabilities are increasingly built into a broad range of equipment, from the routers that route packets on the network to dedicated hardware appliances that are installed inside or on the side of the network. But far from all U.S. service providers have actually implemented the capability. In fact, DPI technology remains in the early stage of deployment across the industry.

Some within the Internet community believe that these DPI functions do not belong in the base routing infrastructure of the network and should only be handled by the computers at the edges of the network, but this view is completely unrealistic in the modern Internet where a comprehensive approach to cyber security is needed. Moreover, differences in design philosophy are common in all fields of technology, and these differing views can often be attributed to differing business and commercial interests. There is no right way or wrong way to implement technology, and the history of the Internet has been to allow both the market and a robust system of self-governance via a host of non-governmental organizations like the Internet Engineering Task Force (IETF) to determine the most appropriate solution for any specific problem at hand.

Applications of DPI

DPI as the immune system of the Internet

The mainstream adoption of the Internet has brought along with it all the good and ills of the physical world. Because everything is globally connected over a common network, the Internet is a massive local community of a billion computers. That means every criminal on the planet has the potential to directly access any computer connected to the Internet. Malicious software such as worms and viruses, which were once the handiwork of curious mischief, are now tools for professional cyber criminals. Any computer connected to today's Internet had best be prepared to defend itself against worms, viruses, spam, Denial of Service (DoS) attacks, botnets⁴, websites hosting malicious code, server hijacking, and a myriad of other threats.

The threat to cybersecurity is huge and growing fast, as the current Administration has recognized.⁵ [Symantec](#) published a report in 2008 that found "the release rate of malicious code and other unwanted programs may be exceeding that of legitimate software applications."⁶ F-Secure, a Finnish anti-virus computer company, estimates that as much malware was produced in 2007 as in the previous 20 years altogether.⁷

Consumers rarely have the means to defend themselves against attacks and they need all the help they can get from the network, search, and application provider. Even innocent-looking documents or websites could contain malicious code in them that take advantage of out-of-date software to compromise computers that merely view the document or website. Even the Internet's Domain Name System (DNS), which acts as the Internet's master "phone book" and translates domain names into IP addresses, has been nearly knocked offline. Every possible avenue of attack that can potentially be attacked will eventually be attacked.

To combat all of these threats, a variety of technical means of must be considered to minimize the threats on the Internet. Security professionals call this multilayered strategy of security "defense in depth." It means that defenses must be put onto the server⁸ infrastructure, the consumer's computers, the software, and even the network. Only when all of these elements are implemented together do we have the highest probability of combating all of the threats on the Internet.

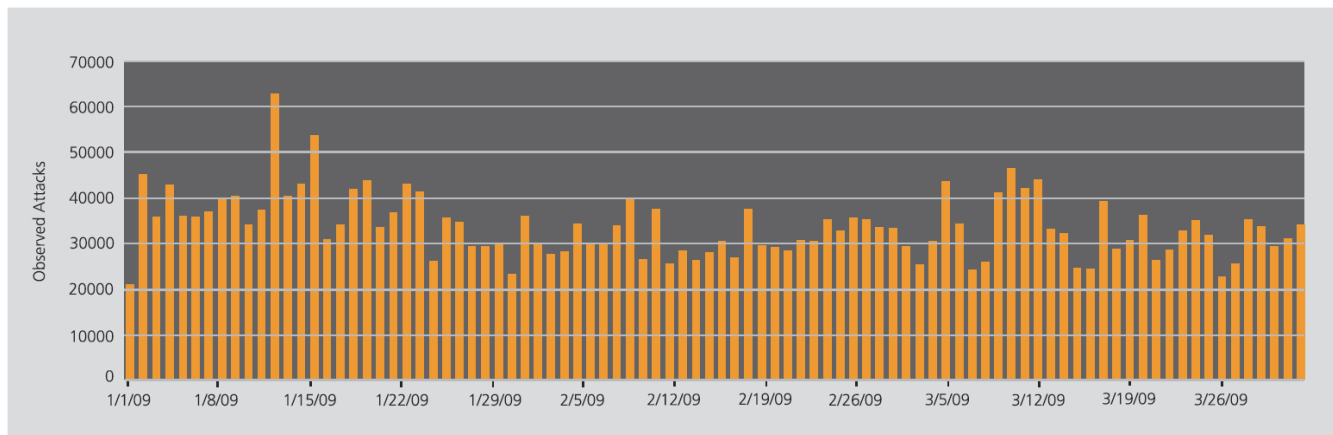
While it's possible to combat spam, viruses, and worms on an individual home or office computer, experience shows that this approach is no match for fast-moving worms and viruses. The vast majority of modern spam protection is also rarely handled on individual computers at the edge of the network. Network-based protections are the predominant way that any corporation or government network manages these security threats. For example, mail gateways or specialized mail appliances will scrub every piece of mail passing through for spam, viruses, or any suspicious payloads. Network Intrusion Detection Systems (NIDS) inspect every packet at great depth to detect hackers, worms, and any other dangerous payloads coming across the network. In modern times, DPI has the potential to become a critical component in the immune system of the Internet.

Network-based DPI can identify Denial of Service attacks and Worms

DoS attacks are a classic example where network operators are best positioned to fight back. A DoS attack occurs when one computer floods another computer on the network through brute force bandwidth saturation or more sophisticated methods that exhausts the victim's resources without exhausting the attacker's resources. Distributed Denial of Service (DDoS) is when multiple computers attack one or more targets overwhelming them with traffic. DDoS attacks are nothing trivial because they can flood networks with

thousands or tens of thousands of megabits of traffic which is the equivalent traffic coming from tens of thousands of broadband enabled homes.

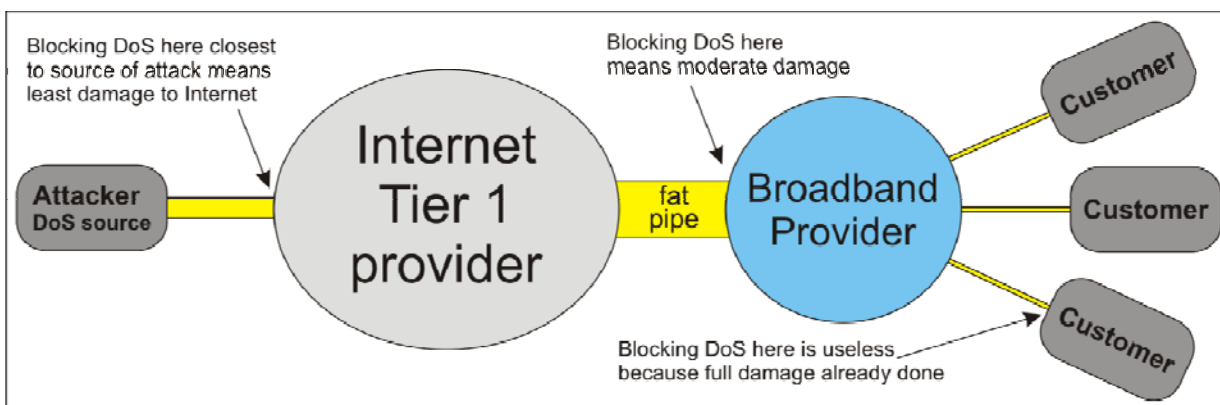
Figure 1 – Observed Denial of Service (DoS) attacks in Q1 2009



Source: Akamai State of the Internet report – First quarter 2009⁹

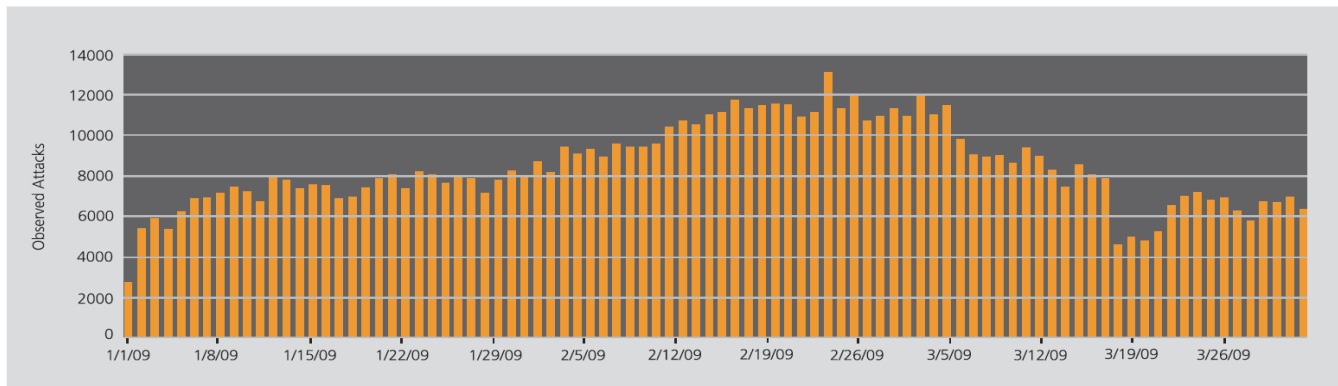
When a DoS attack floods a computer on the Internet, it cannot be stopped at the end-point because the damage is often already done. Network operators routinely block tens of thousands of attacks a day and figure 1 shows how often DoS attacks occur. Network operators often require more sophisticated traffic inspection techniques to minimize false positives (where good traffic is blocked) and false negatives (where bad traffic is let through). Responsible network operators not only block malicious traffic going to their customers, they also block malicious traffic coming from their customers. Figure 2 shows that the closer to the source we block the malicious traffic, the less damage is done to the network.

Figure 2 – Mitigating Denial of Service attacks



Worms like Conficker which have infected up to 15 million computers¹⁰ are another huge problem on the Internet and figure 3 shows how active these worms are on the Internet. The Internet is like a living organism where software worms are like a disease that cannot be cured by only dealing with the problem on the edge of the network. Network based Intrusion Detection Systems (IDS) can effectively detect and block much of this activity by looking for telltale signatures of suspicious activity. Some networks using these systems can quarantine an infected computer trying to infect others until they rid themselves of the worm. SBC (now merged with AT&T) had implemented this type of network IDS system in the past. Comcast announced in October 2009 that it will notify their customers who appear to be part of a botnet infection and offer them tools to rid their computer of the infection¹¹.

Figure 3 – Attacks from Conficker and other worms on port 445



Source: Akamai State of the Internet report – First quarter 2009¹²

DPI in law enforcement and national security

Network based DPI isn't just essential to the Internet's security; it is a key means of complying with U.S. law as provided by the Communications Assistance for Law Enforcement Act (CALEA) of 1994. Telephony providers (traditional or VoIP), which often happen to be the network provider, are required by law to have the capability of providing call records or call tapping. Hunting down cyber criminals often requires the help of Internet Service Providers (ISPs) to track down and provide evidence for conviction.

The Internet has also become the newest international battleground and a new way to engage in corporate and government espionage. Governments and nationalists routinely attack the cyber infrastructure of nations they are hostile to. For many developing nations, acquiring technological knowhow through cyber espionage is as common and acceptable as research and development. Child pornographers distributing their child abusing content often hide behind anonymizer services, and they are almost impossible to track down without the help of DPI fingerprinting mechanisms in the network. All of these law enforcement efforts require very deep and sophisticated network based content inspection, which means DPI is critical to law enforcement and national security and is a critical component of this nation's cyber security.

DPI and targeted advertising

Targeted advertising is just one of the many ways DPI could be used. Yet most of the debate over DPI assumes that DPI and targeted advertising is the same thing. Many assume that targeted advertising is always unethical and a violation of privacy rights when there's nothing wrong with targeted advertising so long as it is disclosed and done with permission.

Debunking Myths About DPI

Myth: DPI is the same as a postal worker peaking inside letters while en route

Critics often cite the example of a U.S. Postal or Fed Ex worker opening your package before delivering it to you when they attack DPI. The analogy is neither apt nor accurate. First, the more apt analogy is with postcards rather than sealed letters and packages. Data packets on the Internet are similar to a postcard, which has its content open and available for all to see. Second, and more important, DPI is not used to "read" the private content in an email – rather, it is used to understand the nature of what's being communicated – voice, email,

or malware, for instance – in order to make appropriate decisions about how best to deliver it. This is similar to how a picture postcard might be handled differently than a text-only postcard in order to avoid damage to the image.

Myth: DPI violates inviolate Internet protocols

Critics have argued that DPI violates long agreed-upon standards and principles of the Internet’s design.¹³ In light of all the legitimate and essential applications of DPI technology we have discussed in this paper, this view of DPI and the Internet bears no resemblance to today’s reality. The Internet is indeed a wonderful system and enables forms of interaction never before possible. But the Internet is simply a very sophisticated machine that continues to evolve and advance. The glue that holds the Internet together is less any particular set of protocols or standards, and much more the set of agreements between operators of Autonomous Systems to meet and share packets at Internet Exchange Centers. These agreements are slowly evolving from a blanket pact to cross boundaries with no particular regard for Quality of Service into a richer system that preserves delivery requirements, protects privacy and provides security. The ability of these agreements to evolve and advance and to incorporate new technologies like DPI is entirely consistent with the structure of the IP packet, the needs of new applications and the history and culture of innovation that has long characterized the Internet’s operation.

Myth: DPI violates privacy

Many in the debate over DPI argue that any network device inspecting user content constitutes a violation of privacy, suggesting that DPI is like an eavesdropper listening in on phone calls.¹⁴ This logic is flawed. If we define any device that scans content as “eavesdropping,” then we would need to outlaw the anti-spam, anti-virus and Intrusion Detection Systems that are crucial to the immune system of the Internet.

Concerns that DPI must necessarily and inherently violate privacy are misguided. An automated computer system that scans for known patterns of malicious software or spam is not a violation of anyone’s privacy; in fact, it’s less invasive than an airport security checkpoint. The airline scanner is manned by human operators who look inside everyone’s suitcases, which may contain embarrassing items for some people. DPI on the other hand doesn’t involve humans looking at the content. To take the airline analogy a bit farther, an airline scanner sees and identifies everything in the suitcase while DPI would see *only* what it has been programmed to see and nothing else. In this context, DPI would function like an airline scanner that only sees items prohibited from being carried on the plane. There is no necessary violation of privacy if the purpose of a DPI content analysis is to protect Internet users or to tailor ads for them with permission. The key distinction here is whether the practice is done with permission and not with the practice itself.

Myth: DPI will start a costly encryption arms race

Some argue that DPI will stimulate an encryption arms race as sophisticated users take steps to prevent ISPs from “reading” their email and that the resulting cost of encryption might raise the bar for doing business on the Internet.¹⁵ A vastly more likely scenario is that individuals and companies conducting legitimate business over the Internet and downloading legal applications and content would see a benefit if DPI were to be applied in judicious and responsible ways. They would not engage in any more encryption than they do today. The real threat that consumers face when websites aren’t encrypted is cybercriminals so widespread adoption of encryption is just as important for Internet cybersecurity as the deployment of DPI enabled network defenses.

Even if encryption were to become a widespread and standard practice, the concern about rising costs is misguided. Given the advances in computing technology the cost of encryption has been negligible for many years. Moreover, the heaviest computational load occurs when the encryption session is being setup during the key exchange phase but that workload is already being handled today during secure login. Leaving the bulk encryption on is barely noticeable because bulk encryption is thousands of times easier to handle than the initial setup. The use of encryption may prevent the caching of web sites (when client computers store repetitive data to avoid retransmissions), but this can be alleviated by leaving public elements of the web site unencrypted.

Conclusion: DPI is Simply a Tool

DPI technology is neutral in and of itself. It is just a tool, and like all tools it can be used for both good and evil. It is the use of DPI that should be debated, not whether DPI in and of itself is a good or bad technology. An example of abuse is when governments censor speech and persecute citizens for their speech on the Internet. But that is an issue with the policies and practices of particular governments, not DPI technology. Citizens of free nations can protest and pressure their governments to change their policy and/or elect new government. The debate over DPI has been shaped by misconceptions and misunderstandings. DPI should not automatically be viewed as a violation of anyone's privacy and while there is some potential for abuse, the applications envisioned for DPI are primarily good and crucial to the survival of the Internet.

¹ A router is a device that routes packets on IP networks such as the Internet to their intended destination.

² A firewall is a network device that sits on the border of an internet networks and the external Internet. It offers some basic border protection to internal networks used by homes, corporations, organizations, and governments.

³ IP addresses are analogous to phone numbers while port numbers can be compared to phone extensions. This allows a single IP address to establish tens of thousands of simultaneous connections using different port numbers.

⁴ A botnet is an army of compromised computers as large as thousands or even millions that have been taken over by criminals. These botnets can perform everything from spam distribution to denial of service attacks.

⁵ http://www.whitehouse.gov/assets/documents/Cyberspace_Policy_Review_final.pdf

⁶ [Symantec Internet Security Threat Report: Trends for July-December 2007 \(Executive Summary\);](http://eval.symantec.com/mktginfo/enterprise/white_papers/b-whitepaper_exec_summary_internet_security_threat_report_xiii_04-2008.en-us.pdf)

http://eval.symantec.com/mktginfo/enterprise/white_papers/b-whitepaper_exec_summary_internet_security_threat_report_xiii_04-2008.en-us.pdf

⁷ F-Secure Corporation (December 4, 2007). "[F-Secure Reports Amount of Malware Grew by 100% during 2007](http://www.f-secure.com/en_EMEA/)".

http://www.f-secure.com/en_EMEA/

⁸ A server on the Internet serves content or applications to Internet users.

⁹ Akamai, "State of the Internet report", <http://www.akamai.com/stateoftheinternet/>

¹⁰ UPI, "Virus strikes 16 million PCs, Jan 26, 2009, http://www.upi.com/Top_News/2009/01/26/Virus-strikes-15-million-PCs/UPI-19421232924206/

¹¹ George Ou, "Comcast heading the right direction on cybersecurity", Oct 9, 2009,

<http://www.digitalsociety.org/2009/10/comcast-heading-th-right-direction-on-cybersecurity/>

¹² Akamai, "State of the Internet report", <http://www.akamai.com/stateoftheinternet/>

¹³ House Subcommittee on Telecommunications and the Internet Hearing, "What Your Broadband Provider Knows About Your Web Use", July 17 2008,

<http://energycommerce.edgeboss.net/download/energycommerce/071708.ti.web.use.hrg.wmv>

¹⁴ House Subcommittee on Telecommunications and the Internet Hearing, "What Your Broadband Provider Knows About Your Web Use", July 17 2008,

<http://energycommerce.edgeboss.net/download/energycommerce/071708.ti.web.use.hrg.wmv>

¹⁵ House Subcommittee on Telecommunications and the Internet Hearing, "What Your Broadband Provider Knows About Your Web Use", July 17 2008,

<http://energycommerce.edgeboss.net/download/energycommerce/071708.ti.web.use.hrg.wmv>