RapidIO™ Interconnect Specification Part 9: Flow Control Logical Layer Extensions Specification

Rev. 1.3, 06/2005

RapidIO Trade Association

Revision History

Revision	Description	Date
1.0	First release	06/18/2003
1.3	No technical changes, revision changed for consistancy with other specifications Converted to ISO-friendly templates	02/23/2005
1.3	Removed confidentiality markings for public release	06/07/2005

Questions regarding the RapidIO Trade Association, specifications, or membership should be forwarded to:

Suite 325, 3925 W. Braker Lane
Austin, TX 78759
512-305-0070 Tel.
512-305-0009 FAX.

Table of Contents

Table of Contents

Blank page

List of Figures

List of Figures

List of Tables

List of Tables

Chapter 1 Flow Control Overview

1.1 Introduction

A switch fabric based system can encounter several types of congestion, differentiated by the duration of the event:

- Ultra short term
- Short term
- Medium term
- Long term

Congestion can be detected inside a switch, at the connections between the switch, and other switches and end points. Conceptually, the congestion is detected at an output port that is trying to transmit data to the connected device, but is receiving more information than it is able to transmit. This excess data can possibly "pile up" until the switch is out of storage capacity, and then the congestion spreads to other devices that are connected to the switch's inputs, and so on. Therefore, contention for a particular connection in the fabric can affect the ability of the fabric to transmit data unrelated to the contested connection. This is highly undesirable behavior for many applications.

The length of time that the congestion lasts determines the magnitude of the effect the congestion has upon the system overall.

Ultra short term congestion events are characterized as lasting a very small length of time, perhaps up to 500 or so nanoseconds. In a RapidIO type system these events are adequately handled by a combination of buffering within the devices on either end of a link and the retry based link layer mechanism defined in the RapidIO Part 4: 8/16 LP-LVDS Physical Layer and RapidIO Part 6: 1x/4x LP-Serial Physical Layer Specifications. This combination adds "elasticity" to each link in the system. The impact of ultra short term events on the overall system is minor, if noticeable at all.

Short term congestion events last much longer than ultra short term events, lasting up into the dozens or hundreds of microseconds. These events can be highly disruptive to the performance of the fabric (and the system overall), in both aggregate bandwidth and end to end latency. Managing this type of congestion requires some means of detecting when an ultra short term event has turned into a short term event, and then using some mechanism to reduce the amount of data being

injected by the end points into the congested portion of the fabric. If this can be done in time, the congestion stays localized until it clears, and does not adversely affect other parts of the fabric.

Medium term congestion is typically a frequent series of short term congestion events over a long period of time, such as seconds or minutes. This type of event is indicative of an unbalanced data load being sent into the fabric. Alleviating this type of congestion event requires some sort of software based load balancing mechanism to reconfigure the fabric.

Long term congestion is a situation in which a system does not have the raw capacity to handle the demands placed upon it. This situation is corrected by upgrading (or replacing) the system itself.

This specification addresses the problem of short term congestion.

1.2 Requirements

The flow control mechanism shall fulfill the following goals:

- Simple - excess complexity will not gain acceptance
- React quickly - otherwise the solution won't work
- Robust - same level of protection and recovery as the rest of RapidIO
- Scalable - must be able to extend to multi-layer switch systems
- Compatibility with all physical layers

1.3 Problem Illustration

The *RapidIO Part 1: Input/Output Logical Specification* defines a transaction request flow as a series of packets that have a common source identifier and a common destination identifier at some given priority. On a link, packets of a single transaction request flow can be interleaved with packets from one or more other transaction request flows.

No assumptions are made on the underlying switch architecture for this discussion of the short term congestion problem. Also for the purposes of this discussion, an idealized output queued switch is assumed, which in literature is also used to compare the performance of a particular switch under study. Packet buffers are associated with the output of the switch. An example switch topology showing output buffers is illustrated in Figure 1-1 below. A point of congestion is therefore associated with an output buffer of such a switch.

The problem that is to be addressed by this specification is caused by multiple independent transaction request flows, each with burst and spatial locality characteristics that typically do not exceed the bandwidth capacity of links or end points. Due to the statistical combination of such transaction request flows, usually

in the middle of multistage topologies, the demand for bandwidth through a particular link exceeds the link's capacity for some period of time, for example, Data Flows a, b, and c for an output port of Switch 3 as shown in Figure 1-1. As a result, the output buffer for this port will fill up, causing the link layer flow control to be activated on the links of the preceding switch stages. The output packet buffers for Switches 1 and 2 then also fill up. Packets for transaction request flows, such as data flow d, in these same output buffers not destined for the output port with the full buffer in Switch 3 are now also waiting, causing additional system performance loss. This phenomenon is known as higher order head of line blocking.

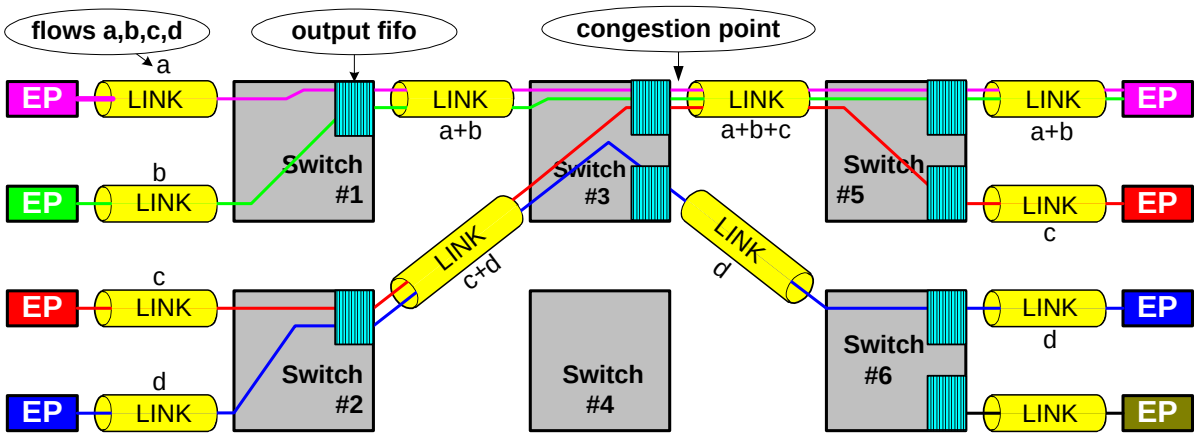

Figure 1-1. Interconnect Fabric Congestion Example

A second problem, less frequently a contributor to system performance loss, occurs when an end point cannot process the incoming bandwidth and employs link layer flow control to stop packets from coming in. This results in a similar sequence of events as described above.

The problem described in this section is very well known in the literature. The aggregate throughput of the fabric is reduced with increased load when congestion control is not applied (see reference [1]). Such non-linear behavior is known as 'performance-collapse'. It is the objective of this specification to provide a logical layer flow control mechanism to avoid this collapse. Research also shows that relatively simple "XON/XOFF" controls on transaction request flows can be adequate to control congestion in fabrics of significant size.

The reason for the described non-linear behavior is illustrated with a saturation tree. The point at which a single transaction request flow that causes link bandwidth to be exceeded and causes buffer overflow is referred to as the root of the saturation tree. This tree grows backward towards the sources of all transaction request flows going through these buffers, and all buffers that these transaction request flows pass through in preceding stages, causing even more transaction request flows to be affected.

An important design factor for interconnect fabrics is the latency between a

congestion control action being initiated and the transaction request flow source acting in response. This latency determines, among other factors, the required buffer sizes for the switches. To keep such buffers small, the latency of a congestion control mechanism must be minimized. For example, 10 data flows contribute to a buffer overflow (forming what is known as a “hotspot”). If it takes 10 packet transmission times for the congestion notification to reach the sources and the last packets sent from the sources to reach the point of congestion after the sources react to the congestion notification, up to 100 packets could be added to the congested buffer. The number of packets added may be much smaller depending on the rate of oversubscription of the congested port.

Reference

[1] “Tree saturation control in the AC3 velocity cluster interconnect”, W. Vogels et.al., Hot Interconnects 2000, Stanford.

Chapter 2 Logical Layer Flow Control Operation

2.1 Introduction

This chapter describes the logical layer flow control mechanism.

2.2 Fabric Link Congestion

In compliant devices, logical layer flow control methods shall be employed within a fabric or destination end point for the purpose of short term congestion abatement at the point in time and location at which excessive congestion is detected. This remediation scheme shall be enacted via explicit flow control messages referred to as transmit off (XOFF) and transmit on (XON) congestion control packets (CCPs) which, like any other packet, require link-level packet acknowledgements. The XOFF CCPs are sent to shut off select flows at their source end points. Later, when the congestion event has passed, XON CCPs are sent to the same source end points to restore those flows.

The method used to detect congestion is implementation specific and is heavily dependent upon the internal packet buffering structure and capacity of the particular switch device. In the example output port buffered switch from “Section 1.3, Problem Illustration” on page 10, congestion occurs when some output buffer watermark is exceeded, but this is not the only way of detecting congestion. Several possible implementation methods are described in Appendix A. These described methods are purely exemplary and are not intended to be an exhaustive list of possible methods.

2.3 Flow Control Operation

The flow control operation consists of a single FLOW CONTROL transaction as shown in Figure 2-1. The FLOW CONTROL transaction is issued by a switch or end point to control the flow of data. This mechanism is backward compatible with

RapidIO legacy devices in the same system.

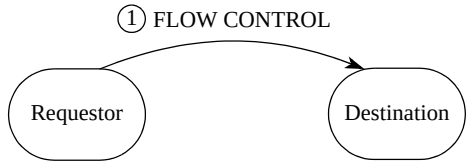

Figure 2-1. Flow Control Operation

2.4 Physical Layer Requirements

This section describes requirements put upon the system physical layers in order to support efficient logical layer flow control.

2.4.1 Fabric Topology

The interconnect fabric for a system utilizing the logical layer flow control extensions must have a topology such that a flow control transaction can be sent back to any transaction request flow source. This path through the fabric may be back along the path taken by the transaction request flow to the congestion point or it may be back along a different path, depending upon the requirements of the particular system.

2.4.2 Flow Control Transaction Transmission

Flow control transactions are regarded as independent traffic flows. They are the most important traffic flow defined by the system. Flow control transactions are always transmitted at the first opportunity at the expense of all other traffic flows if possible. For the 8/16 LP-LVDS and 1x/4x LP-Serial physical layer specifications, this requires marking flow control packets with a “prio” field value of 0b11, and a “crf” bit value of 0b1, if supported. These transactions use a normal packet format for purposes of error checking and format.

Because an implicit method of flow restoration was simulated and found to be impractical for RapidIO fabrics due to lack of system knowledge in the end point, an explicit restart mechanism using an XON transaction is used. In the CCP flow back to the source end point, XOFF and XON CCPs may be dropped on input ports of downstream elements in the event of insufficient buffer space.

2.4.2.1 Orphaned XOFF Mechanism

Due to the possibility of XON flow control packets being lost in the fabric, there shall be an orphaned XOFF mechanism for the purpose of restarting orphaned flows which were XOFF’d but never XON’d in end points. Details of this mechanism are implementation specific, however the end point shall have sufficient means to avoid

abandonment of orphaned flows. A typical implementation of such a mechanism would be some sort of counter. A description of a possible implementation is given in Appendix A. The Orphaned XOFF Mechanism is intended to work with the rest of the XON/XOFF CCPs to handle the short term congestion problem as previously described, and so shall operate such that software intervention is not required or inadvertently invoked.

2.4.2.2 Controlled Flow List

It is required that elements which send XOFFs keep a list of flows they have stopped, along with whatever flow-specific information is needed to select flows for restart, such as per-flow XON watermark level, or relative shut off order. This information shall be stored along with flow identification information in a “controlled flow list”, a memory structure associated with the controlling element. It shall be permissible in the time following the sending of an XOFF CCP for the flow control -initiating element to re-evaluate system resources and modify the flow restart ordering or expected XON watermark level within the controlled flow list to better reflect current system state. It shall not however be permissible to abandon the controlled flow by “forgetting” it, either due to lack of controlled flow list resources or other factors. In the event that limited controlled flow list resources cause the congested element to have insufficient room to issue another XOFF CCP which is deemed more important than a previously-XOFF’d controlled flow, then that previously-XOFF’d controlled flow may be prematurely XON’d and removed from the controlled flow list. The new, more important flow may be XOFF’d and take its place in the controlled flow list.

Details of the controlled flow list are implementation specific, though at the very least it shall contain entries for each currently XOFF’d flow, including flow identification information. It is likely that some state information will be required, such as expected time of flow restart, or per-flow restart watermark levels. The controlled flow list size is selected to provide coverage for short term congestion events only. Remediation for medium and greater -term congestion events is beyond the scope of logical layer flow control as these events likely indicate systemic under-provisioning in the fabric.

2.4.2.3 XOFF/XON Counters

XOFF/XON counters shall be instantiated for some number of output flows at the end point. Since the number of flows may be large or unpredictable, the number of counters and how flows are aggregated to a particular counter is implementation dependant. However, all flows must be associated with a counter. For simplicity, the following behavioral description assumes a single flow associated with a single counter. The counter is initialized to zero at start up or when a new DestinationID and given Priority is initialized. The counter increments by one for each associated XOFF CCP and decrements by one for each associated XON CCP, stopping at zero. Only when this counter is equal to zero is the flow enabled. In no event shall the counter wrap upon terminal count. If the orphaned XOFF mechanism activates, the

counter is reset to zero and the flow is restarted.

2.4.3 Priority to Transaction Request Flow Mapping

When a switch or end point determines that it is desirable to generate a flow control transaction, it must determine the associated flowID for the (non-maintenance and non-flow control) packet that caused the flow control event to be signalled. Maintenance and flow control transaction request flows must never cause the generation of a flow control transaction. For the 8/16 LP-LVDS and the 1x/4x LP-Serial physical layer specifications, the flowID of a transaction request flow is mapped to the “prio” bits as summarized in Table 1-3 of the 8/16 LP-LVDS specification and Table 5-1 of the 1x/4x LP-Serial specification. Determining the original transaction request flow for the offending packet requires the switch to do a reverse mapping.

It is recognized that mapping a particular response to a particular transmission request may be inaccurate because the end point that generated the response is permitted in the physical layer to promote the response to a priority higher than would normally be assigned. Deadlock avoidance rules permit this promotion. For this reason the choice of which flow to XOFF is preferably made using request packets, not response packets, as responses release system resources, which also may help alleviate system congestion.

Additionally, the crf (critical request flow) bit should also be used in conjunction with flowID to decide whether or not a particular transaction request flow should be targeted with a XOFF flow control transaction. A switch may select for shut off a packet with crf=0 over a packet with crf=1 if there are two different flows of otherwise equal importance. Correspondingly, an end point may choose to ignore a flow control XOFF request for a transaction request flow that it regards as critical.

The reverse mappings from the transaction request flow prio field to the CCP flowID field for the 8/16 LP-LVDS and 1x/4x LP-Serial physical layers are summarized in Table 2-1.

Table 2-1. Prio field to flowID Mapping

Transaction Request flow prio Field	Transaction Type	System Priority	CCP flowID
0b00	request	Lowest	A
0b00	response	Illegal	
0b01	request	Next	B

Table 2-1. Prio field to flowID Mapping

0b01	response	Lowest	A
0b10	request	Highest	C or higher
0b10	response	Lowest or Next	A or B
0b11	request	Illegal	
0b11	response	Lowest or Next or Highest	A, B, C or higher

2.4.4 Flow Control Transaction Ordering Rules

The ordering rules for flow control transactions within a system are analogous to those for maintenance transactions.

1. Ordering rules apply only between the source (the original issuing switch device or destination end point) of flow control transactions and the destination of flow control transactions.
2. There are no ordering requirements between flow control transactions and maintenance or non-maintenance request transactions.
3. A switch processing element must pass through flow control transactions between an input and output port pair in the order they are received.
4. An end point processing element must process flow control transactions from the same source (the destination of the packet that caused the flow control event) in the order they are received.

2.4.5 End Point Flow Control Rules

There are a number of rules related to flow control that are required of an end point that supports the logical layer flow control extensions.

1. An XOFF flow control transaction stops all transaction request flows of the specified priority and lower targeted to the specified destination and increments the XON/XOFF counter associated with the specified flowID.
2. A XON flow control transaction decrements the XON/XOFF counter associated with the specified flowID. If the resulting value is zero, the transaction request flows for that flowID and flowIDs of higher priority are restarted.
3. An end point must be able to identify an orphaned XOFF'd flow and restart it.
4. A destination end point issuing an XOFF Flow Control transaction must maintain the information necessary to restart the flow with an XON flow control transaction when congestion abates.
5. Upon detection of congestion within one of its ports, the destination end point shall send required CCP(s) as quickly as possible to reduce latency back to the source end point.

2.4.6 Switch Flow Control Rules

There are a number of rules related to flow control that are required of a switch that supports the logical layer flow control extensions.

1. Upon detection of congestion within a port, the switch shall send a CCP (XOFF) for each congested flow to their respective end points.
2. If a switch runs out of packet buffer space, it is permitted to drop CCPs.
3. A switch issuing an XOFF Flow Control transaction must maintain the information necessary to restart the flow with an XON flow control transaction when congestion abates.

Chapter 3 Packet Format Descriptions

3.1 Introduction

This chapter contains the definitions of the flow control packet format.

3.2 Logical Layer Packet Format

The type 7 FLOW CONTROL packet formats (Flow Control Class) are used by a RapidIO switch or end point processing element to stop (XOFF) and start (XON) the flow of traffic to it from a targeted RapidIO end point processing element. A single transaction request flow is targeted with a CCP. Type 7 packets do not have a data payload and do not generate response packets. The origin of a flow control packet shall set the SOC (Source of Congestion) bit to (SOC=0) if it is a switch or (SOC=1) if it is an end point. The SOC bit is informational only but may be useful for system software in identifying a failing end point.

Definitions and encodings of fields specific to type 7 packets are provided in Table 3-1.

Table 3-1. Specific Field Definitions and Encodings for Type 7 Packets

Type 7 Fields	Encoding	Definition
XON/XOFF	0b0	Stop issuing requests for the specified and lower priority transaction request flows
	0b1	Start issuing requests for the specified and higher priority transaction request flows
flowID	—	Highest priority affected transaction request flow 0b0000000 - transaction request flow A 0b0000001 - transaction request flow B 0b0000010 - transaction request flows C and higher Remaining encodings are reserved for the 8/16 LP-LVDS and the 1x/4x LP-Serial physical layers.
destinationID	—	Indicates which end point the CCP is destined for (sourceID of the packet which caused the generation of the CCP).
tgtdestinationID	—	Combined with the flowID field, indicates which transaction request flows need to be acted upon (destinationID field of the packet which caused the generation of the CCP).
SOC	0b0	Source Of Congestion is a Switch
	0b1	Source Of Congestion is an End Point
rsrv	—	Reserved

Figure 3-1 displays a CCP packet with all its fields. The field value 0b0111 in Figure 3-1 specifies that the packet format is of type 7. Small (tt=0b00) and Large (tt=0b01) Transport Formats are shown in the figure.

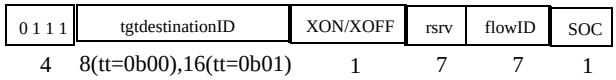

Figure 3-1. Type 7 Packet Bit Stream Logical Layer Format

3.3 Transport and Physical Layer Packet Format

Figure 3-2 shows a complete flow control packet, including all transport and 1x/4x LP-Serial physical layer fields except for delineation characters. The destinationID field of the CCP packet is the sourceID field from packets associated with the congestion event, and is the target of the flow control transaction. The tgtdestinationID field is the destinationID field from packets associated with the congestion event, and was the target of those packets. The tgtdestinationID field is used by the target of the flow control packet to identify the transaction request flow that needs to be acted upon. For all undefined flowID encodings, there is no action required and the tgtdestinationID is ignored. Field size differences for 8 bit address Small Transport Format (tt=0b00) vs. 16 bit address Large Transport Format (tt=0b01) are shown. Note: when tt=0b01 there will be a pad after the CRC.

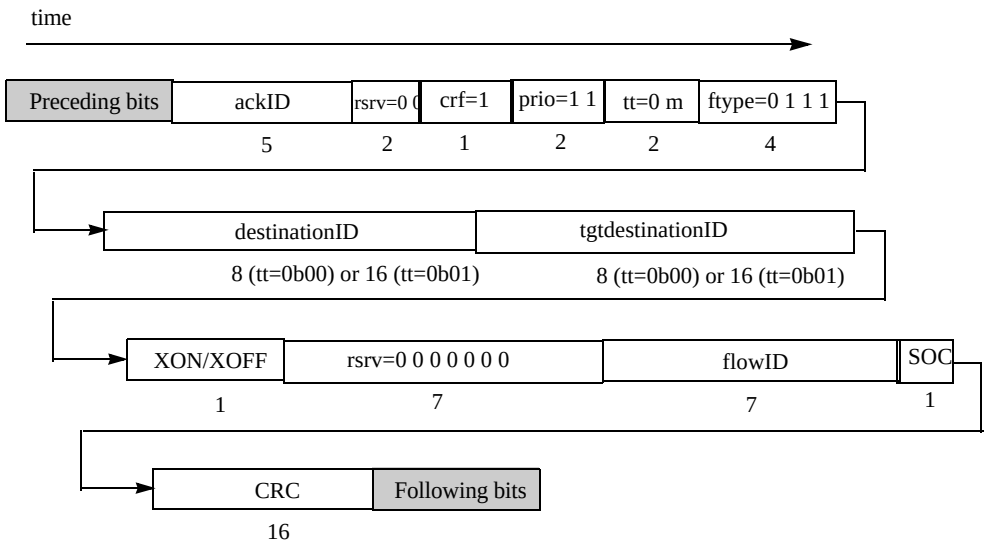

Figure 3-2. 1x/4x LP-Serial Flow Control Packet

Figure 3-3 shows the corresponding 8/16 LP-LVDS physical layer small transport packet.

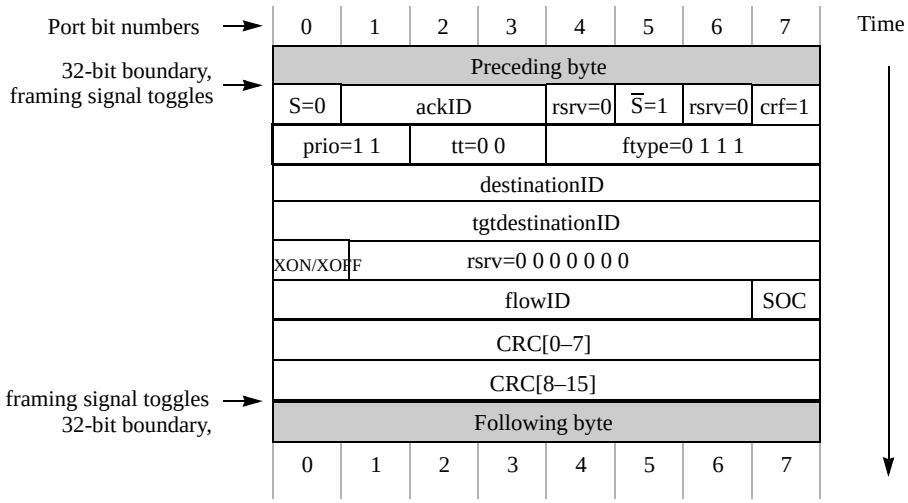

Figure 3-3. 8/16 LP-LVDS Small Transport Flow Control Packet

Blank page

Chapter 4 Logical Layer Flow Control Extensions Register Bits

4.1 Introduction

This section describes the Logical Layer Flow Control Extensions CAR and CSR bits that allow an external processing element to determine if a switch or end point device supports the flow control extensions defined in this specification, and to manage the transmission of flow control transactions for a switch processing element. This chapter only describes registers or register bits defined by this specification. Refer to the other RapidIO logical, transport, physical, and extension specifications of interest to determine a complete list of registers and bit definitions for a device. All registers are 32-bits and aligned to a 32-bit boundary.

4.2 Processing Elements Features CAR (Configuration Space Offset 0x10)

The Processing Elements Features CAR contains 31 processing elements features bits defined in various RapidIO specifications, as well as the Flow Control Support bit, defined here.

Table 4-1. Bit Settings for Processing Elements Features CAR

Bit	Name	Reset Value	Description
0-23	-		Reserved (defined elsewhere)
24	Flow Control Support		* Support for flow control extensions 0b0 - Does not support flow control extensions 0b1 - Supports flow control extensions
25-31	-		Reserved (defined elsewhere)

* Implementation dependant

4.3 Port *n* Control CSR (Block Offset 0x08)

The Port *n* Control CSR contains 31 bits specifying individual port controls defined in various RapidIO specifications, as well as the Flow Control Participant bit, defined here.

Table 4-2. Bit Settings for Port *n* Control CSR

Bit	Name	Reset Value	Description
0-9 (parallel) 0-12 (serial)	-		Reserved (defined elsewhere)
10 (parallel) 13 (serial)	Flow Control Participant	0b0	Enable flow control transactions 0b0 - Do not route or issue flow control transactions to this port 0b1 - Route or issue flow control transactions to this port
11-31 (parallel) 14-31 (serial	-		Reserved (defined elsewhere)

Annex A Flow Control Examples (Informative)

A.1 Congestion Detection and Remediation

The method used to detect congestion is implementation specific and is heavily dependent upon the internal packet buffering structure and capacity of the particular switch device. In the example output port buffered switch from “Section 1.3, Problem Illustration” on page 10, congestion occurs when some output buffer watermark is exceeded. As long as the watermark is exceeded the output port is said to be in a congested state. The watermark can have different levels when entering the congested state and leaving the congested state.

Fabric elements should monitor their internal packet buffer levels, comparing them on a packet by packet basis to pre-established, locally-defined watermark levels. These levels likely would be configurable depending upon the local element's position within the fabric relative to source endpoints and its particular architecture. On the high watermark side, a level should be selected which is low enough that the remaining buffer space is adequate to provide ample storage for packets in-flight, given a worse-case latency for XOFF CCPs to travel back to the source endpoint and shut off the flow in the endpoint. On the low watermark side (if a watermark is used for XON), a yet-lower level should be selected which meets the following criteria;

a) Provides sufficient hysteresis. When considered in context with the high watermark, it should not be so close as to provide a high flow of XON/XOFF CCP traffic back to the source endpoint.

b) Is set high enough that the switch output buffer does not run dry (underflow) in the typical live-flow case (one or more packets are present in the source endpoint output buffer waiting to be sent when the flow is restarted), given the latency of XON CCP travel back to the source endpoint and restoration of the shut-off flow in the endpoint.

The following two examples are provided to show possible methods for detecting and reacting to congestion:

1. Histogram analysis:
   — The switch keeps track of packet quantities for the different transaction request flows for which packets are stored in its output buffer.
   — The switch sorts the transaction request flows according to the number of packets.
   — The switch selects the 1 to 5 transaction request flows with the most

packets stored in the buffers.

— The switch sends an XOFF flow control request to those transaction request flow sources when the watermark threshold is exceeded, as long as flow control transaction routing is enabled on that switch port. Handling of system critical flows intending to bypass the flow control operation is outside the scope of this document.

— The CCP-targeted sources stop transmitting packets for the indicated transaction request flow and all lower priority transaction request flows.

— The switch sends a flow control XON request to those transaction request flow sources when the watermark drops below the threshold.

— The CCP-targeted sources begin to transmit packets for the indicated transaction request flow and all higher priority transaction request flows.

2. Simple threshold:

— The switch sends an XOFF flow control to the source of every new transaction flow it receives as long as the watermark is exceeded, provided flow control transaction routing is enabled on that switch port. Handling of system critical flows intending to bypass the flow control operation is outside the scope of this document.

— The CCP-targeted sources stop transmitting packets for the indicated transaction request flow and all lower priority transaction request flows.

— The switch sends a flow control XON request to those transaction request flow sources when the watermark drops below the threshold.

— The CCP-targeted sources begin to transmit packets for the indicated transaction request flow and all higher priority transaction request flows.

Note that the first method is reasonably fair in that it targets the source of the data flows that are consuming most of the link bandwidth, and that the second method is unfair in that it indiscriminately targets any source unfortunate enough to have a packet be transmitted while the link is congested.

A.2 Orphaned XOFF Mechanism Description

This timer may take the form of a low precision counter in the end point which monitors the oldest XOFF’d flow at any given time. When a flow first becomes the oldest flow (reaches top of an XOFF’d flow FIFO list within the end point) the timer is reset to its programmed value and begins to count down with time. If it is allowed to elapse without a change to the oldest XOFF’d flow, that flow will be presumed to be orphaned due to lost XON CCP and be restarted as if an XON CCP had been received, with the orphaned flow entry removed from the top of the list and the counter reset to count down for the next oldest XOFF’d flow. The length of the count should be long enough to insure that significant degradation of the flow control function does not occur, on the order of several times the width of the fabric expressed in terms of packet transit time, yet not so large that it would fail to elapse

between uncorrelated congestion events. The length of this count shall be programmable through an implementation-dependent register in the end point. The orphaned XOFF mechanism is intended solely as a last-resort mechanism for restarting orphaned flows. It will not be adequate for the purpose of implicit controlled flow reinstatement owing to inherent fairness issues as well as burstyness due to uncontrolled simultaneous multi-flow restart.

Blank page

Glossary of Terms and Abbreviations

The glossary contains an alphabetical list of terms, phrases, and abbreviations used in this book.

C

Congestion. A condition found in output ports of switch and bridge elements characterized by excessive packet buildup in the buffer, when packet entry rate into the buffer exceeds packet exit rate for a long enough period of time.

CCP (Congestion Control Packet). A packet sent from the point of congestion in the fabric back to the source endpoint of particular flows instructing the source to either turn on or off the flow.

Controlled Flow List. A memory structure associated with controlling elements which holds a list of currently controlled flows, used by the element to turn back on controlled flows.

crf. Critical Request Flow. For packets or packets of a given priority, this bit further defines which packet or notice should be moved first from the input queue to the output queue (see *RapidIO Part 4: 8/16 LP-LVDS Physical Layer Specification*, Section 1.2.2 and *RapidIO Part 6: 1x/4x LP-Serial Physical Layer Specification*, Section 5.3.3).

F

flowID. Transaction request flow indicator (see *RapidIO Part 1: Input/Output Logical Specification*, Section 1.2.1).

L

Long Term Congestion. A severe congestion event in which a system does not have the raw capacity to handle the demands placed upon it in actual use.

M

Medium Term Congestion. A congestion event in which a frequent series of short term congestion events occur over a long period of time such as seconds or minutes, handled in RapidIO systems by reconfiguration of the fabric by system-level software.

O

Orphaned XOFF Mechanism. A mechanism in an end point which is used to restart the oldest controlled flow within the end point after a certain period of time has elapsed without the flow being XON'd.

P

Performance Collapse. Non-linear behavior found in non- congestion controlled fabrics, whereby reduced aggregate throughput is exhibited with increased load.

S

Saturation Tree. A pattern of congestion identified within the fabric which grows backward from the root buffer overflow towards the sources of all transaction request flows passing through this buffer.

Short Term Congestion. A congestion event lasting up into the dozens or hundreds of microseconds, handled in RapidIO by Logical Layer Flow Control.

T

Topology. The structure represented by the physical interconnections of a switch fabric.

Transaction Request Flow. A series of packets that have a common source identifier and a common destination identifier at some given priority.

U

Ultra Short Term Congestion. A congestion event lasting from dozens to hundreds of nanoseconds, handled in RapidIO by Link Level Flow Control.

Underflow. A condition within output buffers of switches in which the buffer runs dry.

W

Watermark. A predetermined buffer occupancy level indicating either congestion (high watermark) or abatement of congestion (low watermark).

X

XOFF (Transmit Off). A congestion control packet sent from the point of congestion back to the source of a particular flow, telling the source endpoint to shut off the flow.

XON (Transmit On). A congestion control packet sent from the point of congestion back to the source of a particular flow, telling the source endpoint to restart a controlled flow.

Blank page

Blank page