

Monitoring Load Balancing in the 10G Arena: Strategies and Requirements for Solving Performance Challenges

White Paper



2011 is the year of the 10 Gigabit network rollout. These pipes—as well as those of existing Gigabit networks, and even faster 40 and 100 Gbps networks—are under growing pressure to carry skyrocketing volumes of mission-critical business applications, performance-sensitive video and VoIP traffic, private and proprietary data, and crucial Internet connectivity. Monitoring the network to ensure availability, performance, security, and compliance has never been more important.

Monitoring solutions developed for Gigabit networks are now mature and represent huge investments in capital equipment, process development, staffing, training and experience. This monitoring involves tools such as traffic recorders that capture 100 percent of all network traffic in order to document compliance and conduct forensic investigations. Intrusion prevention systems (IPSs) examine all network traffic to discover and suppress security attacks and malware.

Unfortunately, scaling these security solutions up to 10 Gigabit speeds is not always a straightforward process. To take two examples, traffic recorder performance is limited by the sustained write speeds of disk arrays, and IPSs run enormously complex algorithms constrained by CPU performance. Increasing disk array speeds and CPU performance by an order of magnitude is technologically challenging and results in products with very high price tags. Therefore, simply scaling the performance of monitoring tools is, in many cases, either an unavailable solution or one that is cost-prohibitive. Nor is it acceptable to sacrifice security effectiveness to meet performance requirements, or to throttle performance to achieve required security effectiveness. Both effective security and high performance are essential.

Load Balancing—A Proven Method for Enhancing Performance

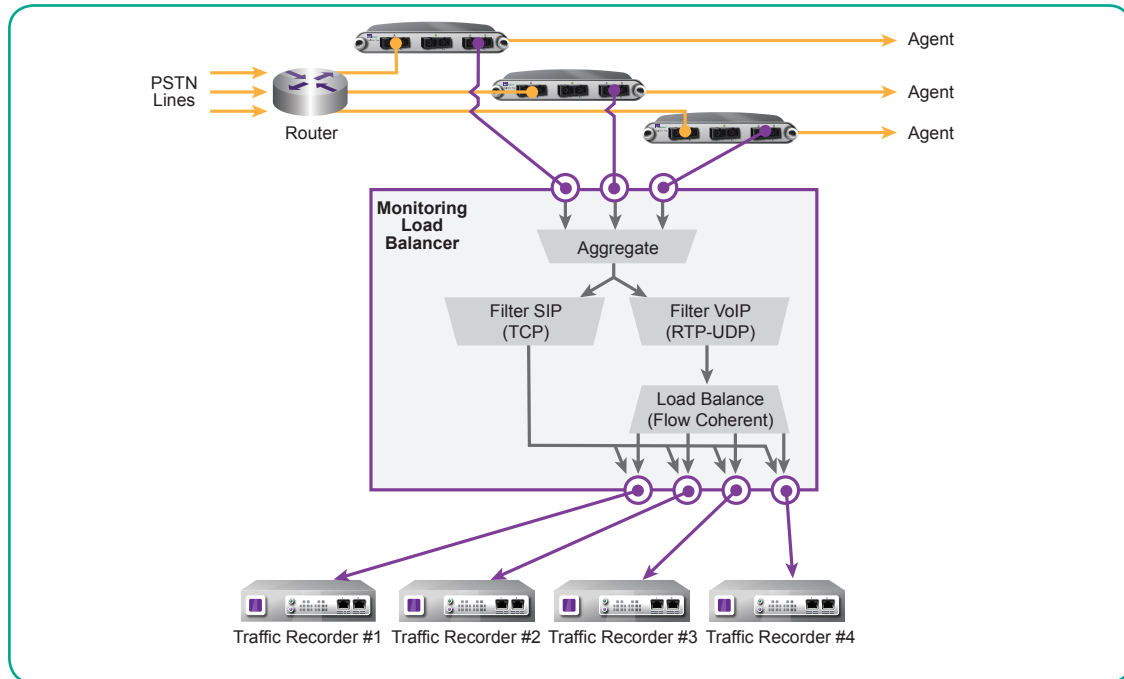
Many network architects and monitoring tool providers are addressing network performance challenges by using load balancing techniques. Load balancing is a well-understood technology that has been used in the past to distribute transactions across pools of servers and to create link-aggregation groups (LAGs) that can utilize parallel WAN links for increased bandwidth. Load balancing monitoring traffic across multiple tools is an obvious extension of the concept. However, monitoring load balancing presents different requirements than do transaction or LAG load balancing.

This paper develops several use cases for monitoring load balancing. In discussing these use cases, it teases out the specific requirements for monitoring load balancing and suggests strategies for meeting them.

Monitoring Load Balancing in the 10G Arena: Strategies and Requirements for Solving Performance Challenges

White Paper

Use Case #1: Call Center—Out-of-Band Load Balancing to Traffic Recorders



Use Case #1: Call Center—Out-of-Band Load Balancing to Traffic Recorders

“Your call may be monitored to ensure quality.” How many times have you heard that sentence from a telephone company, help desk, or other entity? In fact, most call centers today record all calls unless the caller specifically asks not to be recorded. Most call centers have gone digital, so that calls exist within the call center as VoIP traffic on a local area network (LAN), and network traffic recorders are used to record the calls.

As the volume of calls handled by the call center increases, it may reach a level that outstrips the capacity of a single traffic recorder. As a result, many call centers turn to a load balancing solution that deploys multiple traffic recorders operating in parallel. This application can record all of the calls, as long as two key requirements are met:

- For each call, the entire conversation—including both directions of traffic—must wind up on the same traffic recorder, so that it is not necessary to search across multiple recorders to replay a call.
- Call setup traffic (SIP traffic) must be available on the recorder for all the calls on that recorder.

To meet the first requirement, the load balancing algorithm must be “flow coherent,” meaning that traffic flows are kept together on the same monitoring tool. Flows are typically defined by a 5-tuple of header fields in the packets, the 5-tuple being source and destination IP addresses and ports, and the protocol. Packets with the same 5-tuple value are all routed to the same tool, ensuring that entire calls are recorded on a single tool. Moreover, packets with the same 5-tuple but with source and destination addresses and ports swapped must also go to that same tool, in order to capture both sides of the conversation. (The traffic of me talking to you travels in one direction, and when you talk to me, the packets flow in the opposite direction.)

Conclusion: Monitoring Load Balancing Requirement #1: Flow coherency

Monitoring Load Balancing in the 10G Arena: Strategies and Requirements for Solving Performance Challenges

White Paper

Depending on the architecture of the call center network, the 5-tuple may not be a good way to identify flows. In some cases, a single conversation may travel between the two endpoints with given IP addresses, but the ports may vary, depending on the path the packets take. In this case, flows should be defined by a 3-tuple of source and destination IP address and the protocol, but not the ports. In another network, the ports may be fixed but the IP addresses could vary, so that the 3-tuple of source and destination ports and the protocol should be used. The load balancing solution should be able to accommodate different definitions of what constitutes a flow.

Conclusion: Monitoring Load Balancing Requirement #2: Accommodate different definitions of flows

The other key requirement of the call center application is that the SIP traffic which sets up the call be recorded on the same traffic recorder as the voice traffic. One way to accomplish this is to record all of the SIP traffic on each one of the recorders. This creates very little additional load on the recorders because the SIP traffic is very small compared to the voice traffic. In one call center, SIP traffic is transported using the TCP protocol, while the voice is transported using RTP protocol over UDP datagrams. The load balancer uses a filter to select the SIP traffic and replicate it to all traffic recorders. Another filter selects the RTP traffic and load balances it to the traffic recorders. To support these types of solutions, monitoring load balancers should have the ability to select traffic of interest by filtering.

Conclusion: Monitoring Load Balancing Requirement #3: Filtering to select traffic of interest

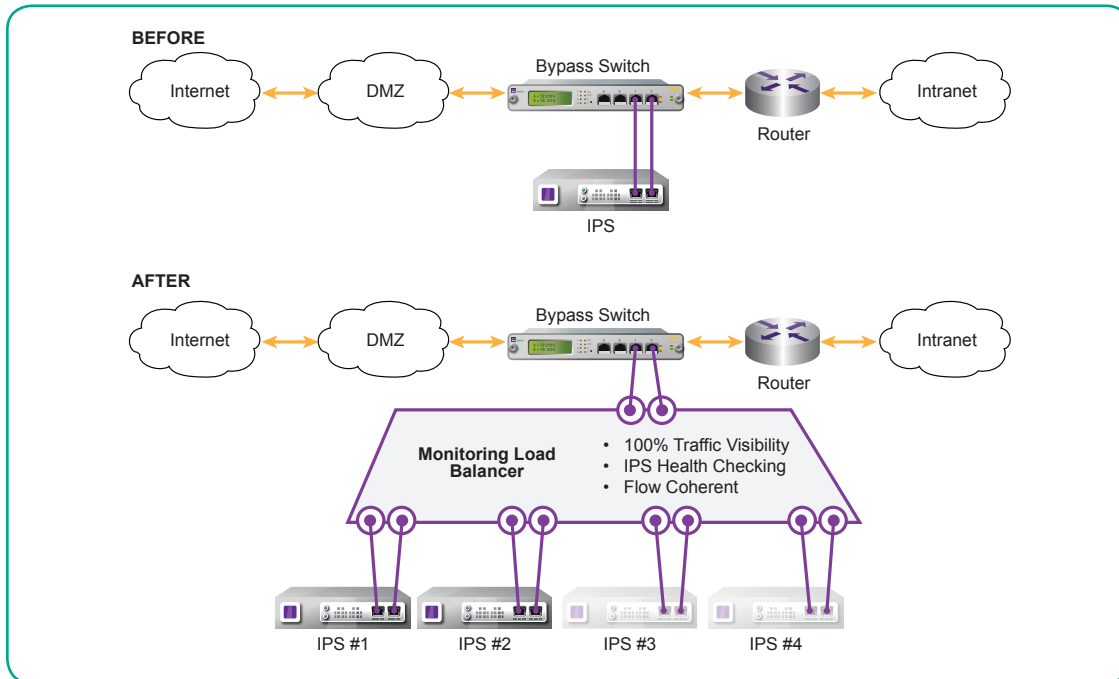
The filtering arrangement described in the last paragraph also implied the ability to regenerate traffic (in this case, the SIP traffic) to multiple outputs, and to aggregate traffic from different sources (aggregating the SIP traffic with the RTP traffic at the outputs.) Aggregation is also needed at the front end because the calls are carried over multiple network links. Therefore monitoring load balancers must have aggregation, redirection, and regeneration functionality.

Conclusion: Monitoring Load Balancing Requirement #4: Include aggregation, redirection, and regeneration capabilities

Monitoring Load Balancing in the 10G Arena: Strategies and Requirements for Solving Performance Challenges

White Paper

Use Case #2: Enterprise Intrusion Prevention—Inline Load Balancing to IPSs



Use Case #2: Enterprise Intrusion Prevention—Inline Load Balancing to IPSs

An enterprise relies on an IPS to keep its intranet secure from intruders and malware on the Internet. The IPS is capable of processing 2 Gbps of traffic continuously, but the link traffic has grown beyond this limit. Rather than dropping some traffic or throttling the speed of the traffic, the enterprise wants to increase the IPS processing bandwidth. They evaluated the latest, high-performance IPS model from their vendor, as well as devices from other vendors. Considering the high cost of the new IPS models along with the learning curve and risks of deploying a different unit, the enterprise realized it would be more cost-effective to purchase a second device identical to their existing IPS and run both in parallel, using a monitoring load balancer to distribute the traffic to the two tools.

The monitoring load balancer has to have enough performance to service the current volume of link traffic as well as enough headroom to support future growth. In this case, the enterprise sized the monitoring load balancer to support the full throughput capability of four IPS units, even though only two were being deployed at this time.

Conclusion: Monitoring Load Balancing Requirement #5: Adequate bandwidth for current needs and future growth

Because the monitoring load balancer, like an IPS, is deployed inline with the link traffic, it is essential that it passes 100 percent of the link traffic without dropping packets. Network switching type load balancers would not satisfy these requirements because they are allowed to take advantage of protocol rules that include TCP retries and dropped datagrams within the performance requirements. Thus, the load balancer would need to be specifically designed for monitoring applications.

Conclusion: Monitoring Load Balancing Requirement #6: 100% traffic visibility

Monitoring Load Balancing in the 10G Arena: Strategies and Requirements for Solving Performance Challenges

White Paper

Inline deployment also demands that the load balancing solution should not introduce a point of failure in the network. In the current deployment, the link was tapped with a bypass switch that would keep the link traffic flowing even if the IPS lost power or had any other type of failure. The architects selected a load balancer that sends heartbeat packets through attached IPSs to continuously validate that the IPSs were passing traffic—the same technique used by the bypass switch. If an IPS fails, the load balancer automatically takes the traffic that was bound to that IPS and redistributes it to the remaining active IPSs. When the failed IPS comes back online, the load balancer returns the traffic to it. Another possible mode of dealing with IPS failures that may be offered by some monitoring load balancers is a port loopback mode. In this situation, traffic simply bypasses a failed IPS as if that IPS were connected through a bypass switch. Yet another mod is N+M tool redundancy where one or more warm-standby IPSs are designated.

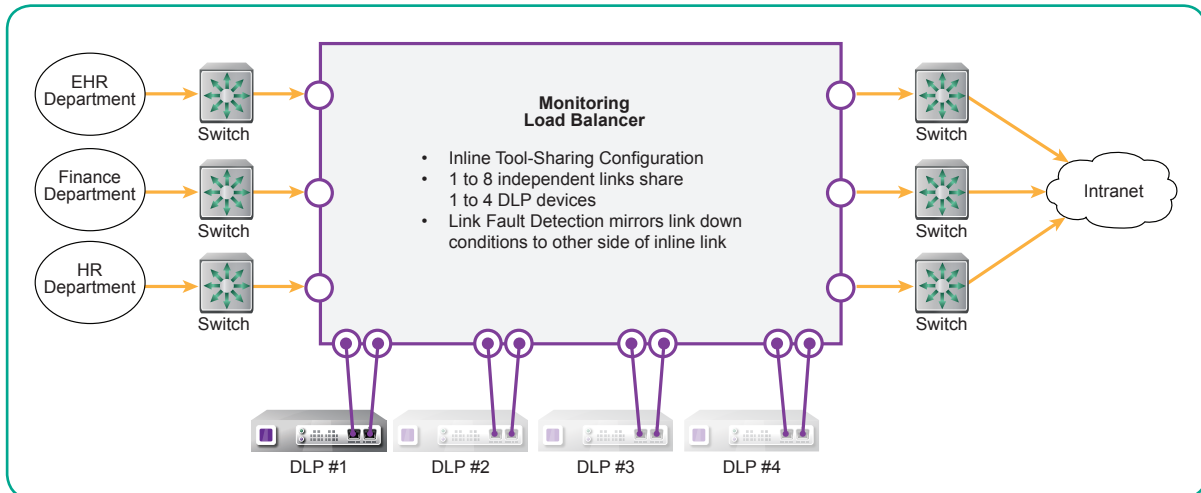
Conclusion: Monitoring Load Balancing Requirement #7: Tolerant of IPS failures

Finally, load balancing must be flow-coherent, because each IPS must see entire flows in order to perform behavioral threat analysis. This requirement is the same as for the previous case.

Monitoring Load Balancing in the 10G Arena: Strategies and Requirements for Solving Performance Challenges

White Paper

Use Case #3: HIPAA—Tool-Sharing Inline Load Balancing



Use Case #3: HIPAA—Tool-Sharing Inline Load Balancing

One component of a healthcare organization's solution to ensure compliance with the Health Insurance Portability and Accountability Act (HIPAA) is deployment of a data loss prevention (DLP) device on the network link that connects to the department responsible for electronic health records (EHR). The DLP device is deployed inline on the link, where it monitors traffic for Protected Health Information (PHI) and blocks certain traffic from being transferred on the link under specific conditions according to configured policies, actively preventing illegal leaking of PHI.

The DLP solution is working so well that the organization decided to deploy it for the Finance and Human Resources departments as well. A purchase order was generated for two new DLP systems, and the amount of the P.O. was quite large. A network engineer who was using a monitoring load balancing solution in another application had a suggestion. He said that his load balancer could share one DLP among multiple links, so if the DLP had enough processing power to handle the traffic on all three links, no additional DLP device would need to be purchased. Moreover, if the traffic on the three links oversubscribed a single DLP in the future, then one, two, or more additional DLP devices could be added onto the load balancer. This would share the processing power of all attached DLP devices among the three links. Seeing the potential of significant saving in both CAPEX and OPEX, the engineer was asked to contact the load balancer vendor and arrange a proof-of-concept demonstration.

Conclusion: Monitoring Load Balancing Requirement #8: Inline tool-sharing capability

This application has many of the same requirements as the previous two cases: 100 percent traffic visibility, flow coherence, flexible flow definition, health-checking of the attached DLP devices, and adequate bandwidth to support future growth. The network architect decided to size the solution to be expandable to as many as eight network links and four DLP devices.

However, the network architect spotted a possible problem in the solution. The link to the Finance department was designed with fault tolerance, so that if the link went down, routers would automatically reroute traffic on a redundant link. But when the load balancer was inserted inline in the link, it seemed that a link could go down on one side of the load balancer, but the switch on the other side of the load

Monitoring Load Balancing in the 10G Arena: Strategies and Requirements for Solving Performance Challenges

White Paper

balancer wouldn't see the down link and therefore the automatic rerouting wouldn't happen. Fortunately, the load balancer included a feature called Link Fault Detect, which makes the inline link behave like a virtual wire: if the link on one side goes down, the load balancer downs the link on the other side to transmit the link down condition to the other endpoint, just like before the load balancer was deployed on the link. This feature eliminates the problem the architect was concerned about.

Conclusion: Monitoring Load Balancing Requirement #9: Fault mirroring across inline links (Link Fault Detect feature)

Implementation

Net Optics xBalancer is an example of a monitoring load balancing solution that is available on the market today. It provides 24 ports of 10 Gigabit / 1 Gigabit load balancing functionality. The following chart evaluates how it meets the requirements we've found in our use cases.

Requirement	Satisfied by xBalancer	Notes
#1: Flow coherency	Yes	xBalancer provides flow coherent load balancing for both out-of-band (one-way traffic) and inline (bi-directional traffic) topologies
#2: Accommodate different definitions of flows	Yes	xBalancer supports flows defined by any combination of the following fields: source and destination IP addresses and ports, protocol, ethertype, vlan, source and destination MAC addresses
#3: Select traffic of interest by filtering	Yes	xBalancer supports Layer 2 through 4 filtering to select the load balancing and monitoring traffic of interest
#4: Aggregation, redirection, and regeneration capabilities	Yes	xBalancer has fully configurable port mapping for aggregation, regeneration, and switching
#5: Adequate bandwidth for current needs and future growth	Yes	xBalancer has 24 SFP+ slots that accommodate 10G SFP+ and 1G SFP transceiver modules, with automatic data rate conversion as needed; the architecture is fully non-blocking, with 480 Gbps and 360 Mpps throughput
#6: 100% traffic visibility	Yes	xBalancer passes all traffic, include layer 1 and 2 errors; no packet drops as long as aggregated traffic does not exceed the monitoring port bandwidth
#7: Tolerant of IPS failures	Yes	xBalancer checks the health of attached inline monitoring devices by periodically sending Heartbeat packets through the devices in both directions; upon tool failure, three modes are supported: allocate traffic to remaining up devices; loopback traffic bypassing the failed device; and N+M redundancy
#8: Inline tool-sharing capability	Yes	xBalancer supports sharing of any number of tools among any number of inline links, subject to the limit of 24 ports
#9: Fault mirroring across inline links (Link Fault Detect feature)	Yes	xBalancer can be configured for Link Fault Detect across any pair of ports to trigger system-level fault recovery



Monitoring Load Balancing in the 10G Arena: Strategies and Requirements for Solving Performance Challenges

White Paper

Net Optics family of monitoring load balancing solutions

	Director™	Director Pro™	Director xStream™	Director xStream Pro™	xBalancer™
Load Balancing					
Static	•	•	•	•	•
Dynamic		•		•	•
Throughput					
Backplane	34G, 54G, 74G	34G, 64G	240G	240G	240G
Max Output Interfaces	34 x 1G	Static: 34 x 1G Dynamic: 16 X 1G	23 x 10G	16 x 10G	23 x 10G
Deployment					
Out-of-band	•	•	•	•	•
Inline		•			•
Features					
DPI as Selection Criteria		•		•	
Expandable	•	•			
Link-state Aware		•			•
Tool Sharing					•
Heartbeat					•



Monitoring Load Balancing in the 10G Arena: Strategies and Requirements for Solving Performance Challenges

White Paper

Conclusion

Load balancing can be a cost-effective solution for dealing with monitoring tool oversubscription due to growing traffic speeds and volumes. Replication of existing tools such as traffic recorders and IPSs can be less costly than upgrading to higher performance tools, and the hidden expenses of deploying unfamiliar new gear and retraining staff should not be overlooked. Load balancing is a mature, well-understood technology, and monitoring access vendors such as Net Optics are offering load balancers that are purpose-built to meet the special requirements of load balancing monitoring traffic. These devices offer 100 percent traffic visibility, flow coherence, and tool sharing capabilities. For organizations facing monitoring challenges as they upgrade data centers to 10G and faster technologies, monitoring load balancing is a solution well worth investigating.

For further information about xBalancer and all of Net Optics' Load Balancing solutions:

<http://www.netoptics.com>

Net Optics, Inc.

5303 Betsy Ross Drive

Santa Clara, CA 95054

(408) 737-7777

info@netoptics.com

Customer First!

Disclaimer: Information contained herein is the sole and exclusive property of Net Optics Inc. The information within this document or item is confidential; it shall not be disclosed to a third party or used except for the purpose of the recipient providing a service to Net Optics Inc. or for the benefit of Net Optics Inc. Your retention, possession or use of this information constitutes your acceptance of these terms. Please note that the sender accepts no responsibility for viruses and it is your responsibility to scan attachments (if any).