2009 Vision for Low-Latency Messaging

When 29West introduced its streaming messaging system, Latency Busters® Messaging (LBM), many in the world of high-performance trading were still measuring its latency in milliseconds. LBM's breakthrough design decreased latencies by more than an order of magnitude, and redefined what could really be called low-latency messaging. With the release of Ultra Messaging® for the Enterprise (UME), 29West brought game-changing low-latency to the persistent and guaranteed messaging market as well.

Far from standing still, we are busy working on several new technologies that will enable even further latency reduction in the message path.

Kernel Bypass: The largest single remaining chunk of latency in the messaging path is the time taken by the OS kernel to run the network stack. 29West believes strongly in kernel bypass technology, which will allow the receiving messaging application to take delivery of incoming messages directly from memory buffers in the receiving NIC. Preliminary tests show this technology delivering latency reduction in the tens of microseconds, perhaps as much as a 50% reduction from current application-to-application messaging latency. 29West is currently working with a number of vendors to perfect the interaction of its messaging software with this kernel bypass technology.

Next-Generation Infrastructure: A natural extension of the desire for low latency is to install faster network infrastructure. The faster switches, routers and NICs used by technologies such as 10 Gigabit Ethernet and InfiniBand do deliver real latency improvements to customers who can deploy them. 29West has designed its products to work seamlessly with these new technologies, and has customers in production with them.

Inter-Process Communication: Moore's Law has continued to hold true even longer than Gordon Moore himself had anticipated. The most recent effect of this rapid increase in processing power has been massive horizontal scaling. Even commodity processors are being produced with 4 cores, and future plans only see this number increase. Enterprise servers routinely hold as many as 64 cores, and businesses are taking advantage of this by consolidating applications that were once spread across many different servers.

Recognizing the need for the fastest-possible messaging between dozens of processes on the same server, 29West will introduce an Inter-Process Communication (IPC) feature in the first half of 2009. This feature will effectively bypass the overhead of network communication through the operating system network stack. Shared memory segments will be used to allow messages to pass between processes with far lower latency and far higher throughput than any network technology can deliver. With absolutely no code changes at all, customers using our memory-sharing IPC transport will see messaging latency slashed down to nearly zero.

We anticipate that many of our most latency-sensitive customers will use the IPC feature to build applications that are far faster than is possible today.

Receive Queue Latency Monitoring: Even in a low-latency system, there are places that messages can be significantly delayed. If a bottleneck exists anywhere after the message has arrived, the associated benefit of the low-latency messaging system is drastically overshadowed. It is very important to monitor receive queue latency to ensure that your application is receiving the full benefit of its low-latency design.

The release of LBM 3.4 and UME 2.1 at the beginning of 2009 will add receive queue monitoring to provide detailed monitoring information unmatched in the industry. Users of 29West messaging will be able to see in real-time how deep the queues in their receiving applications are, as well as - to the microsecond - how long those messages have been enqueued. In conjunction with our SNMP monitoring agent, this level of information can be easily aggregated to a single monitoring point, delivering deep insight into the end-to-end latency of your messaging application.

29West engineers have always been thought leaders in the low-latency race, and we take that position very seriously. We are committed to delivering the best possible messaging solutions to our customers, and continue to strive to reduce latency further wherever possible.