EtherFuse: An Ethernet Watchdog

Khaled Elmeleegy
Rice University

With: Alan L. Cox and T. S. Eugene Ng

1

Ethernet Dependability

- Ethernet is used for mission critical applications
  - Factory automation
  - Hospitals
  - Stock market
  - ...
- Redundancy for fault tolerance

2

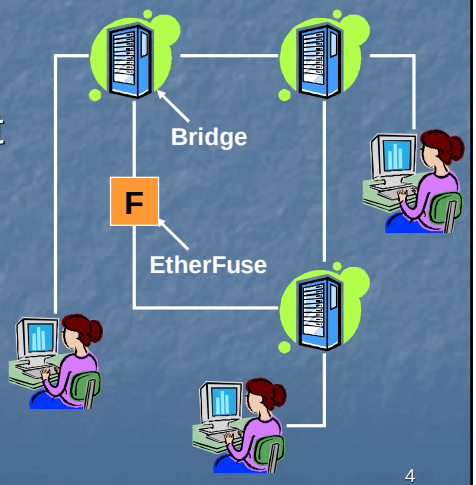
Ethernet Failures

- Ethernet outages happen in real life
  - Difficult to troubleshoot
    - Local failures have network-wide effects
  - Ethernet Meltdown at a Boston Hospital
    (Boston Globe 11/26/02)
    - 3 days of disruption to network operation
    - Fix: Redundancy was eliminated!
- Redundancy paradox

3

Proposed Solution: EtherFuse

- New type of device that can be inserted into existing Ethernet
  - Transparent
  - Mitigates effects of failures
  - Safely allows for redundancy



4

EtherFuse: An Ethernet Watchdog

- Expedites Ethernet's convergence time
- Detects and shuts down forwarding loops
- Avoids Ethernet's packet misforwarding
- More general and addresses problems not handled by other solutions
  - Cisco's Loop Guard
  - UniDirectional Link Protocol (UDLP)
  - RSTP with Epochs

5

Outline

- Ethernet's slow convergence time
- Forwarding loops in Ethernet
- Packet misforwarding in Ethernet
- Evaluating the effectiveness of EtherFuse
- Conclusions

6

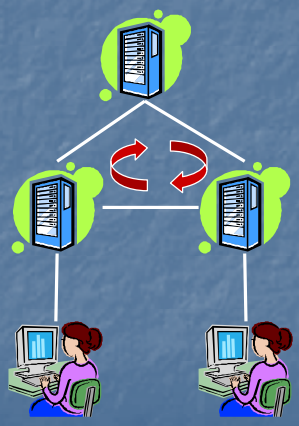
Outline

- Ethernet's slow convergence time
  - Background
  - Cause of slow convergence
  - Mitigation using EtherFuse
- Forwarding loops in Ethernet
- Packet misforwarding in Ethernet
- Evaluating the effectiveness of EtherFuse
- Conclusions

7

Redundancy in Ethernet

- Important for fault tolerance
- Pervasive in real deployments
- Introduces loops



8

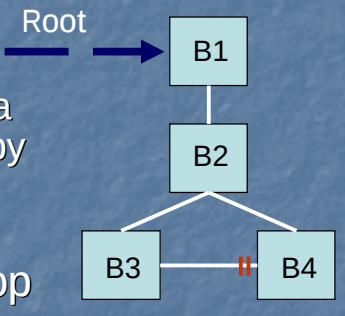
Spanning Tree Protocols

- Build an overlaid loop-free forwarding topology
  - Forwarding topology must be loop free
    - Ethernet uses flooding
    - No TTL in Ethernet packets
    - Packets may persist indefinitely in a loop
- Protocols: STP, RSTP, and MSTP

9

Building The Spanning Tree – An Example

- Elects a bridge as root of the tree
- Builds spanning tree
  - Bridges participate in a distributed algorithm by exchanging protocol messages
- Ports connecting a loop in physical topology are blocked

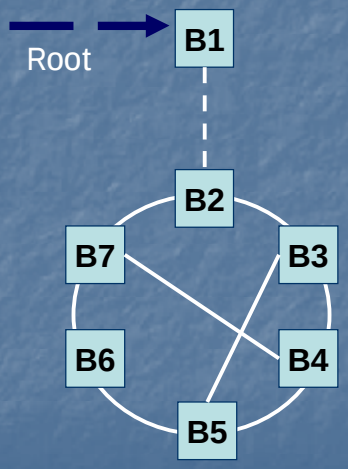


Network of Bridges

10

Ethernet Slow Convergence Time

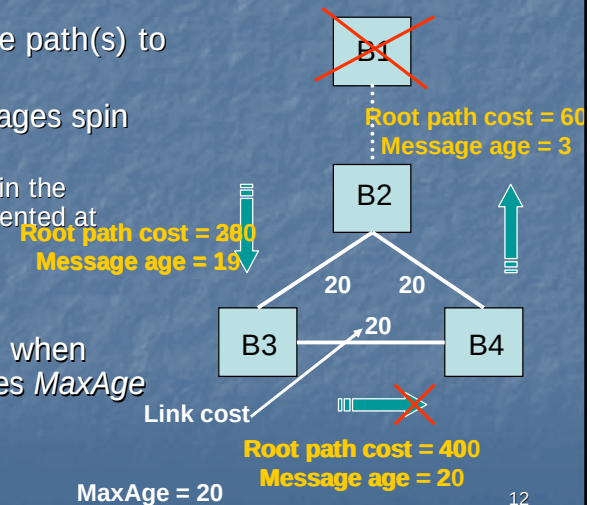
- Due to count to infinity in RSTP and MSTP [Myers05, Elmeleegy06]
  - Count to infinity can occur if:
    - If there exists a cycle in the physical topology
    - Cycle is partitioned from the root



11

The Count-to-Infinity Behavior

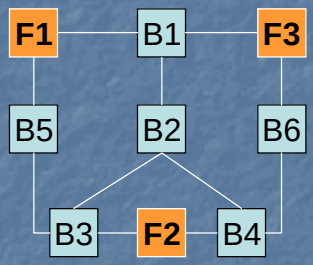
- Bridge(s) cache stale path(s) to the retired root
- Stale protocol messages spin around a loop
  - The following fields in the message are incremented at every hop
    - *Root path cost*
    - *message age*
- Message is dropped when message age reaches *MaxAge*



12

EtherFuse Placement

- One per physical loop



13

EtherFuse: Handling Count to Infinity

- Detection:
  - 2 consecutive spanning tree protocol messages announcing increasing cost to the same root
- Mitigation:
  - Increase message age to MaxAge for all protocol messages until count to infinity ends
    - Expedites expiration of stale information

14

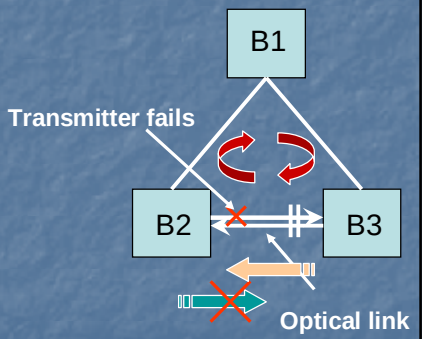
Outline

- Ethernet's slow convergence time
- Forwarding loops in Ethernet
  - Their causes and effects
  - Mitigation using EtherFuse
- Packet misforwarding in Ethernet
- Evaluating the effectiveness of EtherFuse
- Conclusions

15

Causes of Forwarding Loops

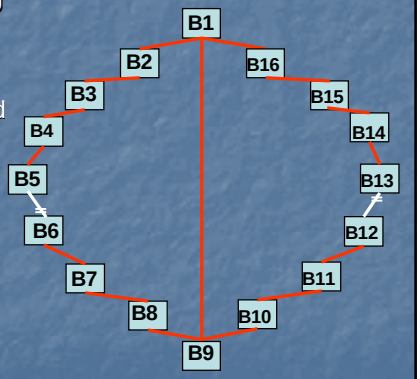
- Due to failures in the spanning tree protocols
  - Protocol message loss
    - Network or control CPU overload
    - Bugs in bridge's firmware
    - Unidirectional links



16

Causes of Forwarding Loops

- Due to failures in the spanning tree protocols
  - Protocol message loss
    - Network or control CPU overload
    - Bugs in bridge's firmware
    - Unidirectional links
    - …
  - Count-to-infinity induced forwarding loops [RSTP, MSTP]
  - Failures driving the network outside the spec [STP]

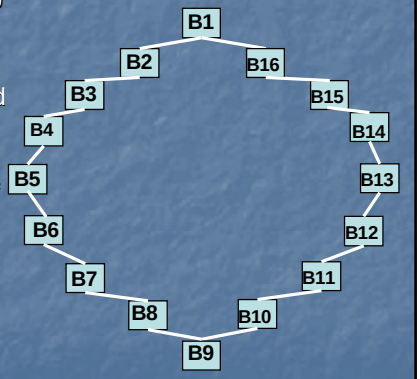


MaxAge = 6

17

Causes of Forwarding Loops

- Due to failures in the spanning tree protocols
  - Protocol message loss
    - Network or control CPU overload
    - Bugs in bridge's firmware
    - Unidirectional links
      - Ex: Failure of transmitter of an optical link
    - …
  - Count-to-infinity induced forwarding loops [RSTP, MSTP]
  - Failures driving the network outside the spec [STP]

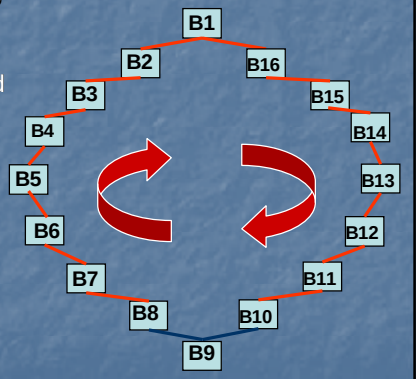


MaxAge = 6

18

Causes of Forwarding Loops

- Due to failures in the spanning tree protocols
  - Protocol message loss
    - Network or control CPU overload
    - Bugs in bridge's firmware
    - Unidirectional links
    - ...
  - Count-to-infinity induced forwarding loops
  - Failures driving the network outside the spec
  - ...
- Temporary or permanent



MaxAge = 6

19

Effects of Forwarding Loops

- Dangerous
  - Frames get trapped in the loop
    - Congestion and instability
    - Flooding end hosts (broadcast storms)
  - Packet misforwarding

20

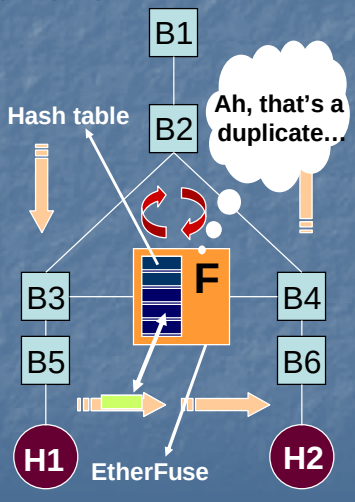
EtherFuse: Detecting Forwarding Loops – Key Idea

- Perfect loop detection requires global view of the network state
  - **à** Can't be done by a single device with local knowledge
- Heuristic:
  - Duplicates are a good indication of a loop
    - Forwarding loop **à** packets spin around the loop to arrive again at the EtherFuse

21

EtherFuse: Detecting Duplicate Packets

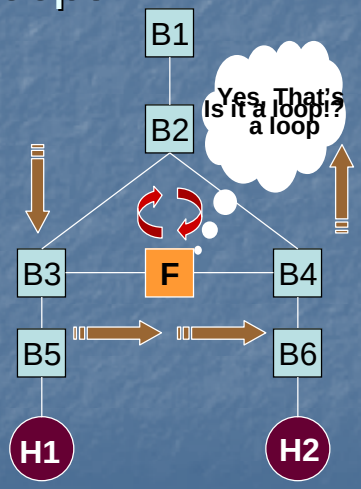
- When a new packet arrives at the EtherFuse, its hash is recorded.
  - CRC used as the hash
- When a packet rearrives at the EtherFuse, its hash exists in hash table
  - Signals a duplicate
- Duplicate detection not conclusive of a loop
  - Hash collisions,…



22

EtherFuse: Detecting Forwarding Loops

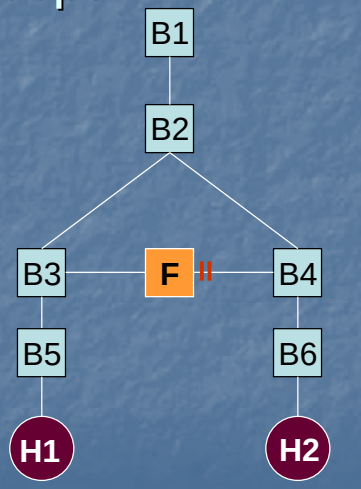
- If duplicate is detected
  - Send a Probe
- If probe arrives at the other port
  - A loop exists



23

EtherFuse: Handling Forwarding Loops

- If a loop is detected
  - Cut it



24

Outline

- Ethernet's slow convergence time
- Forwarding loops in Ethernet
- Packet misforwarding in Ethernet
  - Packet forwarding in Ethernet
  - Mitigation using EtherFuse
- Evaluating the effectiveness of EtherFuse
- Conclusions

25

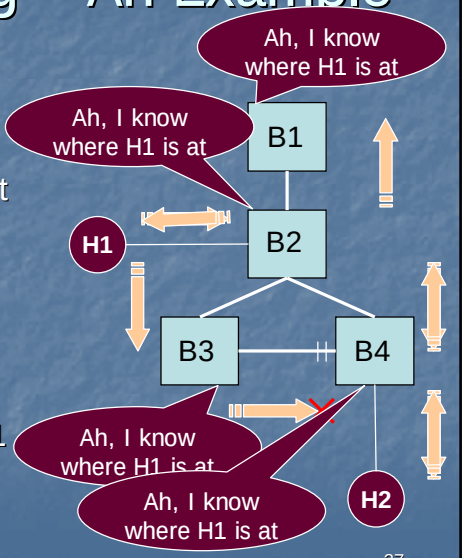
Address Learning

- Ethernet has flat addressing
  - Flooding for address learning

26

Address Learning – An Example

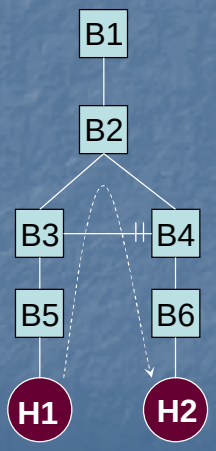
- None of the hosts H1 or H2 has sent any packets
- Host H1 sends its first packet to H2
- H2's location unknown
  - Packet flooded
  - H1's location gets known
  - Packet arrives at H2
- No flooding needed when H1 receives a packet



27

Forwarding Table Pollution

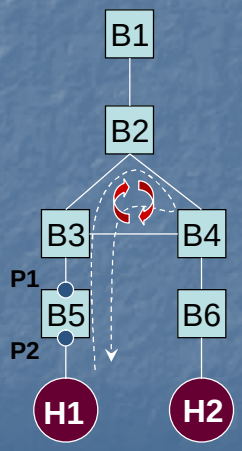
- Under normal operation



28

Forwarding Table Pollution

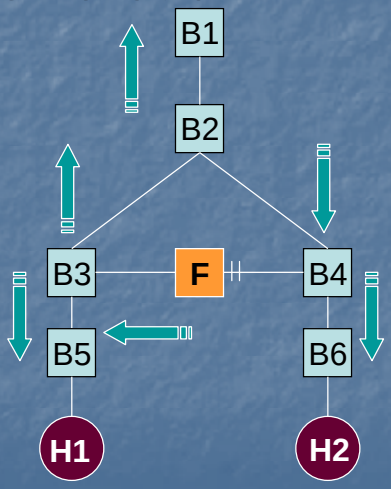
- Under a forwarding loop
  - H1's packet spins the loop to rearrive at B5 at P1
  - B5 thinks that H1 is reachable via P1
  - Packets from H2 to H1 will be dropped by B5
    - H1 cutoff from network



29

EtherFuse: Handling Forwarding Table Pollution

- After a loop is detected and cut
  - Send a topology change message to flush forwarding tables



30

Outline

- Ethernet's slow convergence time
- Forwarding loops in Ethernet
- Packet misforwarding in Ethernet
- Evaluating the effectiveness of EtherFuse
- Conclusions

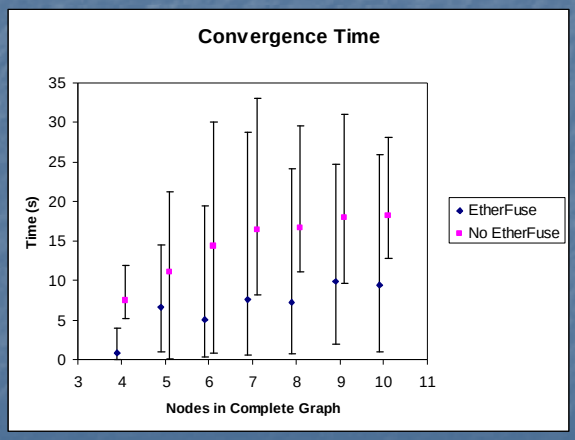
31

Simulator

- Simulates:
  - Ethernet's Rapid Spanning Tree Protocol based on latest IEEE spec for LANs (IEEE 802.1D-2004)
  - EtherFuses

32

Effects of Count to Infinity on Convergence Time

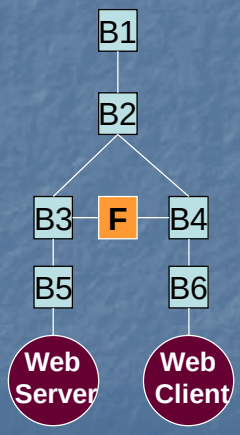


Complete graph topologies under count to infinity

33

Experimental Setup

- Click modular router
  - Software Bridges
  - EtherFuse
- Run on Emulab nodes

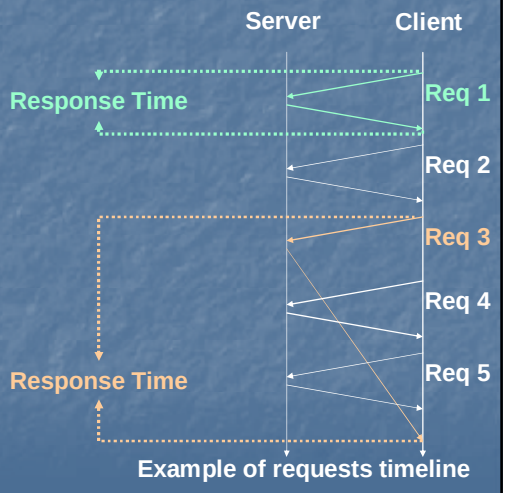


topology

34

Web Experiments

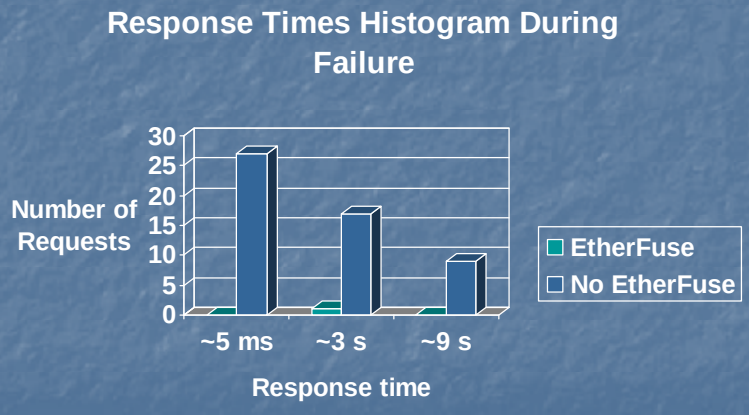
- server:
  - Apache 2.2
- Client:
  - 10 requests/s
  - 44-Byte requests
  - Non-persistent connection



Example of requests timeline

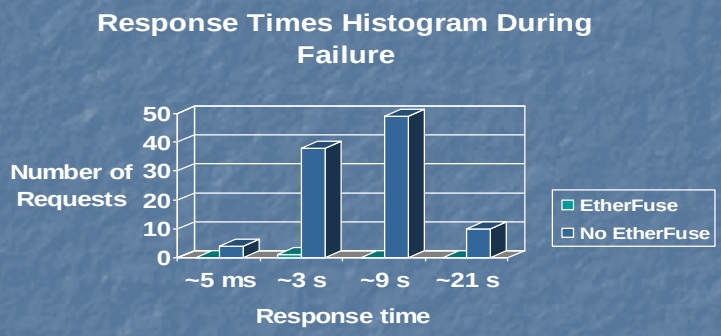
35

Effects of Count to Infinity on Web Requests



36

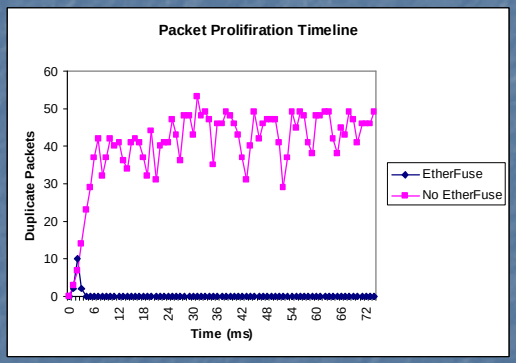
Effects of Forwarding Loops on Web Requests

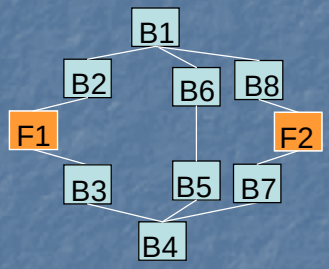


- Count to infinity induced forwarding loop
- Background broadcast traffic of 100 KB/s

37

Effects of Multiple Forwarding Loops





- Single broadcast (ARP) packet injected into the network

38

Conclusions

- Redundancy in Ethernet
  - Important for fault tolerance
  - Leaves Ethernet vulnerable to spanning tree protocols failures
    - Slow convergence
    - Forwarding loops
    - Packet misforwarding
- EtherFuse
  - Safely allows for redundancy
    - Substantially mitigates effects of protocol failures
  - Transparent

39