

## **Building a High-Performance Unified Threat Management appliance**

### **Part I: Hardware issues**

*by Mick Johnson, Sensory Networks*

#### **More Speed, Threats And Functionality**

Network communication is a fact of life for computer users today; working off-line has become almost inconceivable. The Internet pervades our daily life both at home and at work. Of course, the ubiquity of network communication has brought with it correspondingly prevalent threats in the form of viruses, spam, intruders and phishing: the list goes on and on. Whereas it was previously considered sufficient to install a virus-checker on your desktop to consider yourself secure, many businesses today require a multi-layered approach. By providing security checks at both the network gateway and on individual workstations, the risk of a successful attack is much lower. This has led to a huge market for security appliances; a study by IDC research predicted that, by 2007, 80% of all network security solutions will be delivered via a dedicated appliance.<sup>1</sup>

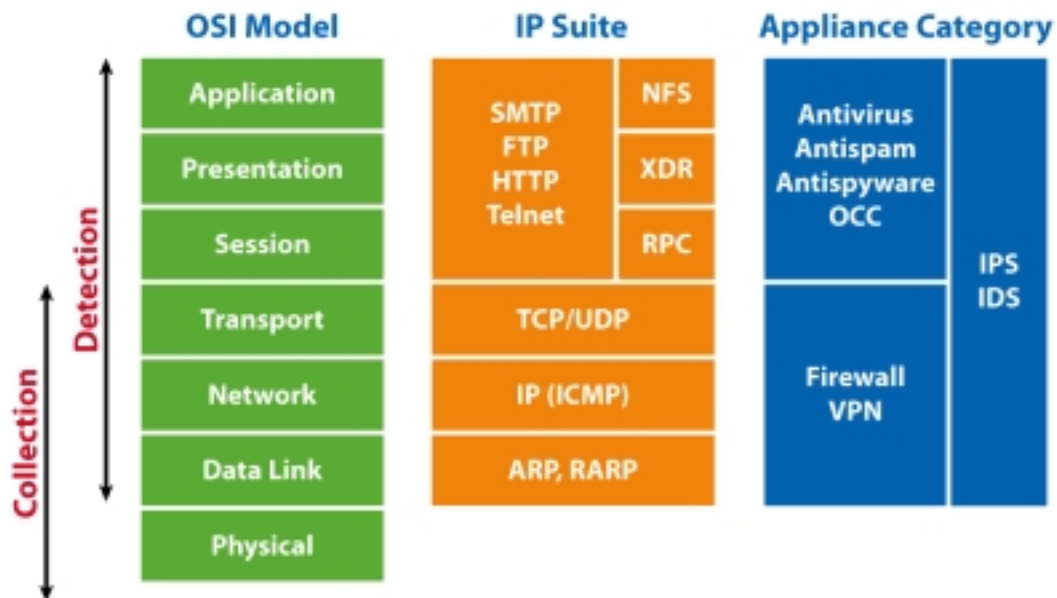
To be effective such devices now have to protect against more and more types of threats. The standard appliance in 2000 delivered primarily firewall and virtual private network (VPN) functions; classifying traffic according to its 5-tuple was considered sufficient for most purposes at the time. As more and more streams of communication open up across the Internet most threats now come higher up the OSI stack. Many applications now tunnel over HTTP or rely on other layer-7 protocols, and any hacker worth his salt can easily split and exploit payloads over multiple packets. Real protection now requires inspecting, decoding and classifying traffic from all OSI layers 2 - 7. Devices that can do so are capable of scanning e-mail and Web traffic as well as processing packet header information. Such devices have been termed unified threat management (UTM) appliances, and the market for such has been predicted by IDC research to overtake that of standard firewall/VPN appliances over the next 5 years<sup>2</sup>, with a compound annual growth rate (CAGR) of 48.7%.

Building a high-performance UTM appliance isn't an easy task however; there are several complicating factors. First, network speeds continue to rise year after year, with Gigabit (G-) Ethernet commonplace in local networks and 10-Gigabit (10G-) Ethernet rapidly becoming so. Second, security threats and their associated costs have also risen; the 2005 FBI Computer Crime Survey estimated that dealing with computer-related crime costs US businesses over \$67 billion a year.<sup>3</sup> Last, each year more functionality must be supported on a single appliance in order to compete in this market. Spyware was largely unheard of before 2004 -- now it's a huge and growing market with several of the existing antivirus vendors attempting to compete with new and upstart companies. VoIP is garnering a great deal of investment in network infrastructure making it an ideal target for both viruses and spam further down the track. With the advent of legislation such as the Gramm-Leach-Bliley, HIPAA and Sarbane-Oxley acts in the US and UK, many businesses are also asking for outbound content compliance (OCC) systems to prevent

the leakage of confidential information. In summary: more speed, more attacks, and more functionality.

### The two tasks: Collection and Detection

The task of scanning traffic at wire speed without missing attacks can be separated into two sub-tasks: collection and detection. Collection involves picking the packets off the wire and processing them through the network stack, reassembling and deciphering packet header information and identifying the relevant payloads. Detection is the task of scanning those payloads for data that signify a particular traffic stream is malicious or unwanted. A given portion of traffic might apply to either collection or detection at different stages: the source IP address must be checked against a set of firewall rules before being used to identify a TCP stream for reassembly and HTTP-level scanning for viruses.



**Fig. 1: Collection And Detection**

These two phases and how they correspond to OSI layers; the various protocols that operate on them and the threat categories that correspond to those protocols are shown in Fig. 1. The term UTM applies to devices that span multiple categories, although other terms exist such as application-layer firewall. For the purposes of this TechNote firewall will refer to a device concerned with classifying and blocking TCP/IP connections based on their 5-tuple.

The three factors identified above; more speed, more attacks, and more functionality, have made the detection phase correspondingly more important. The increase in network traffic has come primarily through more information in packet payloads, while headers have stayed the same size. The increase in attacks has forced designers to spend more time checking different layers of the stack against larger databases. Finally, each added

security function or application in a UTM adds extra workload to the detection phase, irrespective of the amount of traffic. This increased detection workload in today's appliances has led to a massive performance drop-off when any kind of content inspection is turned on. Firewalls that can route packets and block TCP connections at 1 Gbit/s have performance drop to fewer than 300 Mbit/s when deep-inspection is turned on.<sup>4</sup>

## Detection examples: Antivirus, Antispam and IPS

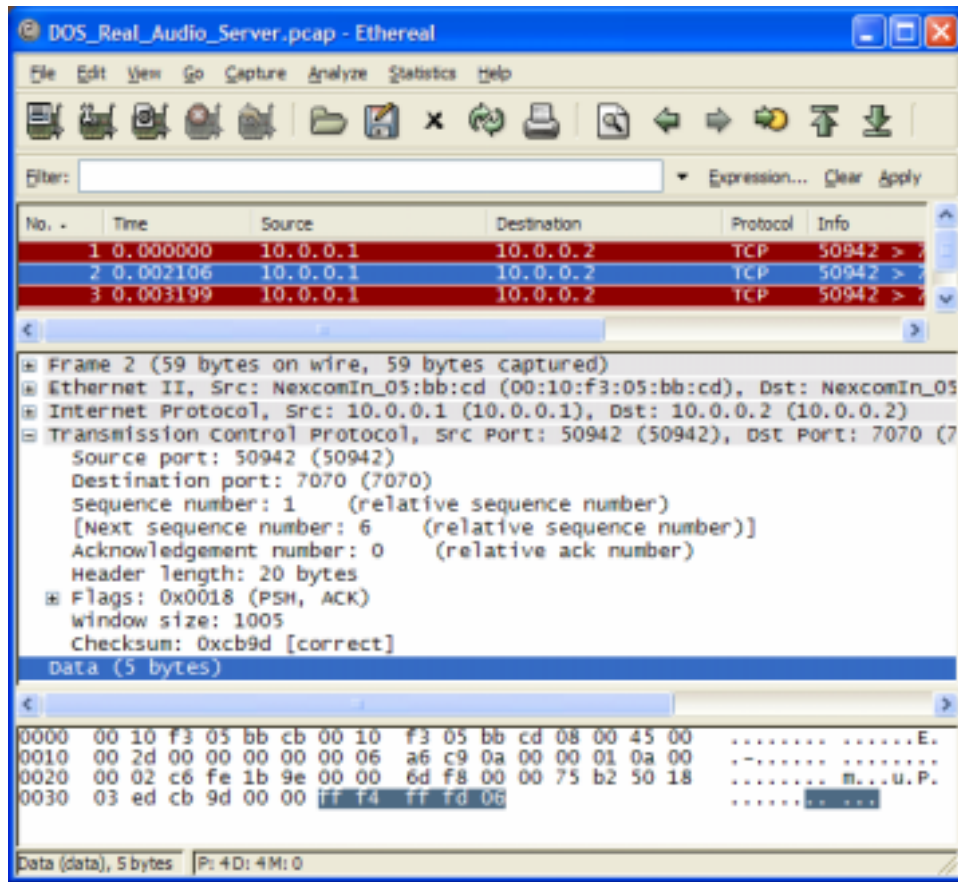
Detecting attacks in traffic is different depending on the type of attack and the OSI layer(s) being checked. A device must at least have antivirus, antispam and IPS/IDS functionality built in to qualify for the category of UTM appliance. Each of these functions operates differently but all use signature database matching as a common task. The popular open-source Snort<sup>5</sup> IPS engine can scan several areas of the network stack for attacks. The following rule is part of the publicly available database and was written to provide defense against denial-of-service (DoS) attacks on a server running Real Audio.<sup>6</sup>

```
alert tcp $EXTERNAL_NET any -> $HOME_NET 7070 (msg:"DOS Real Audio
Server"; flow:to_server,established; content:"|FF F4 FF FD 06|";
reference:arachnids,411; reference:bugtraq,1288; reference:cve,2000-
0474; classtype:attempted-dos; sid:276; rev:5;)
```

The IPS engine performs several checks on packet header information, namely that:

- The source IP comes from outside the network
- The destination IP is inside the network
- The transport layer protocol is TCP
- The destination port is 7070
- The packet is part of an established connection, going to the server

The first 4 conditions deal only with the 5-tuple defining a particular TCP connection, and such checks are well suited to standard firewall architecture. A connection tracking module that keeps track of connection state is necessary to handle the 5<sup>th</sup> condition. However, the rule also specifies that the packet payload must contain **FF F4 FF FD 06** (in hexadecimal). A screen shot of a TCP dump containing such a packet is shown in Figure 2. Processing this payload and scanning it for such signatures is the largest part of the workload for the IPS part of a UTM, and one particularly difficult to maintain high throughput for on a standard hardware architecture. The current Snort library contains approximately 3000 such signatures, and more are added every day.



**Fig. 2: DOS Attack Traffic**

Another key application required for UTM appliances is scanning for viruses and spyware over a variety of protocols such as HTTP and SMTP. This requires packet reassembly and understanding of the application-layer protocol. Like IPS engines, most antivirus systems use signature scanning to detect malicious files. The following is a publicly-available rule from the popular open-source ClamAV project.

```
Trojan.Downloader.Small-140:1:EOF-
73:bb4b104000536a7de82b000000683b10400053e82c0000006a006a00536800104000
6a00e80900000053e80f000000ffd3ccff2510204000ff2504204000ff2508204000ff2
5002040
```

This signature provides protection against a particular type of Trojan downloader by recording a series of bytes, shown again in hexadecimal, common to all known variants of the infected executable. This signature can occur anywhere after the first 73 bytes. There are currently over 30,000 rules in the Clam database, and more are added all the time. The risk associated with virus infection and the ease of distribution makes having an effective antivirus signature database critical for any UTM appliance. However, a single database with tens of thousands of signatures cannot be scanned effectively on standard hardware architecture, due to memory bandwidth and latency constraints.

The final key function required for any UTM appliance is spam-filtering. Today's spam-filters use a combination of statistical and other techniques to classify e-mail messages as spam or not. Below is a standard rule from the 20\_drugs.cf<sup>7</sup> database in the popular open-source Apache SpamAssassin project.

```
body ONLINE_PHARMACY
/\bonline pharmacy|\b(?:drugs|medications) online/i
```

This signature, in regular-expression format, would match on any of the following phrases: *online pharmacy*, *drugs online*, or *medications online*, as long as they began at a word boundary ie *abcbonline pharmacy* would not match. SpamAssassin then combines the weights for all rules triggered by a given message to determine a total weighting. If this weighting is over a given threshold, the message is classified as spam and dropped by the mail transfer agent (MTA). Other techniques include using blacklists for IP addresses and domain names of known spammers. Several such lists are available publicly on the Internet in an attempt to restrict the flow of spam e-mail traffic. A UTM appliance with spam-filtering functionality therefore must look up many large databases including regular-expressions, IP addresses, domain names and ordinary text literals.

Antivirus, antispam and IPS all span both the collection and detection phases, with pattern-matching a common task to all. As more and larger databases are added, this becomes the single largest problem for building a high-performance UTM. The hardware architecture of such a device must therefore treat pattern-matching as one of the most important problems to solve.

### The Key Issues: Memory, Latency, Flexibility

The traditional hardware architecture of a security appliance is much the same as that of a desktop PC. With single- or twin-CPU's, layer-1 (L1) and layer-2 (L2) caches of at most 4 Mbyte, the innards are generally indistinguishable from a workstation apart from several G- or 10G-Ethernet ports and a high-powered Network Interface Card (NIC). The main differences are in the externals such as a rack-mountable chassis and bigger cooling fans. In such a system the NIC captures network traffic, reassembly is handled by the operating system, and content and inspection tasks are handled by user-level applications. This architecture can deliver gigabit performance when the workload is restricted to tasks such as read and write operations on packet headers. However, scanning and pattern-matching on packet payloads has a higher cost in terms of memory bandwidth and latency.

	Size	Access time (cycles)	Access time (ns)
L1	16 kbyte	9	3.16
L2	2048 kbyte	22	7.24
Main memory	3.0 Gbyte	334	111.37

**Fig. 3: Memory Access Times<sup>8</sup>**

Single-core CPU, multi-core CPU and multi-CPU systems all perform best when sequentially-accessing memory or when randomly-accessing purely within L1 and L2 cache (see Fig. 3). A single commercial antivirus signature database can have over 20,000 entries<sup>9</sup> and take up over 5 Mbyte -- too large to fit entirely. Furthermore, lookup tables and databases require random access so a UTM with antivirus, antispware, antispam, IPS and other databases will often be forced to access main memory, stalling the CPU. This inability to fit the necessary databases into fast memory is the one of the key issues in building a high-performance UTM.

The second key issue, system latency, is related to the problem of memory lookups. The vast majority of Internet traffic today uses TCP.<sup>10</sup> The bandwidth and length of Internet infrastructure has led to the exponential back-off nature<sup>11</sup> of TCP becoming increasingly significant in terms of performance. As system latency increases TCP connections back off further, leading to an overall drop in throughput that takes a long time to recover. When pattern-matching is the governing factor in system latency even occasional cache misses will have a disproportionately large effect on throughput for that connection.

Developers of high-performance UTM appliances face 3 key issues: more speed, more attacks, and more functionality. The processing workload of such a device can be split into two phases, collection and detection, of which the latter has become increasingly more important as databases increase in size and number. Three of the key functions for a UTM appliance, antivirus, antispam and IPS, all depend on fast pattern-matching over large and complex signature databases during the detection phase. A high-performance UTM appliance therefore requires a hardware architecture that delivers high-speed memory accesses for multiple massive databases, as well as tight bounds on system latency.

*This was the first part of a two-part series. The second TechNote will focus on the software issues in building a high-performance UTM appliance and strategies for dealing with them.*

## **About the Author**

Mick Johnson is a Product Marketing Manager at Sensory Networks. He has worked in a variety of fields including software engineering, artificial intelligence, and energy-market forecasting. Mick earned a BSc University Medal in Computer Science from the University of Sydney, New South Wales, Australia. He can be reached at [mick@sensorynetworks.com](mailto:mick@sensorynetworks.com)

## **References**

---

<sup>1</sup> Charles J. Kolodgy, Worldwide Threat Management Security Appliances 2005-2009 Forecast and 2004 Vendor Shares

---

<sup>2</sup> Charles J. Kolodgy, Worldwide Threat Management Security Appliances 2005-2009 Forecast and 2004 Vendor Shares

<sup>3</sup> 2005 FBI Computer Crime Survey – <http://www.fbi.gov/publications/ccs2005.pdf>

<sup>4</sup> Deep Inspection Firewalls - Clash of the Titans, Network Computing, 28<sup>th</sup> April 2005, <http://www.networkcomputing.com/showArticle.jhtml?articleID=160910889>

<sup>5</sup> Snort is a registered trademark of Sourcefire, Inc.

<sup>6</sup> Real Audio is a registered trademark of Real Networks, Inc.

<sup>7</sup> This ruleset can be found at [http://spamassassin.apache.org/full/3.0.x/dist/rules/20\\_drugs.cf](http://spamassassin.apache.org/full/3.0.x/dist/rules/20_drugs.cf)  
SpamAssassin is a trademark of the Apache Software Foundation

<sup>8</sup> Tested on a Quad 3.0GHz Xeon machine running a simple CPU loop

<sup>9</sup> Talisker Firewall Antivirus Products - [http://www.securitywizardry.com/av\\_firewall.htm](http://www.securitywizardry.com/av_firewall.htm)

<sup>10</sup> Sean McCreary and KC Claffy, Trends in Wide Area IP Traffic Patterns: A View from Ames Internet Exchange

<sup>11</sup> RFC 793: TCP - <http://www.ibiblio.org/pub/docs/rfc/rfc793.txt>

