

## White Paper

### Overview

One of the current buzz words in Wall Street and in the rest of the financial world is “low-latency” -- minimizing the amount of time it takes for messages to be transmitted from one network endpoint to another. In the financial world this may mean how long a message may take to go from a market data source to a subscriber.

Individual or company fortunes are now being determined by electronic trading algorithms that take as their input company information, analyst recommendations, or security execution orders. In turn, decisions are made at the electronic level that determine an organization’s survival.

Less reliance is placed on the floor trader and the trend is toward more automation using trading algorithms. As a result, a new industry has been created and is primarily driven by the need to send financial messages to their destination, not just at multiples of milliseconds but down to sub-milliseconds.

New revenue streams became possible because of the demand for faster and faster executions. Vendors appeared or reinvented themselves with the focus on the low-latency arms race. Exchanges now provide dedicated server farms that provide closer proximity to the trading floors. With millisecond measurements now giving way to microsecond accuracy service providers and low-latency vendors must find or improve the current methods of quantifying the performance of electronic communication networks (ECNs).

### Low-Latency “Ecosystem”

In the Merriam Webster dictionary an “ecosystem” is defined as “the complex of a community of organisms and its environment functioning as an ecological unit”. We can readily apply this idea to the financial world’s low-latency environment.

A low-latency ecosystem consists of several component types, including (but not restricted to):

- Market Data Sources - providers of financial data such as news or quotes.
- Manufacturers or Vendors – Manufacturers of hardware, software, or a combination of the two that create low-latency systems (e.g. servers) that process, host, or distribute the feeds
- Service Providers - Network service providers used by the two components above that provide the physical transport infrastructure for moving the feed messages to their subscribers
- Users - Subscribers of financial data from Brokerage houses, investment banks, or individual investors
- Server Host - Co-host, co-locations that house servers that execute trades

Some components may belong to one or more types. The health of the ecosystem depends on each component's ability to efficiently do its job and provide services to its immediate neighbor. Therefore, it is of vital importance that "checks and balances" be practiced between them for all to benefit. When things go awry, it is appropriate to ask these questions:

- Is the information provider sending out its data in a timely manner?
- Are the servers transmitting within their published service specifications?
- Is the Service Provider's network adding delay and possibly contributing to packet loss?
- Is the Subscriber actually receiving all intended data in a timely manner?

One technique for answering these questions is to use methods that quantify the performance of each in terms of specific metrics. One such method would be to monitor links between each participant in the low-latency ecosystem (see Figure 1). When due diligence is not practiced, everyone suffers. The consequence?...loss of revenue due to lost trades, complaints, or lost customers.

Being the slowest in the low-latency arms race could mean the difference between keeping a financial organization's doors open for yet another business day or slowly losing its place among its competitors.

## Low-Latency Solution

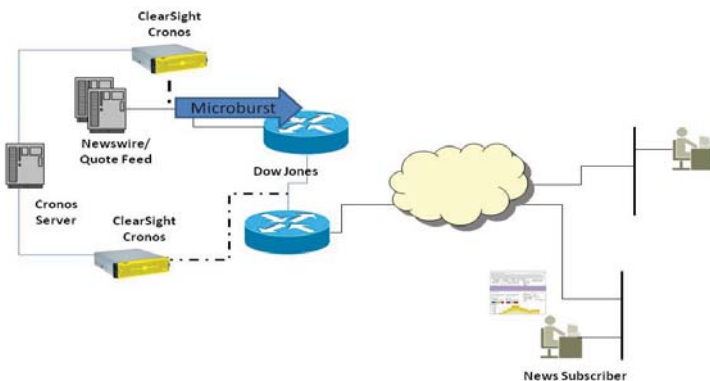


Figure 1: ClearSight Cronos

This paper introduces a tool that gives each component of the low-latency ecosystem the ability to keep honest its closest neighbor in the low-latency ecosystem.

Latency is the delay between the initiation of a network transmission by a sender and the receipt of that transmission by a receiver. Latency is usually the result of an overwhelmed network device such as a switch. "Microburst" storms may result in queued messages at a switch's port or even worse, packet loss. The majority of currently available network monitoring systems measure latency in increments much larger than what is required in order to pinpoint and resolve issues in the low-latency ecosystem. Latency measurements now require sub-millisecond granularity.

In order to obtain the necessary granularity of latency measurements between two points in a network path, synchronization of data collecting agents in each of those two points is required. This usually involves referencing an external time source, private or public. There are a number of synchronization methods with accuracy often being sacrificed in exchange for lower cost and feasibility.

Network Time Protocol (NTP) may be used to synchronize the collecting agents. Accuracy, with respect to a well known public time server accessible via the Internet, will usually be within 50-100 milliseconds at best, as observed in ClearSight's lab environment. An improvement in accuracy was obtained when an NTP server resided in the same network segment as the devices to be synchronized.

Currently, the most accurate solution involves the use of GPS receivers (connected to special GPS antennas) where collecting agents may be synchronized within 50 microsecond accuracy. GPS receivers are able to obtain their times from Earth orbit satellites using microwave signals.

However, physical access to the open sky to deploy a GPS antenna may be a challenge to some organizations. Again in ClearSight's lab, GPS enabled collecting agents synchronized their time with improved accuracy, down to 50 microseconds.

Once two end points are synchronized, the collecting agents are able to accurately timestamp messages.

## Cronos

A new solution by ClearSight Networks, Cronos, is able to monitor and correlate messages from pairs of collecting agents. The physical locations where Cronos agents should be deployed are determined by the network or market feed path or low-latency ecosystem devices under test.

Cronos agents may monitor using mirror ports (from a switch) or network taps. As the Cronos agents are collecting and storing messages (or Ethernet frames) from a pair of Cronos agents vital information is forwarded to a dedicated Cronos Server.

The Cronos Server calculates the latency (see Figure 1) for corresponding packets containing financial messages (multicast flow) traveling from one endpoint (e.g. market feeder) to a second endpoint (i.e. subscriber).

Figure 2 shows the Latency related panes. The top pane displays an aggregated Latency trend graph for all the flows monitored in a user specified time range. There are three lines representing Latency: Max, Mean and Min. All multicast flows represented by the aggregate Latency graph are listed in the middle pane. Selecting a flow in the middle pane will result in the corresponding trending graph at the bottom pane. A user may drill down in either trending graph for higher granularity with respect to time.

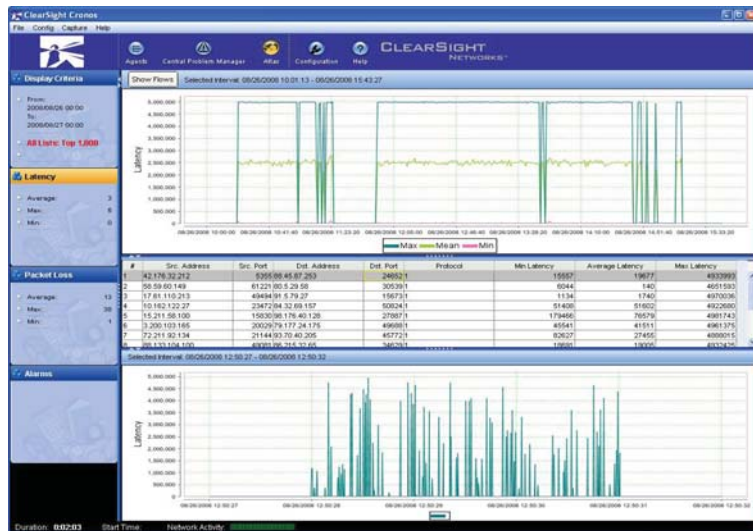


Figure 2: Packet Latency Measurements

## Determining Packet Loss

A potential by-product of latency (due to queuing delays) is packet loss. Packet loss (see Figure 3) leads to costly retransmissions. The Cronos Server will also track the streams for corresponding packet loss in each flow in a user specified time range. As with Latency, the top pane shows aggregated packet loss statistics for all streams. The middle pane lists all flows corresponding to the top pane. Selecting a flow in the middle pane will display a trend graph in which the user may drill down into for higher granularity.

Automatic detection for either metric is available with the use of alarms. These threshold based alarms may be assigned to alerting mechanisms such as sending out an SNMP trap or an email alerting personnel to the detected condition.

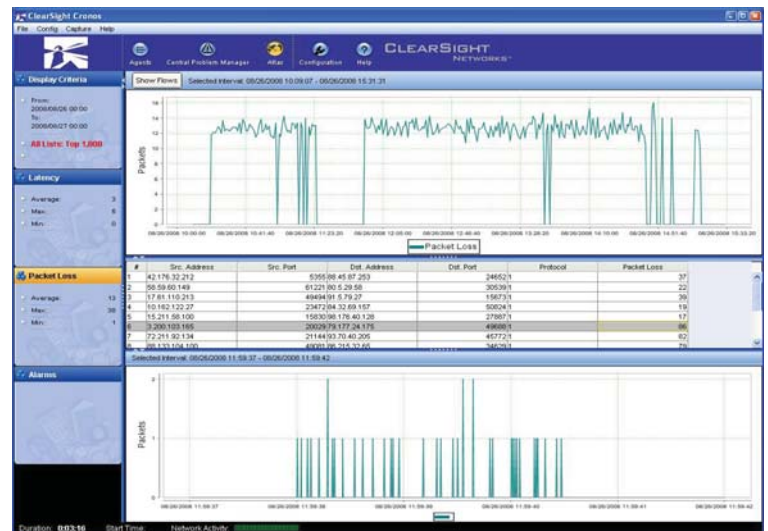


Figure 3: Packet Loss Measurements

## Conclusion

When a “1-millisecond advantage in trading applications can be worth millions of dollars a year”<sup>1</sup> an organization needs to make sure that getting all participants into its low-latency ecosystem is a priority. Current network monitoring solutions are simply not capable of detecting issues in the time granularity required. Current solutions must be modified, or a new breed must be created to fulfill this need. ClearSight Networks provides a solution that allows organization to monitor low-latency deployments for performance related issues.

<sup>1</sup> Data Latency Playing An Ever Increasing Role In Effective Trading. Wall Street’s quest to process data at the speed of light relies on the physical proximity of servers to overcome the technical barriers of data latency. By Richard Martin, InformationWeek. May 25, 2007



**CLEAR SIGHT**  
NETWORKS™

46401 Landing Parkway  
Fremont, CA 94538-6496

For more information, call or email us:  
Telephone (US Toll Free): 1-800-825-7563  
Telephone (International): +1-510-824-6000  
Fax: 1-510-824-6100  
Email: [sales@clearsightnet.com](mailto:sales@clearsightnet.com)  
[www.clearsightnet.com](http://www.clearsightnet.com)