Offload Traffic Failover

Fault Tolerance across Ports and Adapters

Executive Summary

Chelsio's T4 and T5-based adapters offer a complete suite of high reliability and availability features, including port-to-port and adapter-to-adapter failover. This paper discusses both and focuses on the patented Multi-Adapter Failover (MAFO) feature, which ensures all offloaded protocols continue operating in the face of port failure. MAFO allows aggregating network interfaces across multiple adapters into a single logical bonded interface, providing effective fault tolerance.

Introduction

Chelsio is the leader in network protocol offload technology, with specific focus on high reliability, availability and data integrity. Failover is a critical part of Chelsio's adapter capabilities, and benefits from long experience shipping fully offloaded storage and compute networking protocols.

All T4 and T5-based adapters support virtual-port abstractions that provide seamless switchover from port-to-port in the event of a link down. The failover capabilities allow immediate and effortless switching of active offloaded connection traffic from the failed port to other ports within the same virtual port group.

Failover between adapters also benefits from the extensive feature set of the T4 and T5 ASICs. This paper describes how Chelsio's Multi Adapter Failover works, using a bonding driver example in Linux. This elegant and simple capability is unique and covered by at least one issued patent [1].

Failover mechanism

The first step in creating a reliable multi-adapter link is to bind together physical ports of two different adapters and configure them in Active-Backup mode. In Figure 1 below, ports A0 and B0 are ports on adapter A and adapter B respectively that are bonded together in Active/Backup mode, with A0 as active and B0 as backup port.

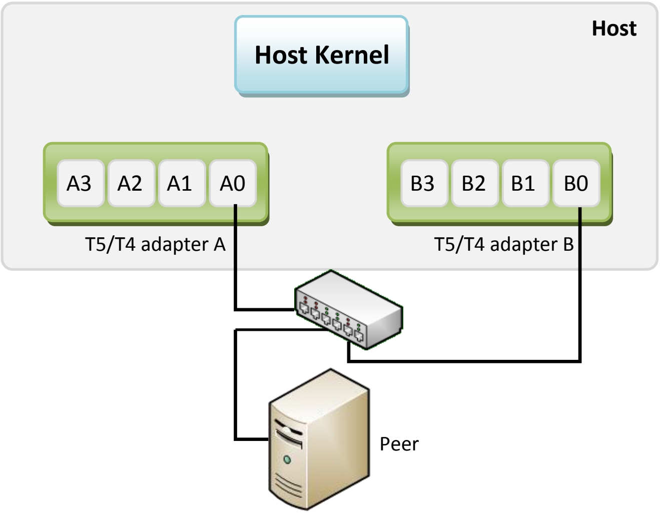

Figure 1 – Active/Backup Bonding Topology

Consider offloaded (e.g. TCP/IP) traffic flowing between port A0 and a peer. In case of a link failure on port A0, the Multi-Adapter Failover implementation will route the offloaded traffic from port A0 through the host to port B0, using the on-board embedded switch in adapter A.

This switching is done in such a manner that is fully transparent to the peer, which remains unaware of the traffic flowing between the adapters. Traffic continues to be tunneled through adapter B to adapter A where it is offloaded. While this can be continued for the rest of the connection, for improved performance, the failover operation suspends passing new payload to the connection on the failed adapter, while it sets up a filter on the backup adapter with the connection's 4-tuple in preparation for handover. The next step is to wait for all outstanding payload to drain.

Drain Tx

Drain all the un-acked Tx payload from Port A0 by looping generated frames back to host kernel and sending them out via Port B0. Similarly, receive all the TCP ACKs coming from the peer on Port B0, and loop them back through adapter A to Port A0.

Drain Rx

Refrain from opening the TCP flow control window as the pending Rx payload is delivered to the host.

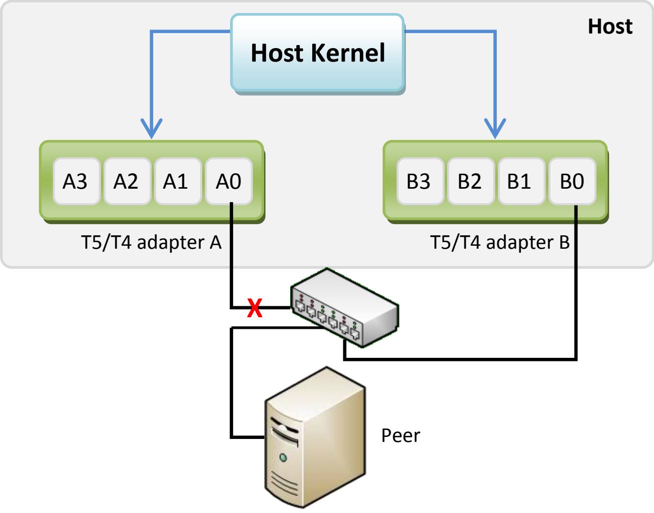

Figure 1 – Multi Adapter Failover Mechanism

Once both Tx and Rx have been drained on adapter A, the host is left with a complete connection state that can be updated into adapter B, before offload processing is enabled and traffic starts flowing from Port B0.

Meanwhile, the link on Port A0 could be back, in which case Port A0 would continue as a backup.

Summary

This paper describes Chelsio's unique mechanisms for offloaded traffic failover between ports and adapters. A reference implementation in Linux is available from Chelsio.

References

[1] United States Patent 8346919, Failover and migration for full-offload network interface devices