FULCRUM
microsystems

Load Balancing in Telecom Servers using FocalPoint

Reduces system cost and improves system flexibility

White Paper

March, 2008

Order Number: WP LoadBal

Introduction

Traditional Telecom Server systems contain standard CPU blades running applications such as intrusion detection, VPN routing and spam filtering. However, although they offer flexibility and ease of programmability, standard CPU subsystems or server blades do not have the required processing power to perform deep inspection of packet headers at 10Gb line rates. This is especially true with back-to-back minimum size packets. To solve this problem, today's high performance Telecom Server systems employ specialized CPUs or NPUs to inspect the packet header on dedicated line cards. These specialized processing cards are much more costly than commodity CPU blades. This paper presents a load balancing method using the FocalPoint FM4000 Series Ethernet L2/3/4 switch devices that enable full line rate processing using standard CPU blades.

System Configuration

An example Telecom Server is shown in figure 1. Here, several line cards are used which contain an NPU subsystem along with a fabric interface chip (FIC). The NPU inspects the packet headers at line rate to determine which services should be applied to the packets. The NPU then routes the packets to the correct service card through the backplane switch fabric. The service cards can be standard CPU blades, but may require specialized daughter cards to connect to the backplane fabric through a FIC. The packets can be routed through the fabric to several service cards before being sent back to the network through the line cards. This configuration requires costly NPU cards for packet classification and a specialized switch fabric connecting the NPUs to the service cards.

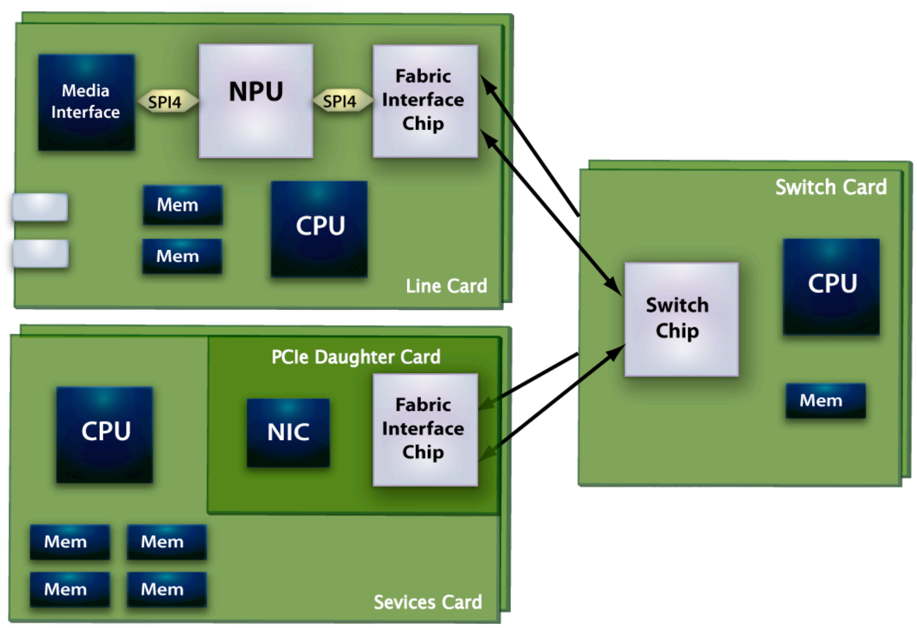

Figure 1: A traditional telecom server block diagram.

An improved system configuration using two members of the FM4000 Series product line, the FM4224 and FM4410, is shown in figure 2. Here, the packets arrive at the FM4224 on the switch card and are load balanced across multiple passive backplane links to other cards in the system. Load balancing can be done such that the deep packet processing function is spread across multiple server blades reducing the processing power required on each blade. Other server blades in the system can be used for further service processing, providing a very flexible deployment of CPU resources. For some applications, specialized services cards will also be used. Even though the packet header processing will need more CPU resources than the number of NPUs required above, the low cost of commodity CPU blades will more than offset this difference. Also, standard blade computer or ATCA chassis and backplanes can be used, further reducing the overall system cost.

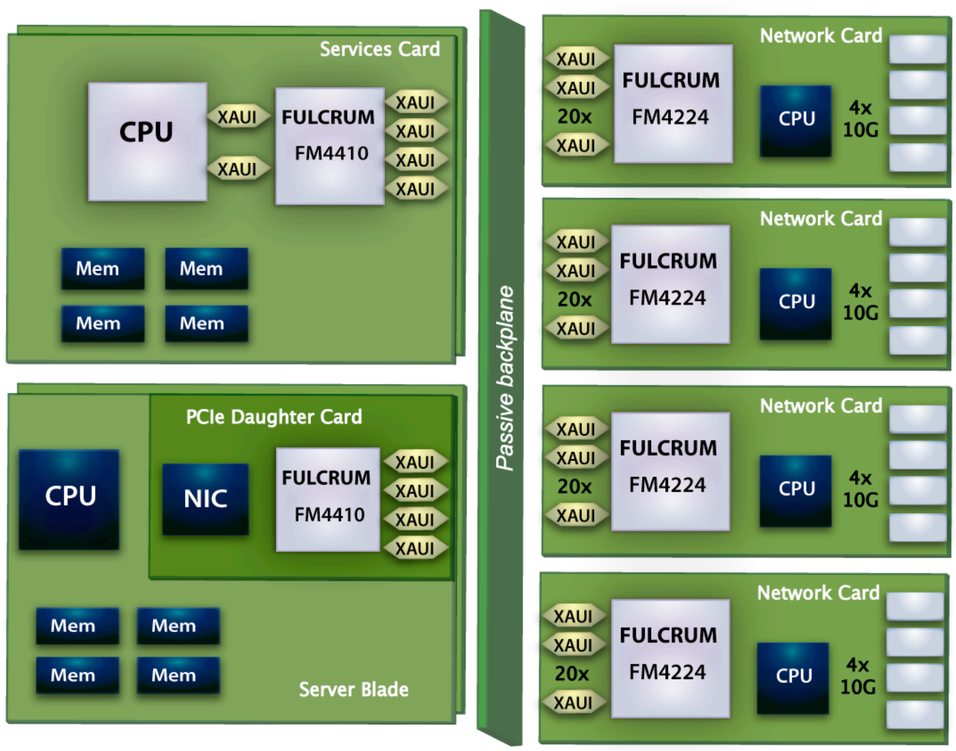

Figure 2: A telecom server using FocalPoint.

Packet Flow

The FM4000 Series devices are feature-rich 10G Ethernet switches that can perform layer 3 routing and can off-load some of the traditional NPU packet processing functions. In addition, the load balancing functions available in the FM4000 Series can be based on a specific set of L2/3/4 header fields or certain fields can be included or omitted if the frame is IPv4/IPv6. More information on this can be found in the FM4000 Series data sheet. This paper will focus on the load balancing functions. Figure 3 shows a simple implementation, which can be used to illustrate packet flow through the system.

Here we assume that link aggregation group (LAG) A connects to network A and LAG B connects to network B. For example, in a Telecom Server application, data may be processed as it flows between these two networks. Also, by using LAGs to distribute traffic across multiple switches, if a switch card fails, active paths will remain through the other switch card(s). A typical system may include up to four switch cards and up to 16 application blades with some of the blades used for packet inspection and some blades used for services.

At the network ingress, a packet will be load balanced by an FM4224 line switch to one of the application blades. This blade can then determine if the packet needs further processing by services running on other blades. If so, the packet can be routed back through one of the line switches for further processing. When packet processing is complete, the packet is load balanced across the line switches by the FM4410 device on the application blade. This in effect, load balances the packet across one of the egress LAG members back to the network. The load balancing uses a modified Persons hash function based on information in the packet header that defines a flow. FocalPoint can be configured to use various L2/L3/L4 header fields. A given flow will always use the same path through the fabric.

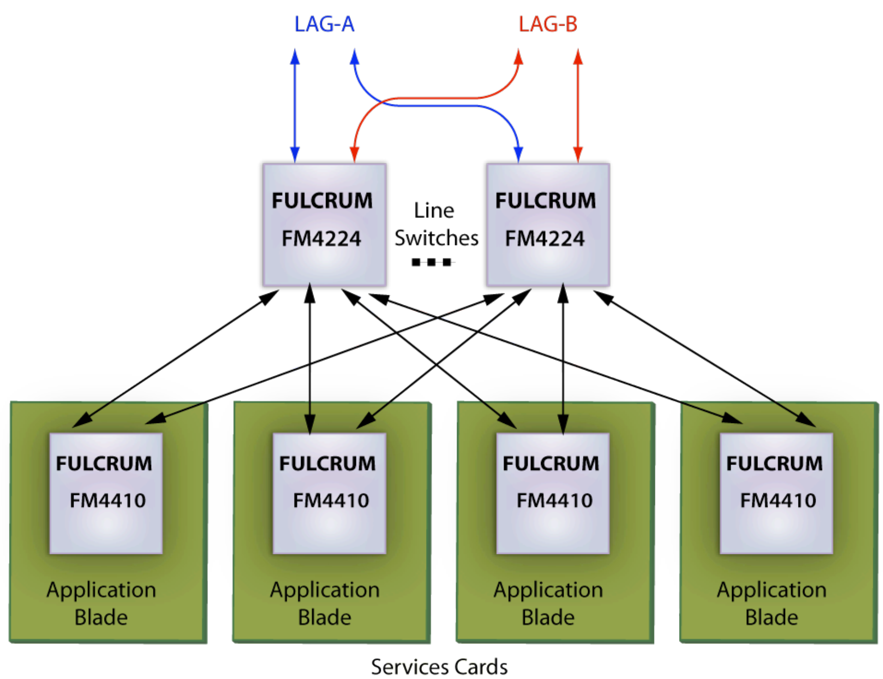

Figure 3: A telecom server load balancing block diagram

above, each blade can process the incoming packets at one quarter the line rate, which greatly reduces the processing requirements and allows the use of commodity CPU blades. Higher performance CPU subsystems can always be used in some blade slots if needed. Of course load balancing will not be perfect and the application blades must be able to accept small variations in packet processing loads.

Load balancing also provides high availability to the fabric allowing continued operation during application blade or switch card failure. If a link failure is detected, software can reconfigure the LAG to load balance across the remaining members.

FM4000 Series Load Balancing Function

Hashing functions vary in their effectiveness in balancing the load among the output ports, and in their cost of implementation in silicon, so the function used must be carefully chosen. FM4000 Series devices use a modified Pearson's hash that is highly effective in load balancing while incurring a modest implementation cost. Various parts of the L2/3/4 header can be used as a source for the hashing function. The source is hashed to a 12-bit value (giving 4096 intermediate bins) and that result is distributed among the output links in the hash group using modulo division. For purposes of load balancing, two hash values can be calculated from the header fields of each frame:

- Layer 3/4 Hash (36 bits)
- Layer 2/3/4 Hash (48 bits)

The keys to these hash functions are constructed in a configurable manner in order to provide the following features:

- Symmetry -- Hash value remains the same when source and destination fields are swapped.
- Static field dependence -- Support for including a specific set of header fields in the hash function.
- Dynamic field dependence, based on frame type -- Certain fields can be omitted or included when a frame is IPv4/IPv6.

Simulation Results

A simulation result presented here highlights the performance of the FM4000 Series load balancing by using highly stressful traffic patterns in a large system with 288 ports and 12 line switches. Actual telecom backplane implementations are likely to present somewhat different patterns to the switching network and may produce slightly different results. The simulation was created using a byte-accurate switch simulator and is cycle-based, with one byte of data being transferred per cycle.

The switch topology and data traffic patterns were simulated as follows:

- 288-port system composed of 24-port 10GE switches.
- 9216 MAC addresses connected to the 288 ports (P1-P288), 32 per port, randomly assigned
- Mesh traffic pattern:
- 1st round: P1 sends to P2, P2 sends to P3...P288 sends to P1
- 2nd round: P1 sends to P3, P2 sends to P4...P288 sends to P2
- Repeated 287 times or more so every port transmits to every other port
- Each round produces 288 ports x 287 frames – 82656 total frames.
- Frame size is 4KB

The load balancing simulation results are shown in figure 4 below. In this simulation, the total number of frames directed to each of the 288 ingress line ports from 9216 random MAC addresses during the whole simulation is counted. For good statistics, the cycle of mesh transmissions was repeated so that 50 frames were transmitted from each port to every other port. The system port utilization rates are irrelevant in this simulation. The loads are normalized so that differences in loading can be easily interpreted in terms of percent. The result is an extremely even load across the 288 line switch ports (12 line switches, 24 ports each).

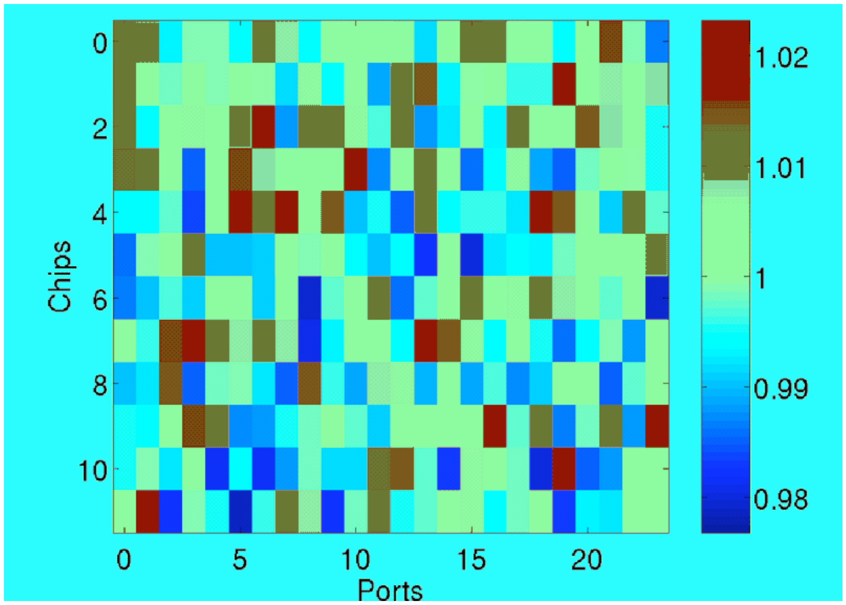

Figure 4: Load balancing simulation results

Conclusion

Today, the commoditization of blade severs has driven down their cost, well below that of specialized CPU blades. In addition, standard blade center chassis have become available with carrier class features and reliability. For high bandwidth packet processing, NPUs have been required to keep pace with the packet arrival rates. This paper has shown how the Fulcrum FM4000 Ethernet switch family can be used to load balance these high bandwidth data flows across multiple low-cost CPU blades with excellent uniformity. This not only reduces overall system cost in Telecom Server applications, but also allows flexible re-deployment of CPU resources as network loads change or new services are added.

Fulcrum Microsystems, Inc.
26630 Agoura Road
Calabasas Hills, CA 91301
818.871.8100
www.fulcrummicro.com