Corvil

Whitepaper

Managing Performance in Financial Trading Networks

Contents

1. Introduction

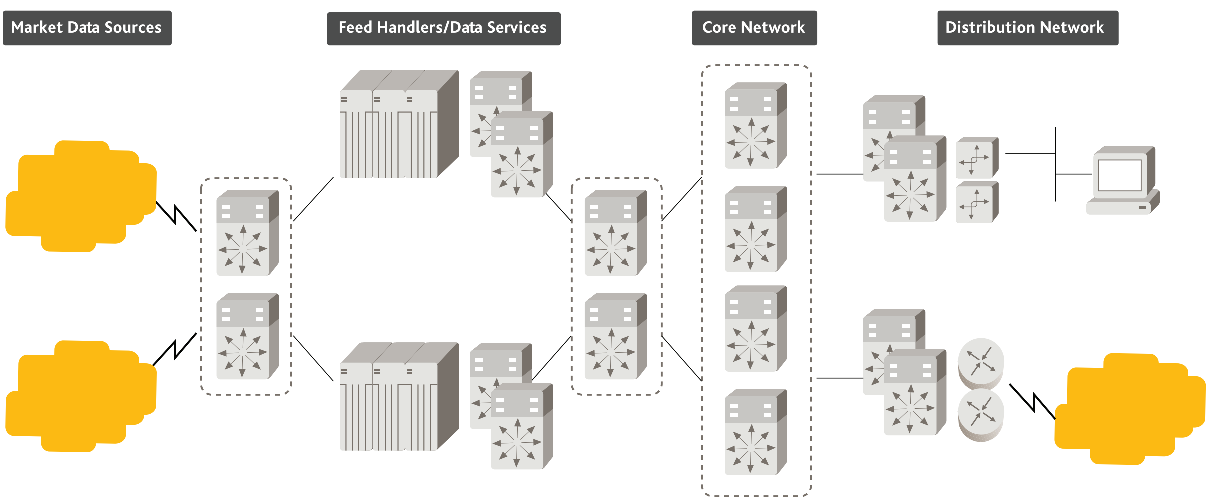
Performance targets in financial trading networks are more stringent than in any other type of network deployed today. Message latency and data loss result in missed trading opportunities and have a direct impact on revenue. The standard of performance needed for success is determined by competition from other trading organizations, and this has led to an 'arms race' in which latency targets have steadily shrunk from tens of milliseconds to milliseconds, and even to microseconds.

CorvilNet is a network monitoring and analysis system by Corvil, designed to provide visibility into network performance at the timescales which matter for financial trading. Traditional network monitoring tools which make measurements over periods of seconds and longer, fail to identify conditions which can add hundreds of milliseconds of latency. CorvilNet addresses this deficiency with a set of new techniques enabling microsecond-level insight into latency and the conditions that cause it. This document describes the benefits that CorvilNet provides with reference to a market-data dissemination environment, depicted schematically in Figure 1. CorvilNet is implemented by deploying measurement appliances at key points within the network that monitor network traffic using either passive taps or Cisco SPAN (Switched Port Analyzer) connections. In the following sections we describe how to use CorvilNet to manage performance in each part of the network where added latency can occur.

2. Performance in the LAN

The biggest network contribution to latency in the LAN typically comes from queuing on switch ports during temporary overload conditions. A badly overloaded 10G switch port can add up to several hundred microseconds of variable queuing delay, and this can rise to several milliseconds at 1G speeds. In the worst case queuing causes packet-loss, introducing

Figure 1
Market data dissemination schematic.



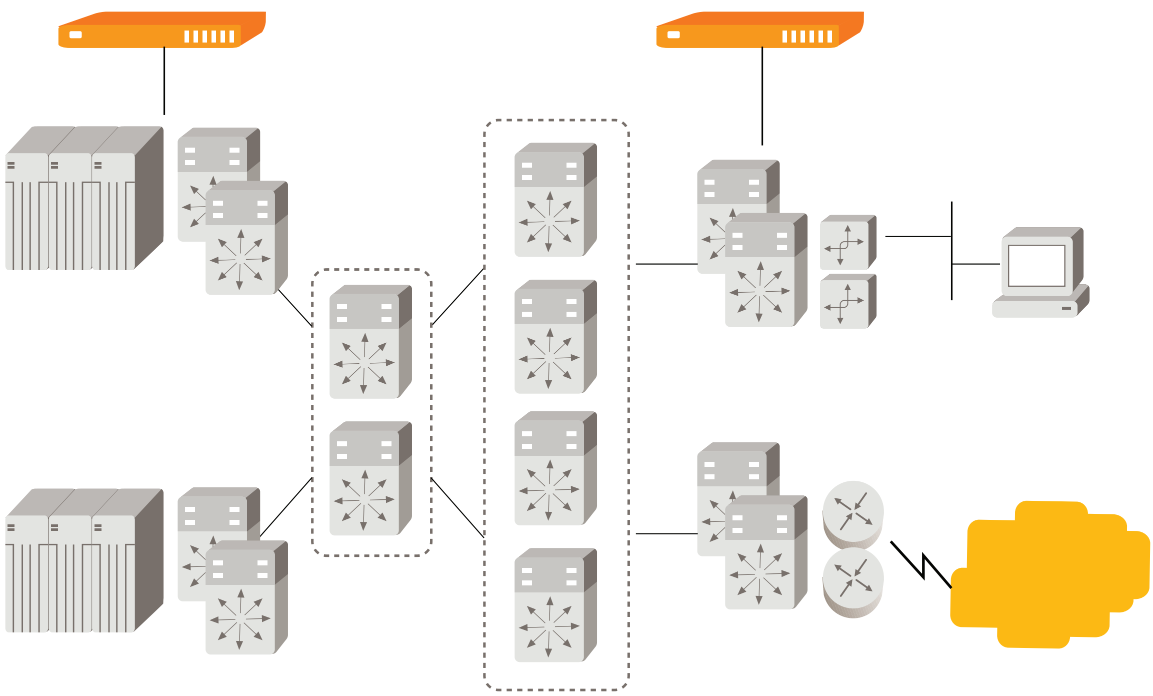
lengthy retransmission delays. The conditions which lead to congestion can be short-lived; they are also intermittent and are often triggered by very fast bursts of traffic. This combination of factors makes LAN congestion difficult to diagnose and trouble-shoot.

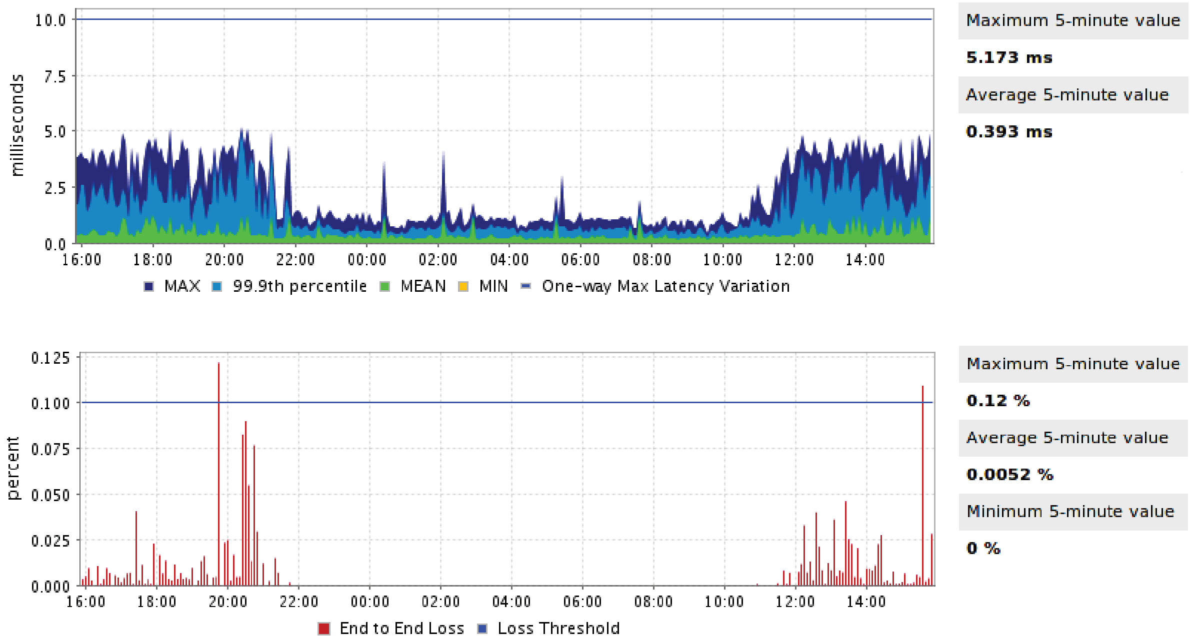
CorvilNet detects LAN congestion using a feature called "Passive Network Quality Monitoring" (PNQM), which measures latency and loss along paths between CorvilNet appliances. Latency measurements from PNQM are accurate to within a few tens of microseconds, which is precise enough to quickly spot sub-millisecond delays incurred at multi-gigabit speeds. Most importantly, PNQM measures latency and loss as they are actually experienced by application packets, which means that it can 'see' transient congestion conditions which occur during short bursts of network activity. These conditions contribute the most to latency, but are easily missed by traditional techniques that measure performance by probing the network with test packets. PNQM avoids doubt by directly measuring the experience of live traffic. PNQM measurements can be produced for every packet traversing the network. As well as tracking IP packets, PNQM can also track messages based on user-defined fields within packets, or in packet or frame headers. Detailed results are made available in near real-time to external applications, and CorvilNet also stores the full distribution of latency and loss measurements internally for up to 60 days with 5-minute granularity (additional external storage is also supported).

A web interface provides visibility into the collected data for humans, and supports statistical summaries such as average, median, and percentiles computed over any timeframe. The system will also monitor these statistics against user-specified performance targets on an on-going basis, triggering alerts when violations occur.

Figure 2

Monitoring performance across the LAN using CorvilNet.





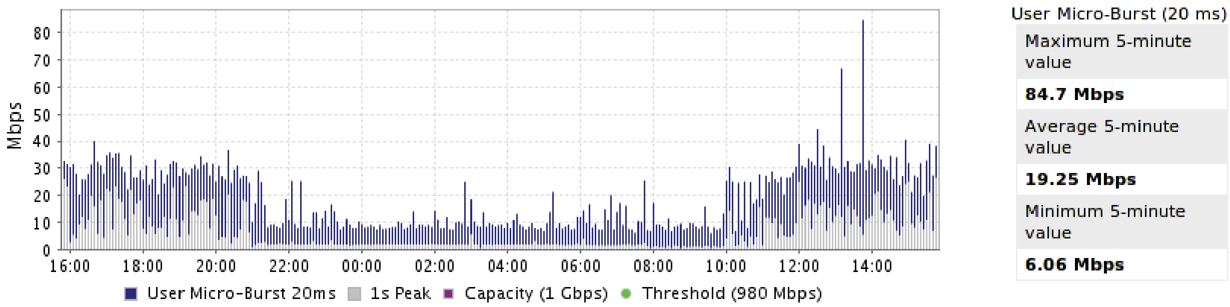
Dealing with congestion problems in the LAN is often a matter of ensuring that links are correctly sized and that load is well-distributed across a sufficient number of paths. Another less-used option is to limit the bandwidth of specific traffic sources which have been identified as problematic. But assessments of traffic load made over periods of seconds or longer don't provide much support for these tasks, given that the aim is to control delays at millisecond timescales. Visibility into short timescale behaviour can be obtained using CorvilNet's Microburst Detection feature, which measures bit-rates over intervals down to 1 millisecond. This technology can make sure that the network is properly dimensioned to handle short, intense traffic bursts which otherwise cause congestion. Microburst Detection also reveals the applications and servers which place the most stress on the network at short timescales – which are not always the ones generating the most traffic on average.

Figure 3

Variable latency (top) and loss (bottom) experienced by application traffic, measured using CorvilNet's PNQM.

Figure 4

Assessing link utilization at millisecond timescales helps make sure the network is dimensioned to avoid latency.



3. Inbound Data Feeds

Connections bringing market data into the trading network add latency in the form of propagation delay, as well as potential queuing delays caused by congestion. Propagation delays can measure up to 100-200 microseconds across a large metropolitan area, but are constant and relatively easy to plan for. Queuing delays will arise if the inbound data rate exceeds the connection capacity, which can happen for short periods of time even if the average bit-rate is low. Queuing can potentially add several milliseconds of variable latency on a 1G connection, or more at lower speeds, and therefore minimizing queuing is a necessary step for maintaining low latency performance.

CorvilNet's Microburst Detection feature is well-suited to identifying links which are saturated with short timescale bursts, even when the link in question is upstream of the measurement point in the data path. Figure 6 shows an example, where measurements made on traffic arriving from a 100 Mbps link indicate that it is saturated at the 20ms timescale, even though it is less than 35% utilized on average. Microburst Detection shows that the link is continuously busy for at least 20ms at a time during bursts, and is likely to be introducing significant traffic latency.

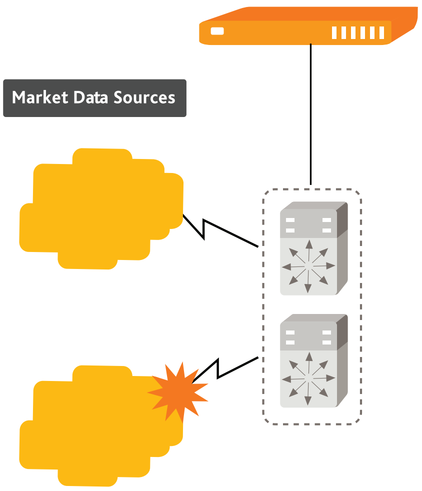
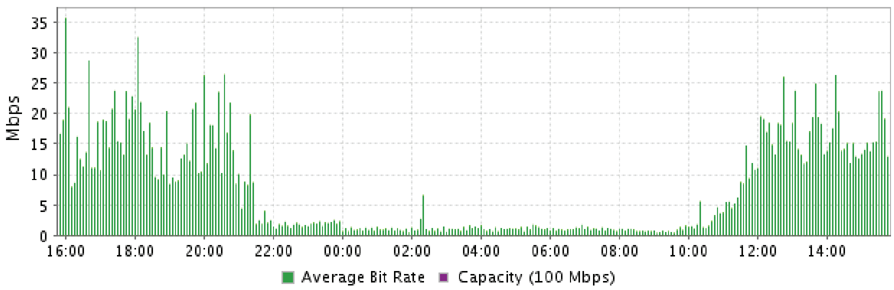


Figure 5

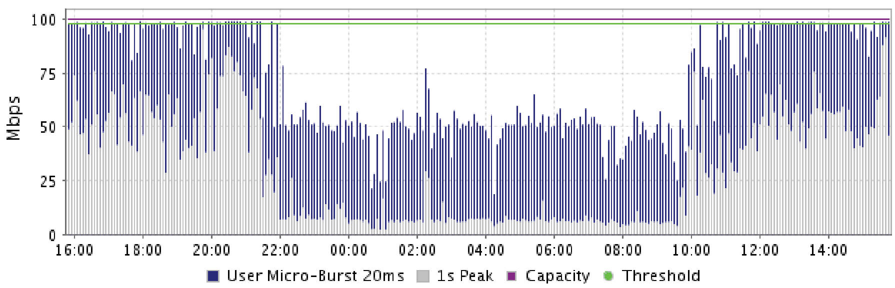
Identifying saturated links at the network edge.

Figure 6

Bit-rate measurements for traffic arriving on a 100 Mbps link. The average bit-rate apparently shows that the link is less than 35% utilized. Microburst Detection reveals that it is in fact saturated for at least 20ms at a time during traffic bursts.



Maximum 5-minute value
35.7 Mbps
Average 5-minute value
7.49 Mbps
Minimum 5-minute value
443 Kbps



User Micro-Burst (20 ms)
Maximum 5-minute value
99.1 Mbps
Average 5-minute value
71.7 Mbps
Minimum 5-minute value
18.8 Mbps

Conversely, re-sizing the link to ensure that 1-millisecond microburst rates never approach the link speed would provide assurance that upstream queuing delays never exceed 1 millisecond.

4. Outbound Data Feeds

In a WAN network connecting a remote market data consumer via a relatively low-speed link, congestion-induced queuing delays are again the largest potential contributor to latency. For example, a 100 Mbps link connecting a consumer located in the same city might have a propagation time of 100-200 microseconds, but could potentially introduce queuing delays of 50-80 milliseconds if it becomes badly congested. At lower link speeds the potential for queuing delay is even greater. These delays can be avoided by ensuring that links are congestion-free.

Deploying a CorvilNet appliance at a remote site allows latency and loss to and from that site to be fully monitored using PNQM. Sometimes however it is not possible to deploy a remote appliance, for example where the end-consumer is a separate organization. Congestion issues can still be analyzed at the edge using a CorvilNet feature called "Expected Quality", which predicts the level of delay and loss which will occur on the remote site bottleneck link.

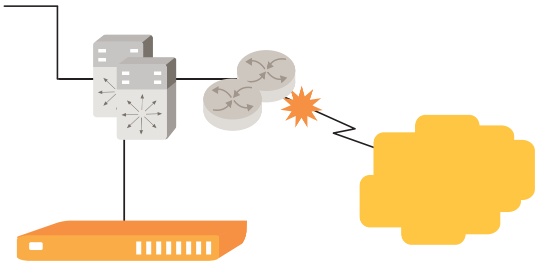
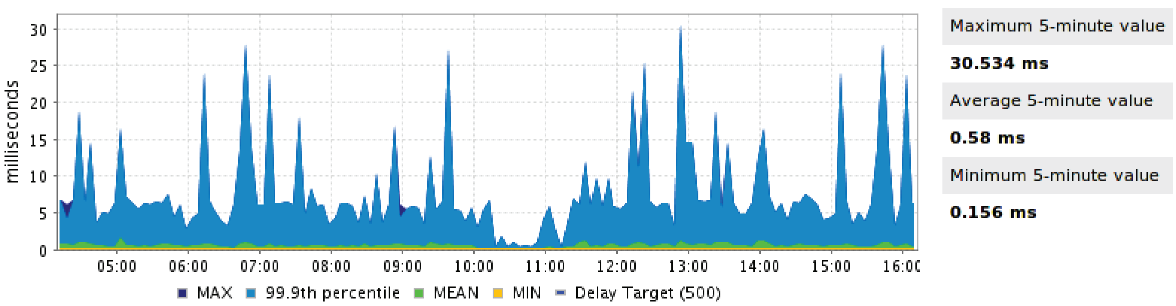


Figure 7
CorvilNet's Expected Quality feature analyzes congestion levels on the bottleneck link towards a remote site.

Expected Quality works by continuously running the observed traffic through a simulation of the link's switch/router interface, taking account of any configured QoS or prioritization mechanisms, and recording the delays and losses experienced. The feature provides valuable insight into congestion at the network edge, which is usually the largest contributor to loss and latency, and is ideal for deciding whether to apply QoS mechanisms or add connection bandwidth. Unlike PNQM, Expected Quality does not measure other possible sources of latency such as a service provider's WAN, or mis-configured/faulty equipment.

Figure 8
Latency induced on a low-speed remote site interface, as predicted using Expected Quality. computed each day from the same underlying data set.



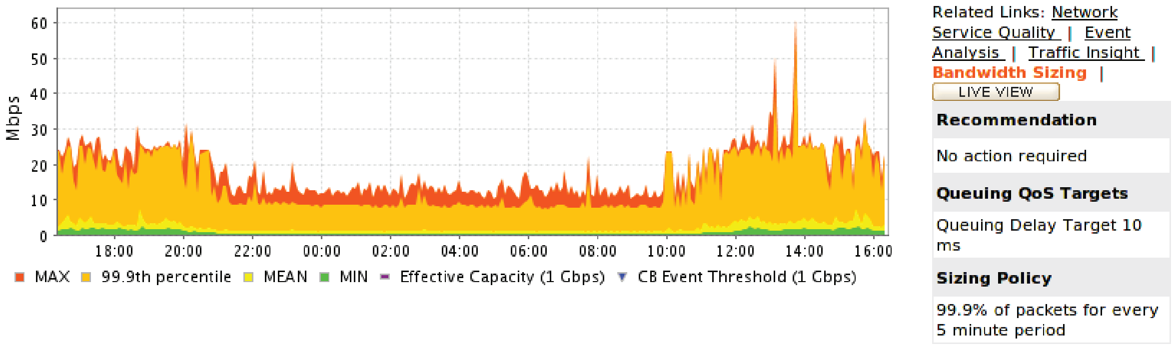


Figure 9

Finding the right interface bandwidth for a bottleneck connection using CorvilNet. The graph shows the bandwidth needed to protect packets from being lost or delayed by more than 10ms on the bottleneck interface.

To assist in sizing connections for remote sites, CorvilNet also provides a Bandwidth Sizing feature which calculates directly the bandwidth needed to maintain a specified latency target. This technology operates by computing – for each individual packet – the bandwidth needed to prevent that packet suffering excessive latency or queuing. Results are then rolled up and summarized over time to reach a Bandwidth Sizing recommendation which will protect a user-specified percentage of packets.

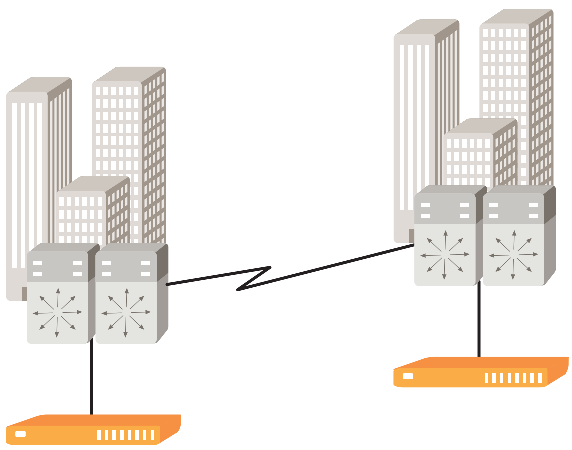
5. Redundant Infrastructures

Where a trading infrastructure is replicated across multiple sites, deploying CorvilNet appliances at each site allows PNQM to monitor performance on all of the inter-site links. There is no limit on the distance over which the technology can operate, and the overhead traffic needed to keep CorvilNet appliances in sync is typically just a few percentage points of the inter-site traffic when every packet is tracked. Utilization levels on inter-site links can be tracked at millisecond timescales using Microburst Detection, and Expected Quality and Bandwidth Sizing can also be employed to ensure that QoS policies are optimized and links sized to achieve specified latency targets.

CorvilNet does not require clock synchronization to operate across multiple sites, even when the sites are widely separated. The PNQM feature includes a sophisticated, continuously re-calibrated model of clock discrepancies between CorvilNet appliances to ensure that variable latency is measured accurately without the need for synchronized clocks. However, path asymmetries in the fixed (propagation delay) component of latency cannot be resolved without synchronization. For example, if the

Figure 10

Multi-site CorvilNet deployments enable full monitoring of performance between sites: loss, latency, millisecond-level utilization, and bandwidth requirements.



propagation delay from site A to site B is larger than the propagation delay in the reverse direction, synchronized clocks will be needed to detect this. For this reason CorvilNet supports clock synchronization via GPS as a deployment option.

In practice, for sites within the same metropolitan area – or even within the same continent – the variable components of latency (due to serialization times and queuing/ congestion delays) are normally found to be much larger than any asymmetries present in propagation delays. The majority of Corvil's customers choose to deploy without GPS synchronization, and we would recommend this option unless path asymmetry is an issue of particular concern.

6. Analyzing Abnormal Events

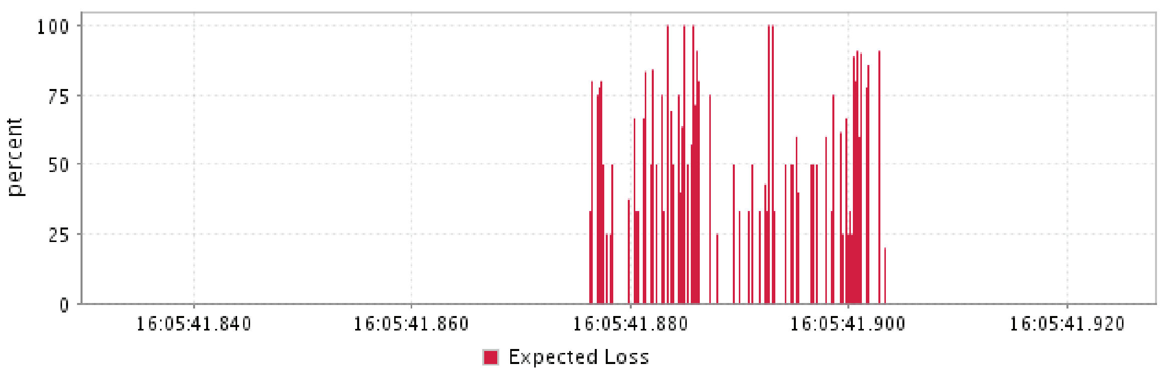
Congestion can build up and disappear again in milliseconds or less, making it hard to diagnose what causes it from a high-level view of traffic activity. On the other hand, continuously capturing detailed activity information is not a scalable approach, given the volume of traffic in today's market data networks. Searching for the cause of an anomaly amid millions of captured packets from an always-on sniffer is no different from looking for a needle in a haystack.

CorvilNet overcomes these difficulties using an event-triggered packet capture facility which automatically records details about just those packets involved in unusual latency, loss, or network load events. This approach enables detailed analysis of anomalous condition, without inundating users with irrelevant data.

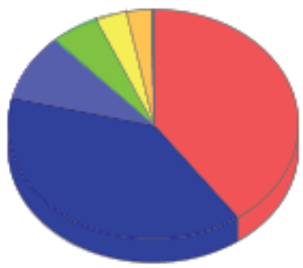
Event-triggered capture retains a configurable amount of data from each packet during an event, as well as a nanosecond-precision timestamp and per-packet measurements including loss, latency, and bit-rate. A drill-down GUI allows the captured data to be visualized in detail. Top-talkers and top-applications (based on a layer-7 application recognition engine) are also available in event inspection mode, to help reveal the services which contribute most to abnormal conditions. And packet captures can also be exported in pcap format to third-party packet decoders.

Figure 11

Inspecting an automatically captured packet-loss event.



Top Applications



	Application	Bytes	Packets
	mcast-OPRA	288 KB	398
	microsoft-SMB	277 KB	308
	mcast-NASDAQ-B	66.8 KB	122
	microsoft-netbios	38.8 KB	69
	mcast-NASDAQ-A	24.1 KB	31
	mcast-Reuters	21.1 KB	45
	Unknown	715 bytes	10
	intraweb	179 bytes	1

Figure 12
Top-talkers and top-applications during an event can help reveal which services contribute to congestion, and which services are most affected.

7. Summary

In all parts of the network, transient congestion is the largest potential contributor to latency. Low-latency performance requires that dynamic congestion conditions can be detected and eliminated. But most network monitoring systems don't analyze the behaviour of traffic or the network at the timescales relevant for low latency. By addressing this deficiency the CorvilNet monitoring system enables network managers to:

- Measure latency and loss as experienced by application traffic with microsecond precision, allowing LAN or WAN congestion problems to be easily diagnosed;
- Size links and distribute traffic load with full insight into the short, intense traffic bursts that generate the most network stress;
- Capture and trouble-shoot high latency events, and quickly identify traffic sources that contribute to poor performance.

With the benefits provided by CorvilNet, network performance management can finally catch up with the world of low-latency financial trading.

Corvil Ltd, 6 George's Dock, IFSC, Dublin 1, Ireland
T +353 1 859 1000 **Tech Support** +353 1 859 1010 **E** info@corvil.com **W** www.corvil.com

CWP-0802-4