Firewall For the Next Generation

Technical White Paper 1.0
Threats in today’s network environment

With network and internet connections becoming more common, network security has also become an important issue and garnered significant public awareness. This is especially when more and more websites and servers are attacked by hackers armed with new and varied hacking techniques.

Online banks are also at the receiving end of threats such as Trojan horses where passwords and account IDs are repeatedly stolen, resulting in financial lost for account holders. The flooding of spam mails not only crippled the whole network economy, programs like malware, ad ware and illegal popup advertisements are also posing endless challenges to network users.

Why are network threats on the rise?

Firstly, more and more broadband services are offering giga-speed lines to end-users, a significant contrast to the past where broadband speed of 2M was considered a premium. Nowadays network backbones are already operating is several G. Unfortunately the significant growth in broadband also brought with it unprecedented growth in network threats, and the resultant damages.

Secondly, network applications that operate above the TCP/UDP layer, such as various IM (chat programs) or P2P software (e.g. EDonkey), are on the rise, and are casually used within the network. These applications could be gateways for network threats such as illegal entry, network invasion, phishing, DOS/DDOS attacks, various malware, spam, etc.
Limitations of the traditional firewall

The traditional firewalls are unable to keep up with the growth of the network and market demands. These firewalls generally pose two shortcomings: low performance and single-function security feature.

**Low performance**

Traditional firewall generally has low performance because all its security functions are accomplished through a single CPU. While today most backbone and enterprise network's broadband are already running in giga-speed, traditional firewalls are at most capable of transacting in mega-speed.

**Single-function**

Traditional firewall equipments are generally single functional, and requires purchase of other network security devices in order to achieve network security: traditional firewall for access control; VPN devices for data encrypt-decryption; IPS products for network invasion and scanning, etc.

Although there are security products in the market that integrates multiple functions, they are software systems that often compromised performance and are unable to deliver a truly integrated solution.

There are three types of firewalls available in the current market: X86-based software firewall, NP-based firewall and the ASIC-based firewall. These firewalls have their individual strength and advantages, as well as obvious shortcomings. Built to cater to specific types of network environments, these three types of firewalls each shares a foothold in the network security market.
The software firewall

As the name suggest, this firewall is a software product based on a single core CPU to accomplish various security functions. To enhance stability, the hardware used is usually a general-purposed CPU on an industrial-strength motherboard, with a standard PCI network card as the external network interface.

In the software firewall, data are transmitted to the CPU via the PCI network card for processing, including performing various security services and related protocols. When the CPU is done, the data are then sent back to the network card via the PCI bus.

When network security products were just on the horizon, software-based firewalls had certain obvious advantages. At that time, users were more concerned with applications and services than performance; to the manufacturer, software firewall’s hardware requirements were simple, low-cost, low fault-rate and the hardware technology involved were less sophisticated. Moreover adding new functions and applications often only involves enhancing or rewriting the original software design while maintaining the same hardware platform, hence development time is short and upgrades are simple affairs.

Disadvantages

But with the explosive growth of the broadband, the shortcomings of the software firewall became more obvious, and the biggest weakness is its processing capability. In some situations the data-processing capability can go up to line-speed, for example, when processing a large packet of 1500 bytes. However when processing small packets of 64 bytes, the performance starts to deteriorate swiftly. Moreover with the increase in the number of security policies to verify against and
the number of hyperlinks to deal with, the software firewall’s performance will take a serious hit. In other words, its performance goes down as the processing load goes up, especially when all these security functions are carried out by a single CPU.

Hence manufacturers begin to use various methods to enhance the software firewall’s processing capability. The simplest and most straightforward method is to use a high performance CPU, e.g. the Xeron chip with server level mainboard, coupled with huge high-speed DDR memory. At the same time, new software techniques can be employed, such as zero-copy data transfer. To enhance key routes, the system could categorize security applications to two levels: Kernel and User Space, where applications in the Kernel level will be given higher priority during operations. However, these methods may ease the situation somewhat, but would not eliminate the problems completely.
The NP-based firewall

**Network processors (NP)**
Since single-CPU cannot solve the performance bottleneck, security product manufacturers begin to look into network processors (NP). Originally, network processors such as Intel’s IXP1200 were developed to solve performance issues between routers and L3 switches during transmission of data. Generally, there are two types of NP. The first type of NP is constructed by integrating multiple CPU in single chip, connected by high-speed bus. When the system is transmitting data, each task is assigned to different CPU as much as possible, achieving a certain degree of load-sharing in order to enhance efficiency. The second type of NP is constructed by integrating multiple microengines and a main CPU core into a single chip. It then utilizes micro-codes to control the microengines for data processing and transmission, while the CPU core is purely used for loading the micro-codes and setting registers. To enhance the NP’s applicability, chip manufacturers often include dedicated accelerating modules to the NP, a dedicated encryption core, etc.

NP-based firewall products can effectively solve the single-CPU performance bottleneck. Performance tests have shown current NP-based firewalls are known to operate in giga speed. However, the introduction of NP into the firewall technology has also introduced new problems.

**Problems with using NP for firewall**
Firstly, NP was developed primarily to solve issues between the routers and L3 switches. When chip designers first developed the NP architecture, the focus was on the router and switch functions. However routers and switches operate on the MAC and IP layer but not the TCP layer. This can be a problem as the
The basic function of a network security product is access control on the TCP/UDP layer. Moreover, security products frequently need to process application layer protocols, and conduct content filtering at the application layer. These were not part of the consideration when designing the NP architecture.

However, this has not stop security products manufacturers from developing firewalls on NP, as they saw the programmability, commonality, and high performance offered by the NP. Still, the first problem faced by firewall developers using NP is in the area of the micro-codes, which are used to control multiple CPUs or microengines. The codes are normally written in C programming language, and then compiled into binary codes using the NP compiler. NP developers usually develop their system micro codes based on the basic standard micro-code supplied by the NP chip manufacturers. These basic codes are usually of very high quality and fully optimized for NP hardware architecture for performance and data transmission. However, as firewall developers are not supplied with such standard codes, they have to develop their own micro codes from scratch, resulting in longer product development cycle. Also, due to the firewall developers’ unfamiliarity with the NP’s hardware capabilities, the micro codes developed are usually inferior and are unable to be fully capitalized on the NP’s potentials. As such, the final security products developed are often inferior, and at times worse off than the firewall utilizing a single CPU.

Furthermore, with multiple CPUs or microengines integrated into one chip, the chip’s surface area is significantly larger, which translate to higher cost and higher difficulty in PCB wiring. These, coupled with the fact that traditional security product manufacturers are software-oriented and thus lack the necessary expertise in hardware development, makes NP-based firewall solutions undesirable.
The ASIC-based firewall

One way to solve the bottleneck issue in the security field is to use ASIC. ASIC are dedicated custom-made chip developed for custom-made security functions. Typically, this security chip takes care of the transmission of data and execution of various security functions, while the CPU takes care of various configuration tasks, exception handling, collection of statistical data, and user interface, etc. With this, ASIC is able to eliminate bottlenecks to achieve full giga speed performance. Data transmission latency is also very short, mostly in microseconds.

Moreover, dedicated security chips are easy to stack, thus enabling performance to double to triple easily. With most functions executed on the chip, PCB wiring are relatively simpler, and the eventual product more stable.

However, as ASIC are custom-made chip, altering it would be extremely difficult. With higher cost and longer development time, ASIC products are unable to adapt to the network security market quickly. Development of ASIC also requires expensive NRE expenses a trained chip design team, which a typical security product manufacturer would not be able to afford.
The next generation firewall

The main issue faced by most current firewall products is: how to provide a wide range of security features while maintaining high performance. NP-based and single-CPU traditional firewall can provide rich security feature, but are seriously lacking in the performance department. On the other hand, ASIC can provide the performance desired, but at the expense of the number of security functions it can provide.

With the limitations of current software, NP-based and ASIC-based firewalls, what the security industry need is a next generation firewall that addresses the issues faced by the current firewall products. The next generation firewall should be capable of high performance yet provides a wide range of security functions. These functions are generally executed the higher layer, such as the application layer.

Data channels and control protocols

Firewall functions can be largely divided into two categories: data channels and control protocols.

Data channels refer to the various data stream that requires services, such as FTP data stream, IPSec encryption, etc. Data channels demands high processing capability and thus requires hardware acceleration. This can be fulfilled with two types of technology, the first being the existing technology in commercial chips that chip manufacturers can provide, such as routers, ARP and MAC. The second type is core security technology such as Session module, IPSec VPN module, etc. In the absence of commercial chips, core security technology can fulfill high performance demand by developing ASIC.
Control protocols refer to the additional protocols required to accomplish data transmission, e.g. IPSec’s IKE protocol, logs, dynamic routing, etc. As control protocols control the transmission of data but not the actual data itself, they do not demand high processing power even though the functions involved may be relatively more complicated. As such control protocols can run on a general-purposed CPU.

**Best of both worlds**

Next generation firewall combines both the hardware and software aspect in its design. Data channels demand for processing power can be fulfilled by placing it in dedicated hardware while control protocols’ flexibility can be accomplished by programming the general-purposed CPU.

Take for example the IPSec VPN.

In the standard Freeswan design, the IPSec VPN is divided into two modules: PLUTO and IKE. PLUTO is responsible for the encryption of data while IKE negotiates IPSec VPN channel’s properties. Hence PLUTO’s encryption tasks requires high performance, and should be accomplished using hardware; while IKE is only negotiating properties, thus not requiring high processing power, and so can be achieved using software.
The next generation firewall uses a proprietary security chip as its core, while utilizes commercial chip and general-purposed high performance CPU. This way, the firewall possess ASIC’s high processing capability and the software’s flexibility, along with some NP-based microengines design. The firewall architecture may look something like this:

The firewall can be divided into two main portions: the system software and the hardware platform. The hardware portion is a combination of the in-house developed ASIC and the commercial chip, along with a general-purposed high-performance CPU.
The proprietary chip accomplishes the main security functions, especially the data applications at the TCP layer and above; the commercial chip accomplishes the transmission of data at the IP layer, and the operation between IP and MAC layer; finally the general-purposed CPU runs the system software. The diagram below illustrates an example of these relationships:

**Programmable chip system - Sentinel**

One of the issues involved when utilizing computer chips to accomplish network security is design flexibility. With network security continuously evolving, the security solutions need also to update themselves constantly. But updating the chips to synchronize with the advancement of security technology can be time and resource-consuming. As such a programmable chip system, such as the Sentinel chip found in the SifoWorks series of firewalls, is an excellent solution.

Sentinel is a highly integrated security chip consisting of four layers of intelligent defense system and an IPSec VPN microengine array. The four-layer intelligent defense system is made up of three layer of programmable hardware and software filtering accelerating engine, a static/dynamic packet filtering
engine and content-matching engine, and a dedicated Tag information mechanism.

The other major module in Sentinel is the multi-protocol VPN engine assembled from the microengine array. As the microengine array has been optimized for IPSec protocol, its processing capability can go up to giga speed.

Conclusion

With the new generation of security technology, such as one utilizing the programmable Sentinel chip in SifoWorks firewall series, it is not impossible to have power, high performance and flexible in one complete firewall solution.
About SifoWorks

SifoWorks is a multi-function firewall system. It combines a core application proxy and a hardware state inspection mechanism to inspect and filter every data packets from the second to the seventh layer. These include data forwarding, classification, route selection, network service classification, security policies and access control, packet signature matching and bandwidth allocation (QoS). To learn more about O2Micro SifoWorks product, visit http://www.o2security.com.

About O2Micro

O2Micro develops and markets innovative power management, and security components and systems for the Computer, Consumer, Industrial, and Communications markets. Since its founding in April 1995, O2Micro has quickly established itself as a leading supplier of specialized devices designed to extend battery operating time, heighten efficiency, and enable secure e-commerce and security.

O2Micro offers ACPI (Advanced Configuration and Power Interface) compliant products. The company designs products compliant with the System Management Bus (SMBus) and Smart Battery System (SBS) specifications, a subset of the ACPI specification (ACPI is an open industry specification co-developed by Intel, Microsoft, and Toshiba).

O2Micro maintains an extensive portfolio of intellectual property, and has numerous trademark Applications and Copyright Registrations. The company maintains offices worldwide including Japan, Taiwan, Singapore, China, Korea and the United States.

Sales office:

O2Micro - California
3118 Patrick Henry Drive
Santa Clara, CA 95054
(408) 987-5920