Error Recovery and Fencing Mechanisms for PCI Express Gen 2 Devices

White Paper

Version 1.0

January 8, 2008

Website: www.plxtech.com
Technical Support: www.plxtech.com/support

Copyright © 2008 by PLX Technology, Inc. All Rights Reserved – Version 1.0
January, 2008

Introduction

Leveraging experience gained with three generations of PCI Express devices, significant enhancements have been made to augment data integrity protection mechanisms in the PLX Gen 2 family of devices. The new devices (PEX 86xx) take into account the lack of ECRC support in the chipsets and implements mechanisms to mitigate error propagation close to the point of origin. The Error Recovery and Fencing mechanisms have been enhanced to block fatal errors from propagating deeper and causing system issues. Further protection is provided to prevent packets with errors from crossing the NT domain. Figure 1 below provides the top level block diagram of a switch and describes the protection mechanisms. The description below assumes the reader is familiar with the PCI Express protocol and understands the error protection mechanisms in the specification. The PLX devices have implemented additional features to address certain weaknesses in the specification while enabling various system usage models, this document addresses those additions.

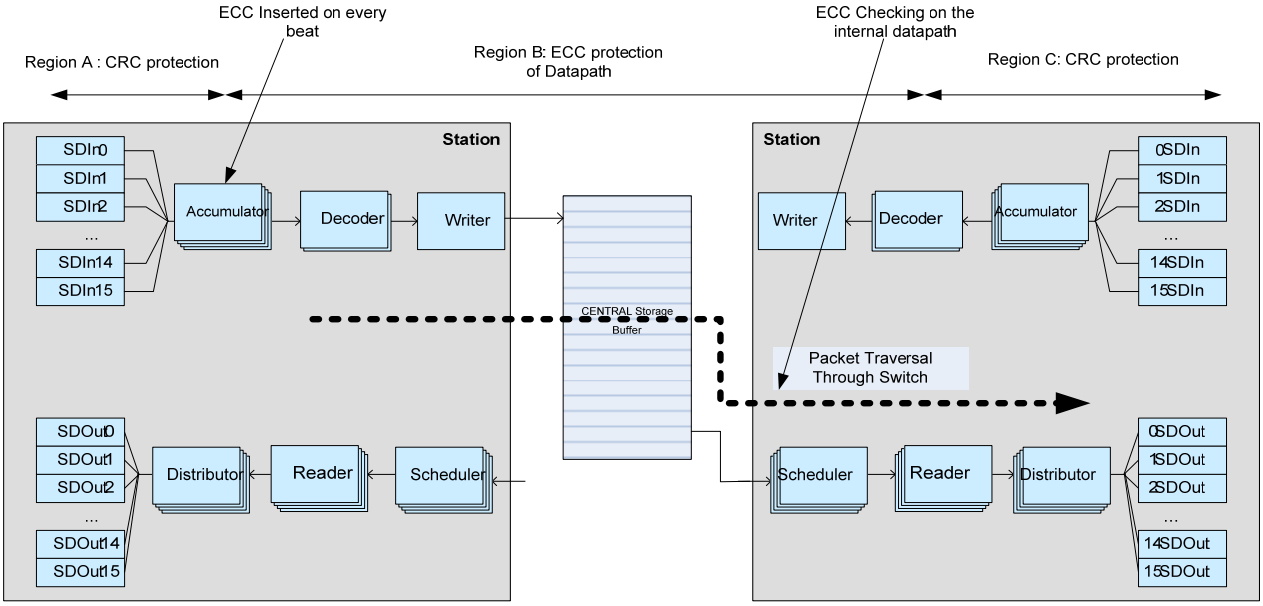

Figure 1. Packet Traversal through Switch

Data Integrity on the data path

Figure 1 illustrates the typical flow of packets thought the switch. As the packets enter in on the SerDes, the physical layer executes several checks:

a) 8b/10b encoding errors
b) Disparity errors
c) Framing errors
d) De-skew Errors

If the packet passes through the PHY layer checks successfully, it gets stored in the accumulator pipe stage and the CRC and Sequence number is checked. The device then inserts an ECC syndrome into every bit of the packet once the CRC is stripped. This protects the packet as it is processed internally in the design. All 1-bit errors are automatically corrected and the 2-bit errors are flagged as fatal errors. The packet is subsequently written into the central RAM by the “writer”. The “reader” reads the packets out of the RAM and does the ECC checking and replaces the ECC on the individual bits with the CRC on the whole packet as it is sent for transmission into the “distributor”.

Copyrighted 2008 by PLX Technology, Inc. All Rights Reserved – Version 1.0

ECRC protection

The switches implement ECRC checking both on the ingress (“decoder”) and the egress (“reader”). This provides traceability of the source of ECRC errors and allows for easier diagnosis of failures

ECC checking on Memories

All memories are ECC protected. The ECC scheme allows for 1-bit detection and correction and 2-bit detection and assertion of fatal errors.

Error Recovery and Fencing Modes

The following modes of error detection and recovery are available.

Mode 1 (Default)

1. When the device receives a packet with a fatal error (Malformed, DataLink Layer Protocol error) from an external device, the device logs the header on the corresponding port, sends out a fatal error message to the host, and asserts the FATAL_ERR# pin.
2. When the device detects an internal fatal error (ECC failure, Credit overflow, Receiver Overflow, Surprise link down) it sends out a fatal Interrupt message to the host and the device specific FATAL_ERR# pin is asserted. In certain situations the delivery of the Interrupt is not guaranteed but the pin is always asserted on a fatal event.

Mode 2 (Generate Internal Reset)

On the detection of a fatal error (internal or external), an internal chip level reset is asserted (equivalent to an in-band reset from the upstream port). No error messages are generated and no attempt is made to block packets in transit.

Mode 3 (Block All Packet Transmission)

On the detection of a fatal error (internal or external), the corresponding port logs the fatal error in the corresponding error status register and asserts the FATAL_ERR# pin. This fatal error detection blocks all the ports from sending TLPs out. No error messages are generated. If a packet is already in transmission, an EDB is inserted to cancel the packet.

Mode 4 (Block All Packet Transmission & Create Surprise Down)

In addition to all the actions in Mode 3 above, the device forces the upstream link to go down thus causing a “Surprise Down” event on the link so that the host is notified.

Port0 configuration space register offset 0x1DC, Bits [13:12] allow for the selection of the modes described above:

00 – Mode 1 (default value)
01 – Mode 2
10 – Mode 3
11 – Mode 4

Copyrighted 2008 by PLX Technology, Inc. All Rights Reserved – Version 1.0

EP & ECRC error handing

Packets with EP or ECRC while not considered fatal could cause system failures if not "fenced" at the source of the faulty device. The following modes describe the options to handle devices generating EP/ECRC packets.

EP/ECRC Mode 1(Default):

When the device receives a TLP with an EP or ECRC error, the packet is forwarded with appropriate error logging.

EP/ECRC Mode 2 (Drop EP/ECRC packet):

When the device receives a TLP with an EP or ECRC error, the EP/ECRC packet is dropped with appropriate error logging and the subsequent packets are forwarded.

EP/ECRC Mode 3 (Drop EP/ECRC & Block Subsequent packets):

When the device receives a TLP with an EP or ECRC error, the EP/ECRC packet is dropped with appropriate error logging and all subsequent packets are also blocked from the source port (thus fencing the affected device). If the source port happens to be the upstream port (or NT-Link upstream port) Type-0 Cfg accesses are still allowed. Once blocked, the block can be removed by writing to the appropriate register.

In Port0 configuration space register offset 0x664, Bit 4 controls the EP fencing and Bit 8 controls the ECRC fencing:

0 – Fencing is disabled
1 – Fencing is enabled

In Port0 configuration space register offset 0x664 Bit 12 selects the EP_ECRC mode of operation:

0 – Mode 2
1 – Mode 3

Error Protection with Non-Transparency:

On systems with an NT connection between two switches, the error protection mechanisms invoked depend on two scenarios:

ECRC enabled: The packet maintains its original ECRC digest until it reaches the egress "reader". When the "reader" extracts the packet from the payload RAM it checks the ECRC and if correct it translates the packet and embeds a new ECRC into the packet. If the original ECRC digest is corrupted, error flags are set (ECRC error etc.,) and the ECRC of the translated packet is corrupted by inverting the digest.

ECRC not enabled: The standard 1-bit and 2-bit ECC check mechanisms are used to check the data integrity of the packet and for errors that get through the ECC checks (due to RAM manufacturing error, etc.) "error fencing" can be enabled on the adjacent device link to prevent subsequent packets from corrupting the rest of the system.

Copyrighted 2008 by PLX Technology, Inc. All Rights Reserved – Version 1.0