HyperIP® by NETEX

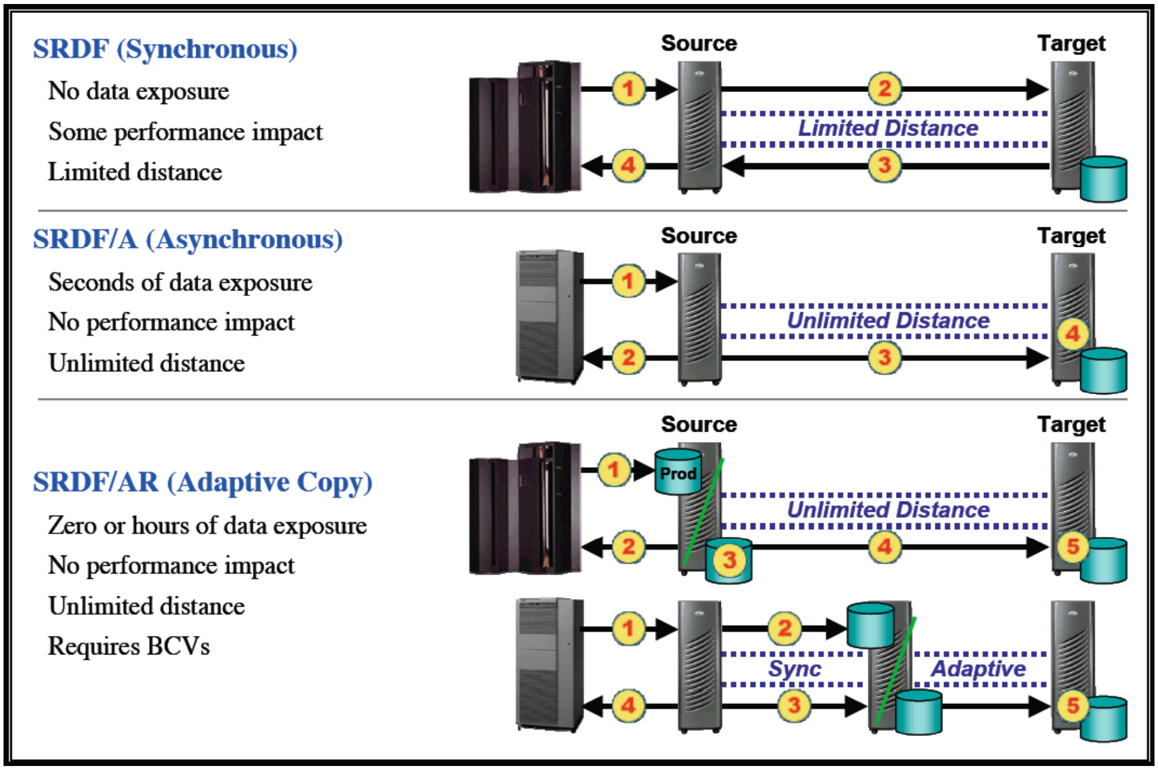
HyperIP®: SRDF Application Note

Steve Thompson, VP of Strategic Planning and Storage Networking, NetEx Software, Inc.

Introduction

HyperIP® is a Linux software application that quantifiably and measurably enhances large data movement over big bandwidth and long-haul IP networks. HyperIP® is a working example of RFC3135. RFC3135 describes techniques used to mitigate TCP performance problems over long-distance wide-area networks. These techniques are called "TCP Performance Enhancing Proxies" (PEP). HyperIP® is the most recent iteration of production-hardened transport, deployed in hundreds of Fortune 1000 accounts, that eliminates the negative impact of jitter, BERS, and network latency on TCP/IP data transfers.

SRDF (Symmetrix Remote Data Facility) is the gold standard for storage-to-storage DR and Business Continuity. It dominates the Enterprise market. There are three primary modes for SRDF: Synchronous, Asynchronous, and Adaptive Copy.



SRDF (Synchronous)

No data exposure

Some performance impact

Limited distance

SRDF/A (Asynchronous)

Seconds of data exposure

No performance impact

Unlimited distance

SRDF/AR (Adaptive Copy)

Zero or hours of data exposure

No performance impact

Unlimited distance

Requires BCVs

SRDF WAN connections typically route SRDF packets either over a Fibre Channel to GigE gateway (FCIP or iFCP) or natively in TCP/IP through the Symmetrix GigE Director. In both cases the SRDF packets are then routed over the standard TCP/IP backbone network. Occasionally, SRDF is routed to the WAN through Fibre Channel to SONET or through ESCON to FAST Ethernet or SONET.

SRDF over TCP/IP through the Symmetrix GigE director is the method most preferred by the market for moving SRDF packets over the WAN. Unfortunately, SRDF throughput on TCP/IP networks over 300 miles and bandwidth at DS3 and above, is dismal. The throughput continues to degrade as distance, BER (bit error rate), congestion, or jitter increases. These conditions also lead to significant packet loss and higher retransmission rates. The net effect is lower SRDF throughput. The fault falls squarely on TCP. HyperIP® is designed to provide SRDF with the throughput it requires by eliminating the limitations of TCP.

TCP high bandwidth long haul issues and limitations

Several characteristics of TCP/IP cause it to perform poorly over high bandwidth and long distances:

□ ***Window Size***

Window size is the amount of data allowed to be outstanding (in-the-air) at any given point-in-time. The available window size on a given bandwidth pipe is the rate of the bandwidth times the round-trip delay or latency. Using a cross-country OC-3 link (approximately 60 ms based on a total 6000-mile roundtrip) creates an available data window of 155Mbps x 60ms = 1,163Kbytes. A DS3 satellite connection (540 ms roundtrip) creates an available data window of 45Mbps X 540ms = 3,038Kbytes.

When this is contrasted with standard and even enhanced versions of TCP, there is a very large gap between the available window and the window utilized. Most standard TCP implementations are limited to 65Kbytes windows. There are a few enhanced TCP versions capable of using up to 512Kbytes or larger windows. Either case means an incredibly large amount of "dead air" and very inefficient bandwidth utilization.

□ ***Acknowledgement Scheme***

TCP causes the entire stream from any lost portion to be retransmitted in its entirety. In high bit-error-rate (BER) scenarios this will cause large amounts of bandwidth to be wasted in resending data that has already been successfully received, all with the long latency time of the path. Each retransmission is additionally subjected to the performance penalty issues of “Slow Start”.

□ ***Slow Start***

TCP data transfers start slowly to avoid congestion due to possible large numbers of sessions competing for the bandwidth, and ramp-up to their maximum transfer rate, resulting in poor performance for short sessions.

□ ***Session free-for-all***

Each TCP session is throttled and contends for network resources independently, which can cause over-subscription of resources relative to each individual session.

The net result of these issues is very poor bandwidth utilization. The typical bandwidth utilization for large data transfers over long-haul networks is usually less than 30% and more often less than 10%.

As fast as bandwidth costs are dropping, they are still not free.

Implications for GigE Director SRDF replication applications on TCP/IP long haul networks

New regulations, business continuity, and disaster recovery has led to a surge of storage-to-storage replication applications over the WAN. Man-made (9/11, blackouts, human error) and natural (hurricanes, earthquakes, tornados, firestorms) disasters have driven demand. TCP/IP has quickly become the preferred SRDF WAN protocol of choice. There are three reasons for this:

1) The market perceives bandwidth is essentially free. This is because the TCP/IP WAN bandwidth already exists for interactive traffic. Conventional wisdom is that SRDF snapshots and replication occur at night or on weekends. This is when the majority of users are not utilizing the network. Thus allowing already existing TCP/IP bandwidth to be leveraged by the SRDF replication applications without negatively impacting current applications.
2) Dedicated, separate SRDF replication WANs are not required. This also eliminates separate WAN management.
3) Additional bandwidth implemented for the SRDF replication applications will be shared by the interactive TCP/IP applications.

The facts show that TCP/IP bandwidth is neither free nor is there typically enough to accomplish the SRDF replication in the window of time allotted. TCP/IP bandwidth utilization and long haul issues are rarely taken into account in calculating bandwidth requirements. The most likely result is a bandwidth shortfall. This means either the SRDF replication cannot complete within the window of time allotted, or the user must buy more bandwidth, otherwise known as a conundrum.

The cost effective solution: NetEx HyperIP®

HyperIP® was designed specifically for large amounts of data over big bandwidth and long distance, to be highly efficient regardless of the BER congestion, or jitter. HyperIP® is a standard TCP/IP network node requiring no modifications to LAN/WAN infrastructures and no proprietary hardware. It provides transparent "acceleration" across long-haul high bandwidth WANs.

HyperIP® provides the following benefits:

- ❑ ***Window size***

The HyperIP® transport protocol keeps the available network bandwidth pipe full. The results are 90+% efficient link utilization. It eliminates the discrepancy between maximum available bandwidth and the results provided by native TCP/IP.

- ***Acknowledgement scheme***

HyperIP® transport protocol retransmits only the NAK'd segments and not all the data that has already been successfully sent.

- ***Slow Start***

Configuration parameters allow HyperIP® to start transmissions at a close approximation of the available session bandwidth.

- ***Dynamic adjustments***

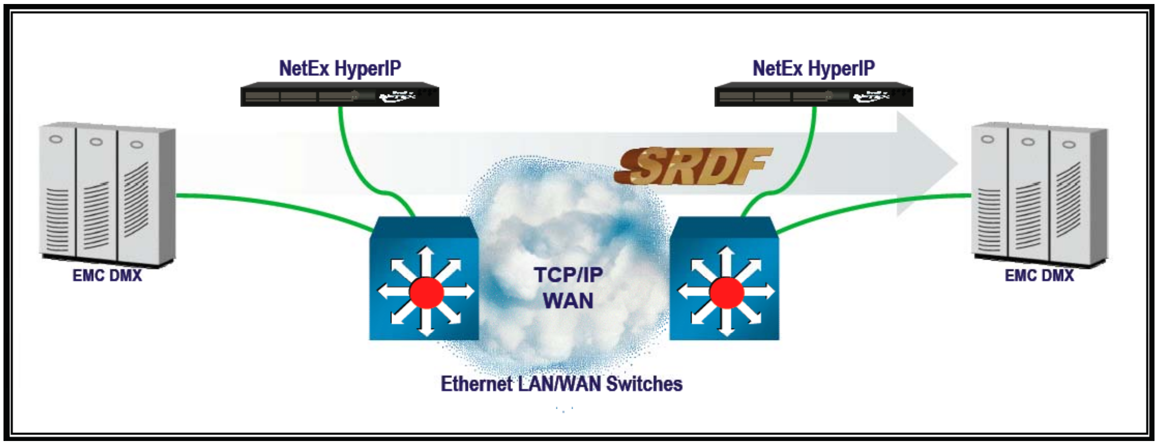
When feedback from the receiver in the acknowledgement protocol is received, HyperIP® quickly "zeroes-in" on the appropriate send rate for current conditions.

- ***Session pipeline***

HyperIP® design allows traffic from multiple TCP sessions to be aggregated over a smaller set of connections between the HyperIP® devices, enabling a more efficient use of the bandwidth and less protocol overhead acknowledging many small messages for individual connections.

HyperIP®'s incredible SRDF test results over long-haul IP WANs

Testing results with EMC SRDF have been outstanding. Although the tests did not include the HyperIP® compression engine, the test results throughput was quite impressive. SRDF achieved bandwidth utilization consistently exceeding 90% from distances of hundreds of miles (with high bit error rates on dirty lines) to as far as geosynchronous satellite distances.



- ***What this means to end-users***

SRDF throughput on native Ethernet TCP/IP fabrics with HyperIP® is now the highest possible throughput option. Windows for SRDF/SNAP and Volume replication can now be met. The promise of free bandwidth for SRDF may just turn out to be real.

HyperIP® software components

- ❑ *IP-Packet Edge Intercept*

This component intercepts IP packets, optimizes for performance, and reroutes over the HyperIP® protocol on the network.

When a message is intercepted and rerouted, the original IP addressing information is retained and sent as additional protocol information. This allows each message to be reconstructed with the original addressing information at the destination side. A pre-built configuration file describing the HyperIP® configuration is processed at initialization.

- ❑ *IP Accelerator*

This component establishes and maintains connections with other HyperIP® nodes on the IP network. This IP Accelerator receives intercepted packets from each of the Edge processes. It aggregates these packets into more efficient buffers, and then passes these buffers to the HyperIP® Transport component, which sends them to the HyperIP® node on the other side of the network.

The remote HyperIP® receives these aggregated buffers on the network and passes them on to the IP Accelerator, which sends the packets from the buffer on to the appropriate Edge Interface process.

- ❑ *HyperIP® Transport*

The transport component provides the transport delivery mechanism between HyperIP® nodes. It receives the optimized buffers from the IP Accelerator and delivers them to the destination HyperIP® node for subsequent delivery to the end destination. It is responsible for maintaining acknowledgements of data buffers and resending buffers when required. It maintains a flow control mechanism on each connection, and optimizes the performance of each connection to match the available bandwidth and network capacity.

Since HyperIP® provides a complete transport mechanism for managing data delivery, it uses UDP socket calls as an efficient, low overhead, data streaming protocol to read and write from the network.

- ❑ *Compression Engine*

The HyperIP® LZO-based software compression engine compresses the aggregated packets that are in the highly efficient IP Accelerator buffers. This provides an even greater level of compression efficiency since a large block of data is compressed at once rather than multiple small packets being compressed individually. Testing has demonstrated compression ranging from 2 to 8 to 1.

Summary and Conclusion

HyperIP® provides the highest possible throughput for SRDF over TCP/IP WANs. It does this with "Production Hardened Transport" that provides:

- Bandwidth utilization that consistently exceeds 90+% regardless of BER, congestion or jitter;
- 2:1 to 8:1 compression;
- Elimination of TCP/IP network latency.

SRDF over TCP/IP on the GigE director with HyperIP® is the solution of choice.