



FC-BB-6 LINK LENGTH DETERMINATION

Brocade: Howard L. Johnson, Dave Peterson

IBM: Lou Ricci, Roger Hathorn

EMC: Erik Smith

T11/11-498v0



Table of Contents

Link Length Determination

- State of the Art
 - Desire for automated determination of buffer resources
- Methods
 - Proactive
 - Reactive
- Behaviors
 - Device responses
 - Service implications
- Concerns
 - Can we improve the existing state of the art?



Link Length

Administration

- Fibre Channel
 - Optimal configuration is multidimensional
 - Link speed, average payload size, and link length
 - Administrative configuration is imperfect
 - Errors result in under utilized links or wasted buffer resources
- Ethernet
 - Optimal configuration is multidimensional
 - Link speed and link length
 - Pause buffering capacity, I/O Consolidation, Convergence
 - Administrative configuration is imperfect
 - Errors may result in wasted buffer resources or potentially lost frames on a congested link
- User Feedback
 - Clamoring for this “little stuff” to be automatic





Link Length

Automated Buffer Resource Determination



Mechanisms

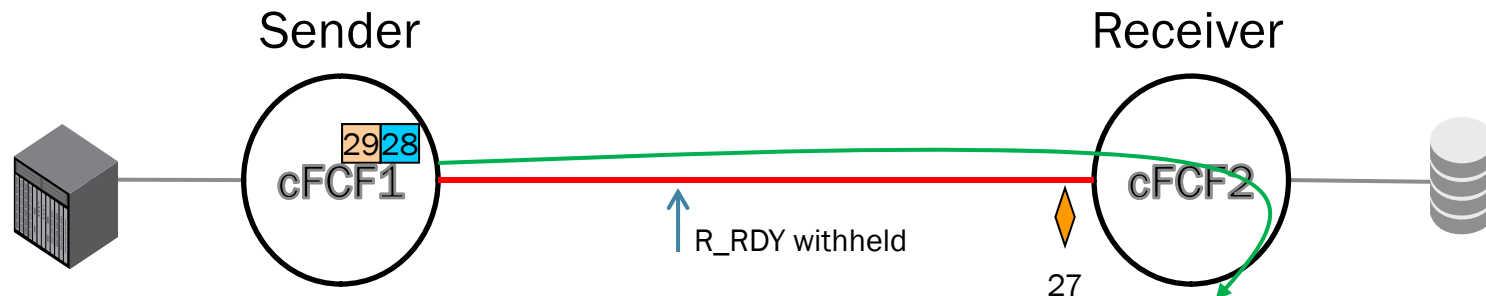
Well known, effective processes

- Proactive method (e.g. credit based)
 - Exchange of buffering capacity at discovery
 - Transmitter responsible for managing flow
- Reactive method (e.g. pause based)
 - Receiver buffer is configured to accommodate bandwidth delay
 - Receiver responsible for managing flow
- Concern is around a particular error scenario
 - Link longer than configured buffering resources
 - Unable to forward frames due to congestion
 - Errors occurring intermittently – just a little bit mis-configured
 - Receiver overrun is a possible behavior of a properly functioning but incorrectly configured device



Proactive Behavior during Error Scenario

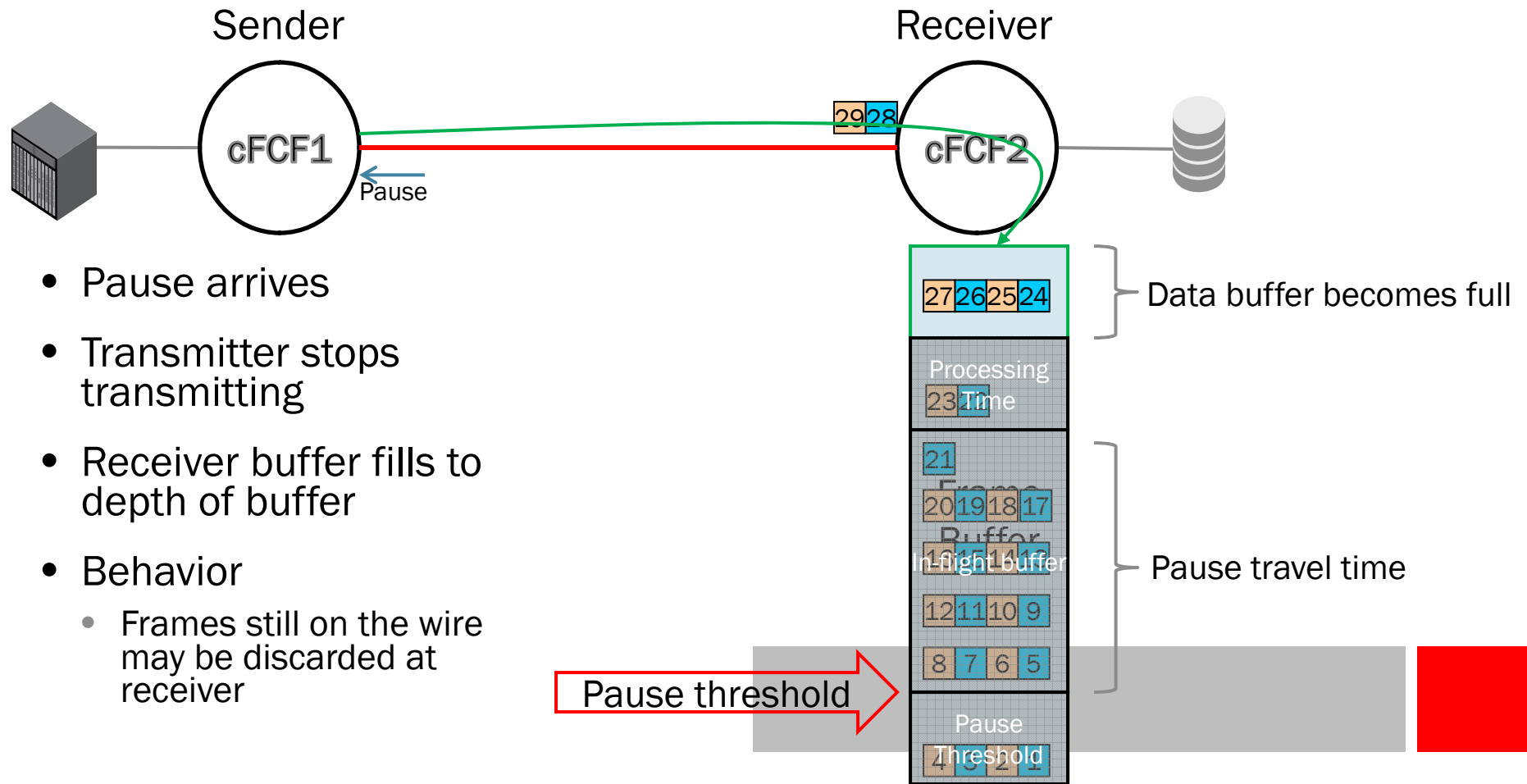
Link is longer than configured credit capacity



- Receiver withholds R_RDY
- Transmitter stops transmitting
- Receiver buffer fills to level of outstanding R_RDY's
- Behavior
 - Link may become under utilized
 - No frames are discarded
 - Consistent complaint among traditional FC users

Reactive Behavior during Error Scenario

Link is longer than configured receive buffer depth



- Pause arrives
- Transmitter stops transmitting
- Receiver buffer fills to depth of buffer
- Behavior
 - Frames still on the wire may be discarded at receiver

Error Scenario Summary

The link is longer than the configured buffer resources support

- Proactive flow control (credits)
 - Transmitter throttles flow
 - Link capacity may be under-utilized
 - Device is not directly aware of performance penalty
 - Must be administratively corrected
- Reactive flow control (pause)
 - Receiver throttles flow
 - Overflow frames may be discarded
 - Device level recovery employed
 - Must be administratively corrected



Fibre Channel: The Once and Future King

A chance to improve the state of the art

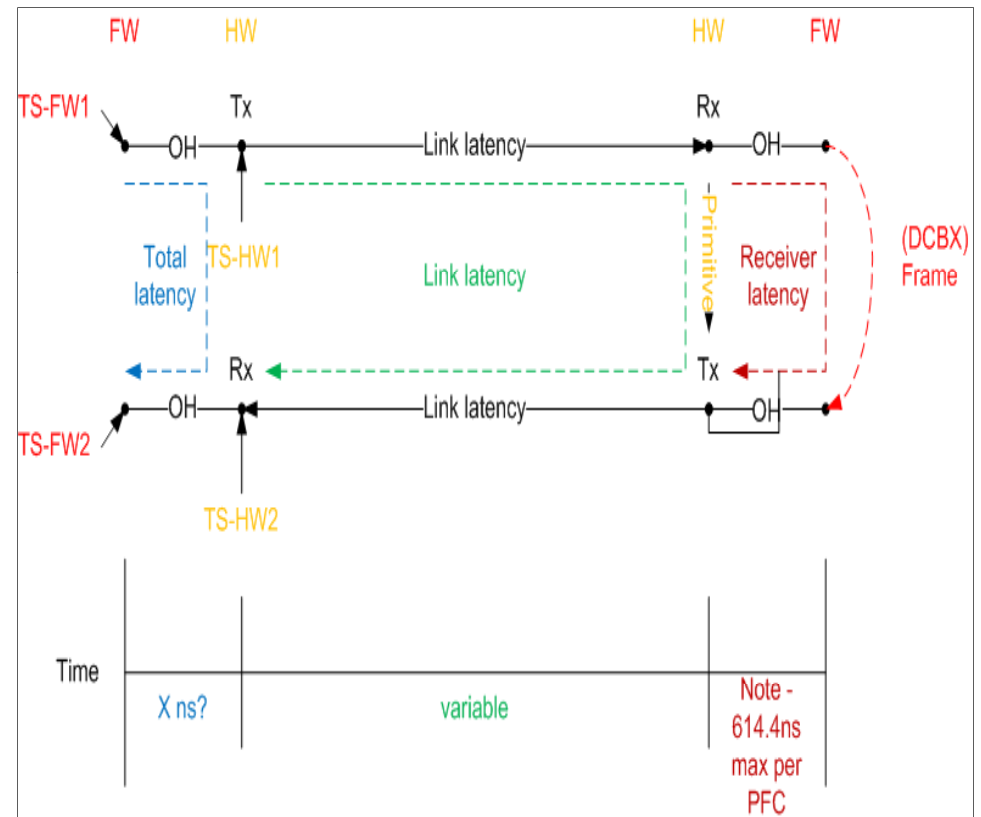
- Concern is with infrequent error scenarios
 - Difficult to discover and correct
 - Dependent on device level recovery
 - In some environments, frame discards are elevated to service calls
 - Under-utilization is even harder to figure out
- FCoE adoption
 - Differences in “corner case” behavior could hinder acceptance of FCoE
 - The invocation of device level recovery could create a negative perception of FCoE
- Opportunity
 - Improve the state of the art by adding link length determination



Proposal

Link Length Determination Objectives

- Use low level mechanism
 - Point to point measurement
 - Consistency for realistic time
- Determine link length
 - The specific method is TBD
- Use results to establish buffer memory resources
 - BB Credits for FC
 - Receiver memory for FCoE
 - Pause/resume thresholds
- Provide notification for unsupportable link lengths



Summary

Improvements for handling long links

- Concern
 - Handling links longer than supported by configured buffer capacity
- An Opportunity to improve the state of the art
 - Provide link length/latency determination and validation
 - Executed at link initialization
 - Exchange of buffer depths at or prior to Fabric Login
 - Provide feedback for incorrectly configured resources
 - Technique
 - Use low level primitive for accurate point to point measurements
- Consideration
 - What link lengths should be supported by FCoE?
- Standards Process
 - Can FC-BB request such changes?
 - Can FC-BB specify a particular solution?





This slide intentionally left blank

Thank You

