TECHNICAL NOTE 4015

Flow Control for FibreXtreme Products

Systran Corporation
4126 Linden Avenue
Dayton, Ohio 45432-3068 USA

FAX: 937-258-2729
Phone: 937-252-5601
or 800-252-5601(U.S. only)

INTRODUCTION

This TECHNOTE describes the use of flow control methods with Systran FibreXtreme and LinkXchange products. Flow Control is the process in which the receiver can control the transmitter output to match its own receive rate.

Systran FibreXtreme products provide a flow control mechanism that should be enabled in almost every application, even when the data source cannot be stopped.

In general, there are two basic arguments for using flow control:

(1) It is difficult to recover under software control from multiple errors caused by FIFO overflows.

(2) It is always better to drop the data at the sending source (as opposed to the receiving destination) if the system experiences a temporary overload. If bad data makes it to the receiver, this bad data must still be read out of the receiver's RX FIFO and handled at the application level.

DISCUSSION

Flow control is necessary in network communications to prevent the transmitting node(s) from flooding the receive node(s) with data. The flow of data from one device to another must be adjusted so the receiving node can handle all of the incoming data. This is particularly important where the transmitting node is capable of sending data much faster than the receiving node can receive it. The objective is to manage the flow of data to maximize throughput.

Topologies Supporting Flow Control

The Serial Front Panel Data Port (FPDP) standard (VITA 17.1) defines the high-speed, low-latency, data-streaming serial communications protocol used by FibreXtreme cards to transfer information across a link (e.g., fiber-optic or copper media).

Flow control can be used in any single-source system that has a return path from the last receiver to the transmitter. When flow control is enabled, the transmitter will not transmit data when a receiver is telling it to back off or the receiver fiber is missing. If flow control is disabled, the transmitter continues to send data even when a receiver signals for it to stop, the link is down, or the receive fiber is missing.

Figure 1 shows a point-to-point topology with data flowing in one direction. Figure 2 shows a point-to-point topology with data flowing in both directions. Flow control works identically in either case. However, for the topology shown in Figure 2, both nodes are capable of sending data and flow control. Figure 3 shows a single-master ring, where any receiver (RCV) can back off the transmitter (XMIT).

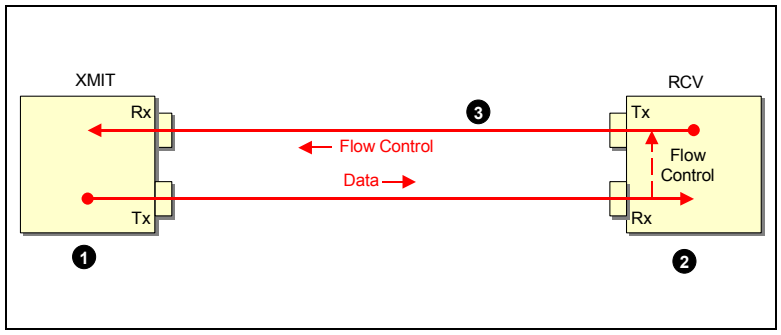

Figure 1. Point-to-Point Topology, Data Flowing in One Direction

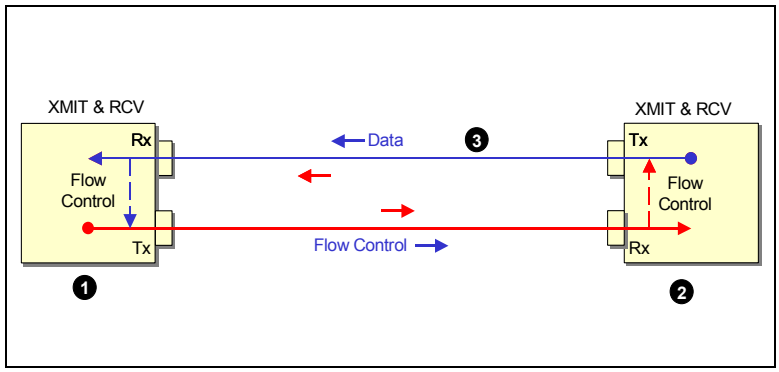

Figure 2. Point-to-Point Topology, Data Flowing in Both Directions

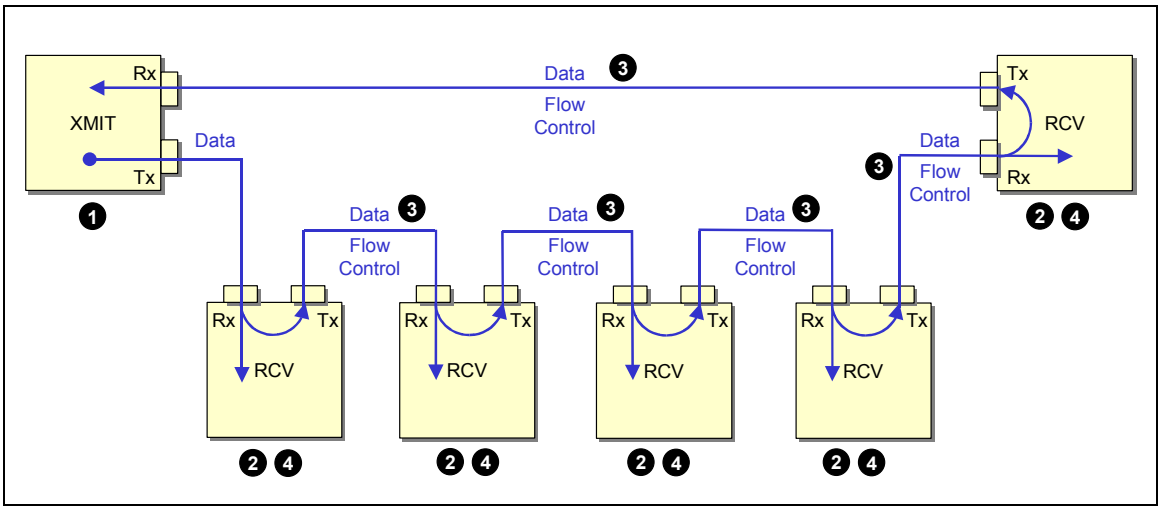

Figure 3. Single-Master Copy/Loop Mode Topology

Flow Control Steps

A system is transferring data in a point-to-point topology with data flowing in one direction as shown in Figure 4. No errors are present and the receiver is removing data as fast as the transmitter is sending it. The data source is an FPDP transmitter and the final data receiver is an FPDP receiver.

The FPDP receiver cannot sustain the system's data throughput. This could be the result of a burst of data that temporarily overwhelms the FPDP receiver, or the FPDP receiver is processing some link errors, which temporarily decreases its receive throughput. The following steps describe how flow control is sent back to the data source.

Systran Corporation
4126 Linden Avenue
Dayton, OH 45432-3068 USA

1. The FPDP receiver sends a suspend signal (/SUSPEND).
2. The FPDP Host Bus Logic stops removing data from the RX FIFO.
3. The RX FIFO fills, which creates a pending overflow condition.
4. The pending overflow signal is sent to the Serial FPDP Logic.
5. A Serial FPDP suspend signal (STOP primitive) is sent across the link to the link transmitter.
6. The link transmitter's Serial FPDP Logic stops sending data across the link and stops removing data from the TX FIFO.
7. The TX FIFO fills, which creates a pending overflow condition.
8. The pending overflow signal is sent to the FPDP Host Bus Logic.
9. A suspend signal (/SUSPEND) is sent to the FPDP transmitter.
10. The FPDP transmitter stops sending data.

Given the elasticity of the FIFOs and delays within the logic, flow control will not always propagate all the way through a system. For example, steps 1-2 shown above occur. However, before the RX FIFO completely fills, the FPDP receiver clears its pending overflow condition, and the RX FIFO empties.

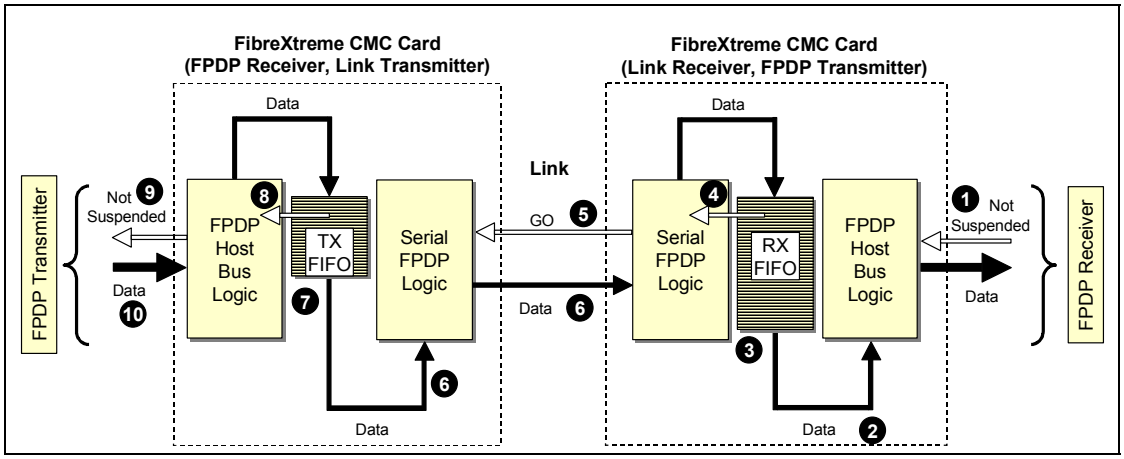

Figure 4. Fully Operational System Going into Suspended Condition

The FPDP receiver clears it overflow condition. The following steps describe how the suspended condition is removed from the system. See Figure 5.

1. The FPDP receiver removes the suspend signal (/SUSPEND).
2. The FPDP Host Bus Logic starts removing data from the RX FIFO.
3. The RX FIFO empties.
4. The pending overflow signal is removed from the Serial FPDP Logic.
5. The Serial FPDP suspend signal (STOP primitive) is cleared, and a Serial FPDP GO primitive is sent across the link to the link transmitter.
6. The link transmitter's Serial FPDP Logic starts sending data across the link and starts removing data from the TX FIFO.
7. The TX FIFO empties.
8. The pending overflow signal is removed from the FPDP Host Bus Logic.
9. The suspend signal (/SUSPEND) to the FPDP transmitter is cleared.
10. The FPDP transmitter starts sending data.

Systran Corporation
4126 Linden Avenue
Dayton, OH 45432-3068 USA

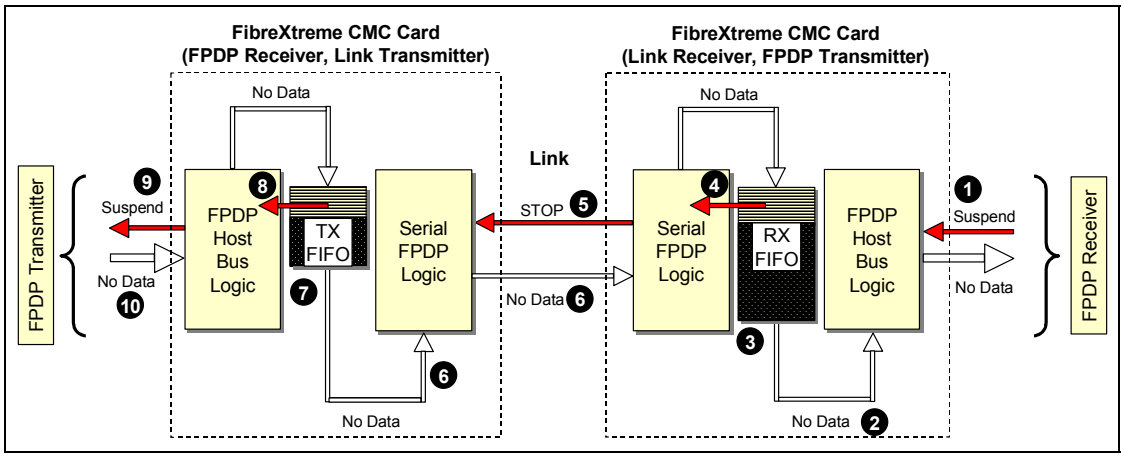

Figure 5. Suspended System Coming out of Suspended Condition

Flow Control Delays in a Point-to-Point Topology

Figure 1 shows a point-to-point topology with data flowing in one direction. Three delays must be considered when determining the additional capacity needed in the receive node's RX FIFO and the maximum cable length possible.

1. Maximum delay between receipt of the Serial FPDP suspend signal (STOP primitive) at the Serial FPDP transmitter and the actual stoppage of the data transmission.
2. Maximum delay for a Serial FPDP receiver to transmit the Serial FPFP suspend signal (STOP primitive) after a pending overflow condition in its RX FIFO.
3. Maximum cable delay from the Serial FPDP receiver to the Serial FPDP transmitter to account for the transmission time of the Serial FPDP suspend signal (STOP primitive).

If the Serial FPDP receiver is also transmitting data (i.e., bi-directional data flow) as shown in Figure 2, the maximum delay for the Serial FPDP receiver to transmit the Serial FPDP suspend signal (STOP primitive) after a pending overflow condition in its RX FIFO increases. The other delays associated with a point-to-point topology remain the same.

Example 1

Question: Will any data be lost if flow control is enabled and a pending overflow condition occurs when two FibreXtreme cards are connected in a point-to-point topology as shown in Figure 2? Assume both cards are saturating the link with data. The maximum data rate per direction that can be transferred is 105 MBps for SL100 and 247 MBps for SL240.

Answer: Assume one of the receiving nodes experiences a pending overflow condition in its RX FIFO. According to the values given in Table 1, the Serial FPDP transmitter will stop transmitting data in 70.814 µs maximum (19.610 µs + 50 µs + 1.204 µs) for an SL100 link and 58.848µs maximum (8.336 µs + 50 µs + 512 ns) for an SL240 link.

A FibreXtreme card is designed to transmit the Serial FPDP suspend signal (STOP primitive) when it has less space in its RX FIFO than the amount of data contained in 20 km of fiber. The number of words in 20 km of fiber is 2657 for an SL100 link and 6250 for an SL240 link.

NOTE: To simplify calculations and to assume a worse case than actual, the total number of 32-bit words in 20 km of fiber is used instead of the actual amount of data in 20 km of fiber. The actual amount of data can fluctuate. There are 6-9 overhead words and 0-512 data words per Serial FPDP frame.

To verify the FibreXtreme card's RX FIFO's pending overflow threshold is set adequately, the number of words in 20 km of fiber must be greater than the number of words that can be transmitted before the Serial FPDP transmitter stops transmitting data after an overflow condition occurs on the receiving node.

70.814 µs / (37.64 ns per 32-bit word) = 1881 32-bit words on the SL100 link

Systran Corporation
4126 Linden Avenue
Dayton, OH 45432-3068 USA

58.848 µs / (16 ns per 32-bit word) = 3678 32-bit words on the SL240 link

Thus, the FibreXtreme card's RX FIFO's pending overflow threshold is set adequately, and no data will be lost when a pending overflow condition occurs.

Flow Control Delays in a Copy/Loop Mode Topology

Figure 3 shows a single-master ring topology. Four delays must be considered when determining the additional capacity need in the receive nodes' RX FIFO and the maximum cable length possible.

1. Maximum delay between receipt of the Serial FPDP suspend signal (STOP primitive) at the Serial FPDP transmitter and the actual stoppage of the data transmission.
2. Maximum delay for a Serial FPDP receiver to transmit the Serial FPFP suspend signal (STOP primitive) after a pending overflow condition in its RX FIFO.
3. Maximum cable delay from the Serial FPDP receiver to the Serial FPDP transmitter to account for the transmission time of the Serial FPDP suspend signal (STOP primitive).
4. Maximum delay between receipt of data by a Serial FPFP receiver and the re-transmission of the data on the next segment of the link.

Example 2

Question: When using a loop topology as shown in Figure 3, what is the maximum amount of fiber-optic cable allowed so flow control will work when one of the receiving nodes experience a pending overflow condition? Assume the transmitter is saturating the link with data. The maximum data rate that can be transferred is 105 MBps for SL100 and 247 MBps for SL240. The maximum number of receive nodes allowed in this topology is four for SL100 and five for SL240.

Answer: Using the values shown in Table 1, the maximum delay per receive node is 22.019 µs (19.610 µs + 2.409 µs) for an SL100 link and 9.360 µs (8.336 µs + 1.024 µs) for an SL240 link.

The maximum delay between receipt of the Serial FPDP suspend signal (STOP primitive) at the Serial FPDP transmitter and the actual stoppage of the data transmission is 1.204 µs for SL100 and 512 ns for SL240.

Neglecting cable propagation delays, the maximum delay between the original assertion of STOP and the actual stoppage of data from the Serial FPDP transmitter is

89.280 µs = 4 * 22.019 µs + 1.204 µs for an SL100 link

47.312 µs = 5 * 9.360 µs + 512 ns for an SL240 link

A FibreXtreme card is designed to transmit the Serial FPDP suspend signal (STOP primitive) when it has less space in its RX FIFO than the amount of data contained in 20 km of fiber. Using 5 ns/m for the speed of light, the propagation delay for 20 km of fiber is 100 µs. The maximum total cable length allowed in the ring is:

(100 µs – 89.280 µs) / 5 ns/m = 2144 meters for an SL100 link

(100 µs - 47.312 µs) / 5 ns/m = 10,537 meters for an SL240 link

Systran Corporation
4126 Linden Avenue
Dayton, OH 45432-3068 USA

Table 1. FibreXtreme Flow Control Delays

	Topology			
Delay*	**Point-to-Point (Uni-directional Data Flow)**	**Point-to-Point (Bi-directional Data Flow)**	**Copy/Loop Mode**	**Delay Calculation***
1	1.204 µs (SL100) 512 ns (SL240)	1.204 µs (SL100) 512ns (SL240)	1.204 µs (SL100) 512 ns (SL240)	1
2	339 ns (SL100) 144 ns (SL240)	19.610 µs (SL100) 8.336 µs (SL240)	19.610 µs (SL100) 8.336 µs (SL240)	2
3	50 µs	50 µs	50 µs	3
4	Not Applicable	Not Applicable	2.409 µs (SL100) 1.024 µs (SL240)	4

* See footnotes at end of this TECHNOTE for explanation of constants and delay calculations

Single-Master Copy/Loop Mode Topology with LX2500

A LinkXchange LX2500 Crossbar Switch can be used to build any of the FibreXtreme topologies discussed so far. For example, Figure 6 shows a single-master ring, where any receiver (RCV) can back off the transmitter (XMIT). This is the same topology as shown in Figure 3 but is built with an LX2500. The LX2500 provides the following advantages in this topology.

1. Provide fault isolation when required so the loss of one node will not break the ring.
2. Provide a means to reconfigure the ring without the need to change cabling.

The flow control delays listed in Table 1 are also applicable when using an LX2500. In addition, the LX2500's port-to-port delay is 5-10ns, regardless of whether an SL100 or SL240 card is plugged into the port.

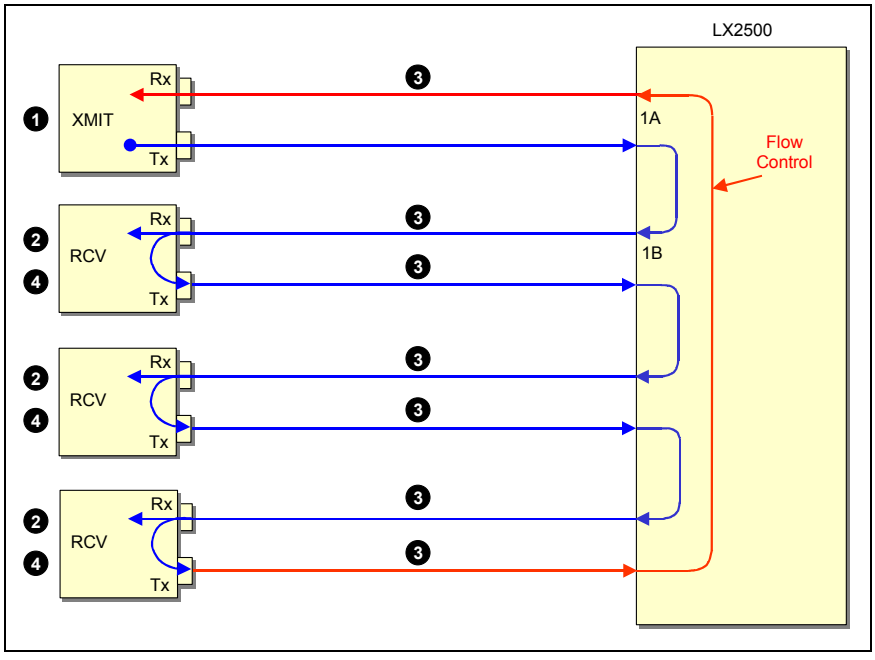

Figure 6. Single-Master Copy/Loop Mode Topology with LX2500

Systran Corporation
4126 Linden Avenue
Dayton, OH 45432-3068 USA

Single-Master Copy Mode Application With Redundant Transmitter

Figure 7 shows a source of data being copied to multiple receivers. RCV 1, the slowest of the receivers is configured to return flow control information to the transmitter. If XMIT 1 stops transmitting (e.g., cable break), the LX2500 automatically switches the connection and Flow Control to XMIT 2 as shown in Figure 8.

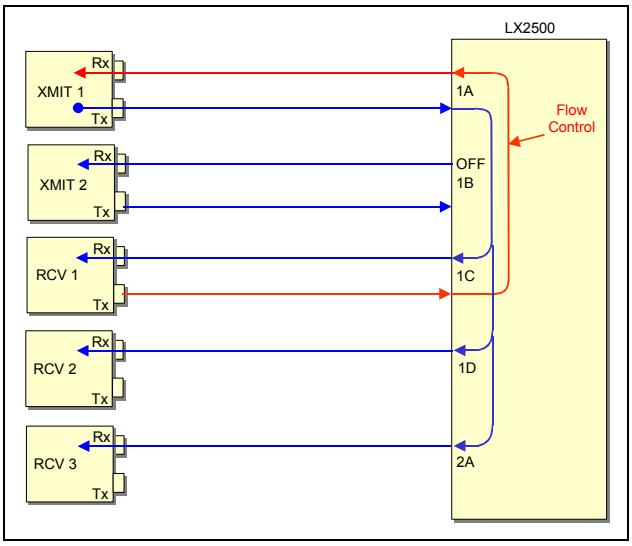

Figure 7. LX2500 Copies XMIT 1 Data to Multiple Receivers

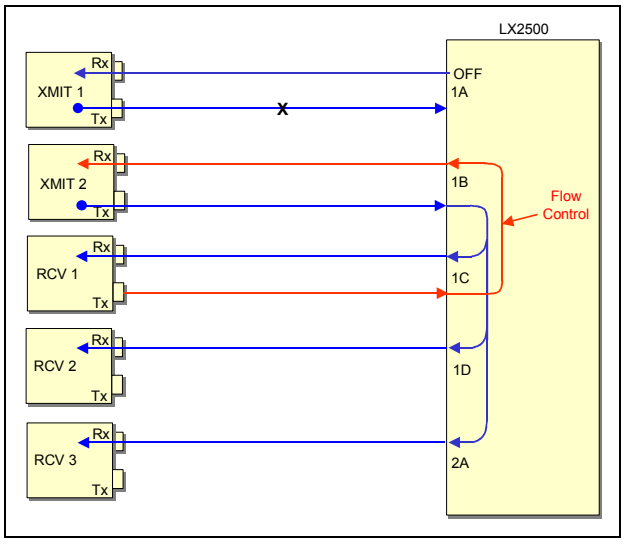

Figure 8. LX2500 Copies XMIT 2 Data to Multiple Receivers

CONCLUSION

When should flow control be used?

Flow control should be enabled in almost every application. Many common applications have a free running data source (e.g., A/D from RADAR). Since the data source cannot be stopped, it is frequently thought flow control is not required. This is not true. Even if the data source cannot suspend its data flow or maximum link throughput must be sustained, flow control information should still be returned to the data source from the receiver node(s). In an overflow condition, some data will be lost. The two choices available in this data-loss situation are

(1) Drop data at the source until the receiver recovers from the overflow condition or

(2) Have the receiver process and discard the link errors and bad data until it recovers from the overflow condition.

It is much better to drop the data at its source as opposed to processing and discarding the link errors and bad data at the receiver. If the receiver's RX FIFO overflows (due to flow control not being used), the receiver must process and recover from all of the link errors and re-synchronize to the data stream. While the receiver attempts to recover from the link errors and re-synchronize to the data stream, the data source continues sending data. The receiver's receive throughput drops drastically while it attempts to recover from the overflow condition. However, the data source continues sending data at its original rate, which makes the receiver's recovery process extremely difficult.

For illustration, an example based on a customer’s actual point-to-point application is given. The data source is transmitting data at an average rate of 240 MBps. The receiver is designed to be faster than the data source and the system works properly with flow control enabled so the customer believes flow control is not needed. After flow control is disabled, the receiver experiences an overflow condition due to a temporary burst throughput that was greater than the expected average of 240 MBps. While the receiver attempted to recover from the overflow condition, its receive throughput dropped to ~2 MBps.

Systran Corporation
4126 Linden Avenue
Dayton, OH 45432-3068 USA

The data source continued to transmit data at an average of 240 MBps, which continued to overwhelm the receiver. The receiver was unable to recover from the initial temporary overflow condition. Based on this experience, this customer saw the benefits of using flow control.

When Should Flow Control Not Be Used?

Some unusual conditions could apply where flow control is not desirable, but they require very careful system planning and should be confirmed with Systran prior to architectural finalization. One exception is for applications that cannot use a duplex fiber-optic link, which means status information (link up and state of flow control) is not available from the remote node. In this circumstance, disable flow control to allow the transmitter to function without the receiver carrier signal.

In general, consider these two basic rules:

(3) It is difficult to recover under software control from multiple errors caused by FIFO overflows.

(4) It is always better to drop the data at the sending source (as opposed to the receiving destination) if the system experiences a temporary overload. If bad data makes it to the receiver, this bad data must still be read out of the receiver's RX FIFO and handled at the application level.

RELATED INFORMATION

- *FibreXtreme SL100/SL240 Hardware Reference Manual for PCI, PMC, and CPCI Cards,* Systran Corporation.
- *FibreXtreme SL100/SL240 Hardware Reference for VME and Rehostable CMC FPDP Cards*, Systran Corporation.
- *LinkXchange LX2500 Crossbar Switch Hardware Reference Manual,* Systran Corporation.
- *Serial Front Panel Data Port (FPDP) Standard, VITA 17.1*

Systran Corporation
4126 Linden Avenue
Dayton, OH 45432-3068 USA

FOOTNOTES

Constants Used in Delay Calculations

- Optical Link Rate = 1.0625 Gbps (SL100), 2.5 Gbps (SL240)
- Optical Link Rate after 8B/10B Encoding = Optical Link Rate * (8 bits / 10 bits) * (1 byte / 8 bits)
- Time for Each 32-bit Word on the Link = 4 / Optical Link Rate after 8B/10B Encoding
- Number of Words in 20 km of Fiber = Speed of Light * 20 km / Time for Each 32-bit Word on the Link
- Maximum Number of 32-bit Data Words per Serial FPDP Frame = 512 for a saturated link, 0 words for an empty link (i.e., no data flow)
- Maximum Number of Overhead Words per Serial FPDP Frame = 6 with Copy Master Mode disabled, 9 with Copy Master Mode enabled (used for all calculations shown below)
- Speed of Light = 5 ns/m
- Maximum Cable Length Between Nodes = 10 km

Delay Calculations

1. Delay 1 = 32 * Time for each 32-bit Word on the Link
2. Delay 2 = Time for each 32-bit Word on the Link * (Maximum Number of 32-bit Data Words per Serial FPDP Frame + Maximum Number of Overhead Words per Serial FPDP Frame)
3. Delay 3 = Speed of Light * Maximum Cable Length Between Nodes
4. Delay 4 = 64 * Time for each 32-bit Word on the Link

Systran Corporation
4126 Linden Avenue
Dayton, OH 45432-3068 USA