

Detecting Duplex Mismatch on Ethernet

Stanislav Shalunov <shalunov@internet2.edu>

Richard Carlson <rcarlson@internet2.edu>

PAM 2005, Boston, 2005-03-31

Talk overview

- What is Ethernet duplex mismatch?
- How do we know it ever occurs?
- Half- and full-duplex Ethernet
- Auto-configuration
- Duplex mismatch
- Model of duplex mismatch
- Manifestation with Poisson streams (prediction)
- Manifestation with Poisson streams (experiment)
- What happens with TCP?
- Network Diagnostic Tool (NDT)
- TCP with duplex mismatch on LAN (experiment)
- Ideas about detection in NDT with Web100 variables

What is Ethernet duplex mismatch?

- On Ethernet, each device has a *duplex* (full or half) setting
- The setting exists for up to 1-Gb/s Ethernet, but (finally!) not for 10 Gb/s
- When two devices on the ends of a physical link disagree about duplex, a duplex mismatch occurs

How do we know it ever occurs?

- Described in a few industry white papers
- Easily recreated in the lab
- Regularly encountered by engineers troubleshooting performance problems
- Goes away when both sides are set to use the same duplex setting
- The cure also causes the disease (more on this later)

Half-duplex Ethernet

- Can work on broadcast *media* (not used anymore)
- Historic default (relic?)
- Mandatory to support
- When two devices try to transmit at the same time, a *collision event* is said to occur
- Carrier Sense Multiple Access with Collision Detection (CSMA/CD) protocol resolves collisions
- Late collision is such a collision that occurs after the first 64 octets (512 octets for 1 Gb/s) are sent – behavior poorly defined

Full-duplex Ethernet

- Made possible by point-to-point media (e.g., 10BASE-T for 10 Mb/s)
- Requires switching
- Each side sends with no blocking on anything
- Receiving is independent of sending
- Performs better than half-duplex
- The preferred modern mode

Auto-configuration

- Before starting to operate, devices attempt to auto-configure with out-of-band signals
- Normal Link Pulses (NLPs) are sent by old hardware (half-duplex only)
- Fast Link Pulse (FLP) burst encodes Link Code Word (LCW) that conveys supported mode
- Each side computes commonly supported modes, and chooses the best of common modes
- In addition, auto-sense protocol not using FLP bursts can be used to detect speed (and media type), but only supports half-duplex

Duplex mismatch

- Protocol is complex
- Operation mode is never communicated explicitly
- *Duplex mismatch* can happen for a variety of reasons
 1. One card is hard-coded to use full duplex
 2. Two cards are hard-coded to use different duplex modes
 3. Auto-negotiation fails (problem can be sporadic)

Model of duplex mismatch

full → **half** Can cause loss, but no extra delay. Loss occurs iff the frame starts arriving at the half side while it's sending.

half → **full** Can cause loss and extra delay. Loss can occur if

1. Collision occurs too late for resending
2. Buffer overruns (the full side keeps sending); sufficient condition: sum of load in both directions exceeds link capacity
3. Excessive collisions

Manifestation with Poisson streams (prediction)

- Poisson streams are easy to analyze (good for validation)
- Analysis is further simplified if both rates are small
- $p_F = \frac{c_H}{c}$
- $p_H = \frac{c_H}{c} \max\left(0, 1 - \frac{c\xi}{m}\right)$
- Experiments conducted (using “thrulay -u”) to validate

What happens with TCP?

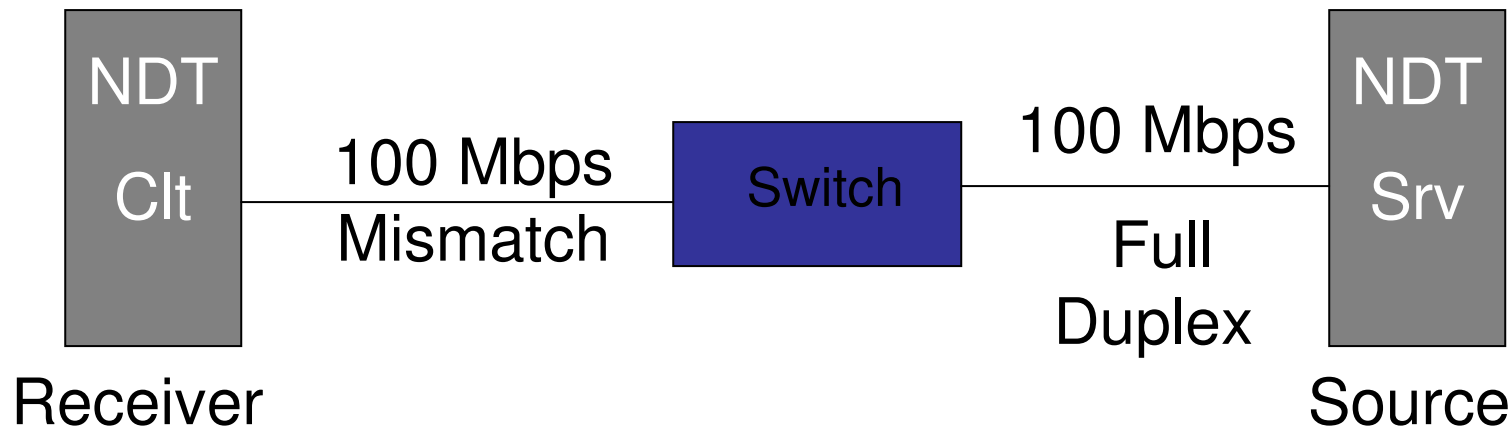
- The case of TCP is considerably more complex
- Throughput is asymmetric
- Depends on delay
- For small delay, one direction isn't too horrible (but still worse than even half duplex); the other direction is severely crippled
- For larger delays, both directions suffer to a greater extent

NDT Background

- Web100-enhanced server handles testing and diagnostic services
- Java-based and command line clients allows testing from any client (local or remote)
- Performance and configuration faults reported back to client
- Drill-down functions provide more details and error-reporting capabilities

Test environment

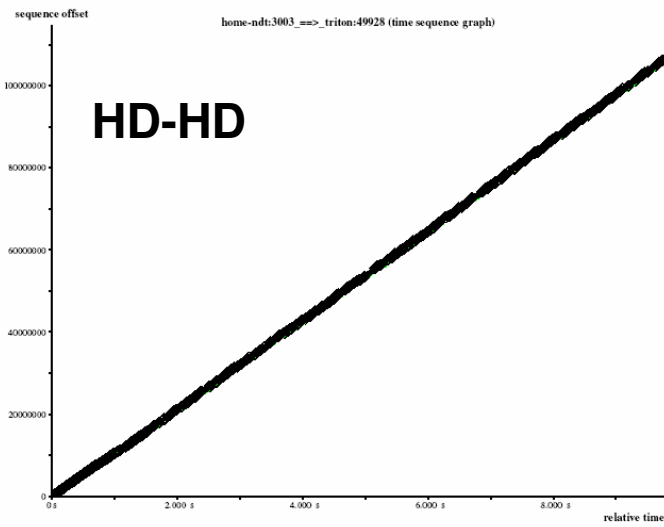
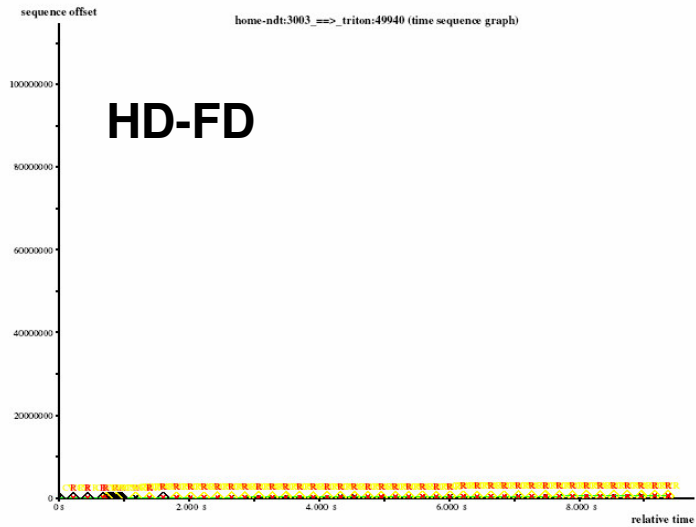
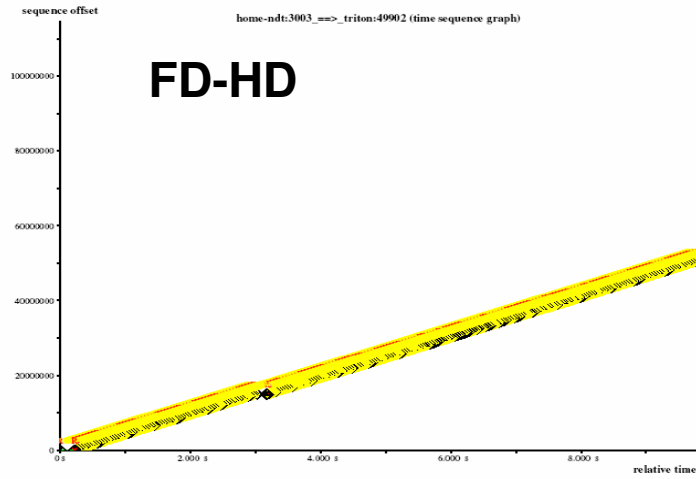
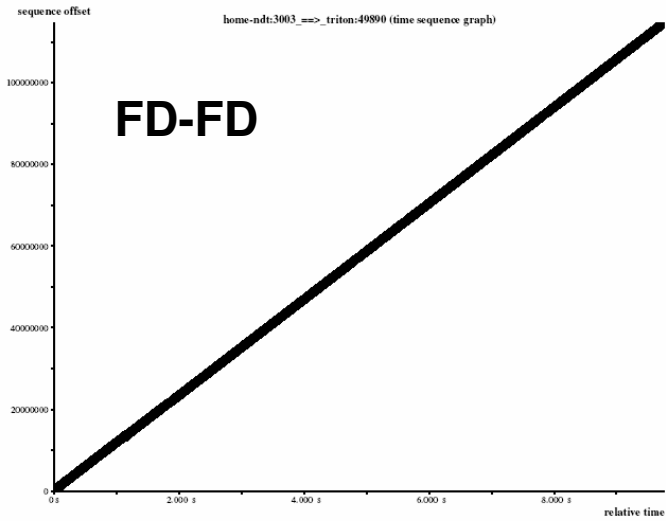
- Receiver is put in various states
 - Switch = full & Host = full or half
 - Switch = half & Host = full or half



TCP Operation on LAN

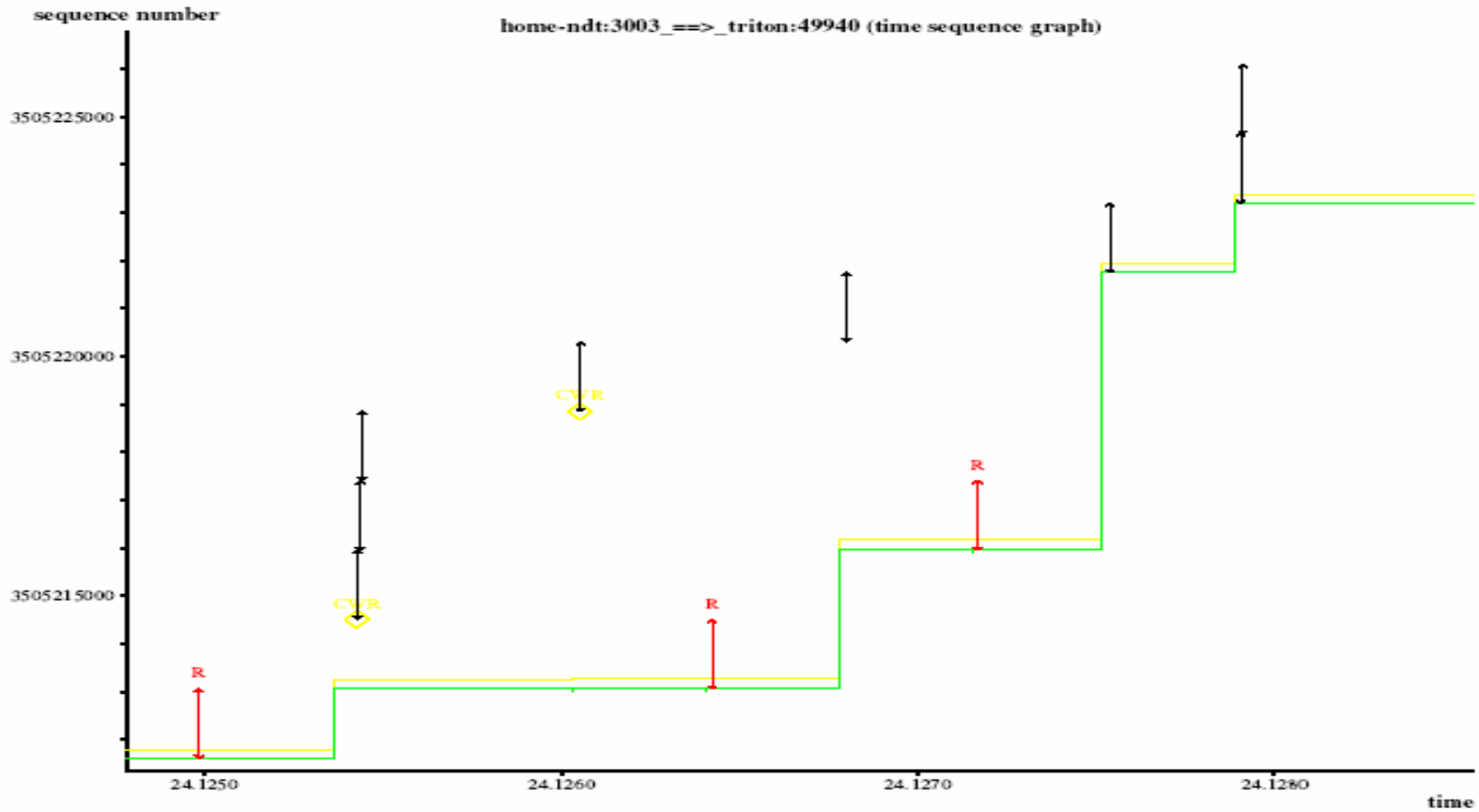
- Observed behavior depends on direction of TCP flow and direction of mismatch
 - Data and ACK packets delivered
 - Data packets lost and ACKs delayed
 - ACKs packets lost and Data delayed
- Losing ACKs has bigger effect than losing Data packets
- Web100 details are only available when ¹⁵ NDT server is source and client is sink

Four Cases of Duplex Setting



Duplex Mismatch

Switch is Half & Host is Full



Tentative Mismatch Detection

- Half to Full mismatch detection
 - Large number of timeouts causes long idle time ($RTO \times \text{timeout value}$)
 - Connection spends majority of the time in CwndLimited state
 - Asymmetric throughput
 - opposite direction is greater

Tentative Mismatch Detection

- Full to Half mismatch detection
 - Large percentage of duplicate ACKs
 - Connection spends majority of the time in CwndLimited state
 - Asymmetric throughput
 - opposite direction is less

- Supplemental slides follow

TCP Mismatch Detection

- Observed behavior depends on direction of TCP flow and direction of mismatch
- Both behaviors should be detected to confirm mismatch condition exists
- Web100 details are only available when NDT server is source and client is sink

Duplex Mismatch

Switch is Full & Host is Half

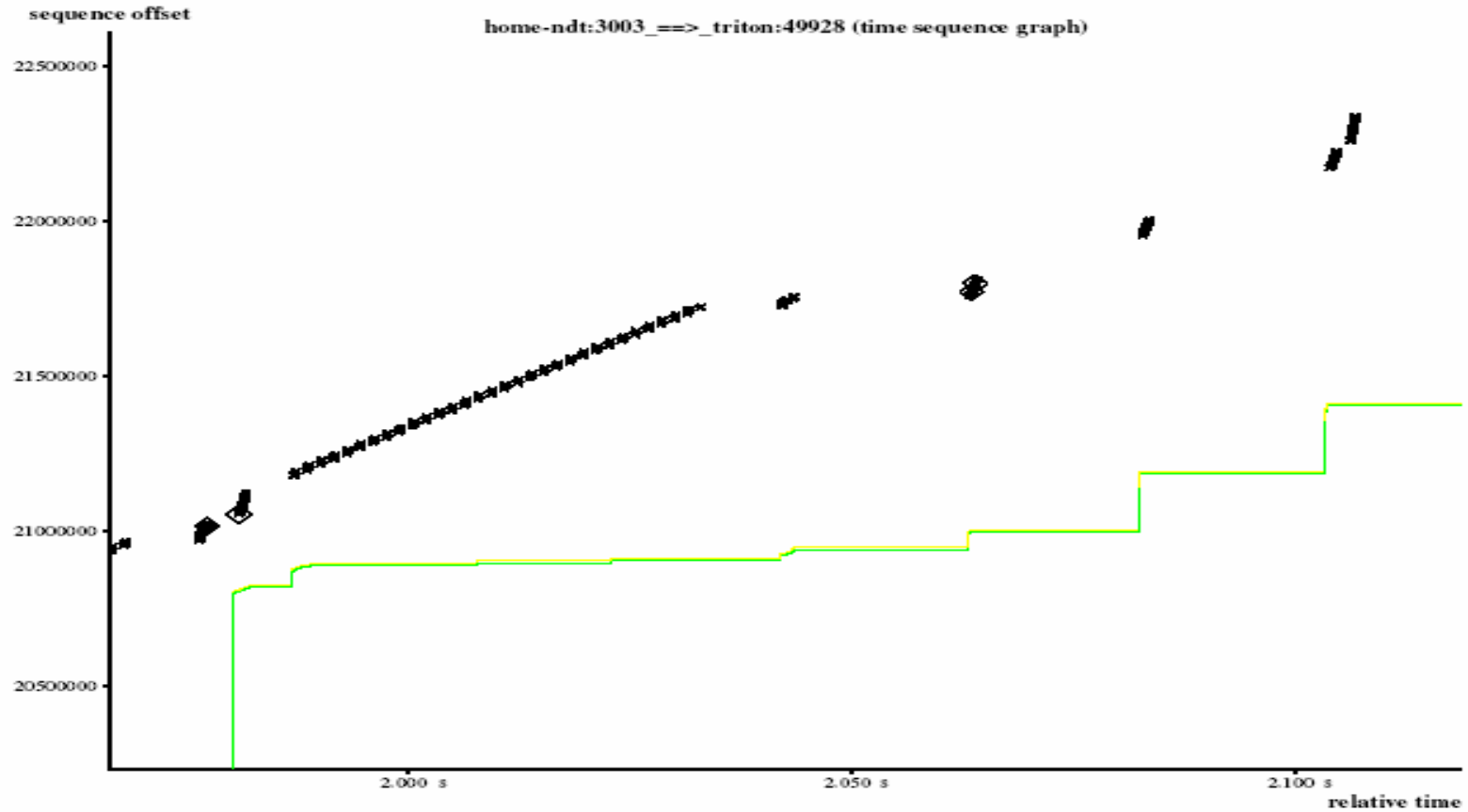
- Connection is limited by CWND collapse
 - Data packet is lost when collision occurs
 - Single data packet lost in 5 RTT periods
 - Large number of duplicate ACKs sent by client
 - Large jump in sequence space as retransmitted packet fills in hole

Duplex Mismatch

Switch is Half & Host is Full

- Connection is RTO limited
 - ACK packets are lost when collision occurs
 - Sender reaches window limit and pauses for ACK
 - Loss of ACKs causes sender to pause for RTO period
 - Duplicate ACKs cause additional collisions
 - Extraneous retransmission generates ACK which starts another burst of packets

Normal Half Duplex Operation



Duplex Mismatch

Switch is Half & Host is Full

