

## **Report**

# **ISP Traffic Management Technologies: The State of the Art**

Prepared by

Graham Finnie  
Chief Analyst, *Heavy Reading*



[www.heavyreading.com](http://www.heavyreading.com)

On behalf of

The Canadian Radio Television and Telecommunications  
Commission (CRTC)

*January 2009*

The views expressed in this report are those of the author and in no way bind the CRTC. The CRTC was not involved in formulating any of the positions contained in this report.

Catalogue No. BC92-68/2009E-PDF

ISBN # 978-1-100-11839-0

## TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY .....</b>	<b>3</b>
<b>I. THE EVOLUTION OF TRAFFIC MANAGEMENT.....</b>	<b>5</b>
<b>II. CURRENT TECHNOLOGIES: THE EMERGENCE OF DPI AND POLICY CONTROL.....</b>	<b>8</b>
2.1 Deep Packet Inspection .....	8
2.2 Deep Flow Inspection And Related Techniques .....	11
2.3 Policy Control and Management .....	11
<b>III. THE FUTURE OF TRAFFIC MANAGEMENT TECHNOLOGIES .....</b>	<b>15</b>

## LIST OF FIGURES

### SECTION I

Figure 1: Selected Vendors in the DPI Area .....	10
Figure 2: Key Standards Organizations and Policy Standards .....	12
Figure 3: Selected Vendors in the Policy Control Area .....	14
Figure 4: A Vision Of Future Service Provision.....	16

## Executive Summary

Over the past few years, the ability of Internet Service Providers, Telcos, Cable TV companies and others to manage Internet traffic running on their networks has improved rapidly. In particular, they can now manage traffic in such a way that certain Internet applications, services or subscribers get preferential access to bandwidth, while others get less or none at all. And they can manage the way that traffic flows to and from specific Internet addresses or subscribers, basing this management if necessary on the applications that subscribers are using.

The purpose of this report is to review the state of the art in traffic management, looking in particular at current and emerging techniques and their potential for improving the ability of ISPs to manage Internet traffic.

In Section I, we briefly examine the history of traffic management, which has its origins in efforts by service providers to improve the performance of Internet Protocol (IP) applications running on large corporate and commercial ISP networks. We also define and explain some basic concepts such as “over-provisioning”, and widely used techniques such as DiffServ and MPLS, and explain why some service providers believe these are no longer adequate in themselves for managing traffic loads. This section also defines some basic concepts in traffic management such as traffic blocking, bandwidth throttling, and shaping.

In Section II, we look at the evolution of current core traffic management concepts and technologies, focusing in particular on deep packet inspection (DPI) and policy control. This section identifies the leading vendors in these areas, looks at the relationships among these various technologies, identifies important technical standards and analyzes the uses to which these technologies are being put in practice.

Finally, in Section III, we look briefly at the future of these technologies and how they are evolving, and identify some key trends that might influence future development. Among other things, these trends include subscriber-centric and mobile-centric traffic management, possible new standards, and integration of existing tools both with each other and into telecoms equipment.

The most important findings of this report are as follows:

- **Traffic management technologies are becoming more intelligent**, allowing relatively fine-grained discrimination among the applications running on IP networks, and allowing service providers if they wish to manage traffic at the level of the individual subscriber.
- **The most important current technology for traffic management is deep packet inspection (DPI)**. DPI is becoming widely deployed because it allows ISPs to identify applications more accurately and because it has a broad range of potential applications, including handling security threats.
- **Compared to other aspects of telecommunications, there are few industry standards for DPI**, making it harder to judge what is possible in this area, but also creating a dynamic competitive environment where the range of options and capabilities available to service providers is growing.

- **Though not yet widely deployed, there is growing interest in and demand for policy servers and architectures, especially among Tier 1 telcos.** Policy servers can handle a broader range of management tasks more flexibly than DPI, and are mostly part of an internationally recognized and standardized architecture. Policy servers tend to focus more on the subscriber than the application.
- **The most important current application for these technologies is management of peer to peer (P2P) traffic and applications.** As well as this, technologies are also often used to give preferential treatment to an ISP's own applications or premium-tier subscribers, to manage the growing range of security threats, and for traffic monitoring and analysis, among other things.
- **Technology development and market demand is shifting from applications management to subscriber management.** Managing at the subscriber level gives service providers more options, and is linked to emerging concepts such as identity management.
- **Development focus is also shifting to meeting the needs of mobile service providers.** As cellular network operators launch flat-rate mobile broadband services, they are seeking more help in managing traffic, applications and subscribers, and vendors are responding.
- **Technologies are continuing to evolve as new ideas such as deep flow inspection are added.** Many of these emerging technologies are focused either on improving the ability to recognize applications, or improving the ability to respond to changes in the behavior of individual subscribers.

# I. The Evolution of Traffic Management

Vendors of traffic management technology often describe the Internet as a “best effort” network, because there is no absolute guarantee that packets transferred over an Internet Protocol (IP) network will arrive in such a way that the underlying application (for example, Web browsing or video streaming) performs consistently or acceptably. The core Transmission Control Protocol (TCP) does, however, include congestion control techniques, defined in the Internet Engineering Task Force (IETF) standard, RFC2581<sup>i</sup>, which can be considered a form of traffic management.

Beyond TCP, the most widely established technique used to ameliorate traffic congestion is “over provisioning”, which means that links are dimensioned so that the bandwidth available exceeds the expected peak or average traffic load by a certain margin. The over-provisioning ratio and the way it is calculated varies widely and depends on a wide variety of factors, including underlying network topology or architecture, the volume of traffic or traffic flows, the number of users using it, the kind of users using it, the mix of applications running on it, historic and anticipated variation in traffic loads, and the link owner’s technology and marketing values, among other things.

In a survey of telco and other service providers conducted in 2007<sup>ii</sup>, Heavy Reading found that over-provisioning is still by far the most widely used technique to achieve acceptable Quality of Service (QoS) in IP networks, but it is gradually being supplanted by other traffic management techniques which may ultimately become more important. In fact, our survey also found that the majority of service providers do not believe that over-provisioning is an adequate long-term solution on its own.

This is largely because of the emergence of Internet applications that have especially demanding characteristics—in particular, voice (and audio, e.g, music), and video (including the service provider’s own IPTV services, “over the top” (OTT) video from sites like YouTube, and large-scale video file downloading, often using peer to peer (P2P) software).

Real-time applications such as telephony do not work well in the presence of jitter (variable delay), latency (delay) or packet loss (all common in congested networks using conventional Internet technologies to control congestion). P2P applications, meanwhile, can result in very large increases in demand for bandwidth and disrupted network planning based on over-provisioning.

These developments mean that the over-provisioning ratio may result in unsustainably high bandwidth costs and make other approaches more cost effective. And the endless multiplication of new Internet applications makes it increasingly difficult for network engineers to plan meaningfully without some capability to control (or at least have insight into) the behavior of specific applications and subscribers. And new issues continue to emerge. For example, some P2P programs are said to create a particular problem because they consume as much bandwidth as they can get. Moreover, over-provisioning is less useful in the access network, where problems are increasing because of the higher volumes of upstream (user-to-network) traffic.

The consequence has been an increasing focus on traffic management at the application or subscriber level, which is the main focus of this report. Parallel efforts to improve bandwidth efficiency via traffic engineering continue, but are not considered further in this report.

Several developments have emerged beyond TCP and over-provisioning that are already widely used to help manage traffic at the level of applications. The evolution of the Internet from a purely academic network to a commercial network used by enterprises and others to deliver services

---

<sup>i</sup> See <http://www.ietf.org/rfc/rfc2581.txt>

<sup>ii</sup> Private multi-client survey of 100 wireless and wireline telcos and cable MSOs, March 2007

with critical technical needs led, for example, (via a number of steps) to the development of two IETF standards called DiffServ and Multiprotocol Label Switching (MPLS).

DiffServ is a relatively simple scheme for classifying traffic into a small number of service or application classes in order to give some traffic types priority. A typical DiffServ classification distinguishes between voice over IP (VoIP), time-sensitive transaction-oriented traffic, and best-effort traffic (eg email).

DiffServ is often used in association with MPLS, which adds a label to packets traversing an IP network. Among other things, this allows different Classes of Service (CoS) to be applied to different applications, so that certain applications (such as voice over IP) get priority if a link is congested.

Both DiffServ and MPLS were created about ten years ago and are now widely used by enterprises in private networks that use the Internet Protocol, and by some ISPs to control the quality of certain services. Variants including MPLS-TE (traffic engineering), and T-MPLS, which is specifically designed to route IP services like telephony.

Meanwhile, a whole range of more generic techniques has emerged and are being applied at various points in the network, from the end user computer, device or gateway through to the core network.

Typical techniques in use include

- **Capping**, which meters and limits the amount of upstream or downstream capacity that an individual user can consume over a specific time period (such as a month).
- **Blocking**, typically applied to traffic that is considered undesirable such as spam, or to applications considered to be consuming too much bandwidth
- **Throttling**, which applies controls to the amount of traffic flowing into a network in a specific period, buffering (storing) the packets or if necessary dropping packets
- **Shaping**, a more complex set of techniques which can control the volume of traffic, the rate at which it is flowing and so on.

The ability to apply these controls on an application-specific basis has been possible for some years, using technologies like DiffServ, but has developed in sophistication over time.

In particular, techniques such as Deep Packet Inspection (DPI) extend the ability to recognize applications and protocols, while policy management creates a structured framework for applying a wide range of policies to subscribers, applications and network flows.

In simple terms, service providers can control traffic along three dimensions:

- **Network**: for example, by bandwidth throttling when traffic reaches a certain volume, with no reference to the underlying application or the source/destination or subscriber
- **Service or Application**: for example, identifying and blocking an application, or giving some applications higher priority than others, especially in periods of network congestion
- **Subscriber**: for example, applying limits to how much bandwidth a particular subscriber can use, regardless of the applications running.

As this description shows, traffic management can be dependent on **or** independent of the application in use, or the individual subscriber using it. Moreover, the relatively sophisticated equipment now available allows these previously separate capabilities to be combined, so that individual subscriber usage of a specific application is controlled, based on the type of subscription or tier or other information.

It is also worth noting that the issues in traffic management, and the solutions, are partly determined by the underlying network and its characteristics. Major telcos, cellular network operators, broadband ISPs and cable TV companies all may have different issues to address. Although there is a lot of commonality in the solutions, the degree to which they have been deployed, and the way they have been deployed, varies.

The use of more complex techniques like DPI and Policy Servers has for example been quite widespread in the cable TV industry. Most cable companies have evolved into so-called cable multi-service operators, or MSOs, offering Internet access and telephony as well as TV. Since cable networks were originally designed just for one-way TV signals, this has required extensive re-engineering of networks, but the basic architecture of a cable TV network, often called “tree and branch” is still different from the architectures used in telco networks, and makes traffic management more of a priority for cable companies. In cable networks, more of the access network bandwidth is shared than in a telco network, creating a greater incentive for cable companies to control access to that bandwidth.

ISPs using telco DSL networks also are making fairly wide use of DPI.

More recently, attention has turned to cellular mobile network operators, largely because these companies face some of the same challenges as cable MSOs (shared or limited capacity links), and partly because most cellular mobile networks are being upgraded to offer high-speed broadband services, including Internet access. Many vendors of the equipment discussed in this report believe this is their biggest opportunity going forward.

Although these technologies are attracting wide attention, they are not universally deployed. In a 2007 survey<sup>i</sup>, Heavy Reading found that about 25% of respondents from about 100 wireline, wireless and cable companies said they were using DPI, with only a small proportion using policy servers—though in both cases there was strong intention to deploy more of this equipment in future, and use has probably spread since then.

---

<sup>i</sup> Private multi-client survey of 100 wireless & wireline telcos and cable MSOs, March 2007

## II. Current Technologies: The Emergence of DPI and Policy Control

In this section, we look at the evolution of several current traffic management concepts and technologies, focusing in particular on DPI and policy control. This section identifies the leading vendors in these areas, looks at the relationships among the various technologies, identifies important technical standards and analyzes the uses to which these technologies are being put.

### 2.1 Deep Packet Inspection

Internet traffic takes the form of discrete packets of data which contain not only the content (eg voice, email etc) but also other information that identifies where the packet has come from and where it is going to, among other things.

DPI as a concept first emerged around the year 2000, but because it is not a standardized technique, its meaning is somewhat elastic. At one level, inspection of packets can be said to be “deep” if it achieves the basic objective of recognizing the underlying application that the packet is carrying. How it does this is usually proprietary and often confidential, and is a core source of differentiation among vendors. However, the term DPI is often related to the seven layer Open Systems Interconnection (OSI) model for communications, which divides the task of interconnecting computer systems into seven layers.

According to this definition, “shallow” data inspection only looks at easy-to-examine information such as source and destination IP addresses in Layer 3 headers. Deep packet inspection looks at Layer 4-7 headers, as well as relevant information in the packet payload itself, and often looks at a sequence of related packets in order to form a more complete picture. This allows for better identification of the underlying application. In fact DPI is often referred to as a “Layer 7” technology because Layer 7 is the “Application” layer in the OSI model.

In sum, DPI equipment inspects the contents of packets traveling across an IP network. It can more or less accurately identify the application or protocol in use by examining the source and destination IP address, the port number, and packet payload. Port numbers are a basic means for identifying applications; for example, email using the Simple Message Transfer Protocol (SMTP) uses port 25. Packet headers include this information, along with source and destination address and other data including DiffServ class information where relevant. The packet payload itself (eg part of a Web page) may be examined to look for strings in the protocol that identify it (eg “kazaa”, which appears in one of the fields used to handle Kazaa requests). Equipment may also look for telltale signs of an application, such as the length of the packet payload.

Putting all this information together, DPI equipment can identify applications with varying levels of accuracy. DPI is a black art in which both false positives and false negatives are unavoidable, but in which the benefits still outweigh the drawbacks for many buyers.

DPI was originally used mainly as an offline traffic monitoring and planning tool which analyzed traffic to help service providers understand what applications were consuming the bandwidth, and how that was changing. This was the core application, for example, for equipment supplied by early entrants such as P-Cube (now part of Cisco) and Sandvine.

Today, however, DPI is probably most widely used to identify and control the bandwidth available to certain applications—in other words, for real-time traffic management. Most importantly, it is used to identify peer to peer (P2P) protocols such as Bit Torrent, so that it can if necessary be blocked, throttled or shaped in some way.

However, this is by no means the only application of DPI. Other applications include identifying and blocking malicious applications and security threats; improving the performance of “critical”,

premium or paid-for applications; applying parental controls on a subscriber by subscriber basis; provision of tiered services; personalized advertising; and so on. And some equipment comes with an applications development environment that can be used by its purchasers to develop their own applications.

Roughly speaking, the use of DPI is evolving along the following timeline:

- Off-line tool to analyze network traffic
- Identify and block or shape P2P traffic
- Handle security threats or nuisances such as Distributed Denial of Service (DDoS) attacks
- Service tiering and premium service control
- Parental control & URL filtering
- Personalized advertising, targeted service offers etc
- Third party service management

DPI is a dynamic technology field, and the range of equipment and capabilities available is wide. Heavy Reading recently identified 12 vendors that were selling self-developed DPI equipment to service providers and telcos (there are others that sell only to enterprises building their own networks). Figure 1 identifies some key vendors, and shows that for the time being small specialists dominate the field.

A Light Reading report published in 2008 valued the 2008 worldwide DPI market at about \$450m, and projected it to grow to just over \$1 billion by 2012.<sup>i</sup> It argued that much of the growth would come from the use of DPI by cellular mobile network operators.

Vendors of DPI equipment compete on a range of technical criteria, including:

- The number of application or protocol “signatures” they have compiled—usually many hundreds in a typical signature “library”. They also compete on their ability to recognize protocols that have been deliberately disguised by Web sites or applications—something that is discussed in more detail in the next section on DFI— as well as the speed at which they update libraries in light of changing applications and traffic patterns. Many vendors offer a continual updating service based on software releases, analogous to that offered to users of security or virus detection software.
- Whether equipment can be placed “inline”, making it more valuable for real-time traffic management. Inline DPI equipment is deployed directly in the bitstream and acts immediately to block or throttle applications in real time— rather than off-line, in which the device takes data off-line, examines it, cleans it up if necessary or otherwise alters it, and re-injects it into the bitstream. In-line DPI is more difficult to do, but potentially more valuable where speed and low latency are key requirements.
- Size of links or number of subscribers or flows supported, with some vendors claiming to be able to handle traffic on links running at 80Gbit/s or more and to manage networks with millions of customers.
- Computer processing power, which affects the speed at which the DPI equipment identifies and act on information. Cutting-edge computer technologies such as multi-threading and parallel processing are being applied by some vendors.
- Ability to handle security threats such as Distributed Denial of Service (DDoS) attacks.

---

<sup>i</sup> See <http://www.lightreading.com/insider/>

**Figure 1: Selected Vendors in the DPI Area**

<b>VENDOR</b>	<b>COMMENTS</b>
Allot Communications	Relatively broad family of DPI products including centralized server-like product and distributed edge enforcement devices; also now includes security appliance. Customers are mostly smaller ISPs using it for traffic management but also include some larger mobile operators
Arbor Networks	Originally specialized in handling security threats for carriers; broadened out into mainstream DPI with acquisition of DPI vendor Ellacoya in 2008. Majoring on in-line devices, Ellacoya is one of the longest-established and larger DPI specialists
Bivio Networks	Smaller DPI vendor that recently began targeting telcos; focused on high-speed distributed computer processing as the core technology platform, and on programmability of appliances.
Cisco Systems	DPI product is called SCE, originally developed by P-Cube, which Cisco acquired in 2004. Mostly used for monitoring, but now used by some for real-time traffic management as well. New version handles 10Gbit/s links; being integrated into Cisco routers
Cloudshield	Smaller DPI specialist that emphasizes open applications development, especially in security area, based on its own OS and language; available as a blade in IBM BladeCenter. Has a handful of Tier 1 customers using equipment mostly for traffic management and security.
Procera Networks	DPI/DFI specialist that recently started to target Tier 1 telcos; emphasis on self-developed recognition language with more than 800 signatures currently claimed. Customer base mostly small ISPs and municipal telcos using equipment for fair use management.
Radware	Relatively large and well-established DPI & security specialist using both DPI and DFI; primarily focused on controlling quality of telco services and controlling OTT applications, but recently added "behavioural" threat control. Claims many large Tier 1 customers.
Sandvine	Leading DPI specialist that is now broadening products and technology away from core applications monitoring and P2P traffic control; customer base originally focused on north American cable MSOs, but now shifting to wireline & wireless telcos, and other regions.

Initially, DPI equipment was mainly deployed in major aggregation and peering points, but is now being distributed more widely and closer to the customer, which improves the ability to apply controls per subscriber or per service.

Another potentially important trend in DPI is the integration of DPI capabilities into existing devices such as edge routers. Some telecoms equipment vendors, such as Starent and Zeugma Systems, make a virtue of the fact that their equipment has had DPI capabilities from the start; Starent has successfully established a range of core and gateway network platforms for mobile operators that include DPI, while start-up Zeugma make an edge router for network operators that includes DPI.

There are fairly strong linkages between security and DPI that are also noteworthy. Most DPI equipment has some ability to handle security threats, and some are specialists in the area. For example Arbor Networks, the major provider to carriers of flow-based inspection devices for detecting security threats, bought a leading DPI vendor, Ellacoya, in order to integrate the two fields more effectively. Another, Adaptive Mobile, specialized in helping telcos to handle spam, viruses and similar problems, but has broadened this to include any so-called “blended” threats that eat up network resources in unanticipated ways. Others such as Allot Communications have moved in the opposite direction: formerly a DPI specialist, Allot bought a security specialist, Esphion, in 2008 in order to improve its ability to meet telco security requirements.

## **2.2 Deep Flow Inspection And Related Techniques**

Conventional DPI has some significant shortcomings in its ability to recognize applications, and this has led both existing and new vendors to look for ways to augment DPI with new approaches that are more accurate. These new approaches have various names, but for simplicity we bracket them as deep flow inspection, or DFI.

DFI can be thought of as a technique that complements DPI by more accurately identifying underlying applications and protocols. DFI infers the application (or threat) from the behavior of the flow of packets, instead of looking for protocol signatures or port usage in the packet itself. This is important because more and more traffic is encrypted or tunneled through the network, and more and more applications disguise themselves by, for example, using the “wrong” port, making it difficult for conventional DPI to identify the traffic. Some applications have so-called port agility and are not associated with any particular port number; others masquerade as HyperText Transfer Protocol (HTTP) traffic (port 80) but are actually (for example) voice over IP. More and more P2P traffic is encrypted, making it difficult to identify using conventional DPI.

Unlike DPI, DFI is not a standard industry term and there is a range of related techniques variously called behavioral analysis, heuristic analysis, pattern recognition, and so on. But the principle is generally the same: to infer the identity of applications that cannot otherwise be detected by DPI because they have been disguised in some way. For example, one technique is to look at a packet length histogram and compare it to a library of packet length histograms: P2P control layer packets are said to be much shorter than pure HTTP packets.

Almost all vendors are looking towards adding these techniques, or have already done so. Vendors adding or highlighting these capabilities currently include Allot, Anagran, Ericsson, Nokia Siemens Networks, Procera and Radware.

Anagran, a specialist in this area, characterizes its approach as “behavioral traffic management.” Anagran looks at flow characteristics such as rate, shape, duration and size, and uses this along with port numbers, source/destination address and protocol to improve identification.

However, most orthodox DPI vendors are also adding this kind of technology. For instance, Radware includes both DPI and DFI capabilities in its Network Delivery Controller product.

## **2.3 Policy Control and Management**

In principle, policy control is a broader set of techniques than DPI that applies controls to Internet traffic flows (among other things) within a structured and standardized architecture. It has strong appeal for larger telcos and service providers, but is not yet as widely deployed as DPI.

Policy tools are in some ways competitive with DPI equipment, and in others ways complementary. Policy servers can be used alongside DPI, and often are used this way by larger equipment vendors. At the same time, some DPI vendors, such as Allot and Sandvine, are adding new ca-

pabilities, sometimes by acquisition, and re-branding themselves as suppliers of policy tools to broaden their appeal.

Policy control is necessarily a broad concept because it 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, customers and applications. For example: *"If customer A subscribes to the Gold Tier package, and if it is the weekend, then customer A may download unlimited numbers of music videos."* This set of conditions can clearly be extended by simply adding other terms – for example, information on the age of the customer, or on how much he or she has previously downloaded. And dynamic information (e.g., where the customer is, what device is in use, or network conditions at the time) can be added to the rules invoked in a particular case.

As this description implies, an important feature of most policy tools is that they have (or can have) links into subscriber and billing databases, making policy equipment a potentially valuable means for providing more customized services to customers. It also means that policies are often related to subscribers rather than applications; hence one policy specialist, Camiant, emphasizes in its description of its "fair use management" tool that it is "application agnostic" and designed to be used at the subscriber level.

Policy architectures have been standardized over the past 3-4 years by several important international telecommunications standards organizations. These include 3GPP and 3GPP2, which creates standards for cellular mobile network operators; ETSI, which created a policy architecture for wireline telcos as part of its TISPAN architecture for next-generation telco networks; and CableLabs, which created a policy architecture for cable multiservice operators (cable MSOs).

**Figure 2: Key Standards Organizations and Policy Standards**

ORGANIZATION	MAIN OUTPUTS
3GPP	Release 7 (current) PCRF/PCEF architecture; Release 8 now in preparation
ETSI TISPAN	Although TISPAN is now integrated into 3GPP, ETSI initially developed a distinctive policy architecture for Tier 1 telcos deploying wireline NGNs
CableLabs	Within PCMM architecture, Policy Server is defined, and CMTS acts as PEP
OMA	PEEM has developed and defined reusable policy mechanisms for OMA enablers such as presence
Broadband Forum	WT-134 Policy Control Framework now under preparation. Well-established TR-069 auto-configuration spec also relevant here.
IETF	Key AAA standards, in particular Diameter; also developed the Common Open Policy Service (COPS) protocol to support policy control on QOS signaling protocols.

These architectures typically envisage two elements—a Policy Decision Point, or PDP, which is usually (but not invariably) a highly intelligent and compute-intensive centralized device that is usually (but not invariably) associated with a policy rules engine, and the Policy Enforcement Point, or PEP. The job of the PDP, as its name implies, is to make policy decisions on behalf of less intelligent devices. Because rules engines are generic and flexible entities, policies can be invoked to handle an indefinitely wide range of conditions, triggered by the presence of a particular application, a particular subscriber, a destination URL, or indeed any data point that can be identified and effectively acted upon.

There are fewer vendors in the policy server (PDP) area, but they include several well-funded specialists as well as—and importantly—most of the major telecommunications equipment vendors. Figure 3 lists some of these vendors.

In our basic generic architecture, policies are enforced by PEPs, which unlike PDPs are usually (but not invariably) distributed devices closer to the subscriber. A PEP is essentially any piece of subscriber equipment that is capable of enforcing a policy decision – a wide range of equipment that includes broadband-remote access servers (B-RASs), gateway GPRS support nodes (GGSNs), DPI appliances, media gateways, session border controllers (SBCs), and so on. These PEPs vary primarily in the types of enforcement they can perform, from relatively basic (GGSN) to relatively complex (DPI), but the basic principle is always the same.

As discussed earlier, policy standards also include interfaces to equipment such as the Home Location Register (HLR) or Home Subscriber Server (HSS) used in 3G mobile networks. These are repositories of subscriber data, and by referring to subscriber data, policy servers can make policy decisions which directly relate to individual users.

A related area here that is beginning to converge with policy is Authentication, Authorization and Accounting (AAA), which ensures that subscribers are correctly identified, and get only the resources and services to which they are entitled; it may be applied both statically and dynamically, or associated with real-time resource availability. This is important to properly securing networks and differentiating among subscribers based on their type of subscription (e.g., how much bandwidth they have contracted for, with what volume or time limitation, and other criteria.) The more sophisticated the AAA tools, the more differentiation among subscribers is possible.

In fact, the full name of the relevant 3GPP policy standard is the “Policy Charging and Rules Function”, which emphasizes the importance of the ability to charge where appropriate after a policy decision has been made.

As this discussion makes clear, policy tools are more subscriber-centric than DPI tools. In fact, one of the leading specialists in this area, Bridgewater Systems, tags its offers as “subscriber-centric policy management”, emphasizing the company's origins in the AAA area. Bridgewater majors on the ability to understand in real-time a subscriber's “state” before making a policy decision, including, for instance, whether to allocate bandwidth resources. Another specialist, Broadhop, makes a rules-based policy management appliance, which is primarily aimed at simplifying and automating per-subscriber service provisioning and deployment.

One other noteworthy development in this area is Resource Admission Control (RAC), a concept defined in the ETSI TISPAN. Its main purpose in the standard is to allow telephone companies that are implementing all-IP networks to offer a telephone service that emulates orthodox telephone services when calls are set up—that is, when a request to make a call is received, it can deny access with a busy signal if the network is considered to be too congested to carry the call.

This principle of admission control could of course be applied in a range of other contexts, and is beginning to be more widely used.

**Figure 3: Selected Vendors in the Policy Control Area**

<b>VENDOR</b>	<b>COMMENTS</b>
Bridgewater Systems	Subscriber-centric policy management, largely focused on mobile operators, with focus on matching AAA functionality to an understanding of subscriber "state"; wide range of features and functionality supported
Broadhop	Primarily focused on provisioning issues, and offers a wide range of supporting applications such as parental control and plug and play VOIP
Camiant	Specialist provider of centralized policy server that supports all 3GPP and PacketCable standards for PDPs; its core market has been cable MSOs, but it is now targeting wireless and wire-line telcos as well
Ericsson	Most aggressive of telecom equipment manufacturers in entering policy space, with self-designed products in both PDP and PEP area; SAPC supports 3GPP Release 7 spec and interfaces; DPI product both standalone and add-on to products including Red-back routers
Juniper Networks	High priority development area for Juniper, which is integrating teams working on network-oriented and subscriber-oriented product programs; latter includes strong emphasis on identity, seen by Juniper as core to policy; security also seen as key
Nokia Siemens Networks	Converged policy server for both fixed and mobile networks that supports 3GPP Release 7; Flexi ISN is a DPI-based PEP; Nokia Siemens also has HSS and identity products through its acquisition of Apertio
Openet	Focused on mediation and billing for mobile operators, especially content billing; new Policy Manager is the basis for wider entry into policy server space; based on 3GPP Release 7

### III. The Future of Traffic Management Technologies

As Section II has indicated, policy management in general and traffic management in particular is a fast-evolving and dynamic sector. In this section, we look at some of the public plans of major vendors in this area, and relate that to known ISP requirements in order to speculate about the likely role and shape of future techniques.

#### **Making Traffic Management Tools More Subscriber-Centric**

A lot of traffic management today does not refer to information about the subscribers themselves. However, as a generalization, most telcos and cable MSOs are tending to move away from application-specific controls to subscriber- or client-specific controls, in part because technologies are more widely available to do this. As we noted in the last section, policy tools tend to be more subscriber-centric, while DPI vendors are also focusing development in this area.

Subscriber-oriented vendors envisage subscribers picking and choosing from a wide range of options and tuning applications to meet their specific needs – meaning that services are potentially more valued and therefore potentially more valuable. The emphasis shifts from controlling bandwidth costs to retaining and upgrading customers.

Most vendors, both on the DPI and policy side, have moved development in this direction. For instance, DPI vendor Allot offers a so-called “Subscriber Management Platform”, said to allow real-time identification of subscribers (not simply one-off look-ups, Allot emphasizes), rather than just identification of IP addresses. Others such as Ellacoya (now Arbor) have focused on back-end integration with Operational Support Systems (OSS) and Business Support Systems (BSS) using RADIUS, DIAMETER and other Internet AAA standards. This is in line with Arbor's belief that all DPI and policy will move to per-subscriber control.

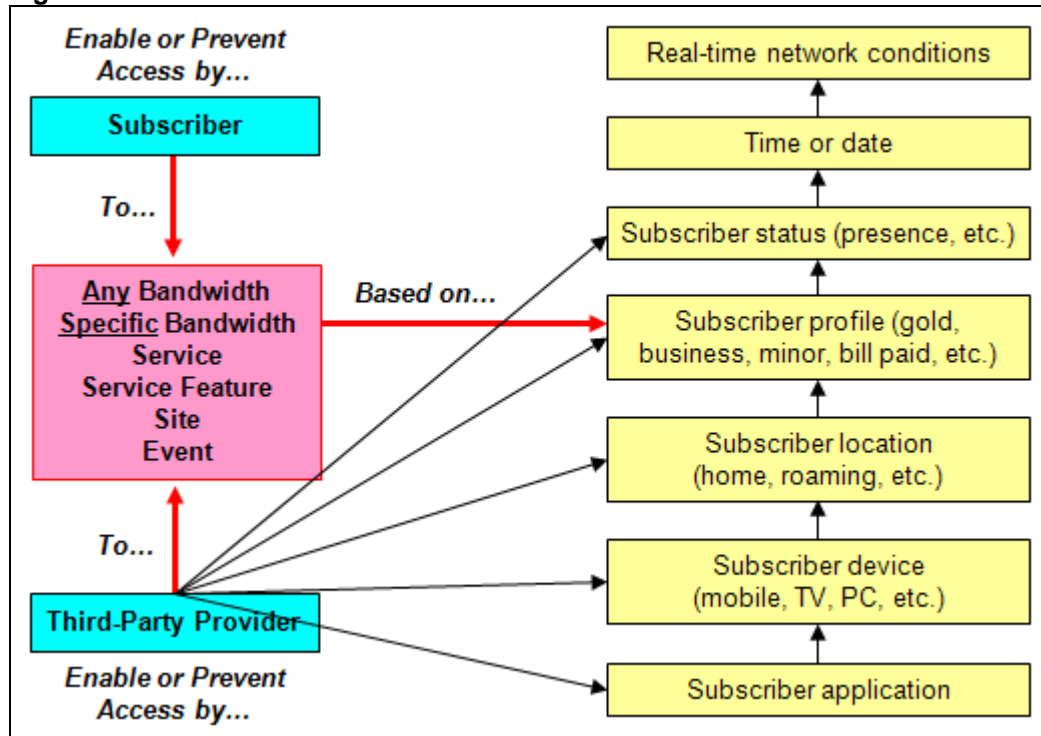
As noted in the last section, policy specialists have often been focused on the subscriber side from the start. Bridgewater's core concept is “subscriber-centric policy management” based on both static profiles and subscriber “state.” State information might include, for example, whether a subscriber is on- or off-network, what kind of network the subscriber is on, and so on. The company emphasizes that it handles both “state” and historical subscriber data on a common core platform.

This trend is to some degree being influenced by work in a related but separate field called identity management. The basic idea in this nascent field is that if a customer has been identified by, say, a mobile network operator as customer A, that information can be federated (under closely controlled conditions) to a third party such as a Web e-commerce site without the need for the customer to identify himself a second (or third) time. Often called “single sign-on,” federated identity typically restricts the 3<sup>rd</sup> party to only the information that is required in a particular case, protecting the user's private data as well as their real identity in most cases.

Many of the ideas here have not been deployed in networks, but do suggest the directions that things might move in. Figure 4 is a vendor-inspired vision of how broad-ranging these capabilities could become, based on currently available technology. As the figure shows, one aspect of this effort is to try to create more constructive relationships with third party Web sites and developers as customers of networks.

The enormous growth in mobile broadband services—and traffic— is the single most significant trend in cellular mobile networking today, and is leading to a re-think of approaches to traffic management.

**Figure 4: A Vision Of Future Service Provision**



### **Making Traffic Management More Mobile-Centric**

Although uses cases may be converging, the technical challenges posed by traffic management in mobile networks are different from those in fixed networks. Most obviously, cellular subscribers move around, creating AAA challenges and increasing the need for, at the least, centralized databases. Another big difference is that bandwidth is divided equally among a larger number of users, which means that handling a large number of subscribers is relatively more important than the size of the link that is controlled. And several vendors note that applying policy control on radio links requires some expertise in the area.

This, along with an interest in fixed-mobile convergence, access-agnostic networking and so-called “three-screen” service provision (across a TV, a PC and a mobile device) is leading most vendors to move toward more converged solutions that work in a variety of contexts. But some caution that the particular problems of working in a wireless radio access network context could trip up some wireline specialists, and vendors vary in the degree to which they are moving in this direction.

### **More Accurate Identification of Applications**

Driven in part by the ongoing battle with botnets, spam and other threats, there is continuing investment in technologies that enable applications to be more accurately identified. This is a classic technology arms race in which there is no ultimate “winner”, but most of the specialists as well as the larger players are investing significant sums in this area. The trend towards behavioral and heuristic techniques is likely to continue, and means that DPI, DFI and related techniques will

likely continue to prove valuable to service providers for the basic task of applications identification, despite widespread efforts to evade identification.

Hostility to DPI, especially among P2P applications providers, may also lead researchers to successfully develop compromises that bridge between the interests of the two communities. For example a German DPI vendor, Ipoque, has developed an application that allows customers to “white list” Bit Torrent files and sites deemed to be legitimate, and other similar efforts could emerge.

### **Potential Impact from Further Internet Standards or Revisions**

Similarly, controversy over net neutrality and handling of P2P traffic has led to renewed interest among Internet researchers in finding solutions that have the approval of the “community”, perhaps in the shape of a new IETF RFC. For example, two IETF Birds of a Feather (BoF) ad hoc group called Transport for Advanced Network Applications (TANA)<sup>i</sup> and Application Layer Traffic Optimization (ALTO)<sup>ii</sup> are looking at different ways to improve the way P2P applications consume Internet bandwidth.

Another project, Proactive network Provider Participation for P2P (P4P)<sup>iii</sup>, aims to allow P2P protocols to communicate with network management systems. The underlying aim here is to create a more constructive relationship between Web applications developers and network providers. P4P has support from a number of major telcos including AT&T, Telefonica and Verizon, as well as Web content developers including Bit Torrent, Joost and Limewire.

However, it is too early to say whether these and other efforts will bear fruit and affect the use of the traffic management tools already in use.

### **Greater Hardware and Software Integration**

We already saw in previous sections that the boundaries between previously distinct categories of equipment and types of vendors in areas including policy, DPI, AAA and charging have begun to blur. That trend will probably continue over the next year or two, perhaps leading to tools that have wider utility.

A related trend is a move towards blade-based equipment rather than stand-alone appliances.

For example The SCE is currently an appliance design, but Cisco says that the long-term direction is a blade-based architecture, with SCE blades added to a router chassis. Cisco believes this will improve performance and scalability, and says that its customers are pressing all suppliers, including Cisco, for fewer boxes.

Hence, DPI functionality is a key new feature of Redback's SmartEdge Multi-Service Edge Router 1200, launched in 2007 (although DPI is only being deployed commercially in the second half of 2008). The 1200 has DPI capabilities built into the box and can handle in-line tasks such as P2P traffic control and threat mitigation at line speeds of 10 Gbit/s.

Some vendors argue that other classes of equipment can handle simple policy decisions directly, meaning that a telco may be able to delay deployment of a sophisticated and more expensive

---

<sup>i</sup> See <http://www.ietf.org/internet-drafts/draft-shalunov-tana-problem-statement-01.txt>

<sup>ii</sup> See <http://tools.ietf.org/html/draft-marocco-alto-problem-statement-00>

<sup>iii</sup> See <http://www.pandonetworks.com/p4p>

policy server solution. For example, Nextpoint claims its SBCs and security gateways can provide a kind of "PDP Lite," handling at least some decisions in lieu of a true PDP. And Cisco noted that big improvements in semiconductors and computing power was making it easier to push the compute power needed for DPI and other policy tasks out to the edge.

**In summary,** this is a highly dynamic field of technology research in which we can expect to see rapid and potentially unexpected development, and continuing improvements in the ability of telcos and service providers to integrate traffic, application and subscriber management.

Appendix A: About the Author

**GRAHAM FINNIE**

**CHIEF ANALYST, *HEAVY READING***

Graham Finnie has been researching telecommunications for over 20 years. He joined *Heavy Reading* in 2004 after a ten-year tenure at the Yankee Group. Among other things, Finnie has been responsible at Heavy Reading for a series of reports on new control technologies including IMS. Most recently, he authored the Heavy Reading report *Policy Control & DPI: The New Broadband Imperative*, September 2008. Finnie has also written widely on next-generation broadband technologies. He was appointed Chief Analyst at Heavy Reading in February 2007. Finnie is based in the U.K. and can be reached at [Finnie@HeavyReading.com](mailto:Finnie@HeavyReading.com).