California Fault Lines: Understanding the Causes and Impact of Network Failures

Daniel Turner
Kirill Levchenko,
Alex C. Snoeren,
Stefan Savage

UCSDCSE
Computer Science and Engineering

Why Study Failure

- Failure is a reality for large network
- Achieving high availability requires engineering the network to be robust to failure
- Designing mechanisms to effectively mitigate failures requires deep understanding of real failures

What do we Know

- Big Failures generate news stories

What do we Know

- Big Failures generate news stories

Slashdot NEWS FOR NERDS. STUFF THAT MATTERS.

Log In Create Account Help Subscribe Firehose

Slashdot
Apple
Ask Slashdot
Book Reviews
Games
Hardware
Idle
Interviews
Linux
Mobile
Science

Transatlantic Cable Fault Disrupts Internet In UK

Posted by timothy on Wed Nov 26, 2003 12:32 PM
from the electrify-the-atlantic dept.

An anonymous reader submits

> "Web traffic between the U.S. and Europe has been hit after an undersea cable developed a major fault on Tuesday. Because the TAT-14 cable network is shaped like a ring, it should be able to cope with one such failure -- but unfortunately the consortium that owns it hadn't fixed an earlier problem, just off the U.S. coast. Just shows how systems with build-in redundancy can still go badly wrong...."

What do we Know

- Big Failures generate news stories

Slashdot NEWS FOR NERDS. STUFF THAT MATTERS.

Log In Create Account Help Subscribe Firehose

Slashdot
Apple
Ask Slashdot
Book Reviews
Games
Hardware
Idle
Interviews

Transatlantic Cable Fault Disrupts Internet In UK

Posted by timothy on Wed Nov 26, 2003 12:32 PM
from the electrify-the-atlantic dept.

An anonymous reader submits

> "Web traffic between the U.S. and Europe has been hit after an undersea cable developed a major fault on Tuesday. Because the TAT-14 cable network is shaped like a ring, it should be able to cope with one such

Fiber Cuts Slash Silicon Valley's Internet Arteries

By Stephen Lawson, IDG News

Print Digg Twitter Facebook More...

Cuts in fiber-optic lines early Thursday at two locations near Silicon Valley shut down two IBM facilities and affected an organization in charge of Internet domain names.

What do we Know

- Big Failures generate news stories
  - Rarely contain useful details
  - Most networks failures are not catastrophic

Fiber Cuts Slash Silicon Valley's Internet Arteries

By Stephen Lawson, IDG News

Print Digg Twitter Facebook More...

Cuts in fiber-optic lines early Thursday at two locations near Silicon Valley shut down two IBM facilities and affected an organization in charge of Internet domain names.

Why do we know so little?

Why do we know so little?

- Collecting comprehensive failure data is difficult
  - Lightweight techniques are limited
  - Special purpose monitoring is expensive
- Access to network data is limited data
  - A few publicly available studies
    - [A. Markopoulou ToN '08] [C. Cranor SIGMOD 03]
  - Many networks consider data proprietary
- Some networks can't invest time or capital

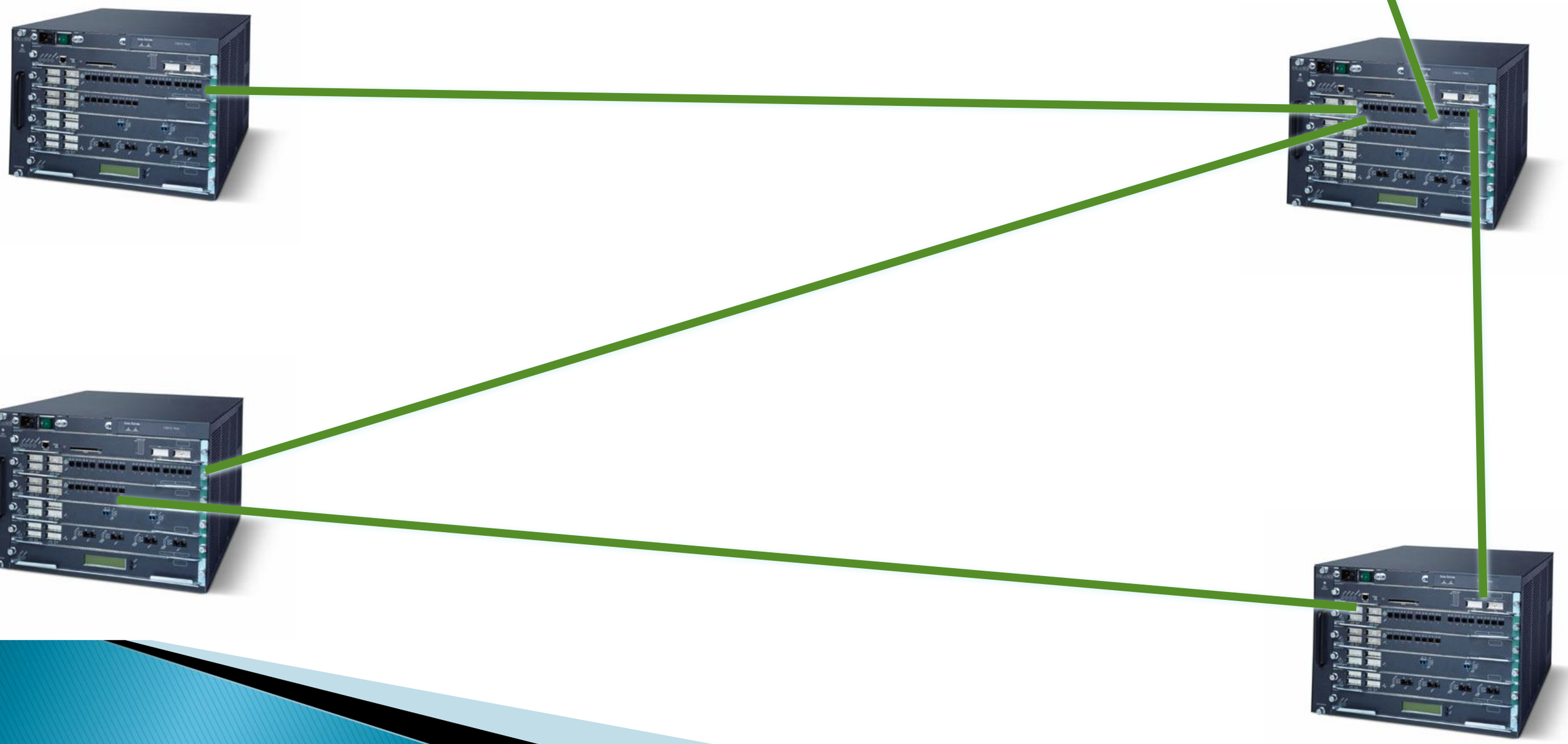
Contributions

- Methodology to reconstruct failure history of a network
  - Using only commonly available data
  - No need for additional instrumentation
- Analyze a production network

Network Failure History

- A time series of Layer-3 failure events
  - I.e, for each link a set of state transitions between up and down
- And, where possible, annotated with:
  - What caused the failure?
  - What was the impact of the failure?

Failure Example



Failure Example

interface GigabitEthernet1/1
ip address 137.211.22.8 255.255.255.254

Failure Example

interface GigabitEthernet0/2
ip address 137.211.23.2 255.255.255.254

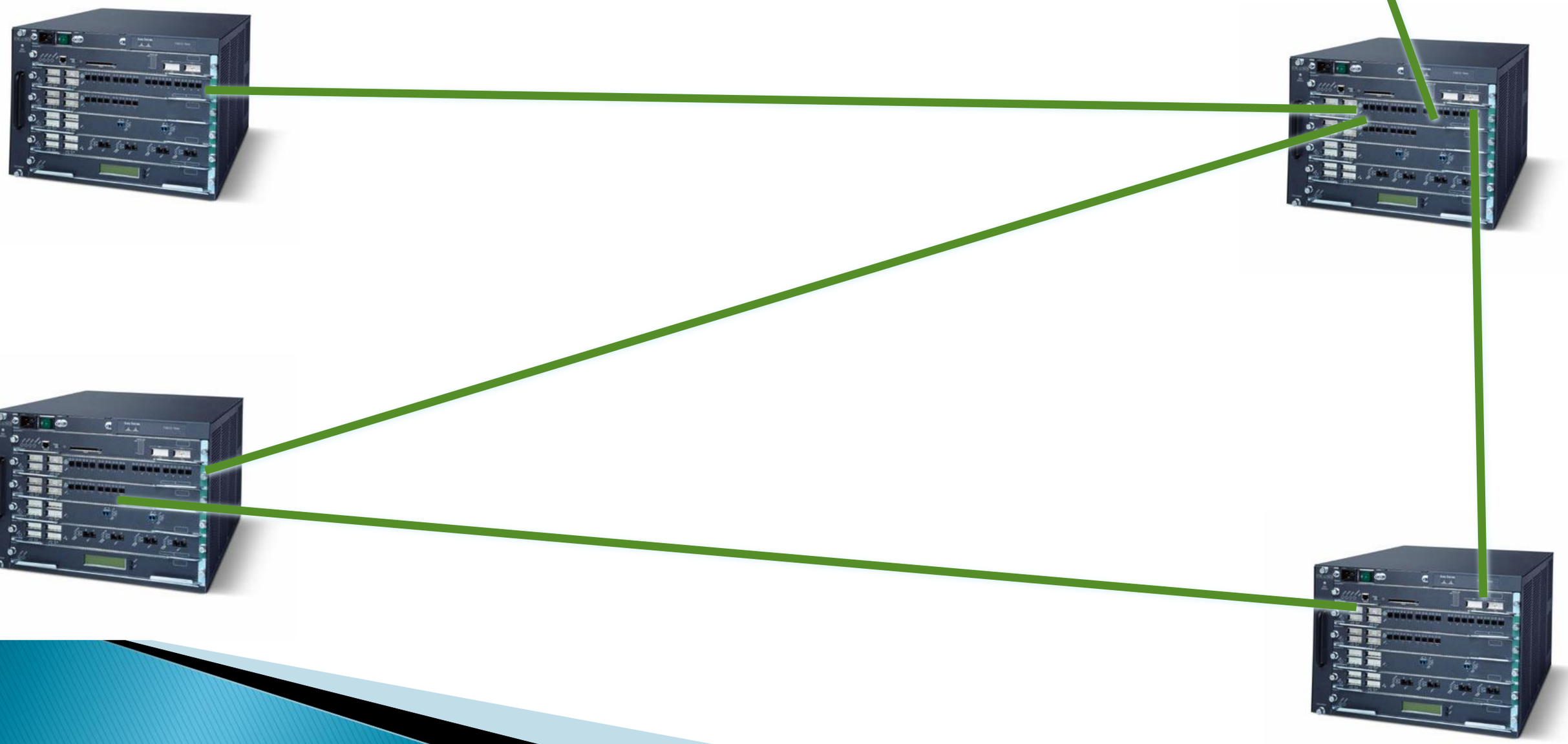
Failure Example

```
interface GigabitEthernet1/1
ip address 137.211.22.9 255.255.255.254
```

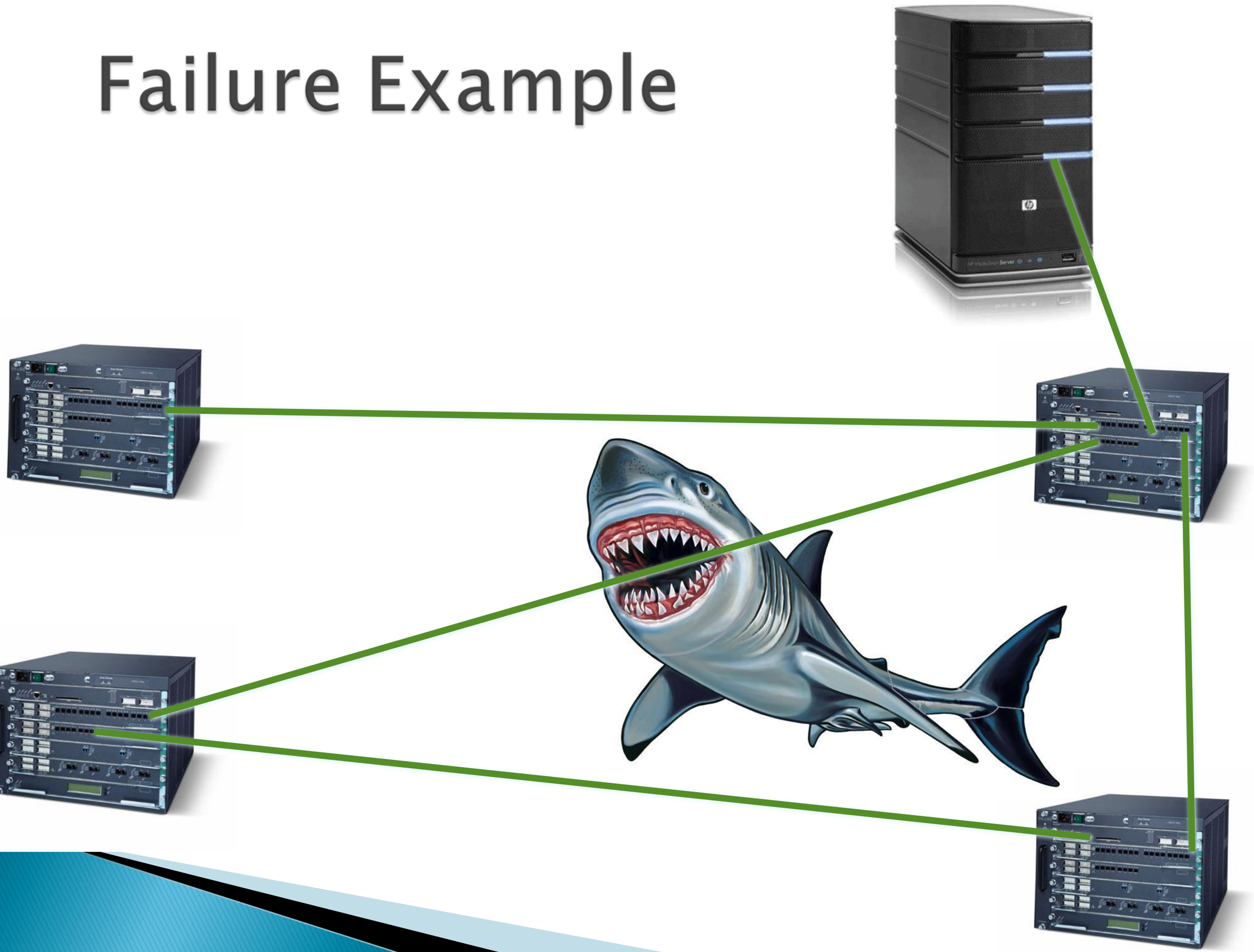
Failure Example

interface GigabitEthernet3/2
ip address 137.211.25.9 255.255.255.254

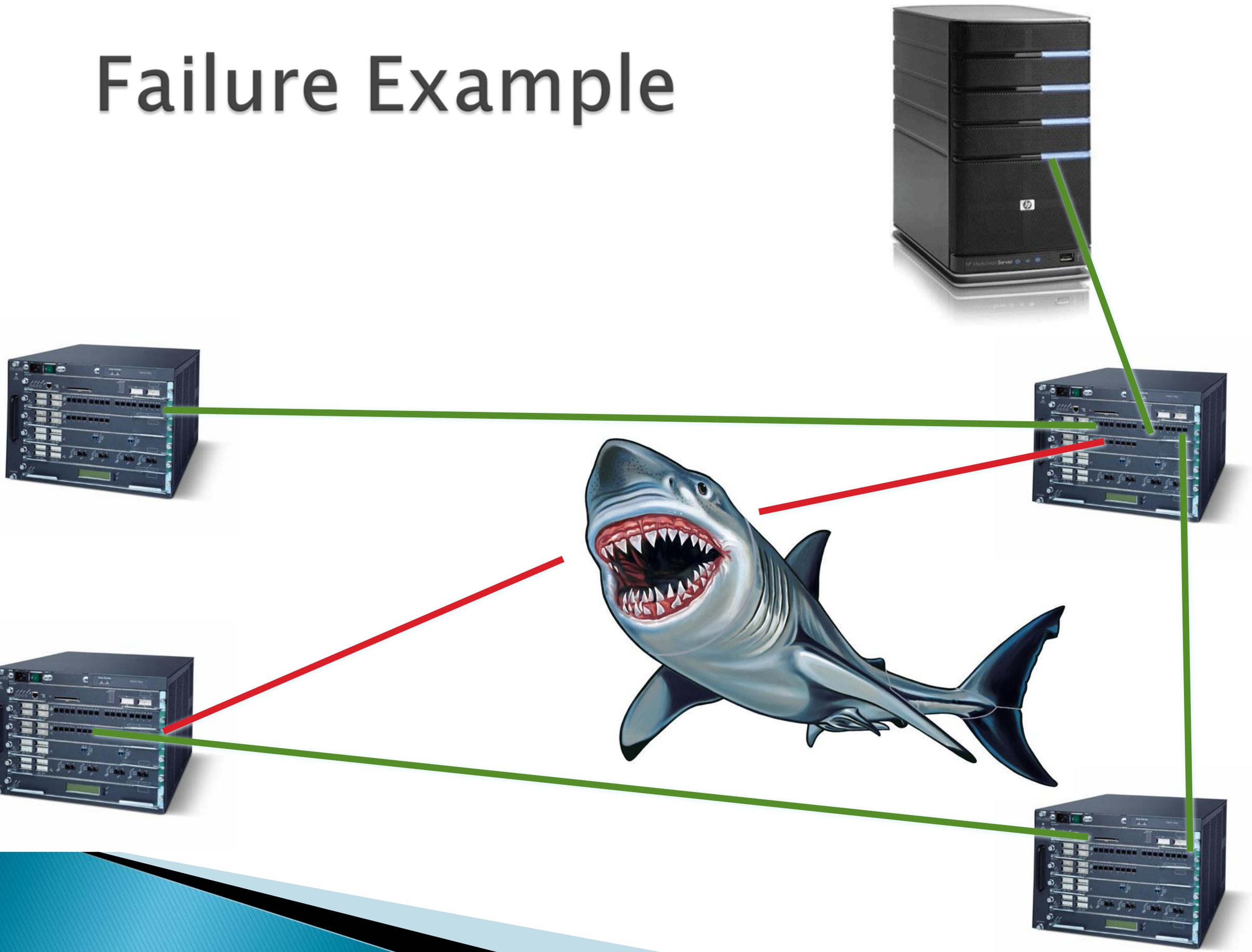
Failure Example



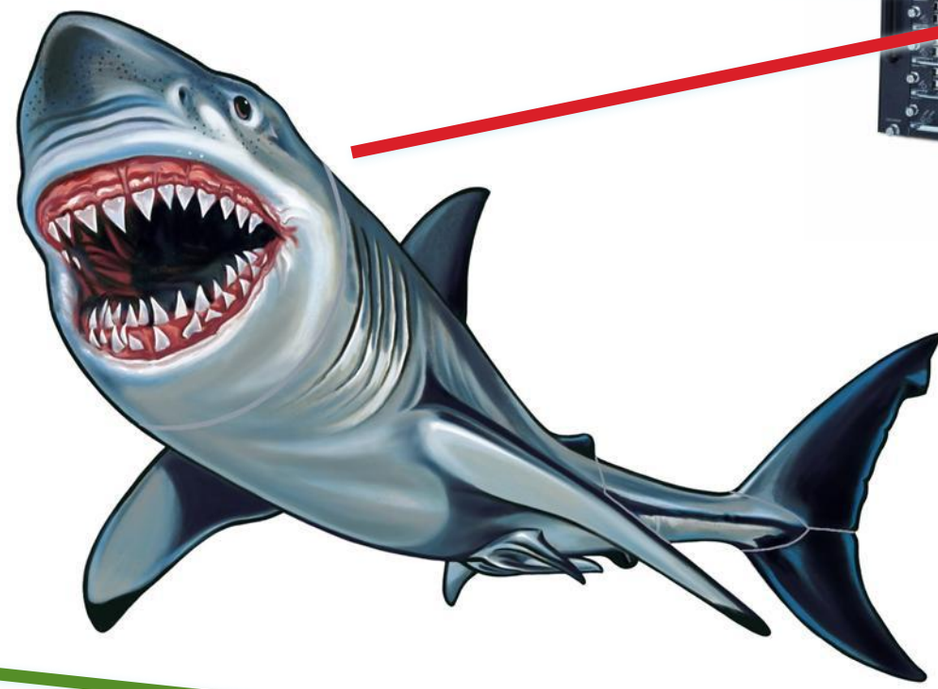
Failure Example



Failure Example



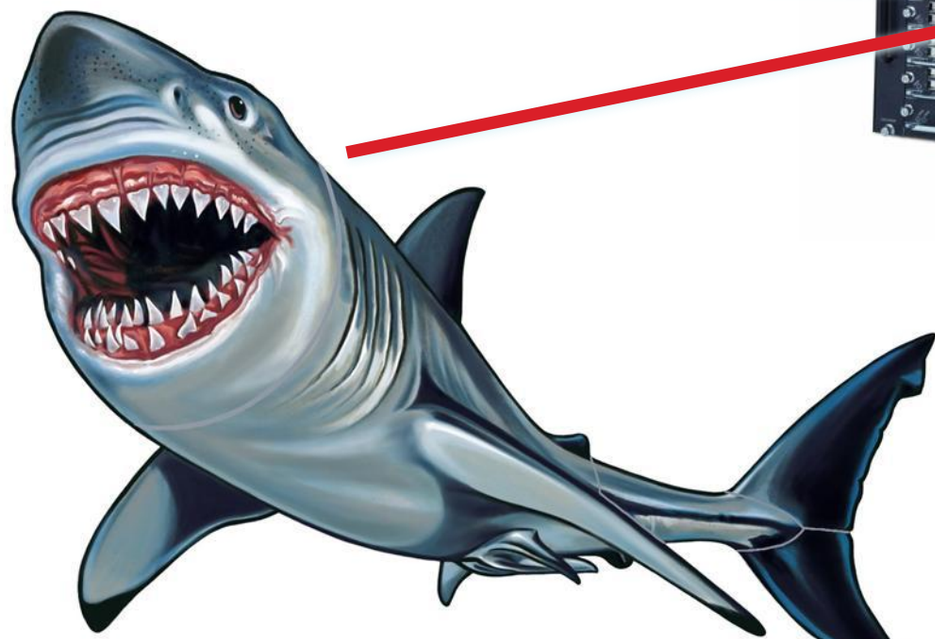
Failure Example

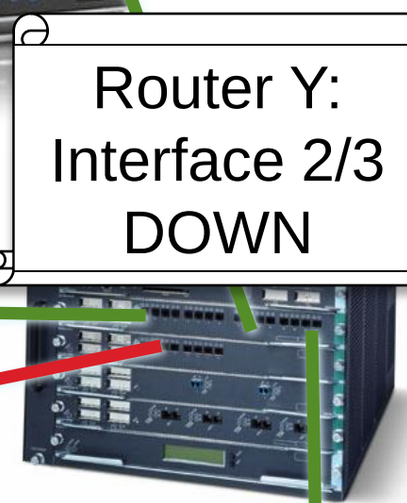


Router x:
Interface 1/1
DOWN

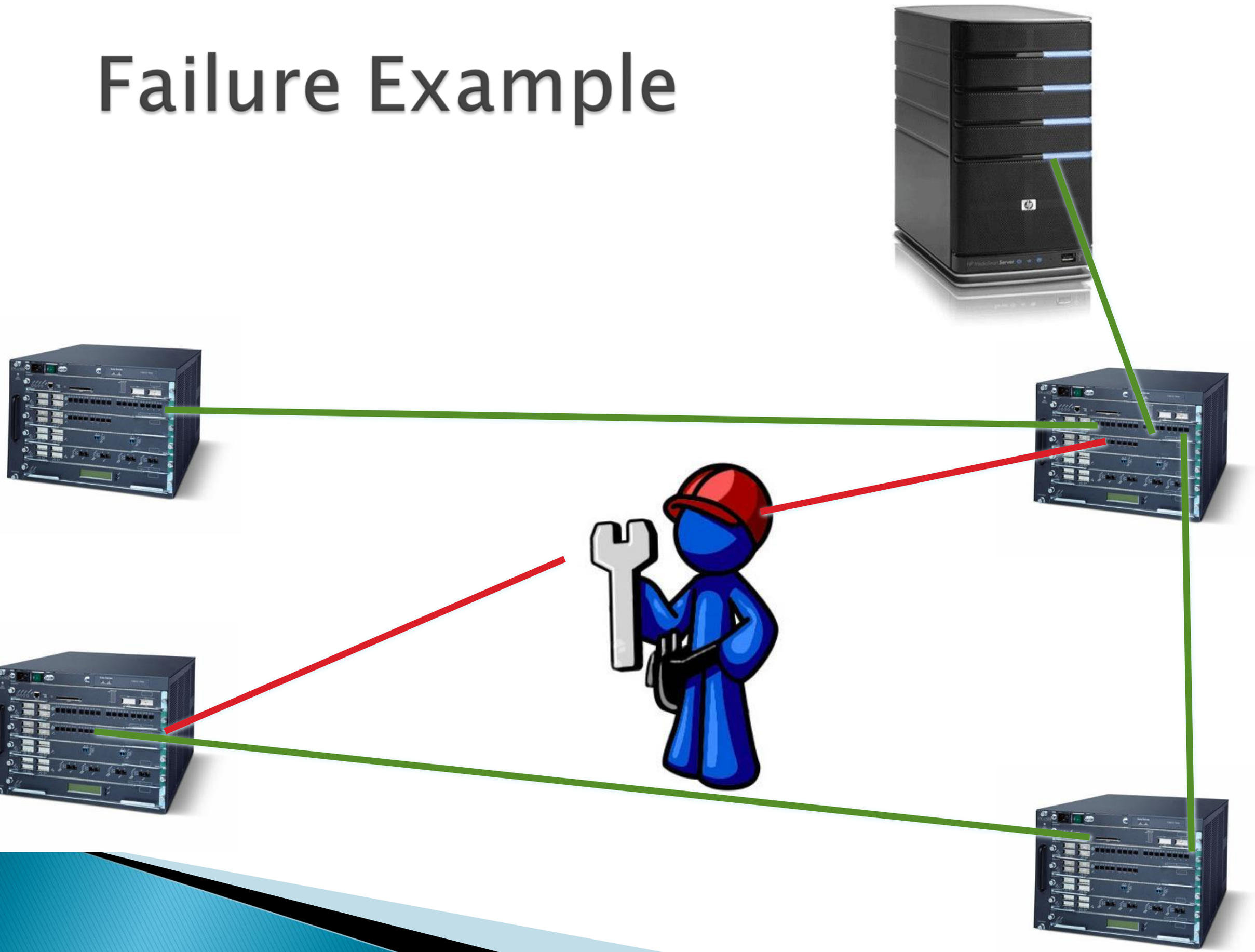
Failure Example

Router Y:
Interface 2/3
DOWN





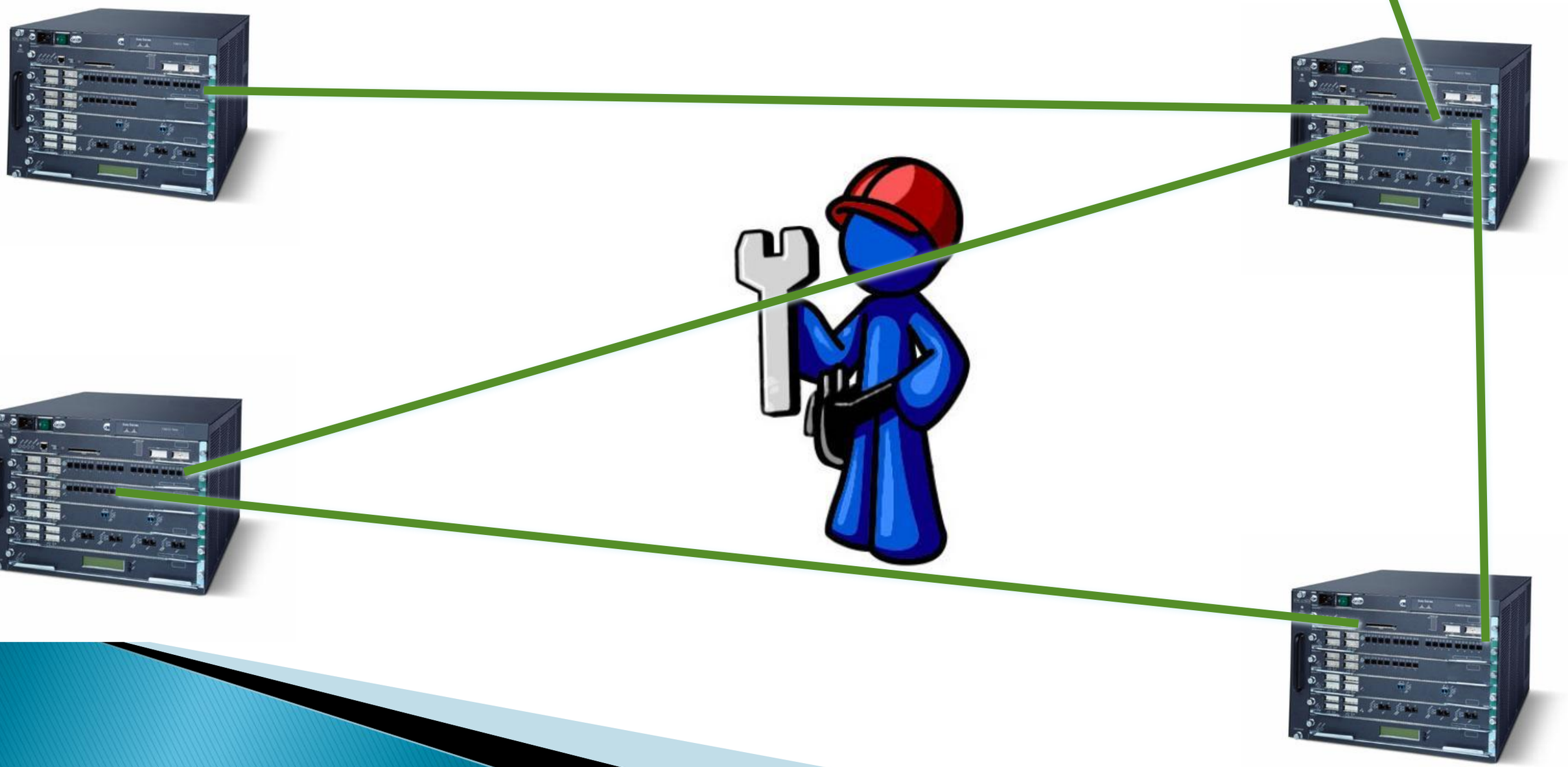
Failure Example



Failure Example

Router x:
Interface 1/1
UP

Failure Example



Failure Example

Router Y:
Interface 2/3
UP

Failure Example

This message is to alert you that the CENIC network engineering team has scheduled an emergency repair

Start 0001 PDT, FRI 9/02/06
End 0200 PDT, FRI 9/02/06

SCOPE: Shark bites through cable

IMPACT: Loss of redundancy between San Francisco and Los Angles

COMMENTS
It left behind a tooth

Reconstructing a Failure

- How can we reconstruct a failure 4 years later?
  - Syslog
    - Describes **interface** state changes
  - Router Configuration Files
    - Maps interfaces to Links
  - Operation announcements
- Caveat: data not intended for failure reconstruction

Reconstruction Example

Reconstruction Example

?

?

?

?

?

?

?

?

interface GigabitEthernet1/1
ip address 137.211.22.8 255.255.255.254

Reconstruction Example

?

?

?

?

?

?

?

interface GigabitEthernet1/1

ip address 137.211.22.8 255.255.255.254

Reconstruction Example

?

?

?

?

?

?

?

?

interface GigabitEthernet1/1

ip address 137.211.22.8 255.255.255.254

Reconstruction Example

137.211.22.9

interface GigabitEthernet1/1

ip address 137.211.22.8 255.255.255.254

Reconstruction Example

137.211.22.9

?

?

interface GigabitEthernet0/2
ip address 137.211.23.2 255.255.255.254

?

?

?

interface GigabitEthernet1/1
ip address 137.211.22.8 255.255.255.254

Reconstruction Example

137.211.22.9

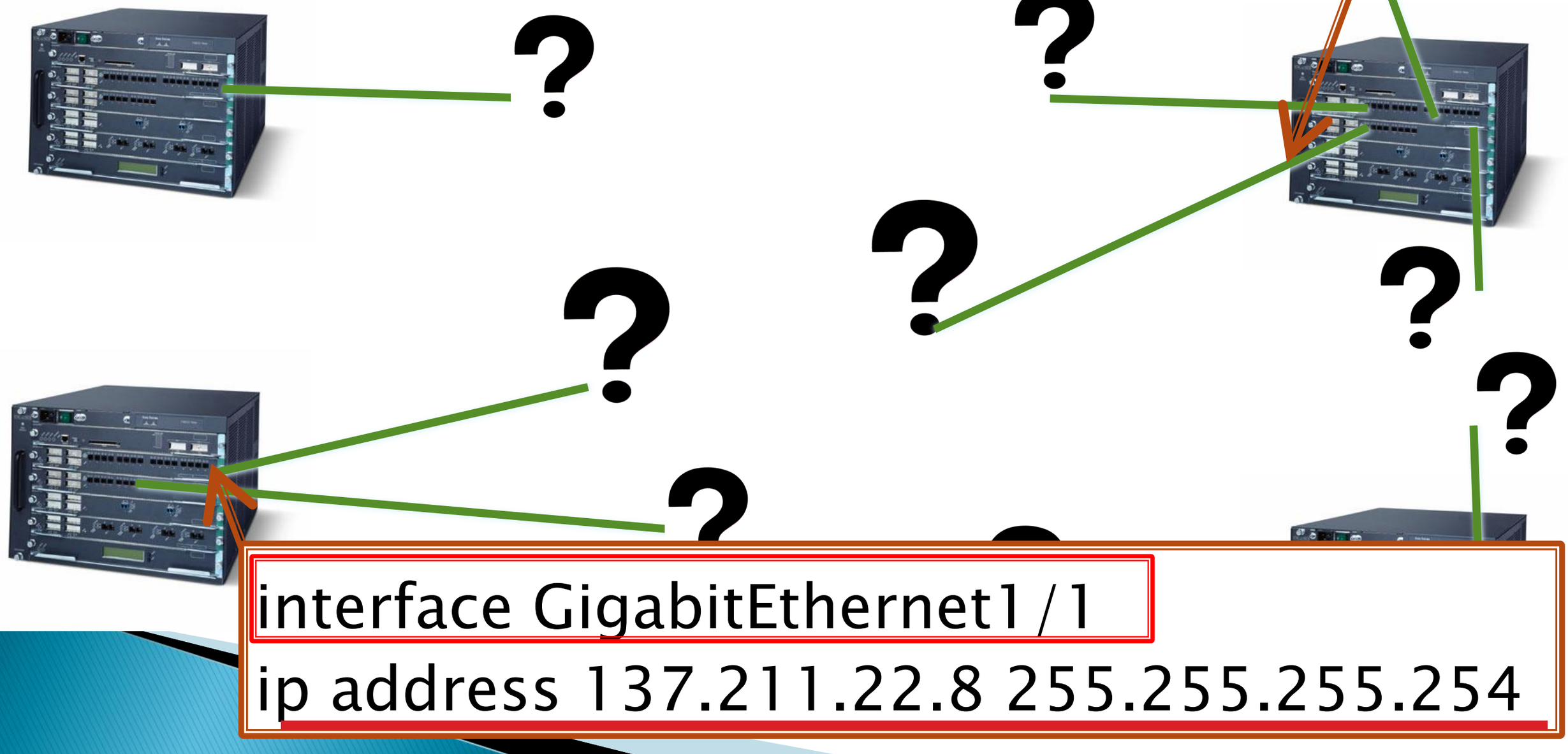
interface GigabitEthernet1/1

ip address 137.211.22.8 255.255.255.254

Reconstruction Example

137.211.22.9

interface GigabitEthernet1/1
ip address 137.211.22.9 255.255.255.254

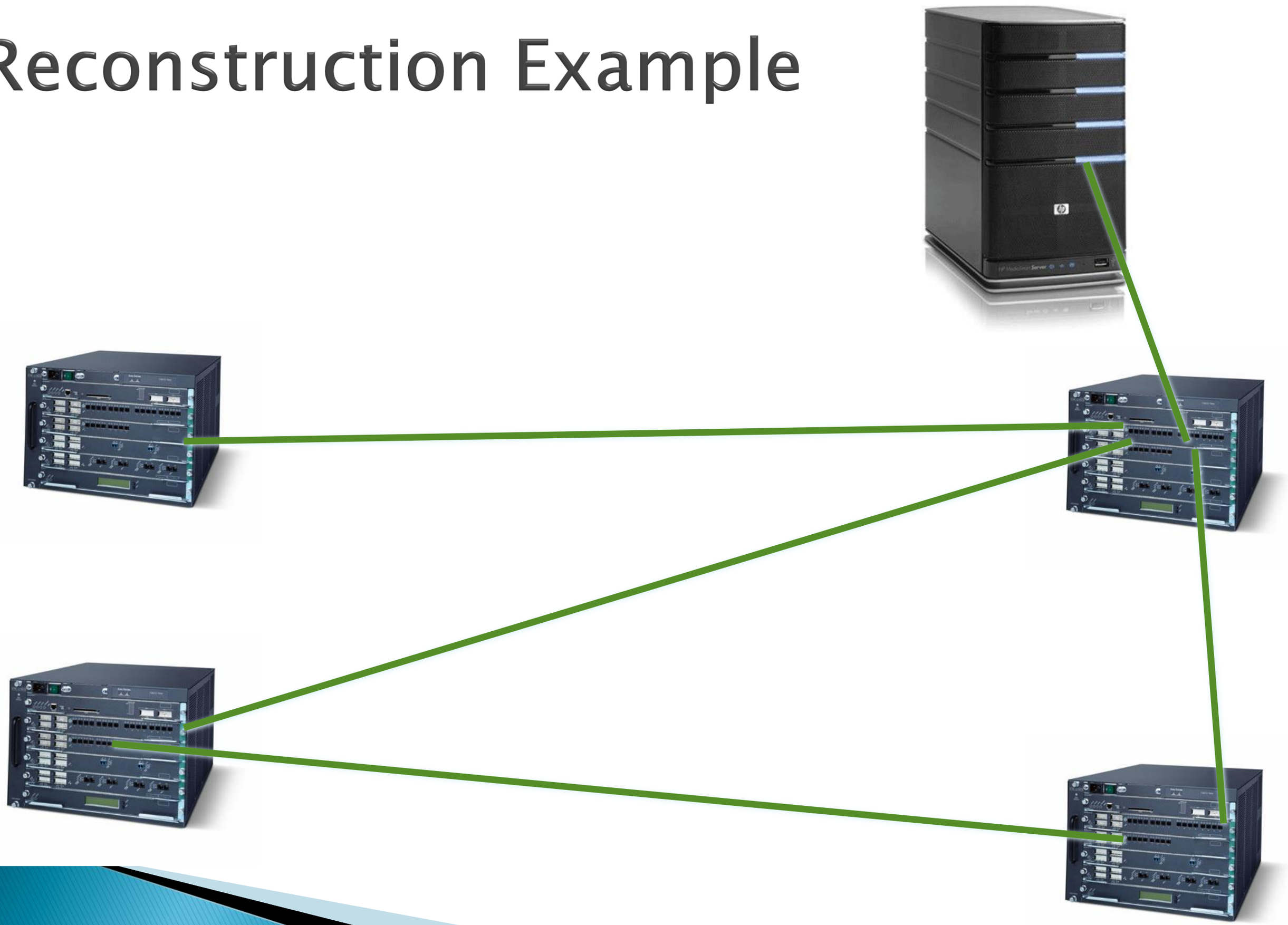


interface GigabitEthernet1/1
ip address 137.211.22.8 255.255.255.254

Reconstruction Example

?
?
?
?
?
?
?

Reconstruction Example



Reconstruction Example

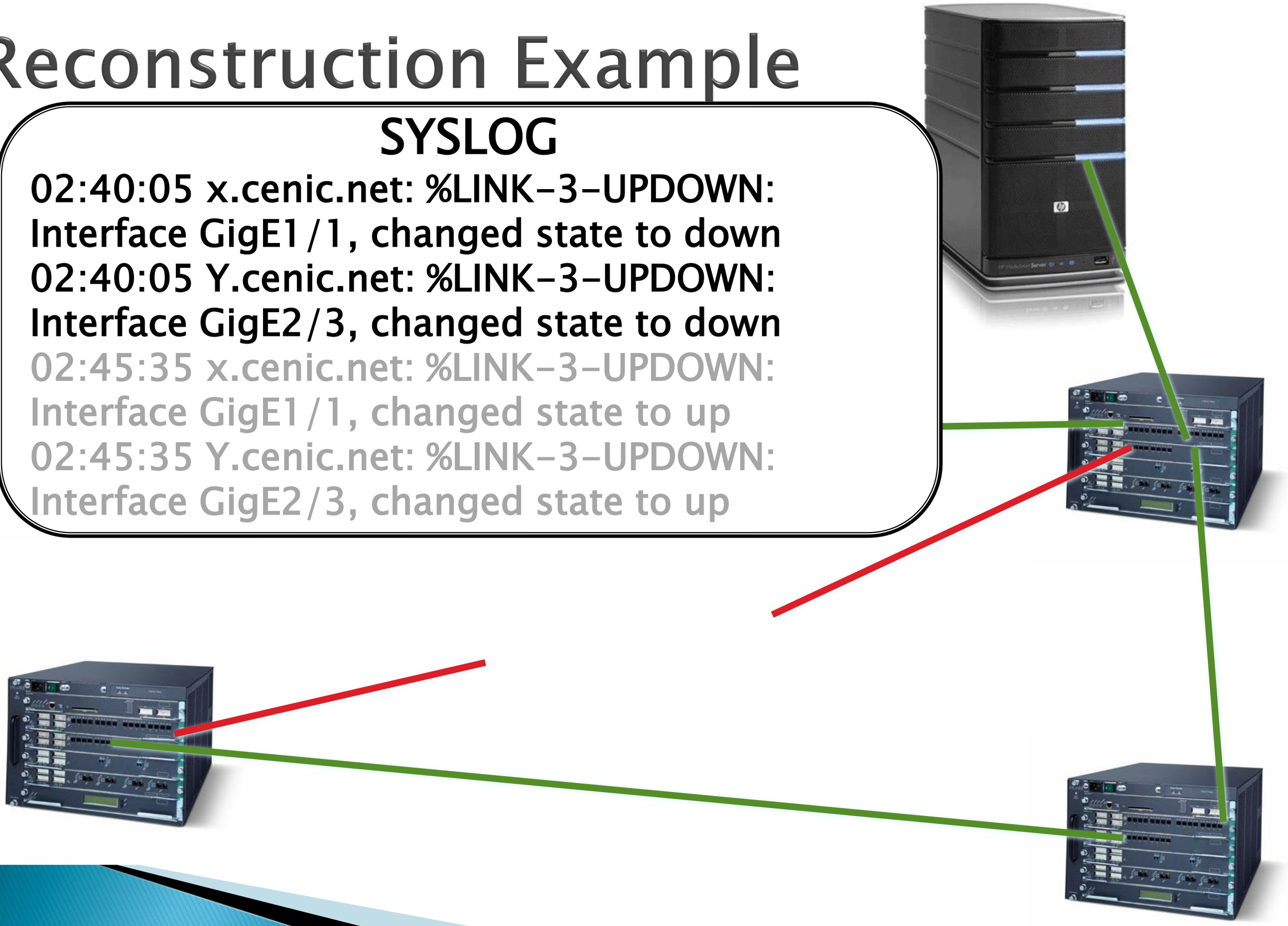
SYSLOG

02:40:05 x.cenic.net: %LINK-3-UPDOWN:
Interface GigE1/1, changed state to down
02:40:05 Y.cenic.net: %LINK-3-UPDOWN:
Interface GigE2/3, changed state to down
02:45:35 x.cenic.net: %LINK-3-UPDOWN:
Interface GigE1/1, changed state to up
02:45:35 Y.cenic.net: %LINK-3-UPDOWN:
Interface GigE2/3, changed state to up

Reconstruction Example

SYSLOG

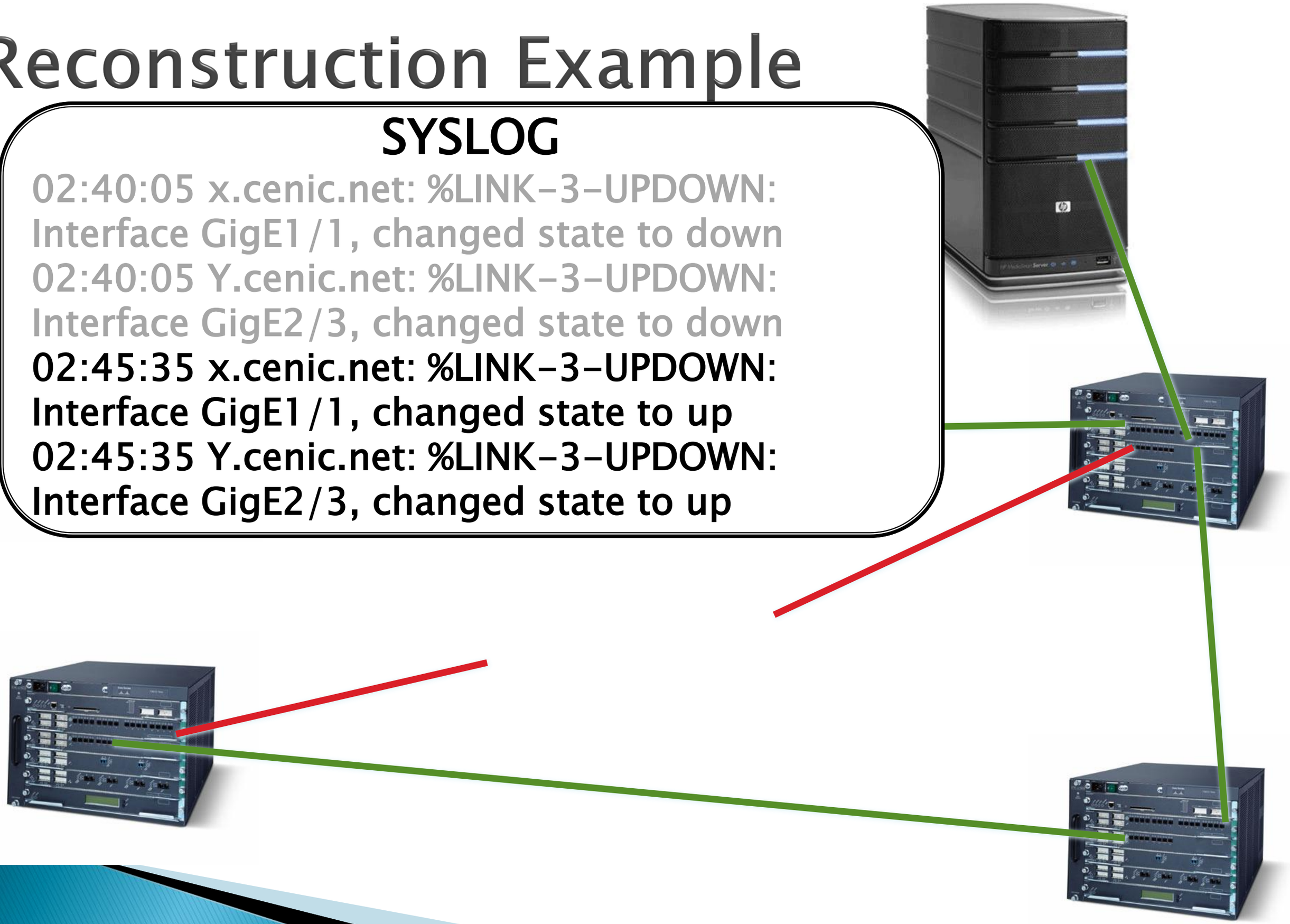
02:40:05 x.cenic.net: %LINK-3-UPDOWN:
Interface GigE1/1, changed state to down
02:40:05 Y.cenic.net: %LINK-3-UPDOWN:
Interface GigE2/3, changed state to down
02:45:35 x.cenic.net: %LINK-3-UPDOWN:
Interface GigE1/1, changed state to up
02:45:35 Y.cenic.net: %LINK-3-UPDOWN:
Interface GigE2/3, changed state to up



Reconstruction Example

SYSLOG

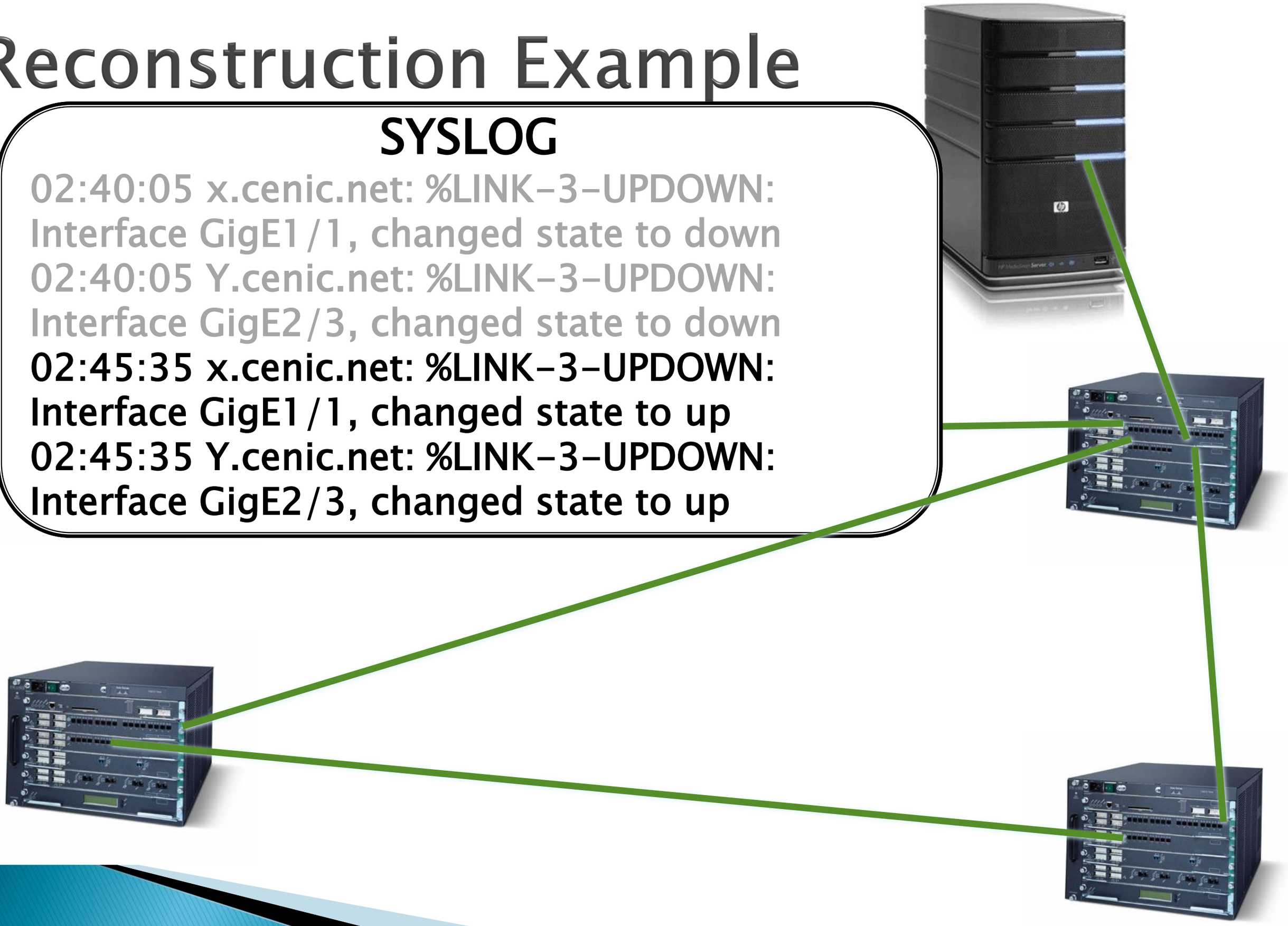
02:40:05 x.cenic.net: %LINK-3-UPDOWN:
Interface GigE1/1, changed state to down
02:40:05 Y.cenic.net: %LINK-3-UPDOWN:
Interface GigE2/3, changed state to down
02:45:35 x.cenic.net: %LINK-3-UPDOWN:
Interface GigE1/1, changed state to up
02:45:35 Y.cenic.net: %LINK-3-UPDOWN:
Interface GigE2/3, changed state to up



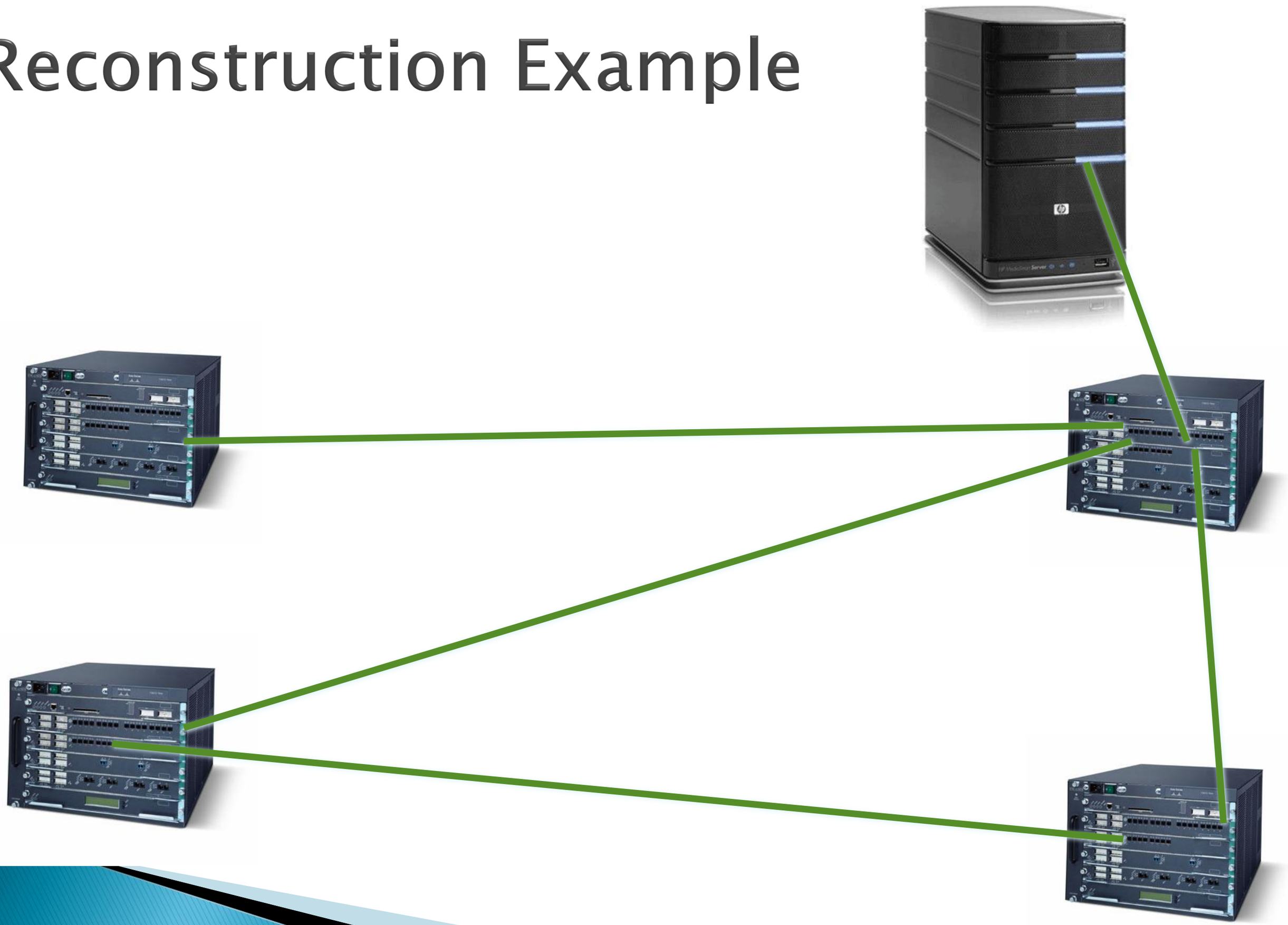
Reconstruction Example

SYSLOG

02:40:05 x.cenic.net: %LINK-3-UPDOWN:
Interface GigE1/1, changed state to down
02:40:05 Y.cenic.net: %LINK-3-UPDOWN:
Interface GigE2/3, changed state to down
02:45:35 x.cenic.net: %LINK-3-UPDOWN:
Interface GigE1/1, changed state to up
02:45:35 Y.cenic.net: %LINK-3-UPDOWN:
Interface GigE2/3, changed state to up



Reconstruction Example



Reco

This message is to alert you that the CENIC network engineering team is performing an emergency repair

Start 0001 PDT, FRI 9/02/06
End 0200 PDT, FRI 9/02/06

SCOPE: Shark bites through cable

IMPACT: Loss of redundancy between San Francisco and Los Angles

COMMENTS
It left behind a tooth

Reco

This message is to alert you that the CENIC network engineering team is performing an emergency repair

Start 0001 PDT, FRI 9/02/06
End 0200 PDT, FRI 9/02/06

SCOPE: Shark bites through cable

IMPACT: Loss of redundancy between San Francisco and Los Angles

COMMENTS
It left behind a tooth

Reco

This message is to alert you that the CENIC network engineering team is performing an emergency repair

Start 0001 PDT, FRI 9/02/06
End 0200 PDT, FRI 9/02/06

SCOPE: Shark bites through cable

IMPACT: Loss of redundancy between San Francisco and Los Angles

COMMENTS
It left behind a tooth

Reco

This message is to alert you that the CENIC network engineering team is performing an emergency repair

Start 0001 PDT, FRI 9/02/06
End 0200 PDT, FRI 9/02/06

SCOPE: Shark bites through cable

IMPACT: Loss of redundancy between San Francisco and Los Angles

COMMENTS
It left behind a tooth

Reco

This message is to alert you that the CENIC network engineering team is performing an emergency repair

Start 0001 PDT, FRI 9/02/06
End 0200 PDT, FRI 9/02/06

SCOPE: Shark bites through cable

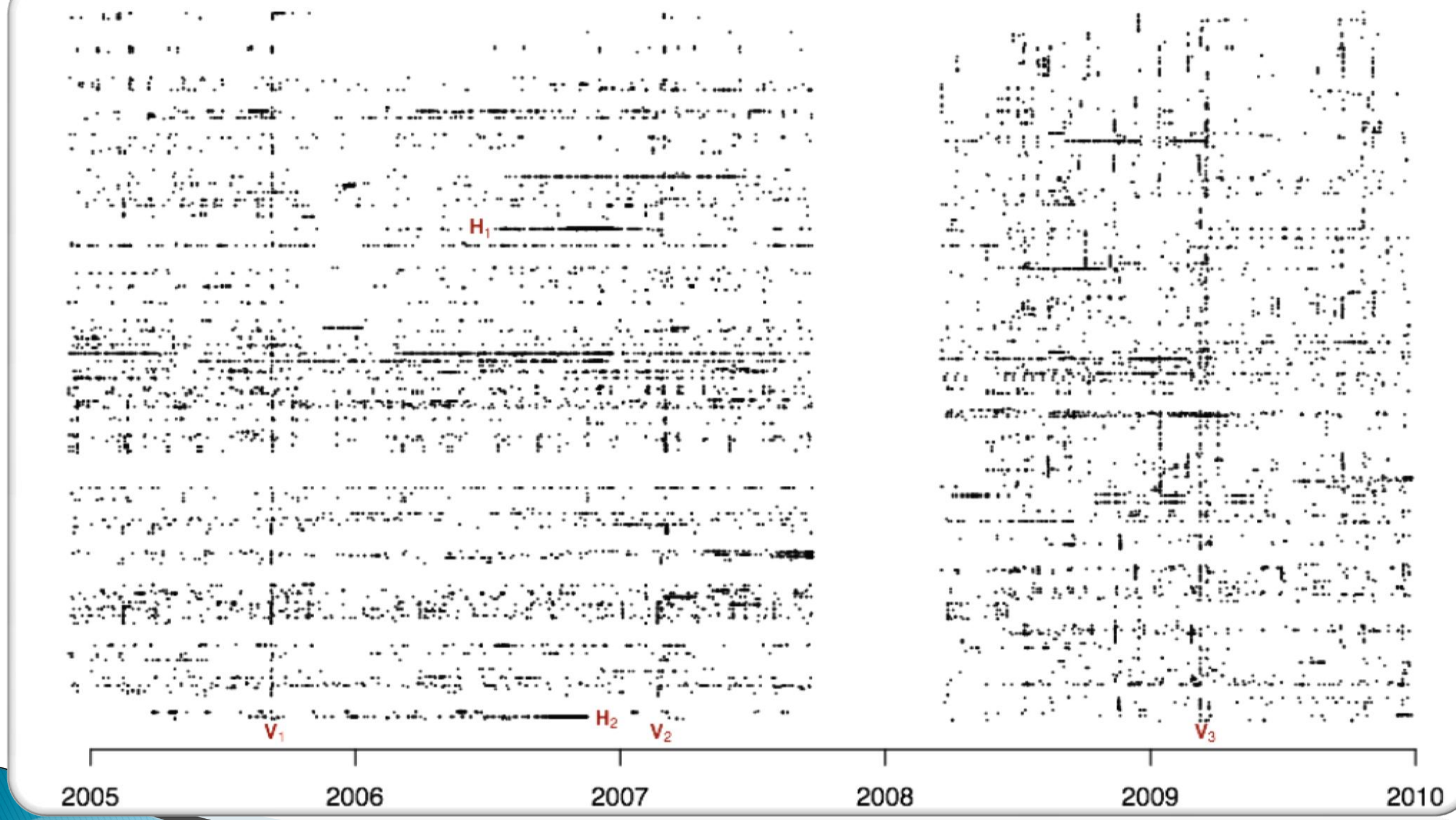
IMPACT: Loss of redundancy between San Francisco and Los Angles

COMMENTS
It left behind a tooth

CENIC Network

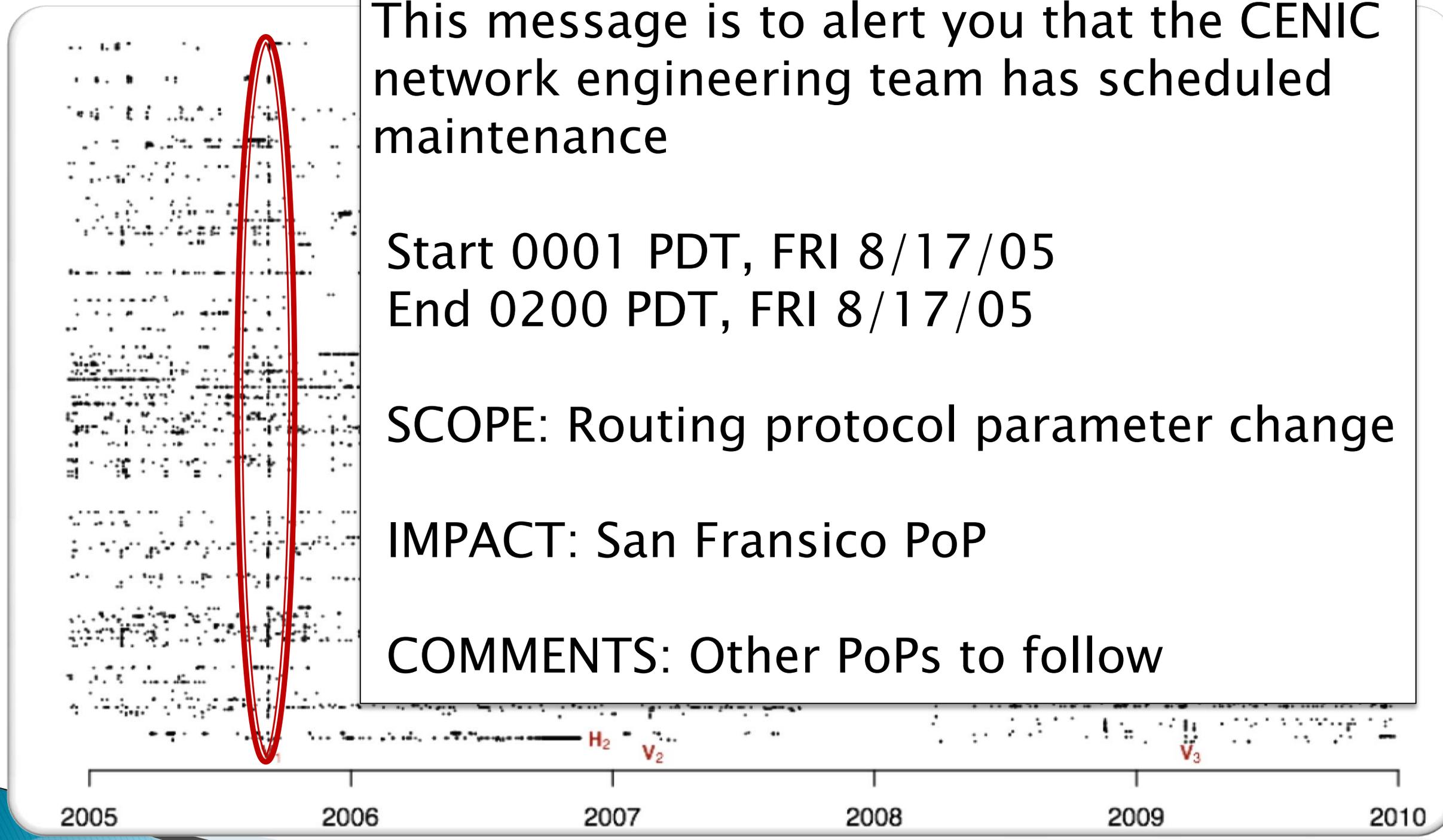
- Serving California educational institutions
- Over 200 routers
- 5 years of data

Failures in CENIC



Failures in CENIC

Failures in CENIC



This message is to alert you that the CENIC network engineering team has scheduled maintenance

Start 0001 PDT, FRI 8/17/05
End 0200 PDT, FRI 8/17/05

SCOPE: Routing protocol parameter change

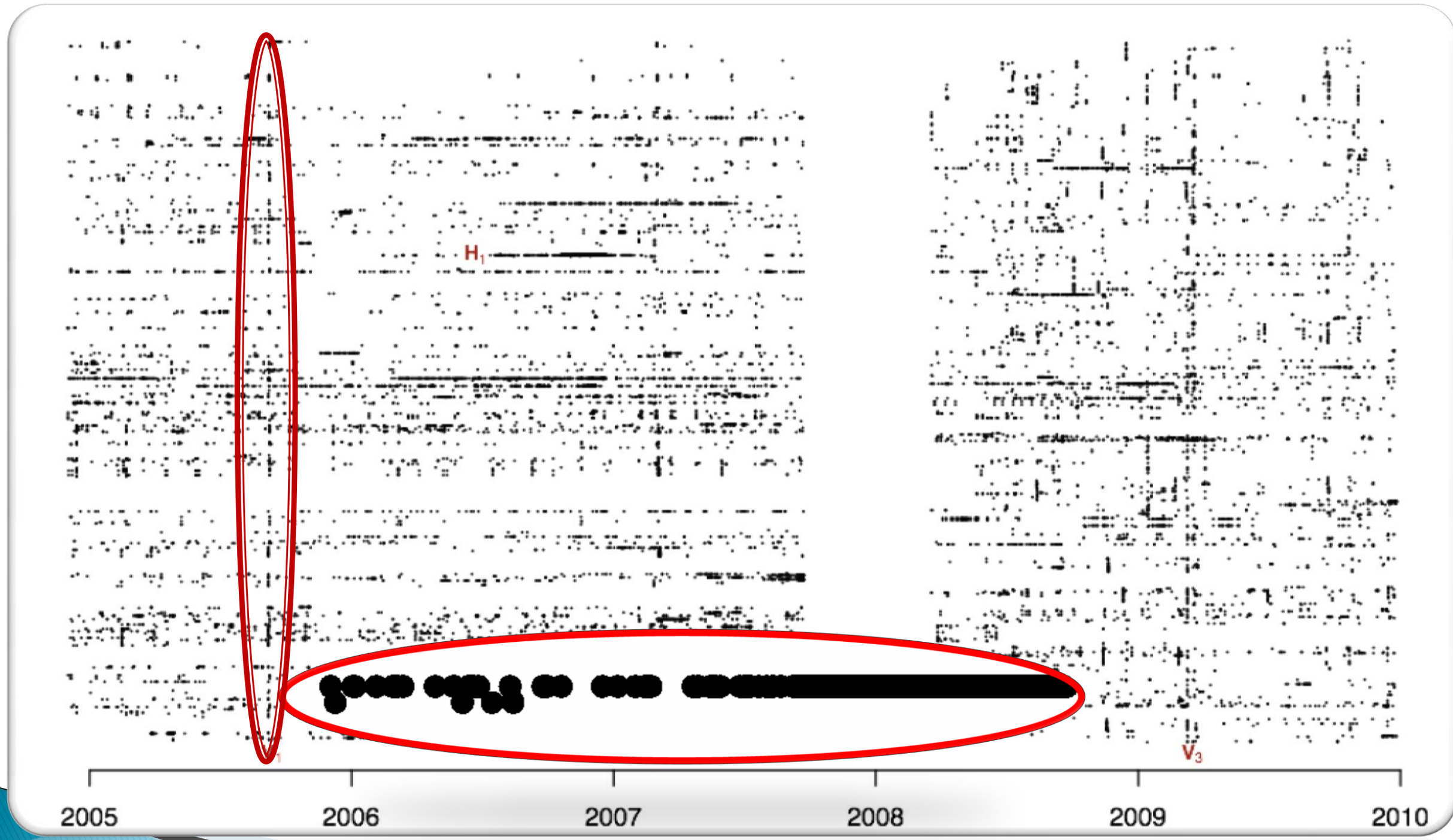
IMPACT: San Fransico PoP

COMMENTS: Other PoPs to follow

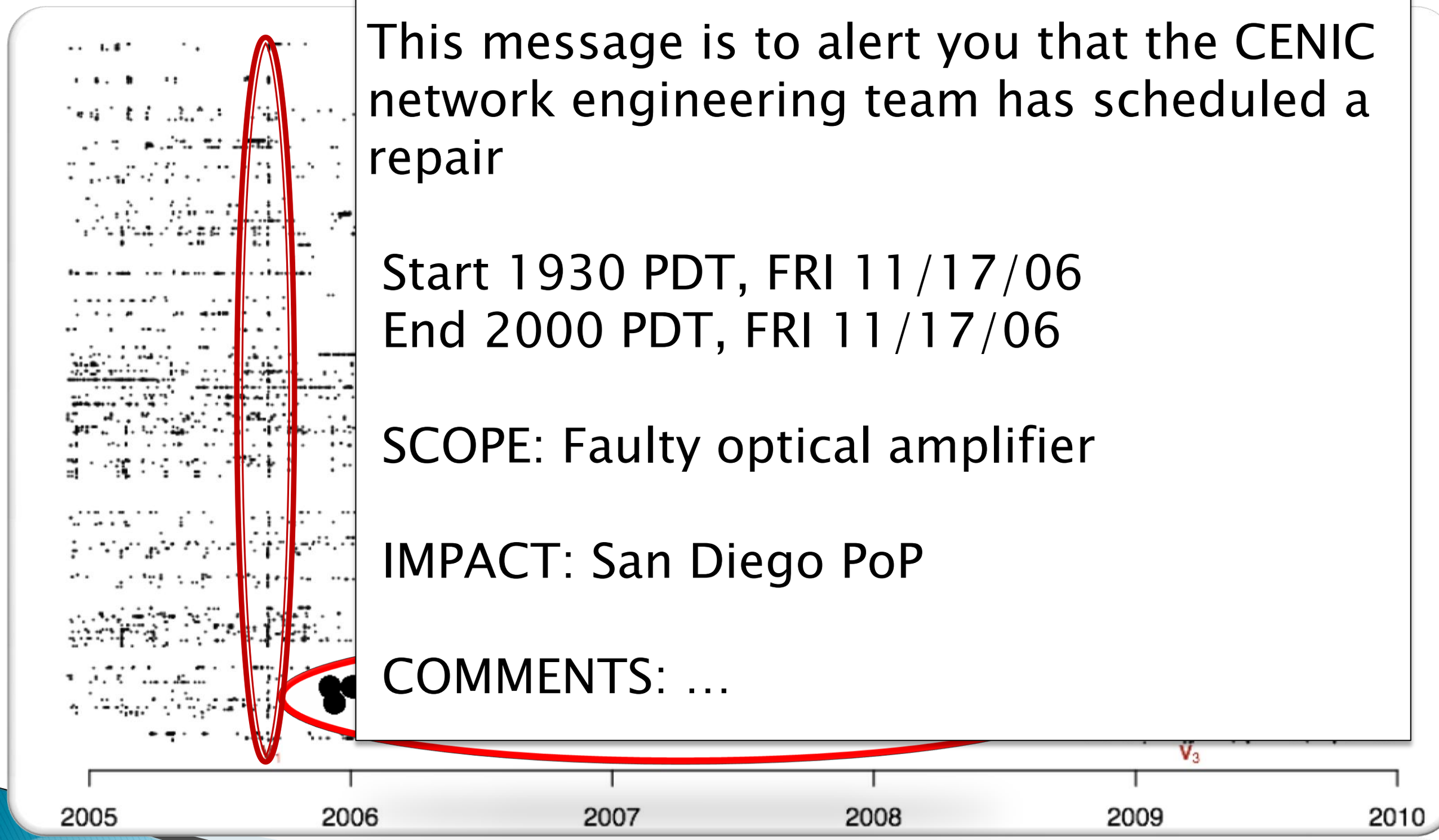
Failures in CENIC

Failures in CENIC

Failures in CENIC



Failures in CENIC



This message is to alert you that the CENIC network engineering team has scheduled a repair

Start 1930 PDT, FRI 11/17/06
End 2000 PDT, FRI 11/17/06

SCOPE: Faulty optical amplifier

IMPACT: San Diego PoP

COMMENTS: …

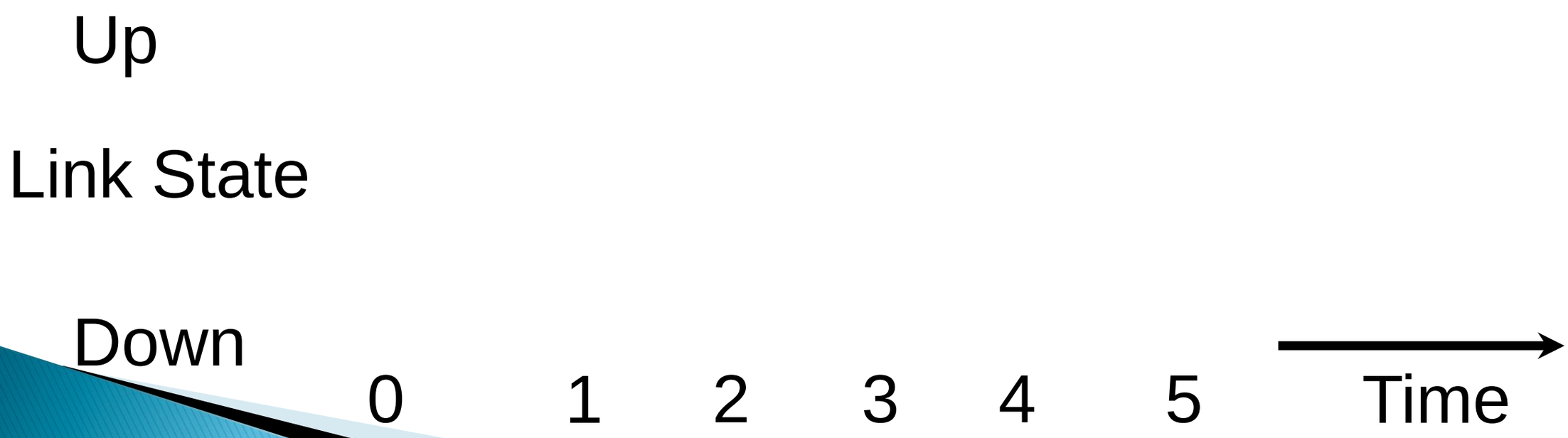
Outline

- Motivation
- Methodology
  - **Limitations**
  - Validation
- Findings in the CENIC network

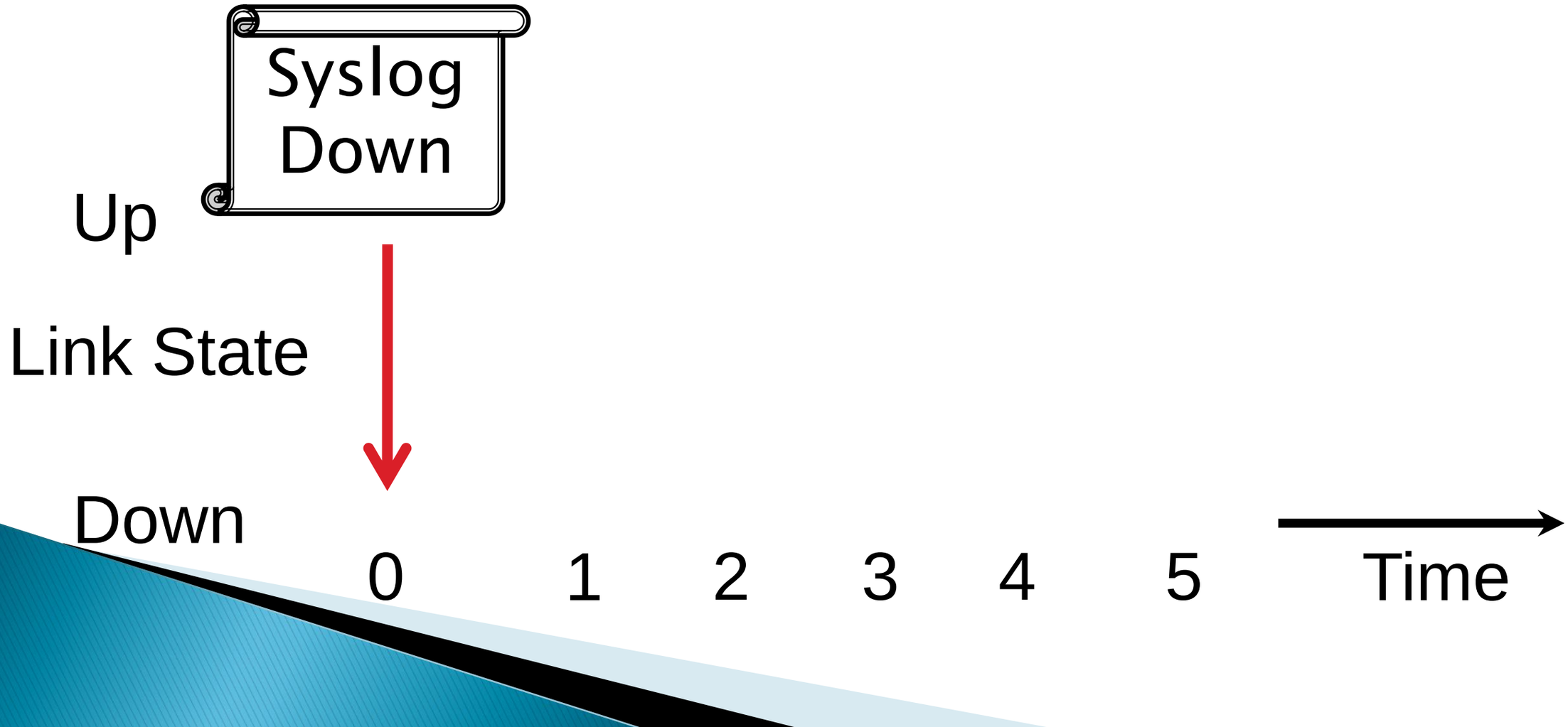
Syslog Limitations

- Syslog messages are sent from routers to a central server
  - Using UDP
- Messages are lost

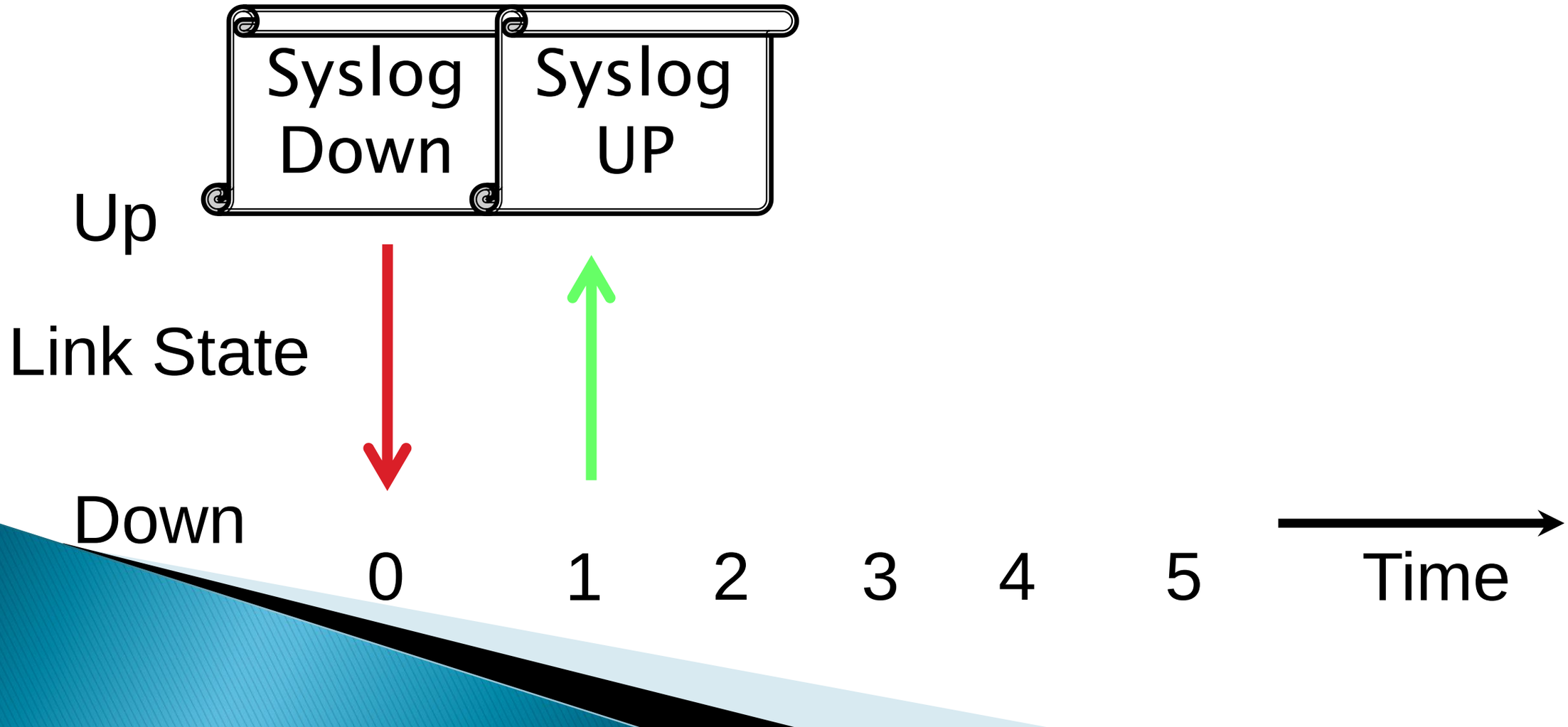
Reconstructing Link State



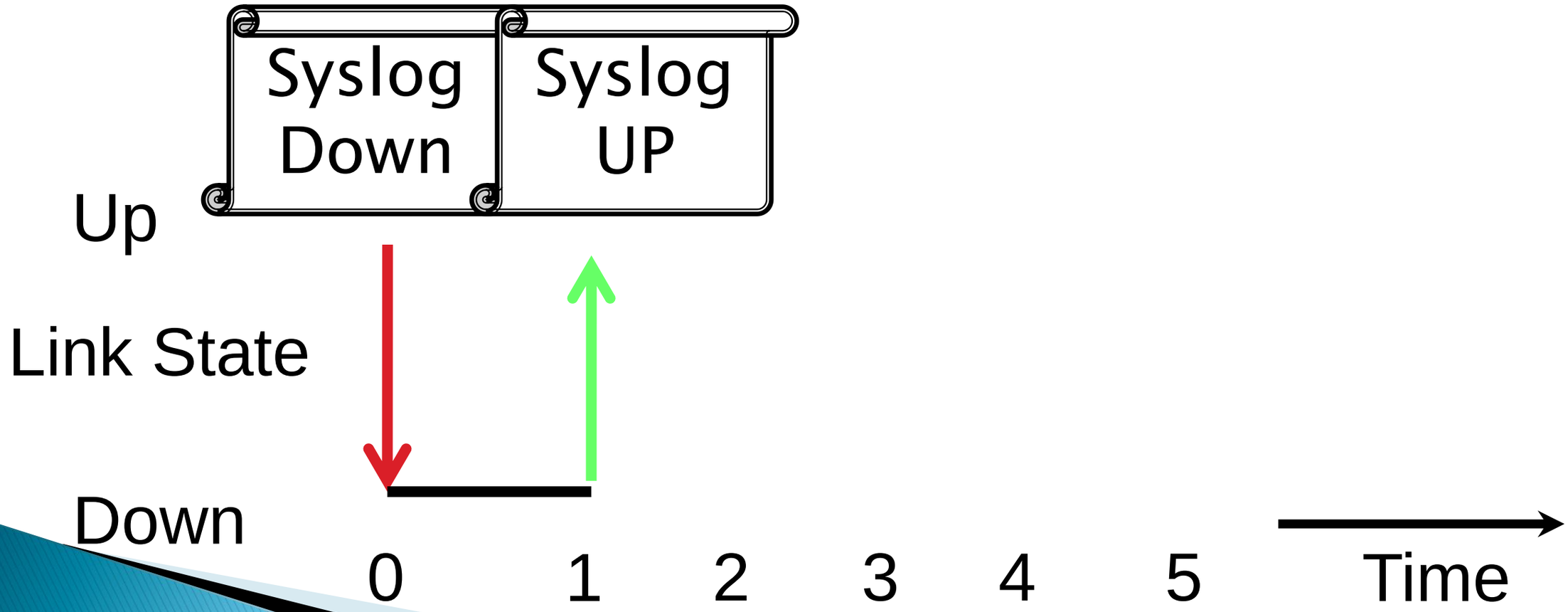
Reconstructing Link State



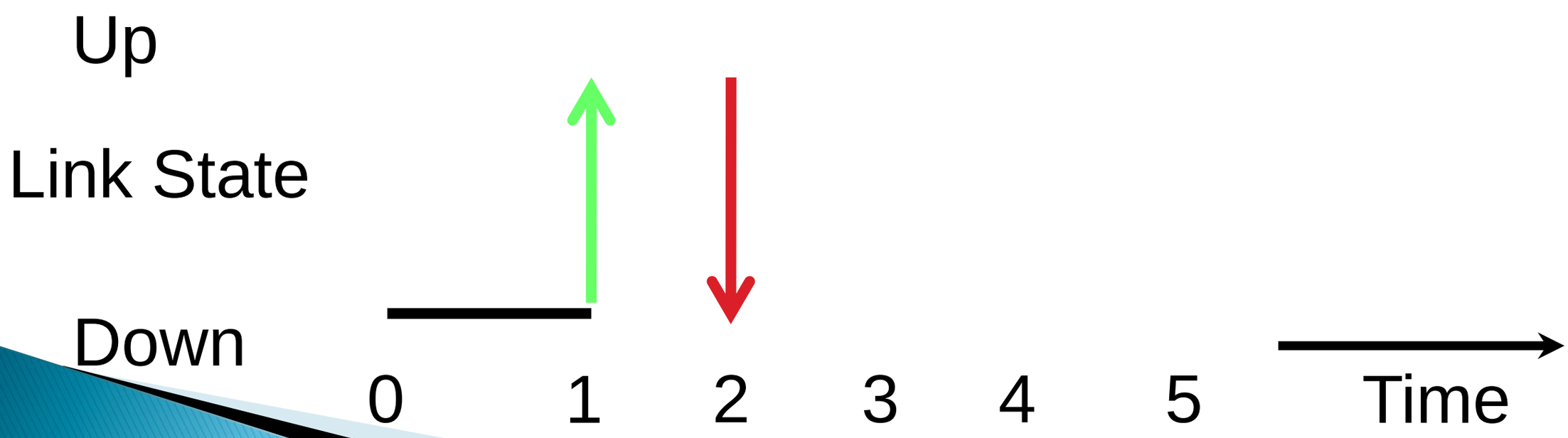
Reconstructing Link State



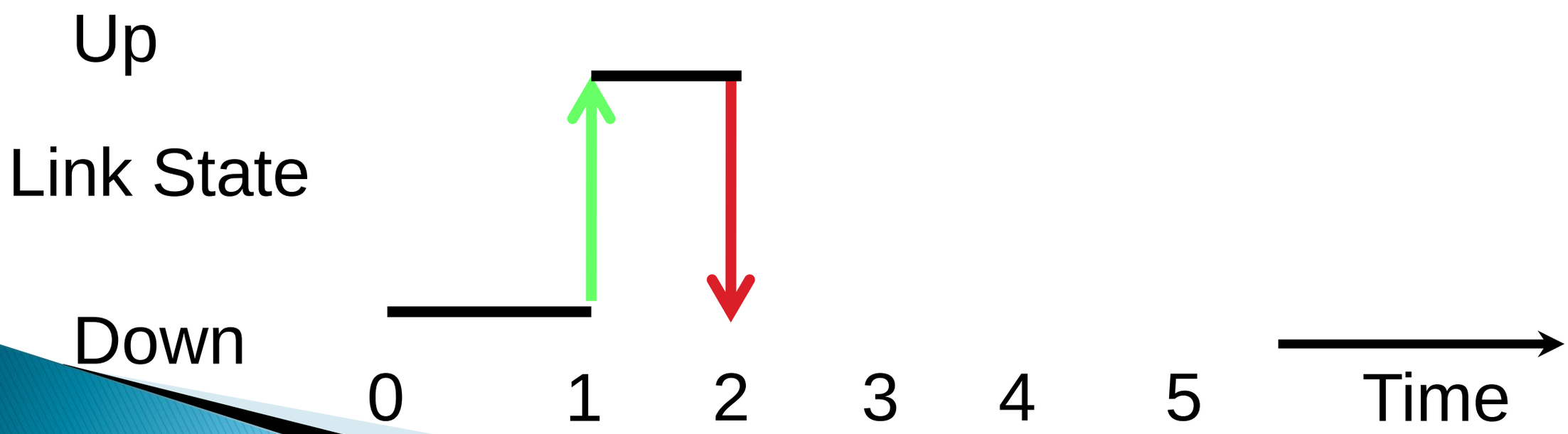
Reconstructing Link State



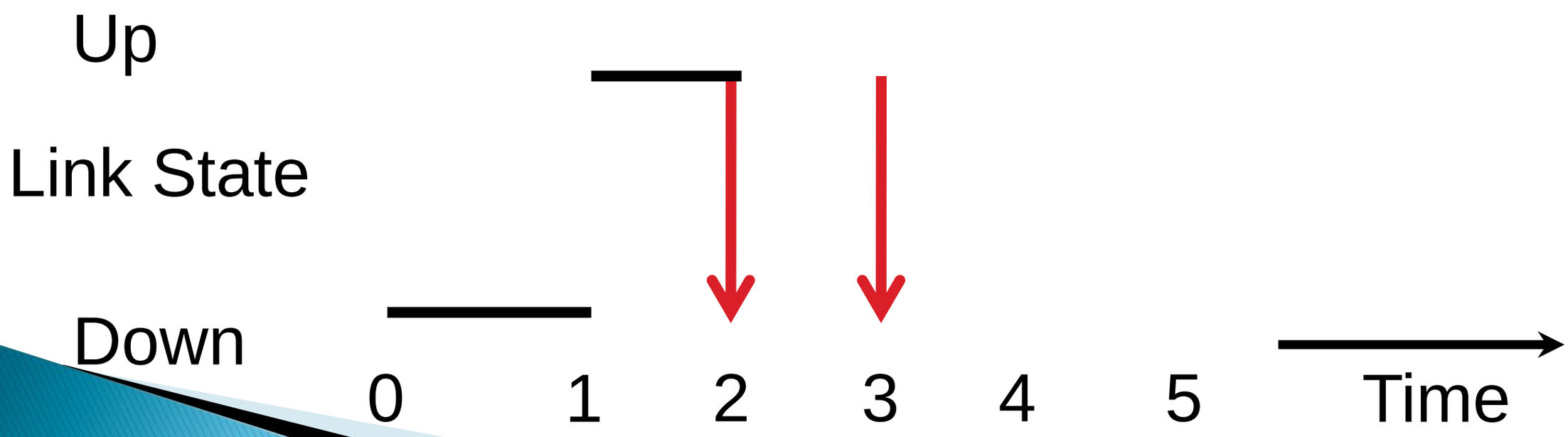
Reconstructing Link State



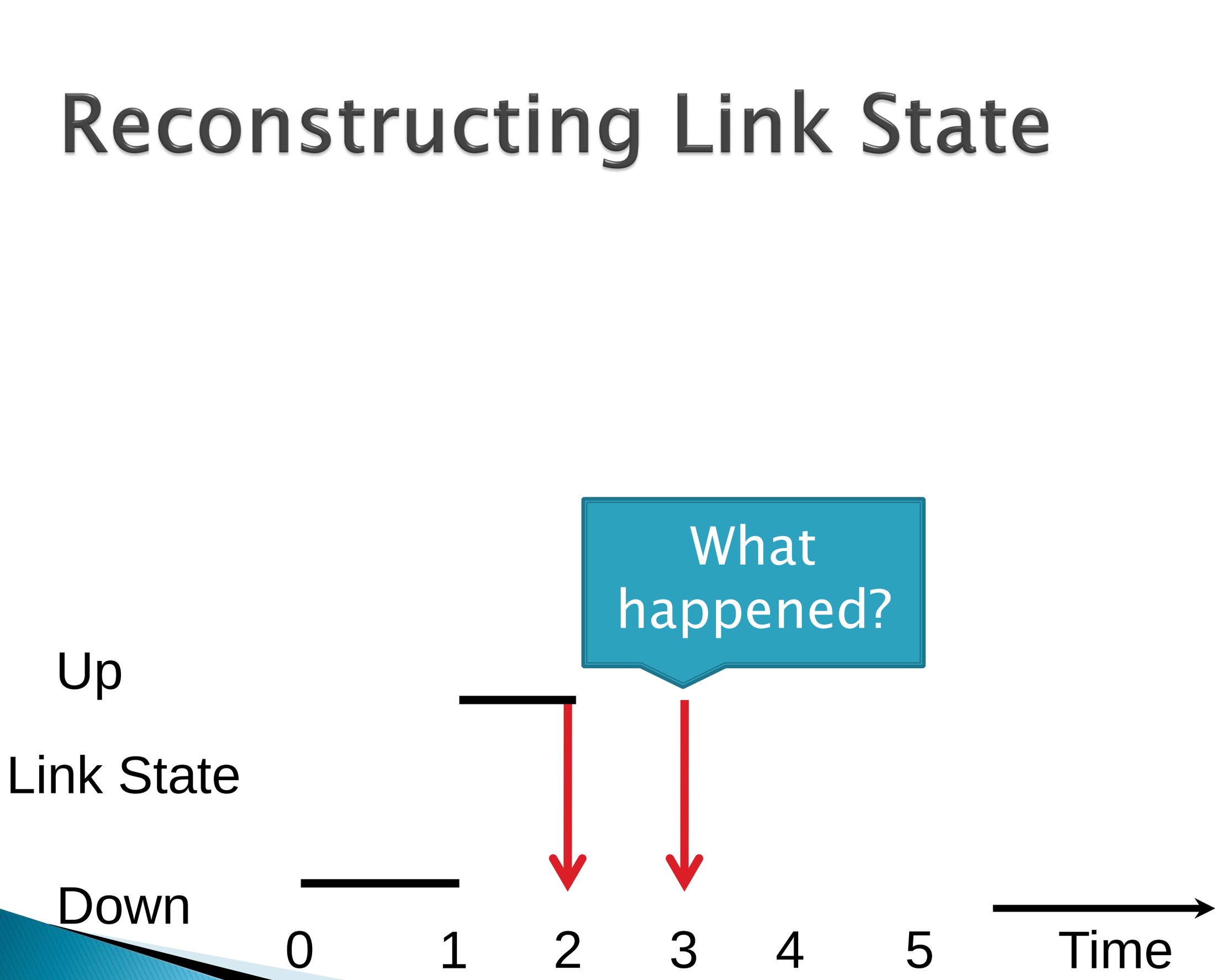
Reconstructing Link State



Reconstructing Link State



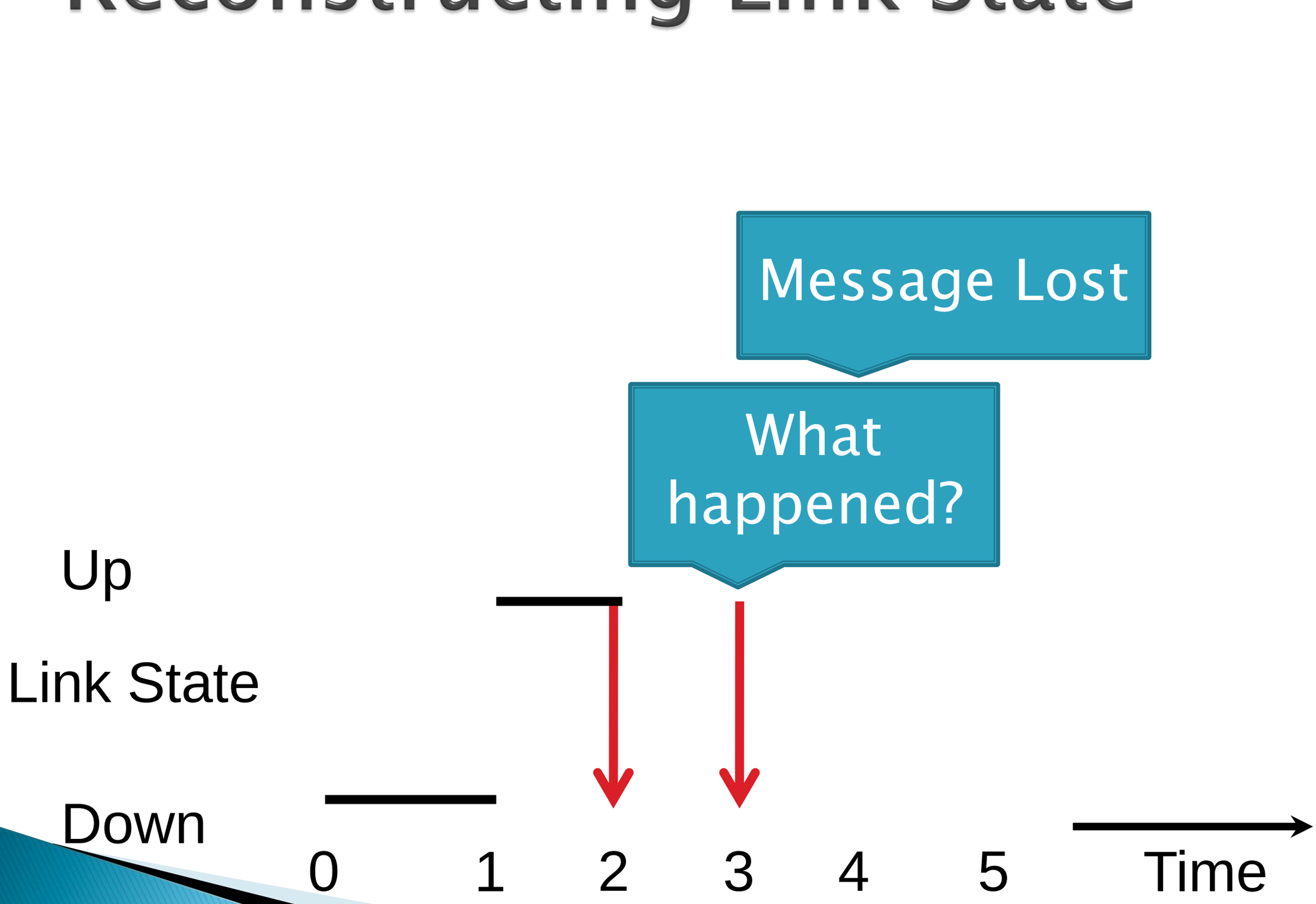
Reconstructing Link State



Reconstructing Link State

Message Lost

What happened?

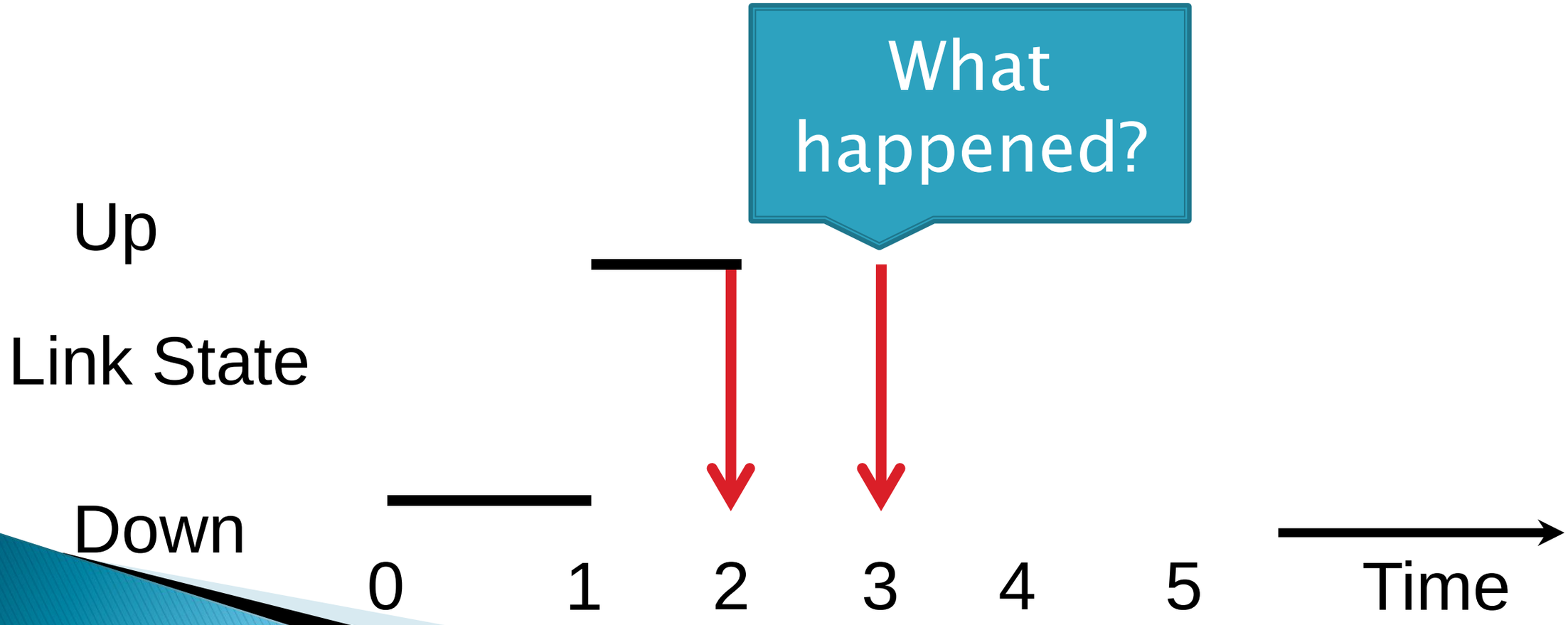


Reconstructing Link State

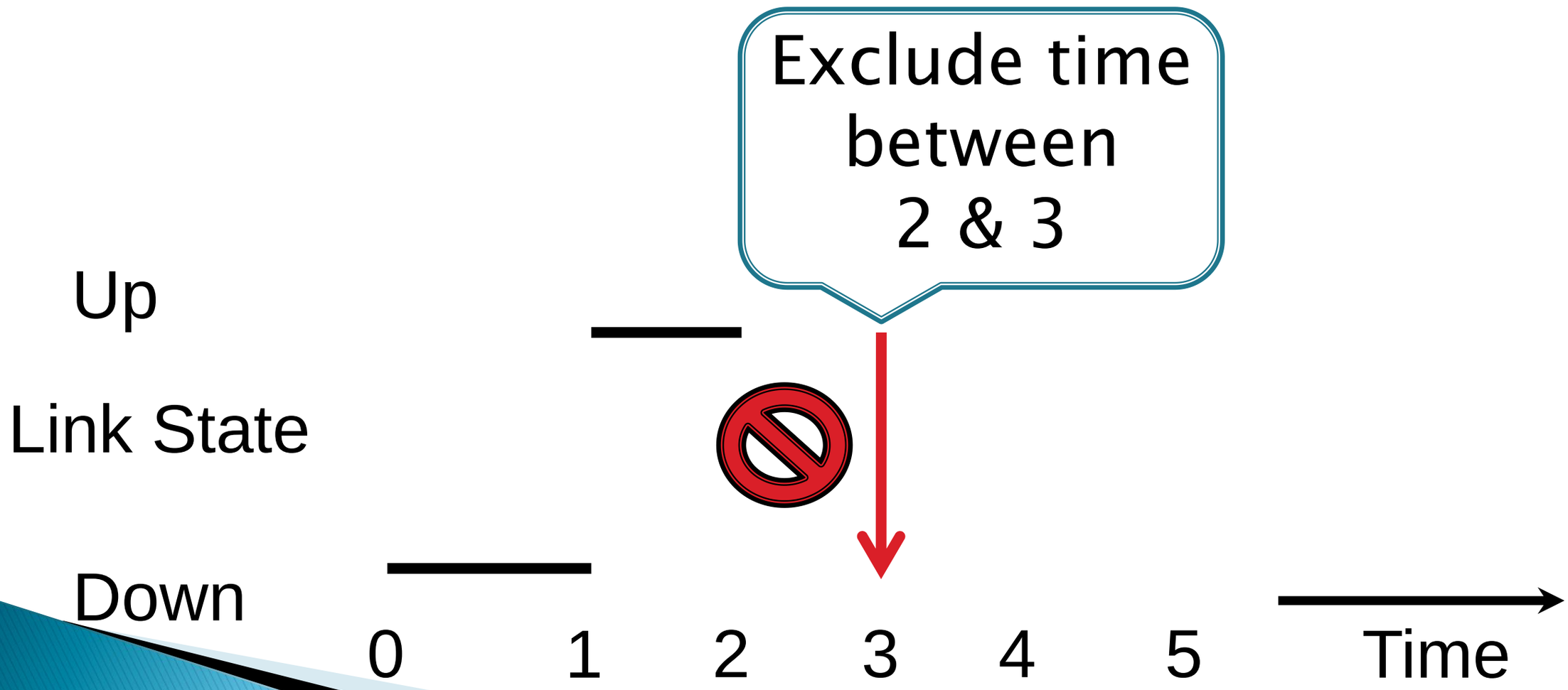
Spurious Message

Message Lost

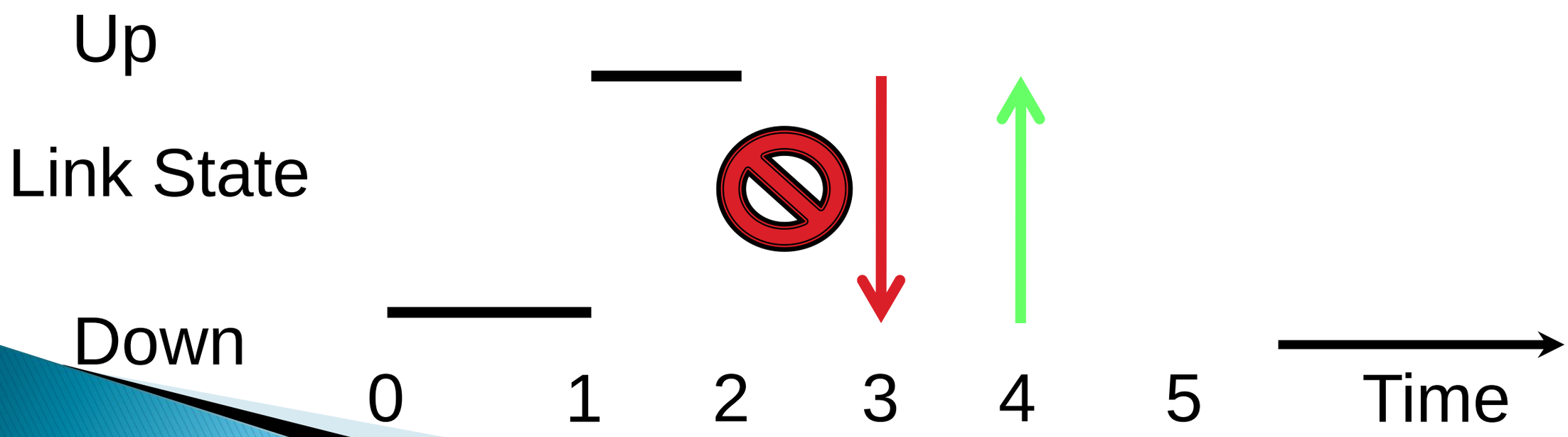
What happened?



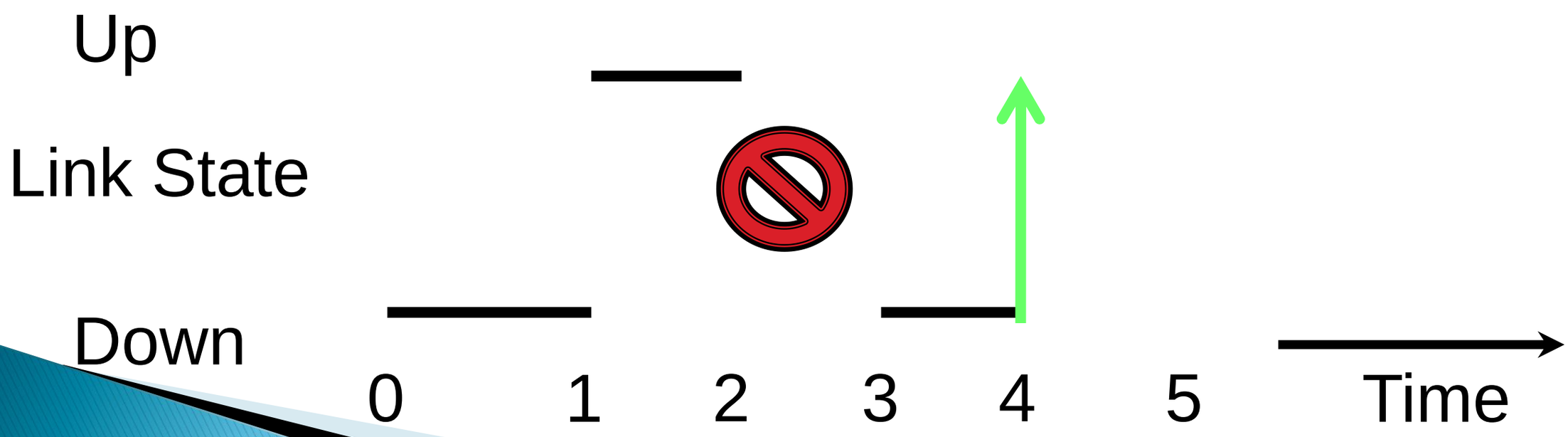
Reconstructing Link State



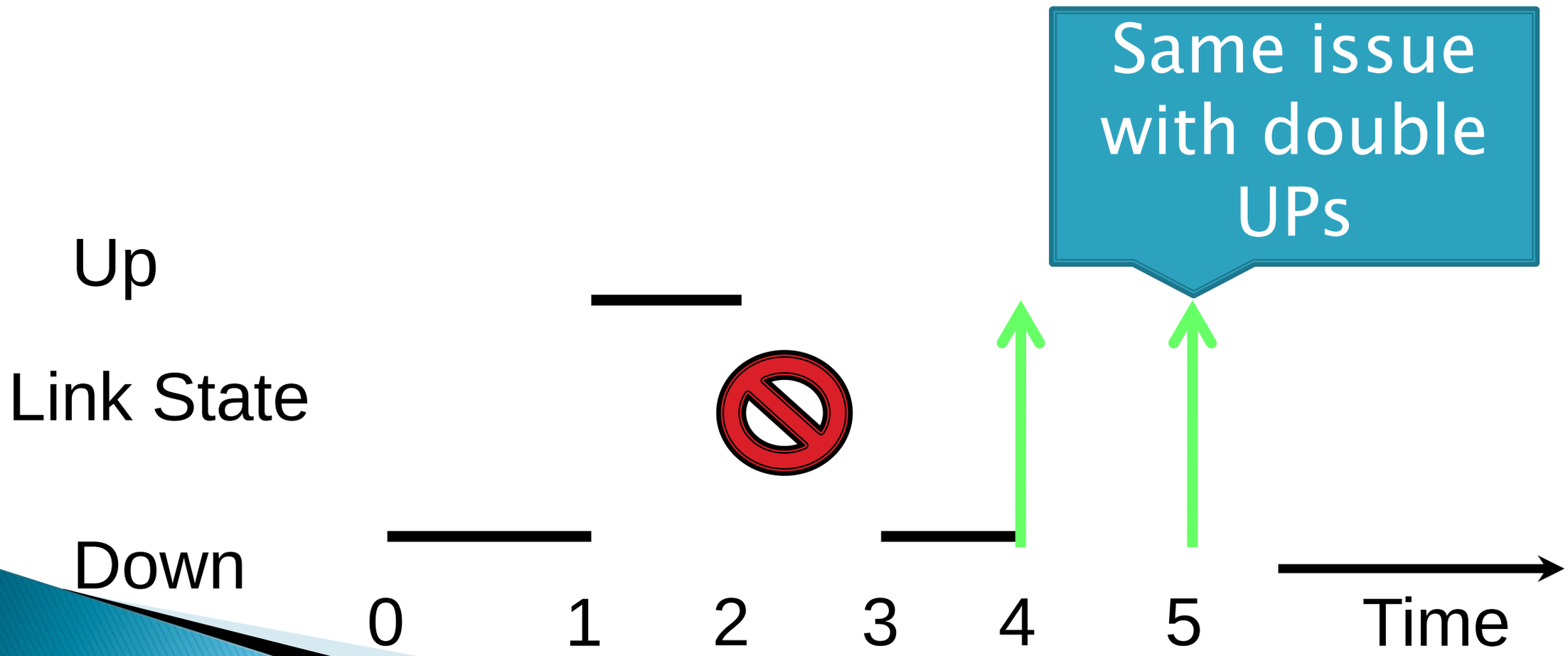
Reconstructing Link State



Reconstructing Link State



Reconstructing Link State



Configuration Limitations

- Configuration files are logged intermittently
- Configuration files do not describe layer 2 topology

Limits of Operational Announcements

- Operational announcements are written by humans
  - Selection bias
- Categorization is subjective

Testing Comprehensiveness

- Are there events mentioned in announcements that aren't in syslog
  - Manually checked random 1% of announcements
    - 97% of events were confirmed

Validating Up States

- How do we know syslog is accurate?
- CAIDA Skitter project (now Ark)
  - Traceroutes to every /24 on the Internet
  - 75 Million probes over 6 months traversed CENIC
    - confirmed no traffic over any interface that we thought was down

Validating Down States

- Can we verify links were down?
  - Routing protocols aim to mask failures
  - Isolation is externally visible
    - BGP updates are sent
- Route Views project records BGP traffic
  - Verified 105 out of 147 isolation events

Outline

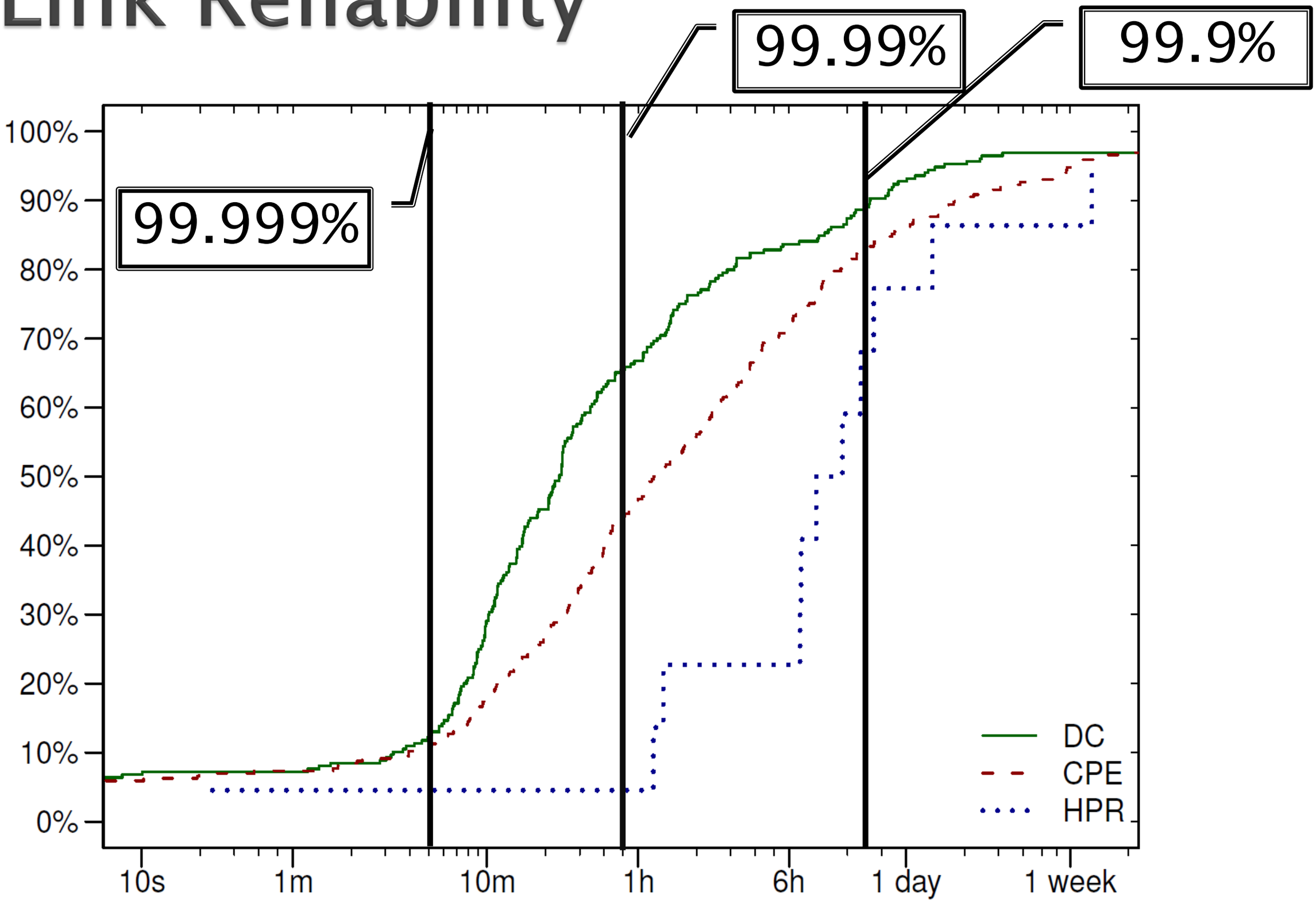
- Motivation
- Methodology
  - Limitations
  - Validation
- **Findings in the CENIC Network**

CENIC Network

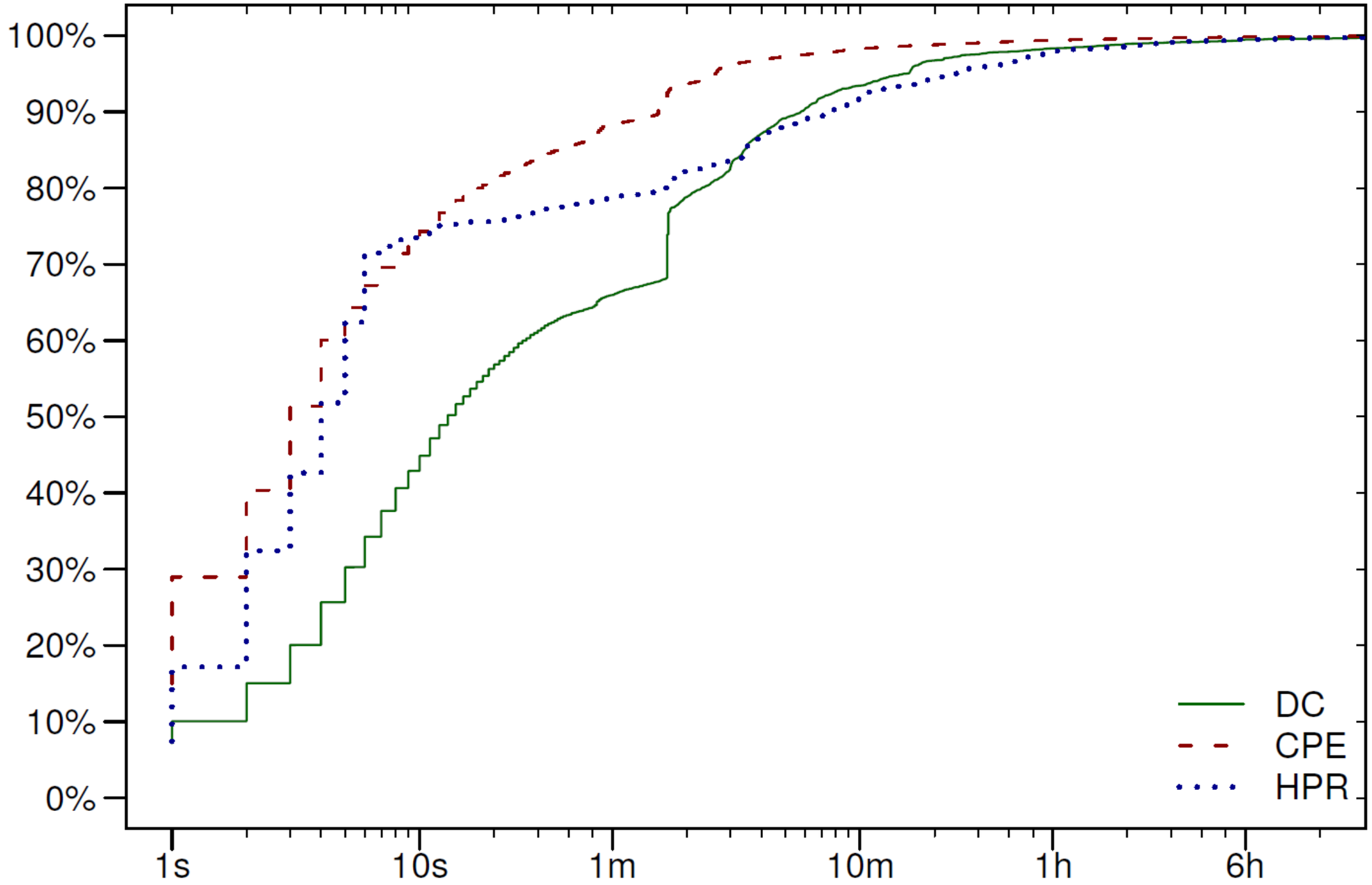
- Three Types of Links:
  - Backbone
  - Customer Access
  - High Performance Backbone

(backbone)

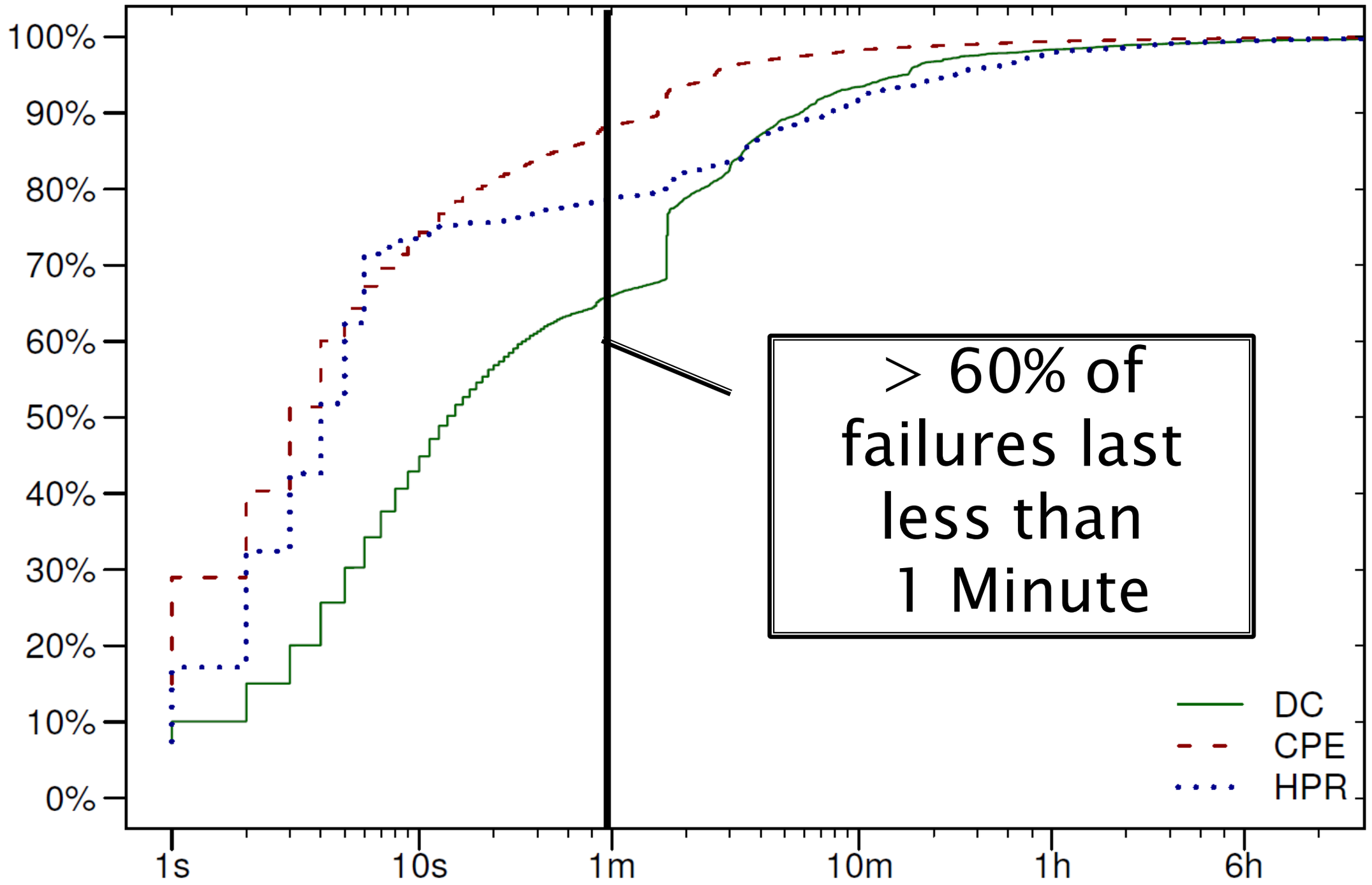
Link Reliability



Failure Duration



Failure Duration



Determining Causes of Failure

- 7,000 email announcements
- 3,000 events
- 28% of events describe a failure
- 18% of observed failures are explained

Cause of Failure

Cause	Notices	Failures
Hardware	25%	20%
Power	20%	6%
External	15%	15%
Software	12%	32%
Other	20%	18%
Configuration	8%	9%

Cause of Failure

Cause		
Hardwa		
Power		
External		
Softwar		
Other	20%	18%
Configuration	8%	9%

Other

* Machine room flooded
* DoS attack
* Construction crews demolished a manhole with active cables
* Or unsolved

Cause of Failure

Cause	Notices	Failures
Hardware	25%	20%
Power	20%	6%
External	15%	15%
Software	12%	32%
Other	20%	18%
Configuration	8%	9%

Cause of Failure

Cause	Notices	Failures
Hardware	25%	20%
Power	20%	6%
External	15%	15%
Software	12%	32%
Other	20%	18%
Configuration	8%	9%

Event Durations

		Time to repair	
Cause	Events	Avg	Med
Hardware	20%	95 m	5 m
Power	6%	93 m	18 m
External	15%	61 m	4.6 m
Software	32%	10 m	4 m
Configuration	9%	5 m	1 m
Other	12%	46 m	6 m
Unknown	5%	52 m	6 m

Event Durations

Cause	Events	Time to repair	
		Avg	Med
Hardware	20%	95 m	5 m
Power	6%	93 m	18 m
External	15%	61 m	4.6 m
Software	32%	10 m	4 m
Configuration	9%	5 m	1 m
Other	12%	46 m	6 m
Unknown	5%	52 m	6 m

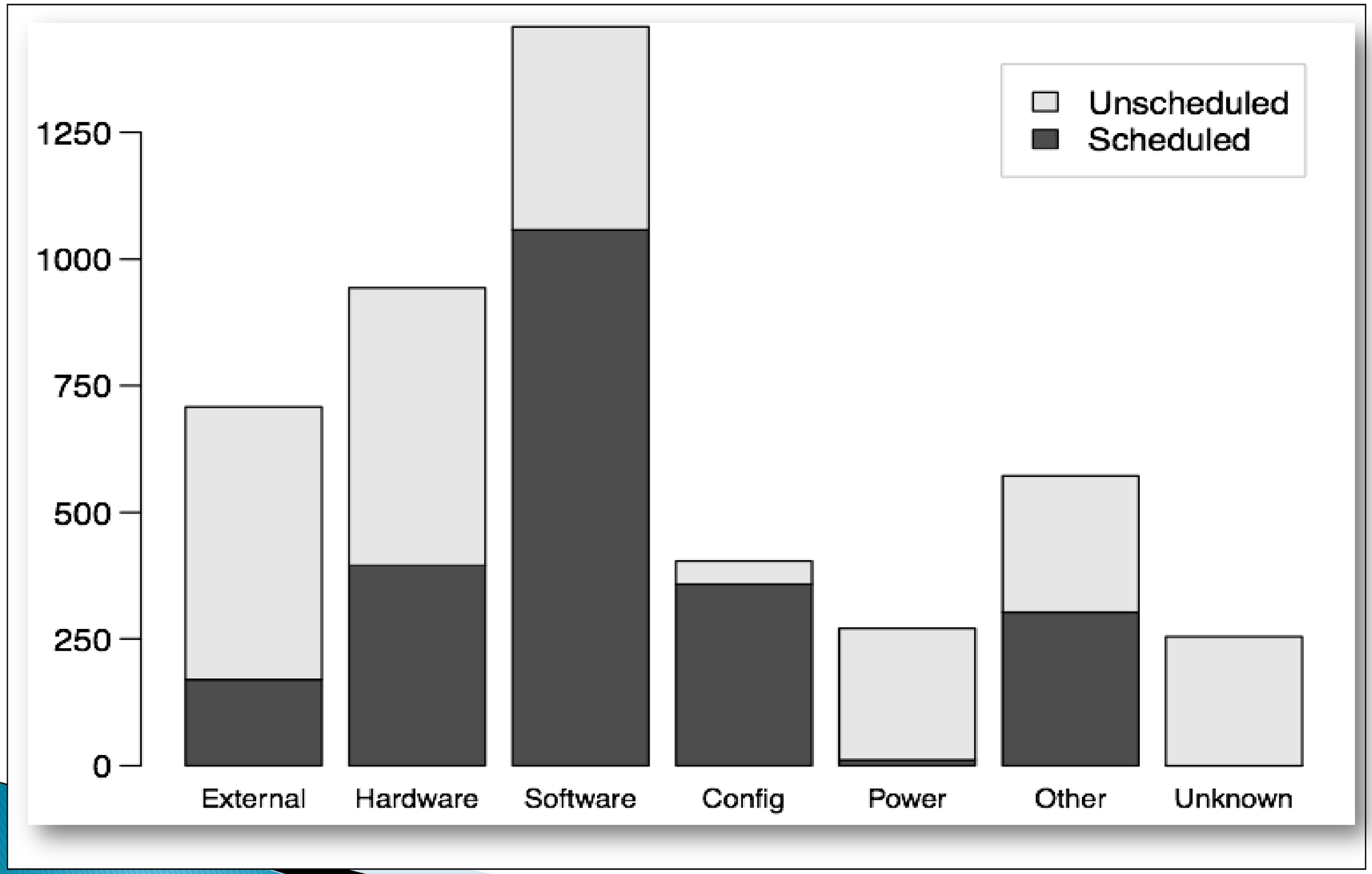
Event Durations

Cause	Events	Time to repair	
		Avg	Med
Hardware	20%	95 m	5 m
Power	6%	93 m	18 m
External	15%	61 m	4.6 m
Software	32%	10 m	4 m
Configuration	9%	5 m	1 m
Other	12%	46 m	6 m
Unknown	5%	52 m	6 m

Considering Failure Impact

- Not all downtime is equal
  - Some failures are unexpected
    - Fiber cuts
  - Some failures are scheduled
    - Software upgrades

Considering Failure Impact



Impact

- Scheduled vs. Unscheduled
  - Simple metric to evaluate impact
- Difficult to gauge impact of most failures
  - Only 18% of failures are covered by an email
- Customer isolation events have a clear impact
  - Recall, BGP traffic makes these easy to spot

Impact

		Downtime duration		
Cause	# of Events	Avg	Med	95%
Power	10	5 h	20.6 m	33 h
Hardware	13	8.2 h	32 m	3.7 d
Software	8	6 m	2.7 m	13.9 m
N/A	75	8 h	32 m	3.7 d

Impact

		Downtime duration		
Cause	# of Events	Avg	Med	95%
Power	10	5 h	20.6 m	33 h
Hardware	13	8.2 h	32 m	3.7 d
Software	8	6 m	2.7 m	13.9 m
N/A	75	8 h	32 m	3.7 d

Conclusion

- Engineering for failure requires real data
  - Data has historically been difficult to obtain
- Methodology to perform historical failure analysis with low-quality data sources
- Shared our findings in the CENIC network
  - Reliability of individual components
  - Causes of failures
  - Impact of failures

UCSDCSE
Computer Science and Engineering