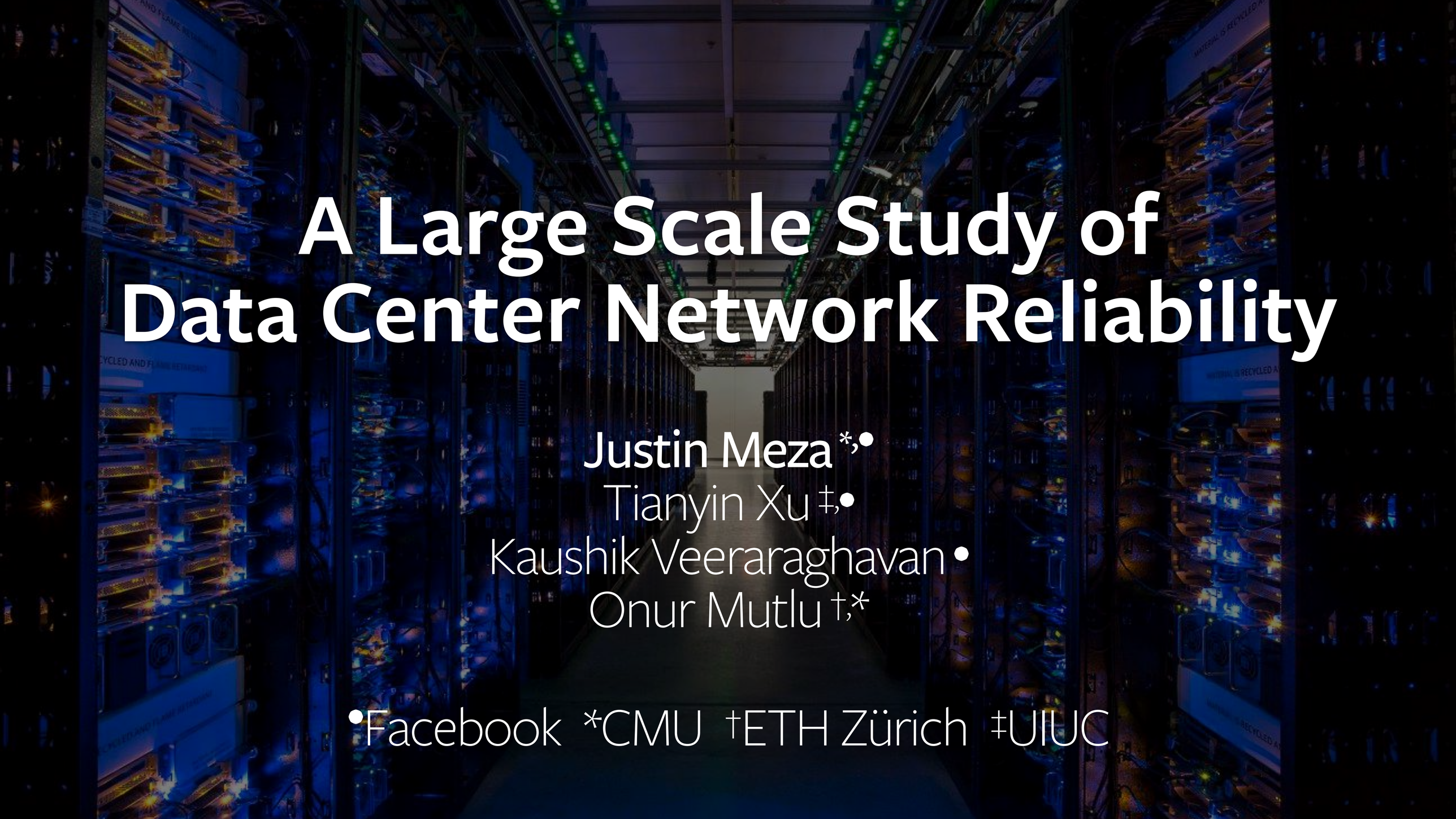
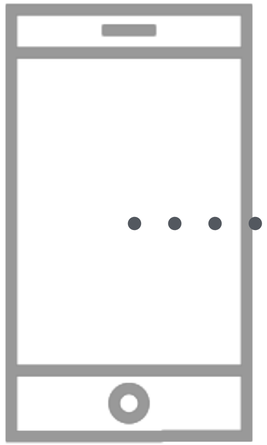
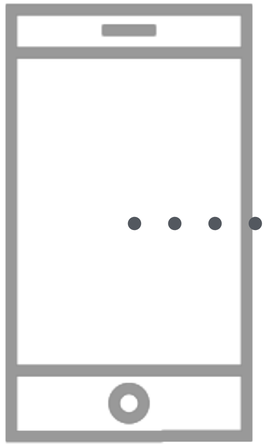


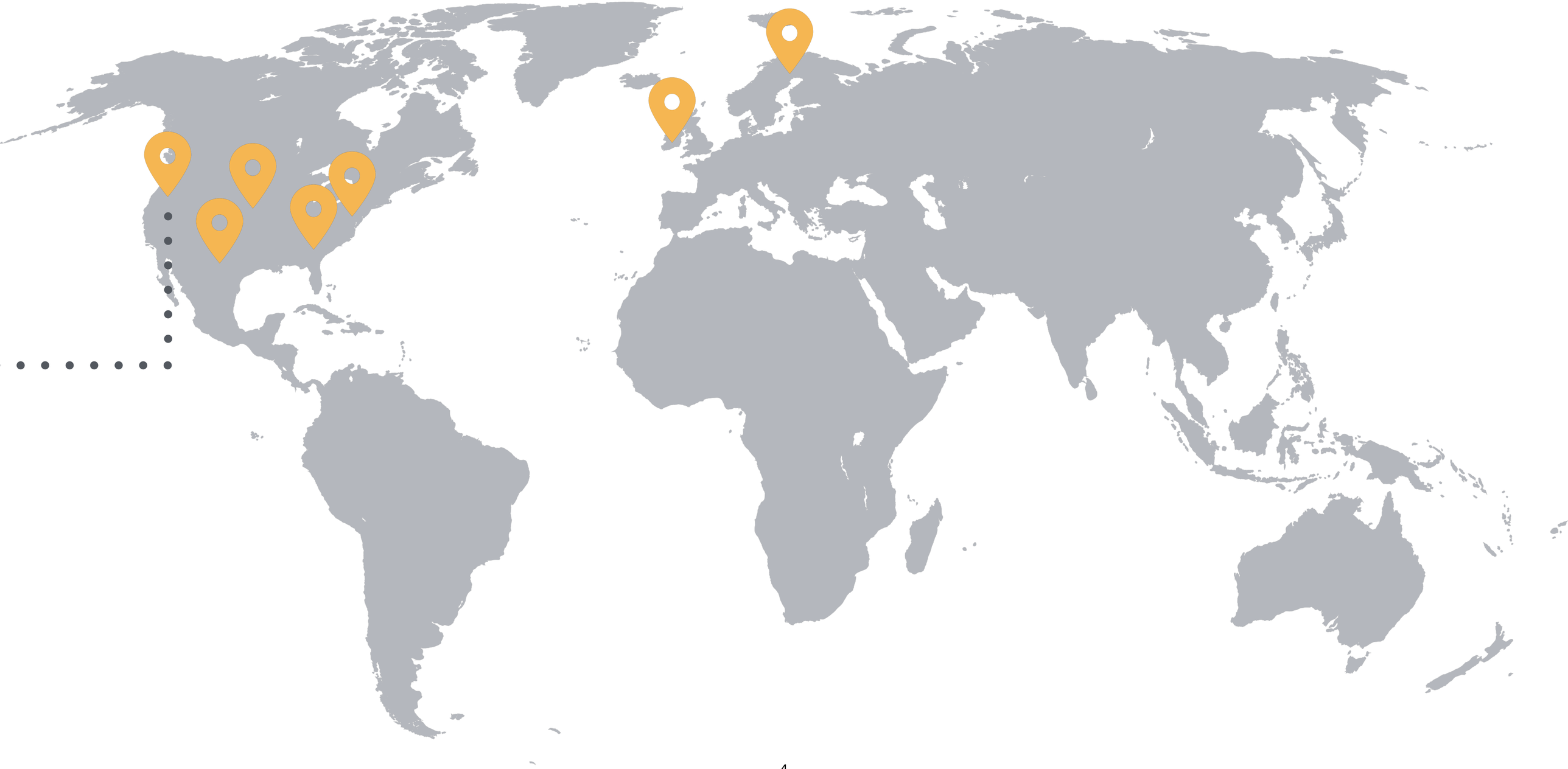
A Large Scale Study of Data Center Network Reliability

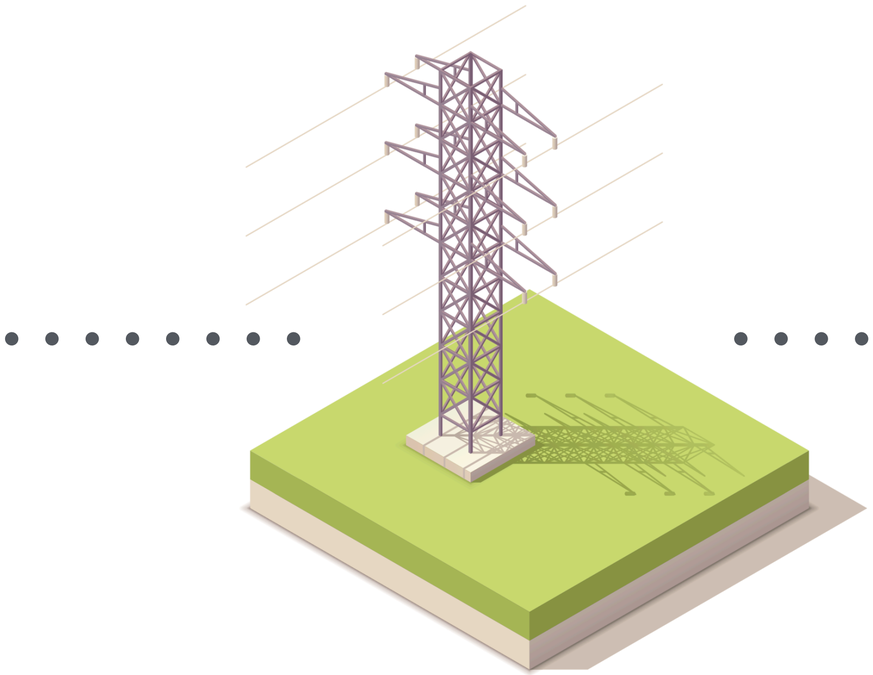
Justin Meza [*,•]
Tianyin Xu [‡,•]
Kaushik Veeraraghavan [•]
Onur Mutlu [†,*]

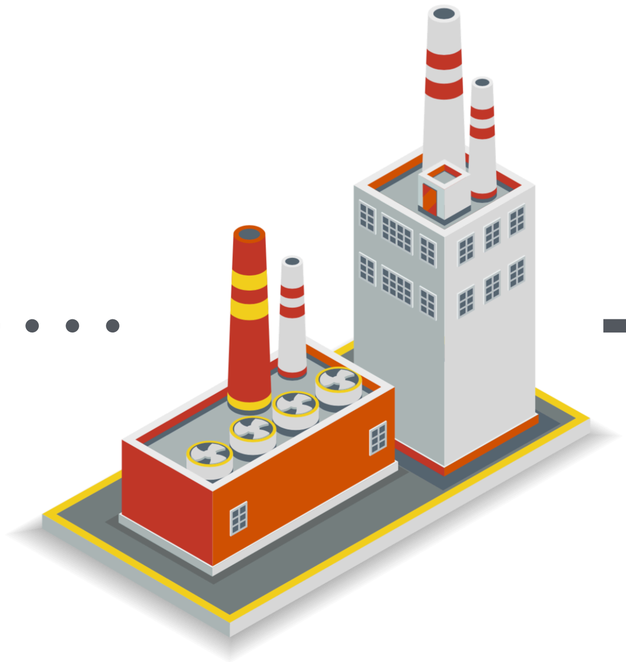
[•]Facebook [*]CMU [†]ETH Zürich [‡]UIUC









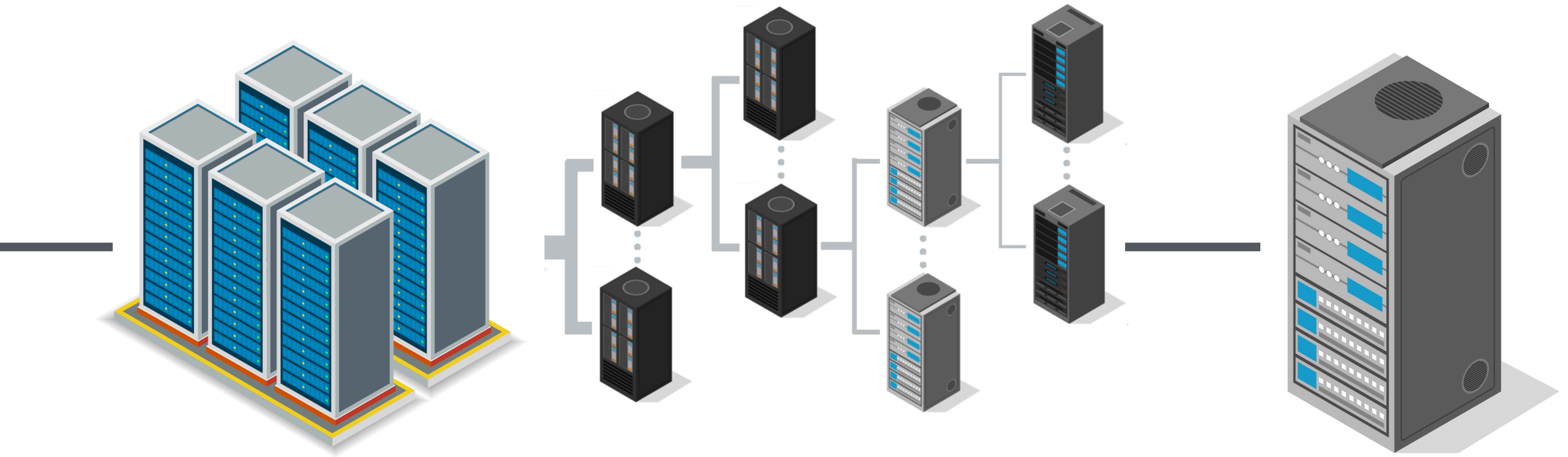


WAN

Internet

ISP

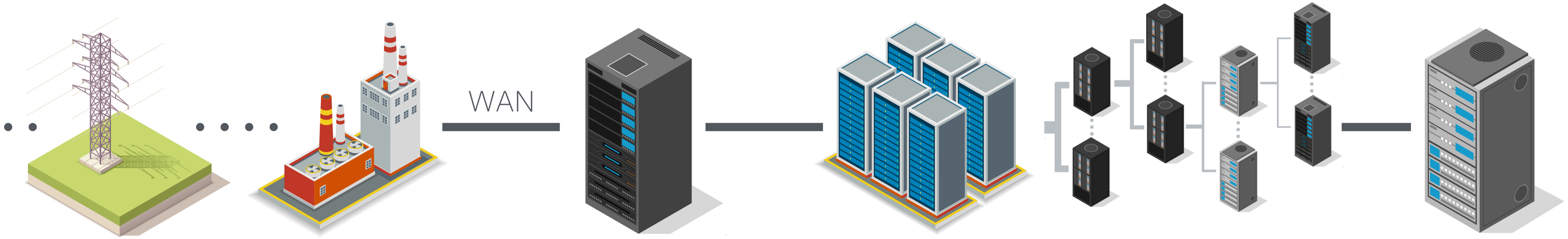
Edge Node



Core Switches

Data Center Fabric

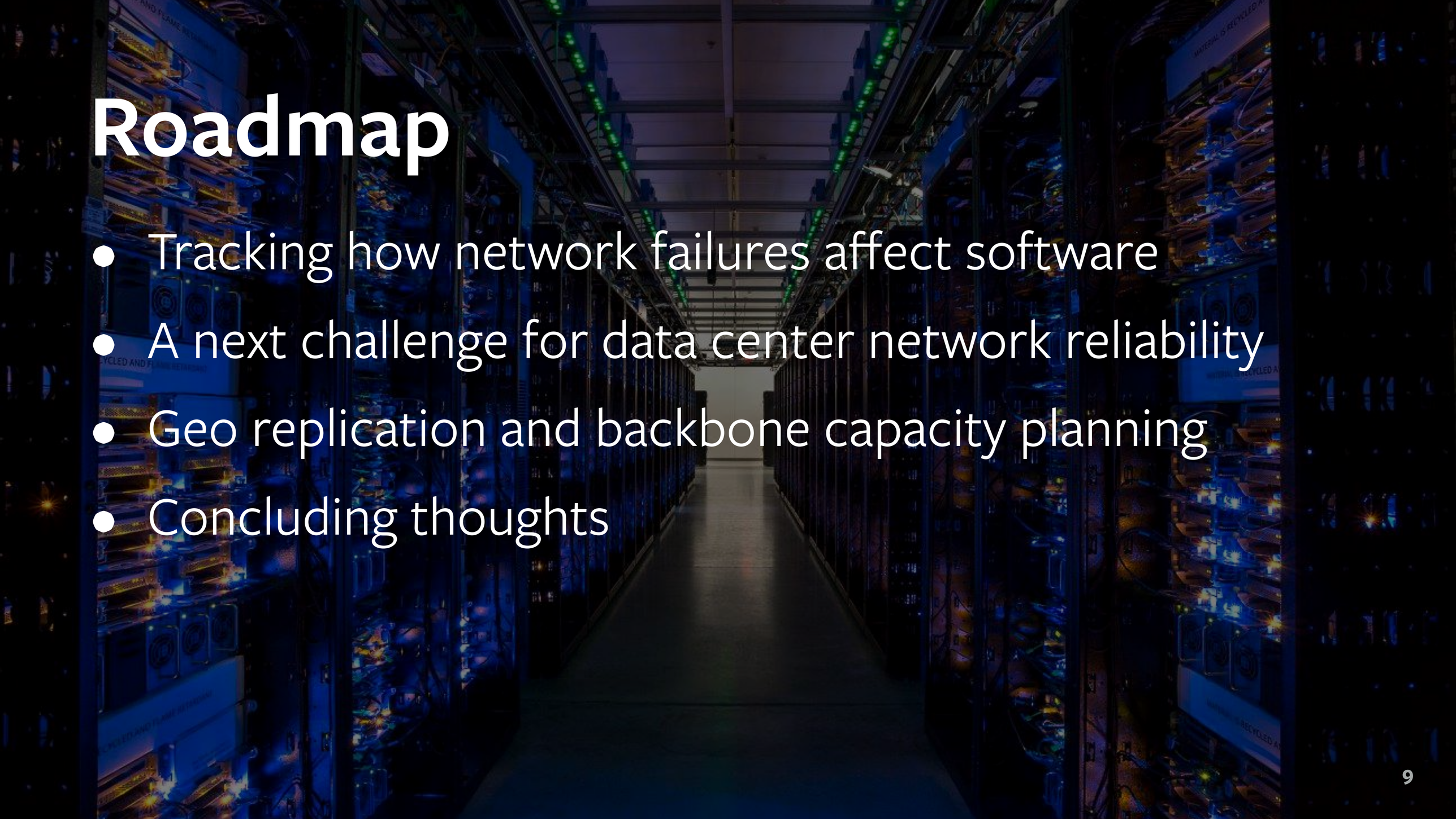
Top of Rack Switch



Internet ISP Edge Node Core Switches Data Center Fabric Rack Switch

Key takeaways

- As DC networks become more software managed, the next challenge is first and last hop reliability
- Growth in geo replicated software drives the need for reliable backbone network capacity planning



Roadmap

- Tracking how network failures affect software
- A next challenge for data center network reliability
- Geo replication and backbone capacity planning
- Concluding thoughts

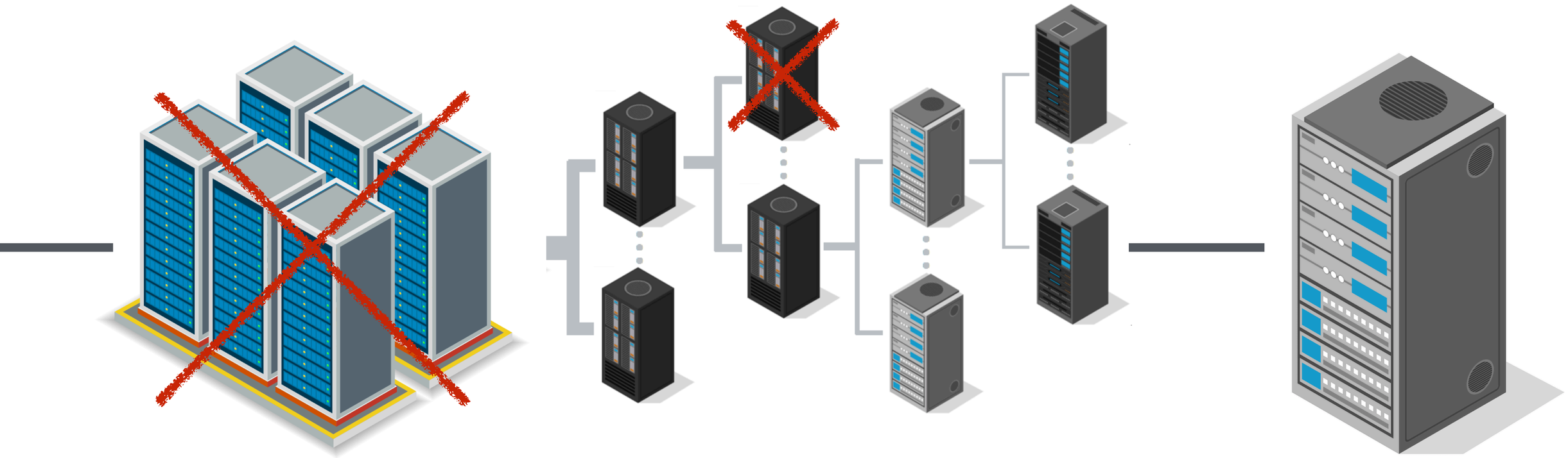
Network Incidents

cause

Software Failures

that result in

Site Events (SEVs)



Core Switches

Data Center Fabric

Top of Rack Switch

Automated repair

Device	Repair Ratio	Avg Priority / Wait / Repair Time
Core	75%	0 (highest priority) / 4 m / 30.1 s
FSW	99.5%	2.25 / 3 d / 4.45 s
RSW	99.7%	2.22 / 1 d / 2.91 s

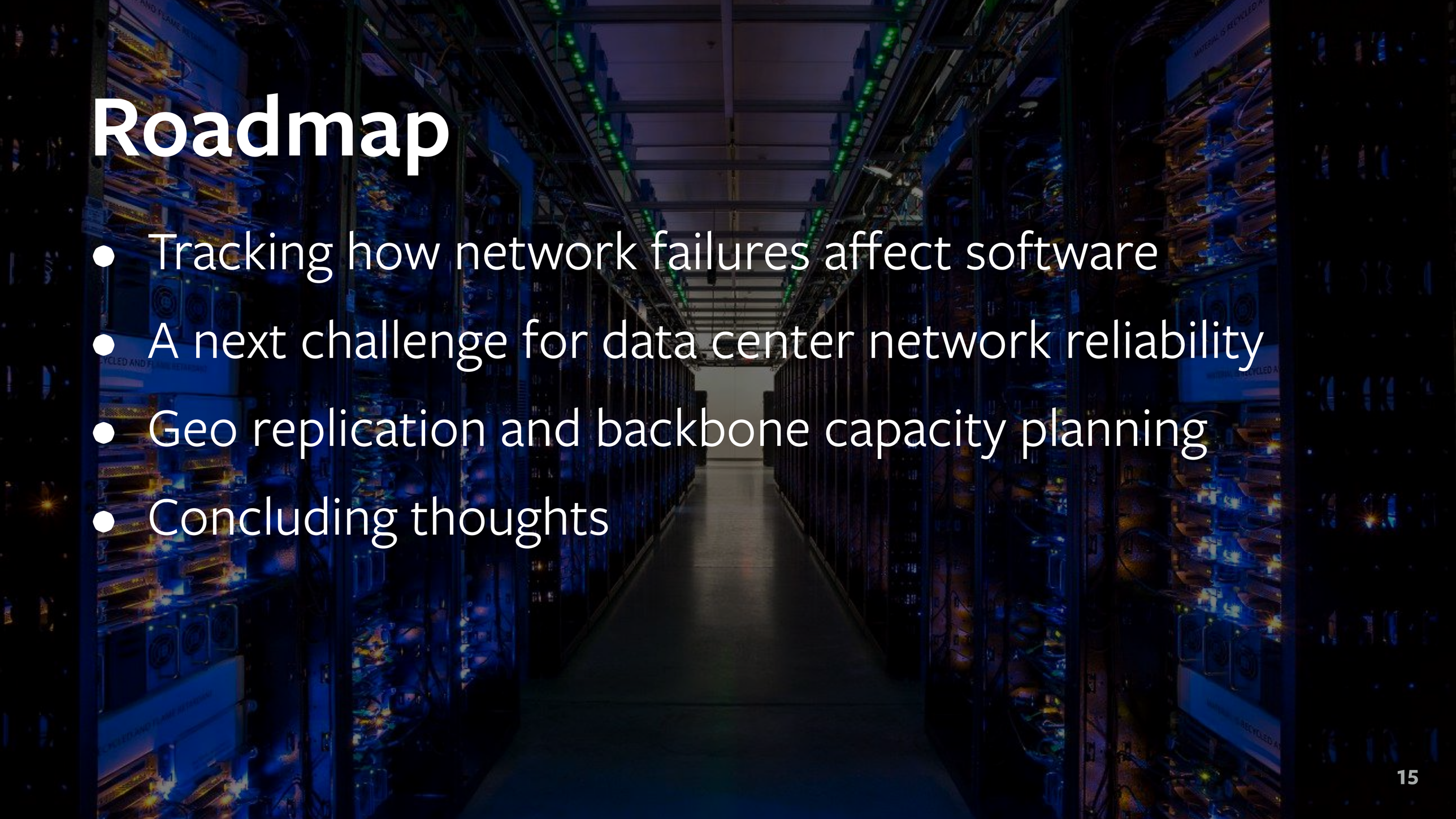
- Not deployed on all devices, but highly effective
- Top fixes: port failures, config files, switch ping failures

SEV reports

- Filled out by engineers who fixed the problem
- Contain metadata (switch ID, switch type, ...)
- Classified based on severity
- Continually audited for accuracy and completeness

Network incidents, not events

- Want to know software impact and severity
- Events alone don’t provide enough context
- Often masked by redundancy and automated repairs
- We examine a different class of failures
  - Software failures resulting from network events
  - Top-line impact on software reliability

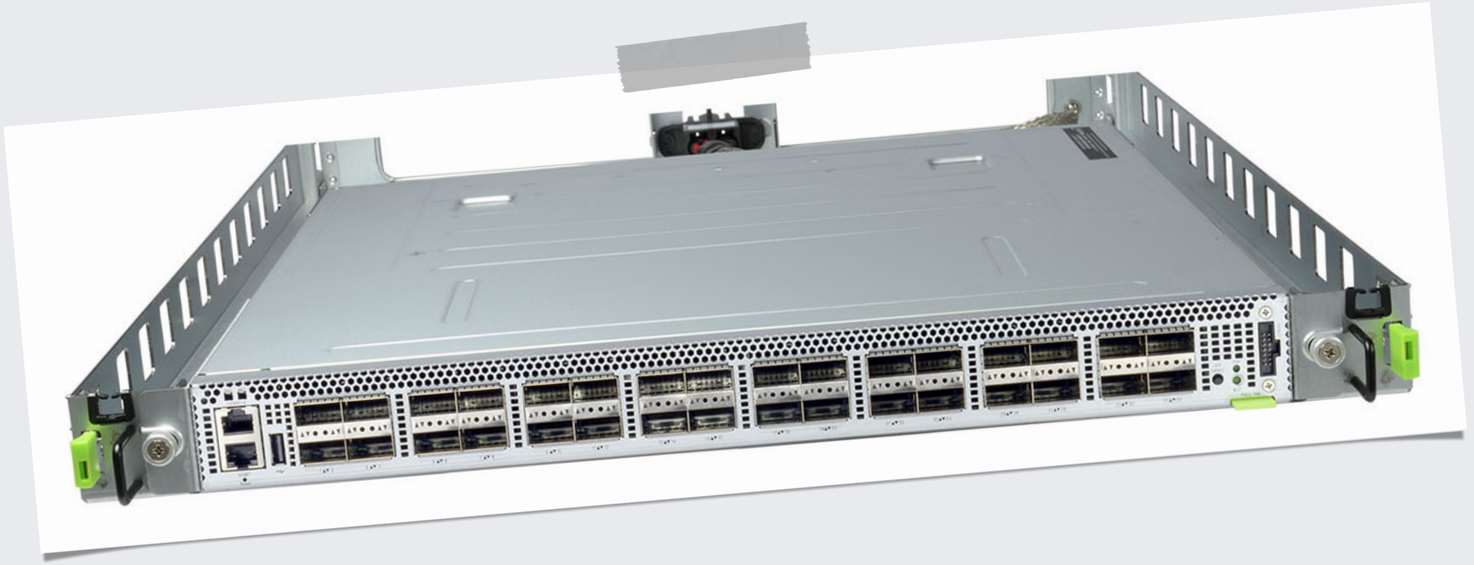


Roadmap

- Tracking how network failures affect software
- A next challenge for data center network reliability
- Geo replication and backbone capacity planning
- Concluding thoughts

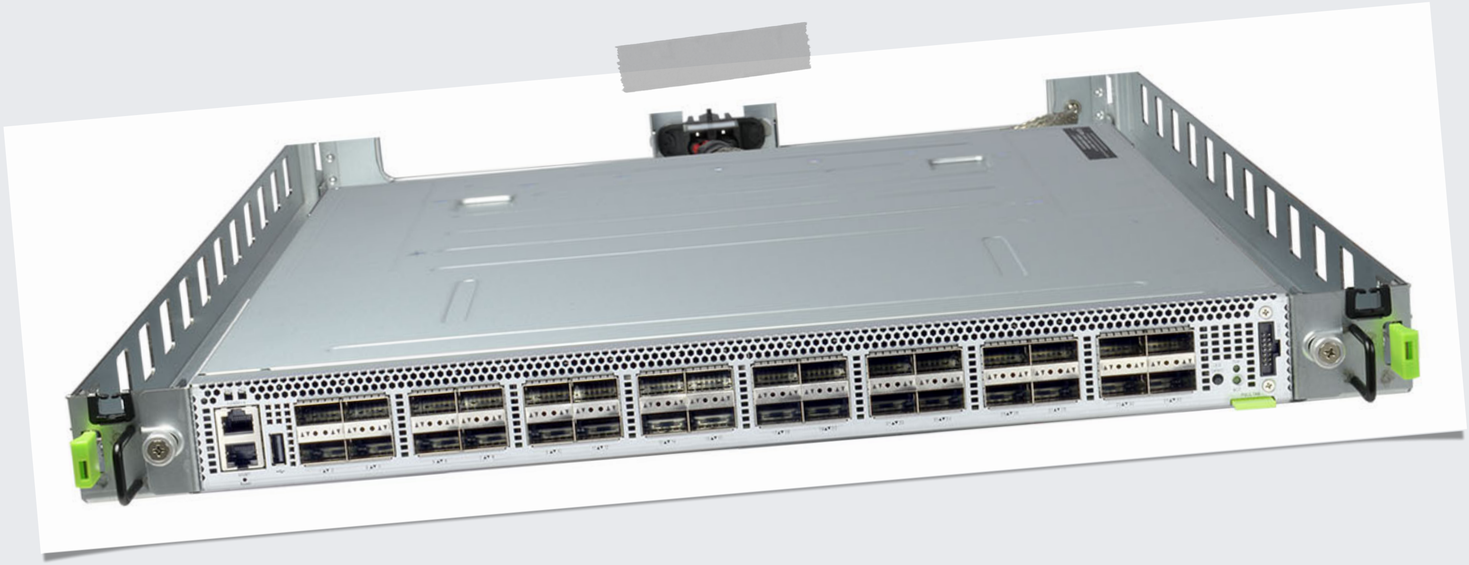
Data center network trends

- Simple, custom switches
- Software-based fabric networks
- Automated repair

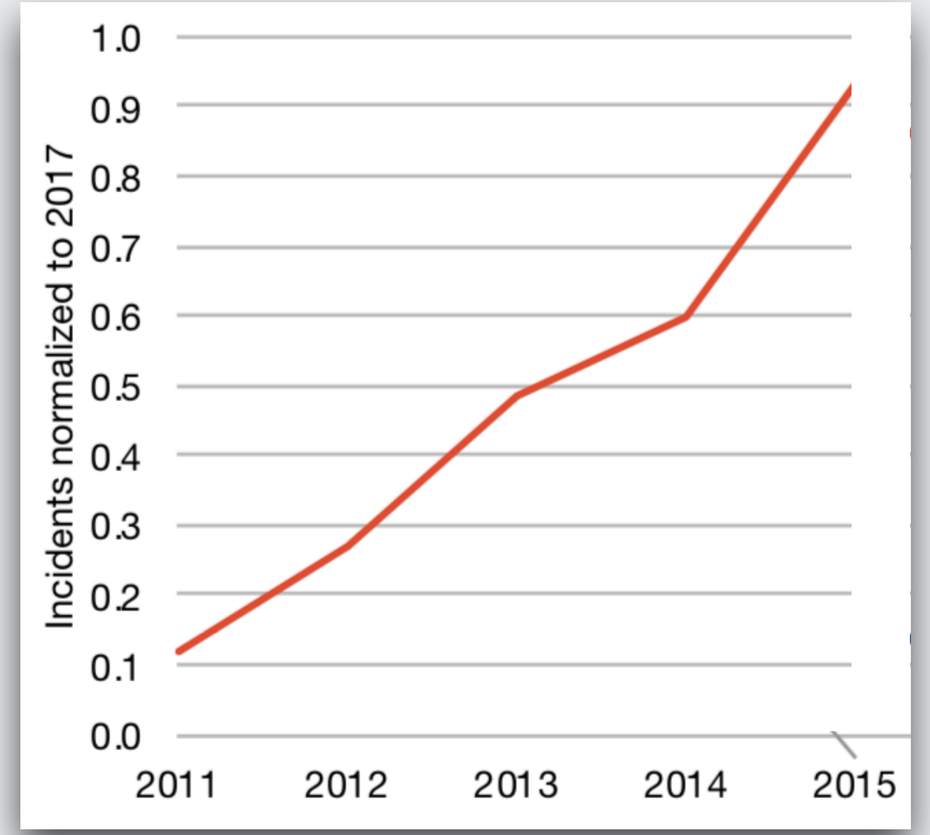


Data center network trends

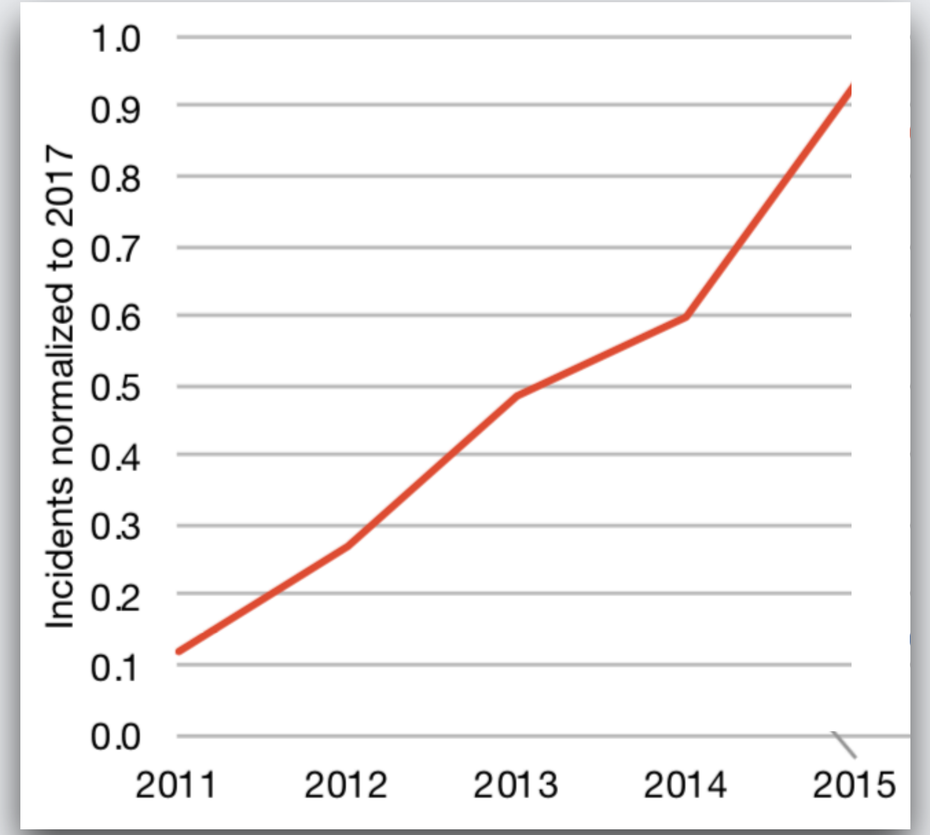
- Simple, custom switches
- Software-based fabric networks
- Automated repair



Older cluster-based design

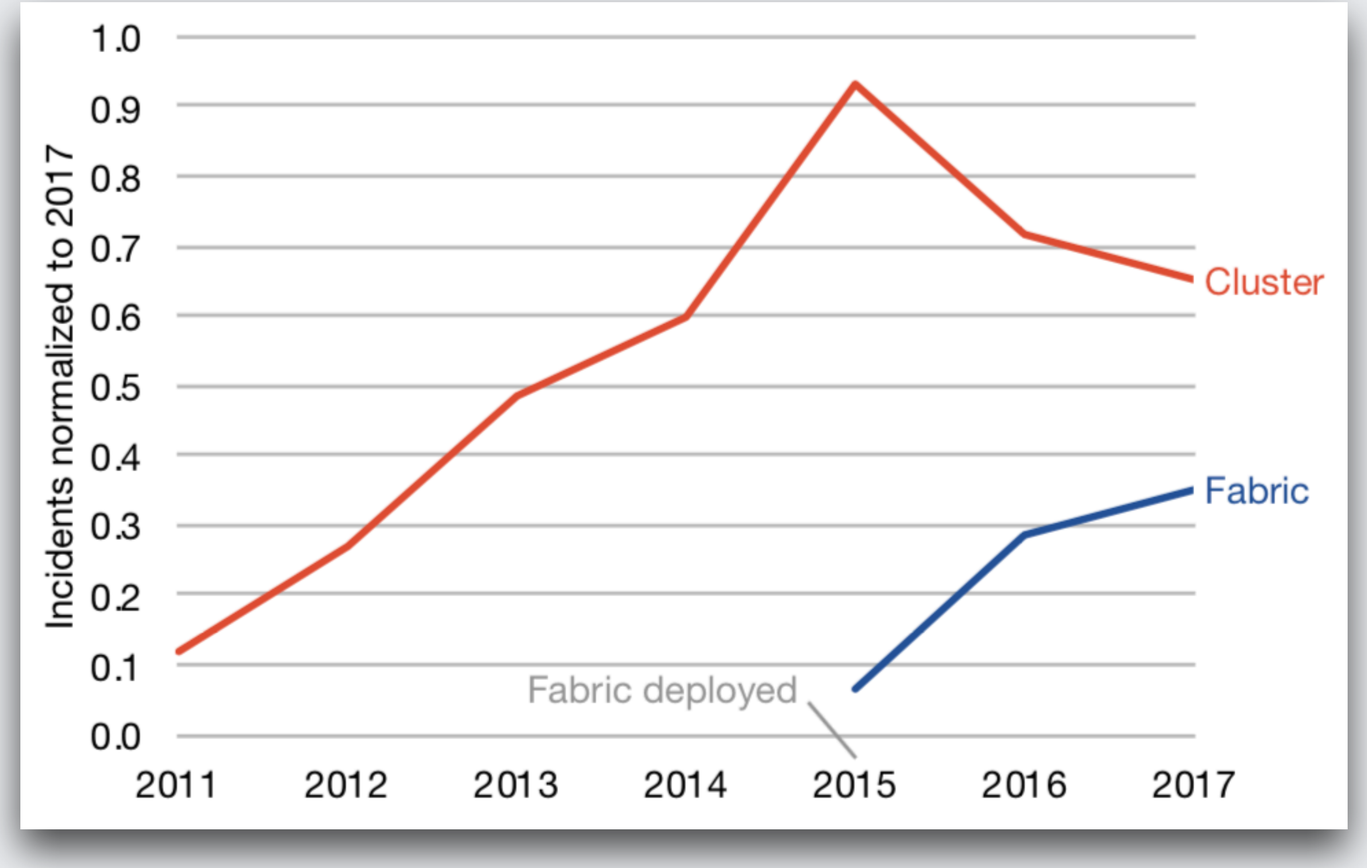


Older cluster-based design

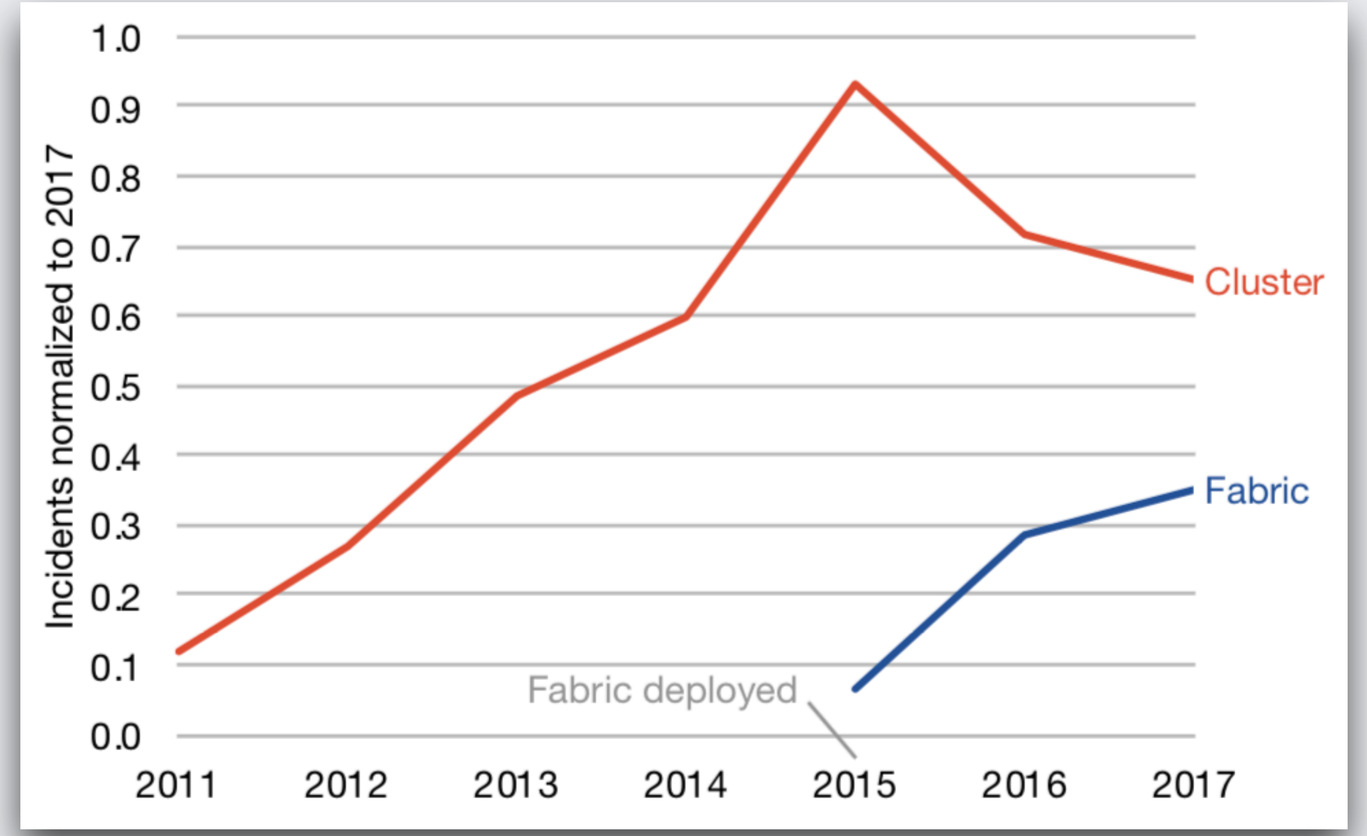


- Cluster network incidents increased 9x over 4 years

Cluster versus fabric designs

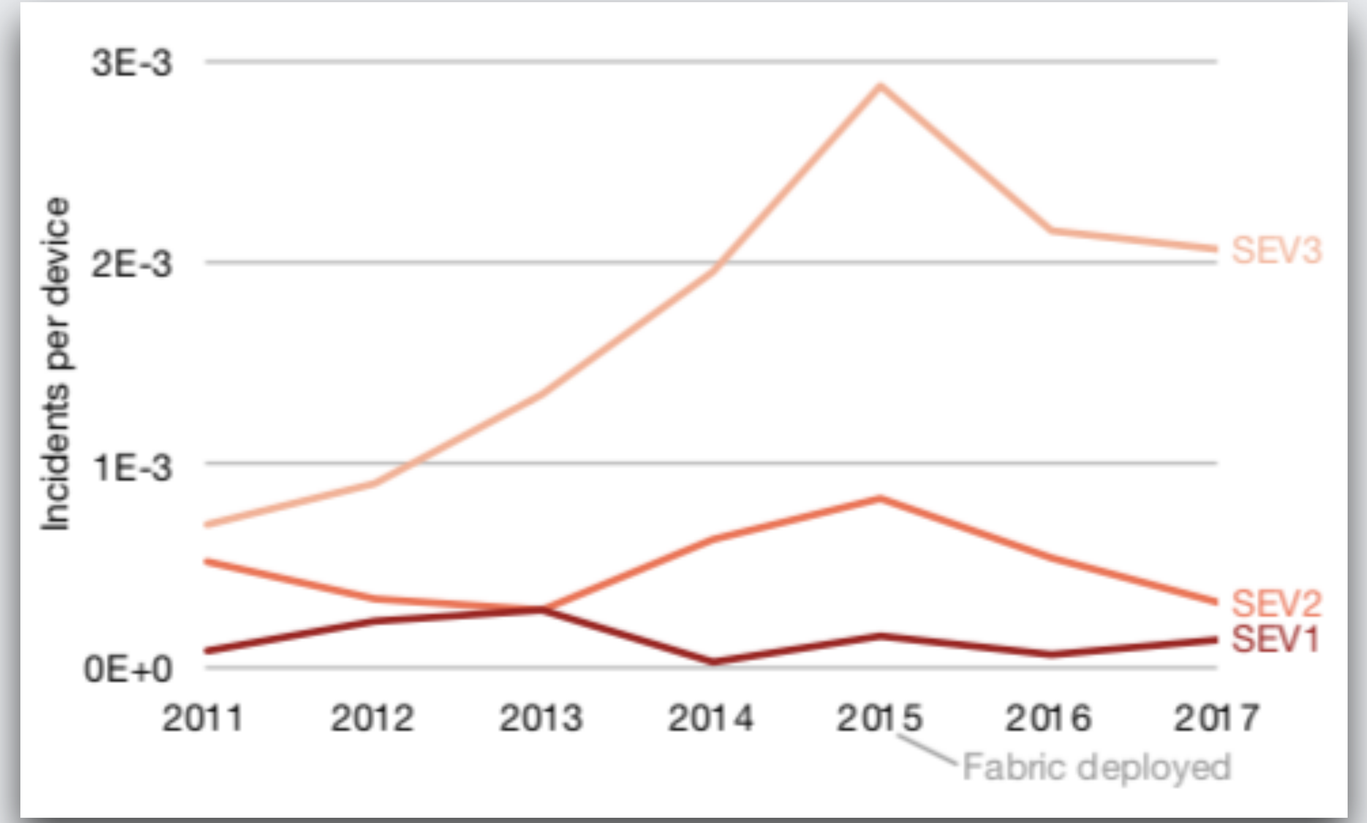


Cluster versus fabric designs



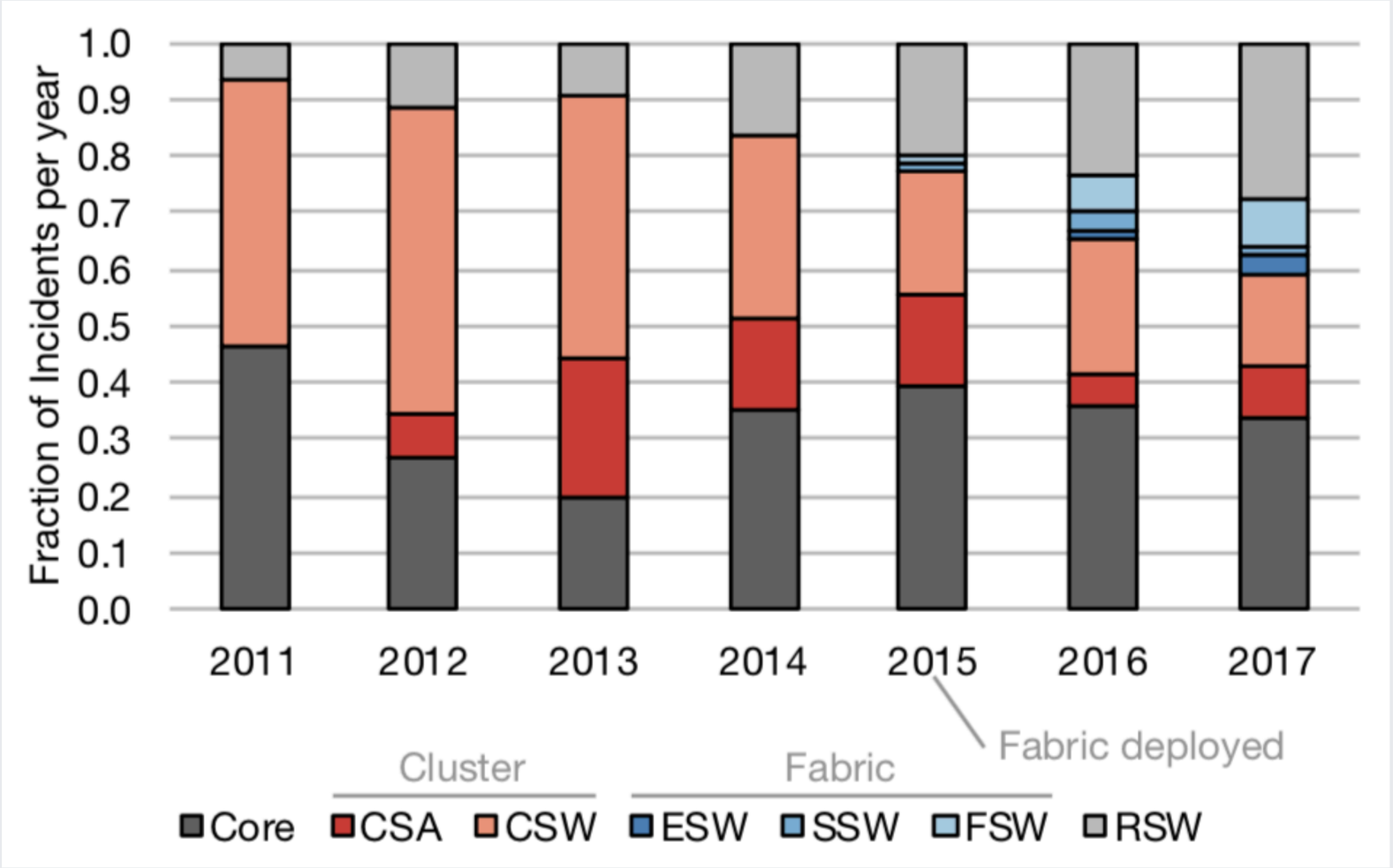
- Cluster have 2X total incidents & 2.8X on a per-device level

DC fabric has fewer incidents

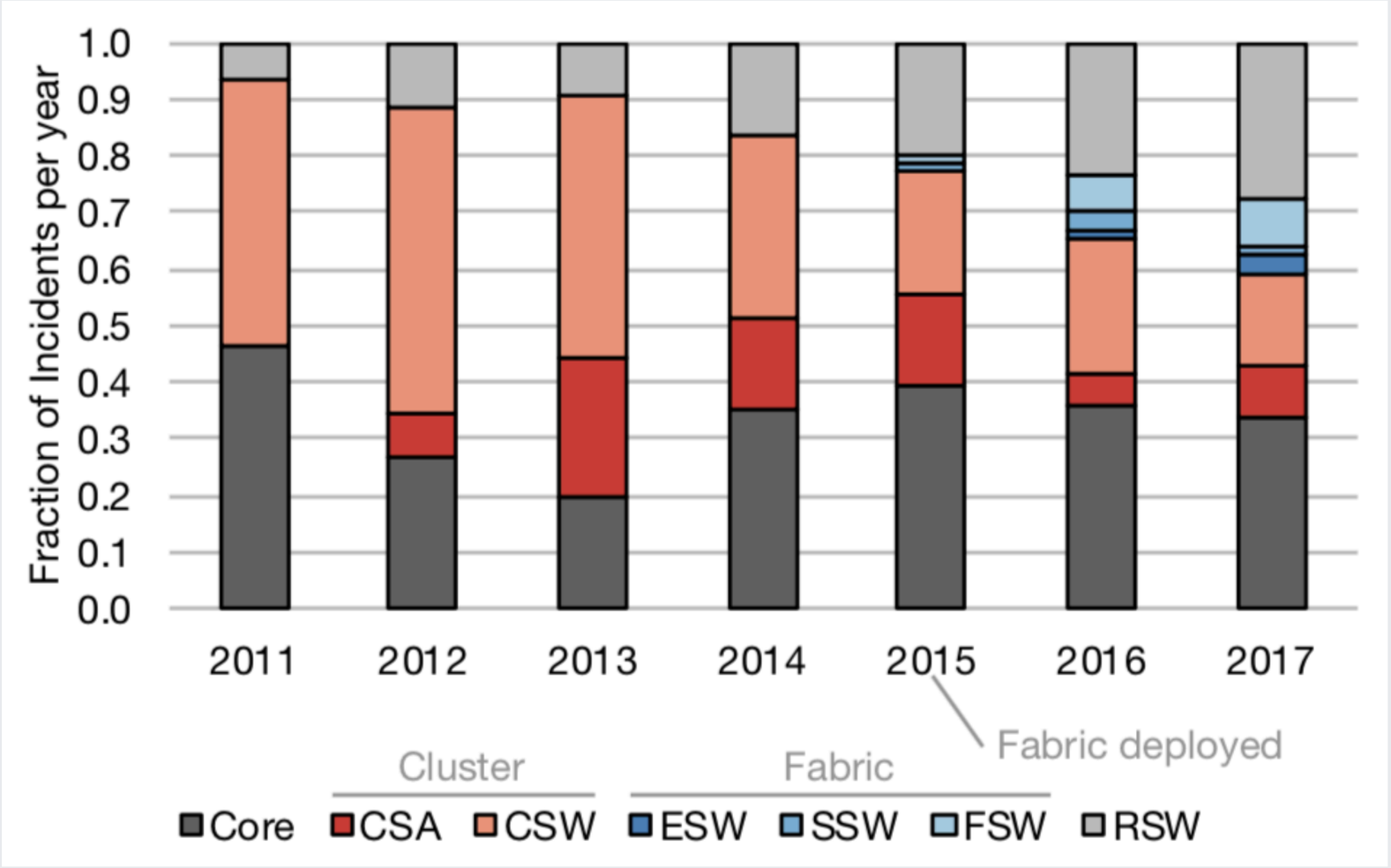


- Reversing the negative software-level reliability trend

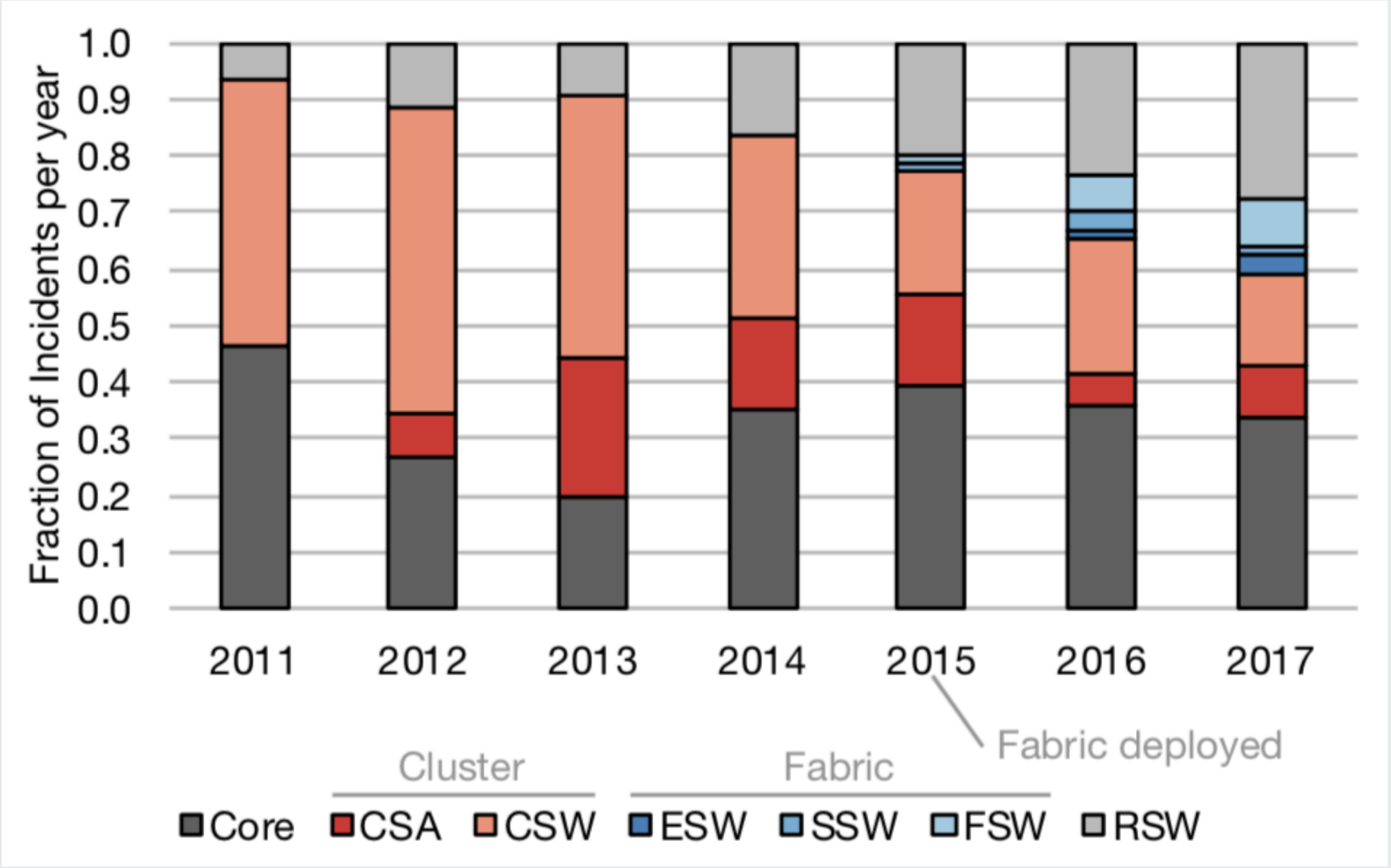
First and last hop reliability



First and last hop reliability



First and last hop reliability

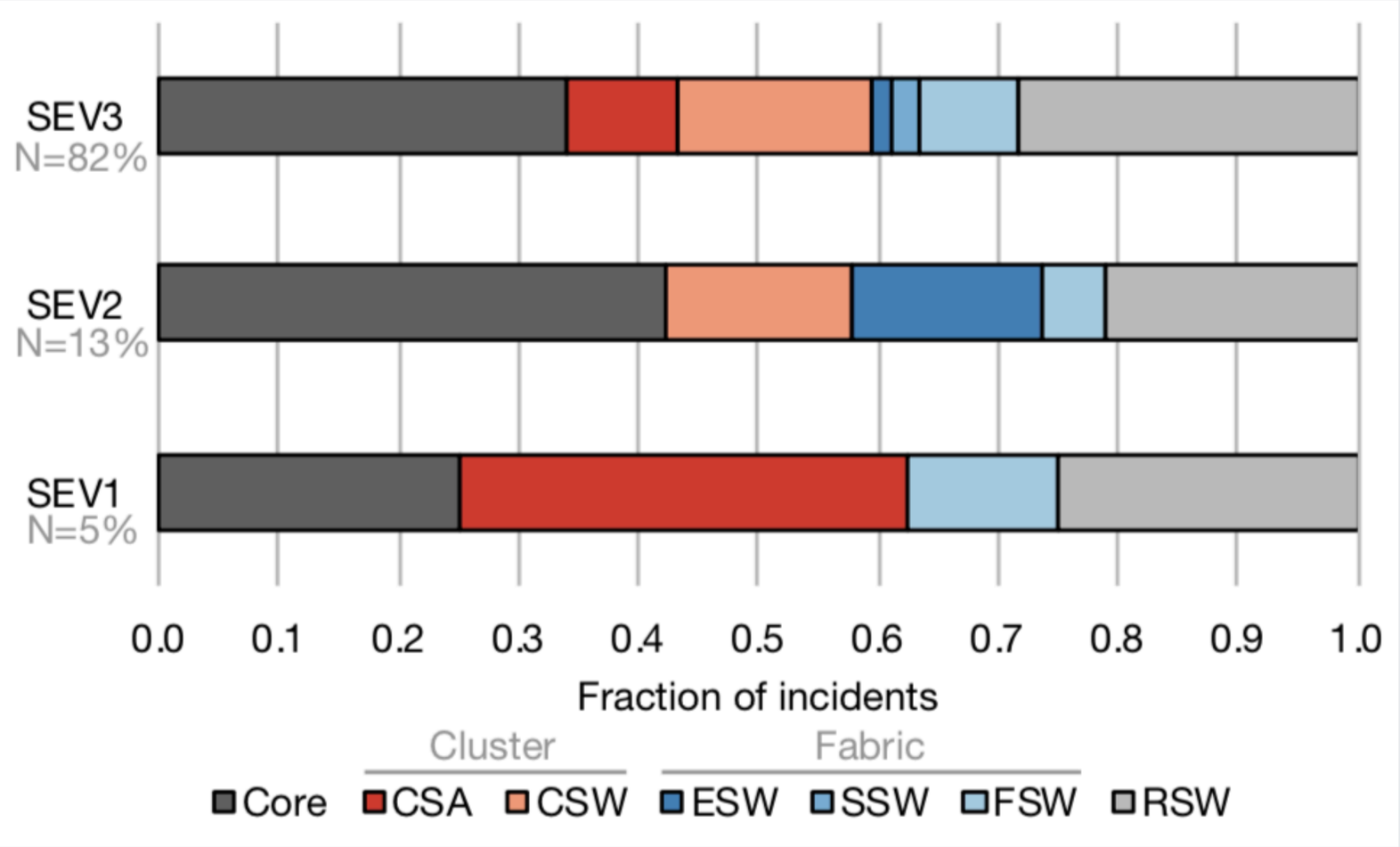


rack switches make up

82%

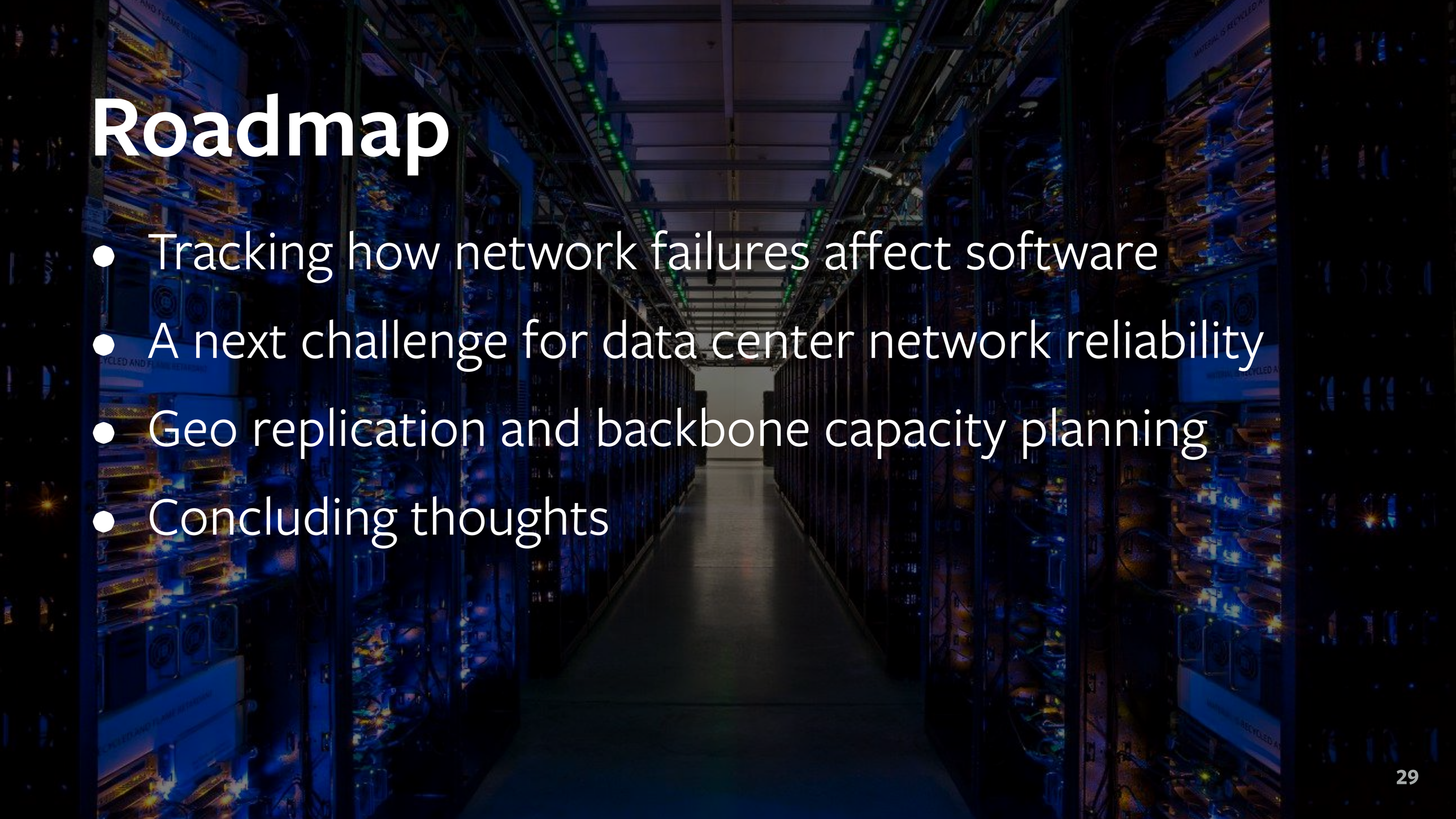
of network devices

Main cause across all severities



Implications for DC networks

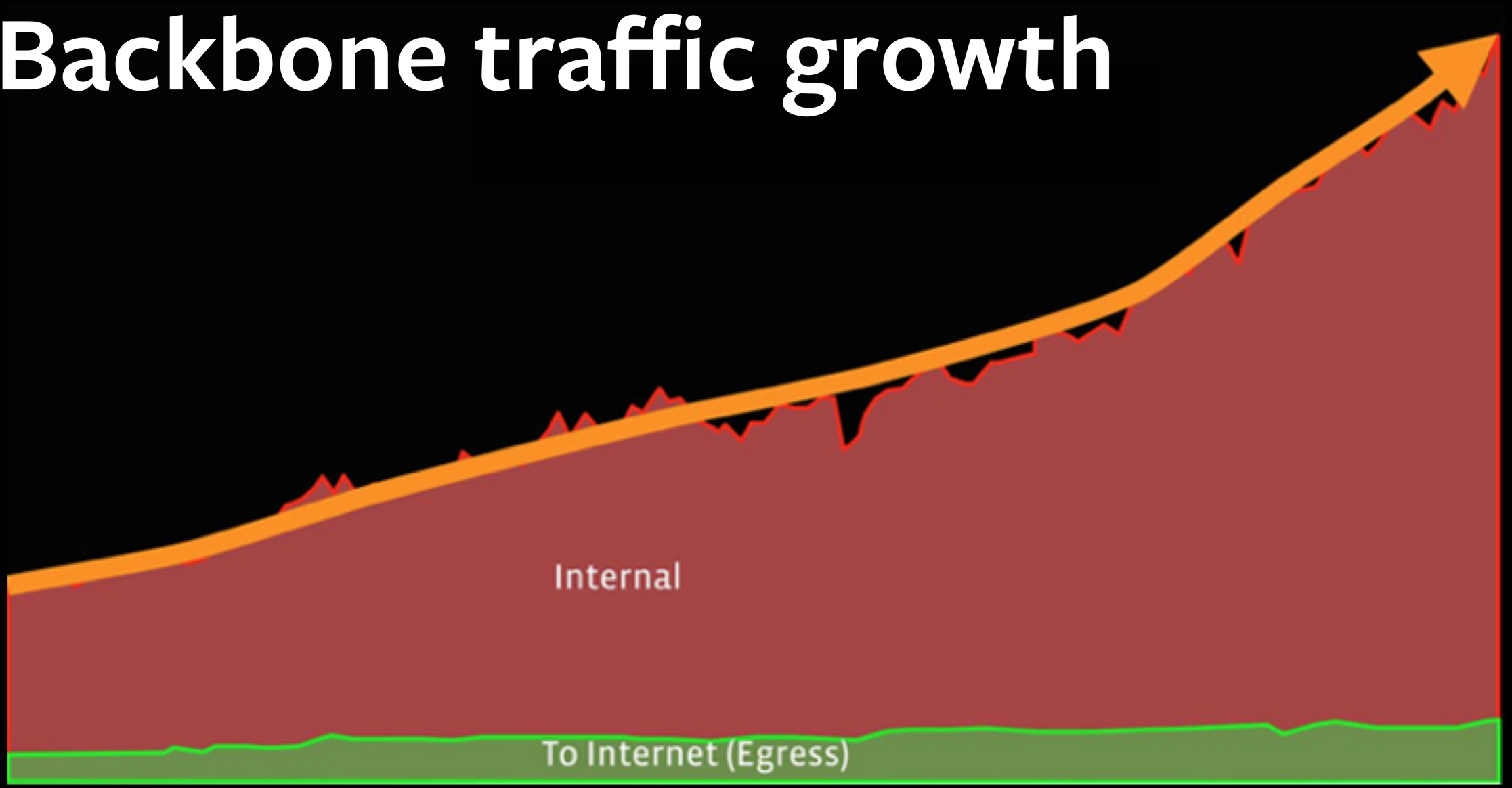
- More redundant switches is one approach
- Make software more resilient
- More aggressive automated repairs



Roadmap

- Tracking how network failures affect software
- A next challenge for data center network reliability
- Geo replication and backbone capacity planning
- Concluding thoughts

Backbone traffic growth



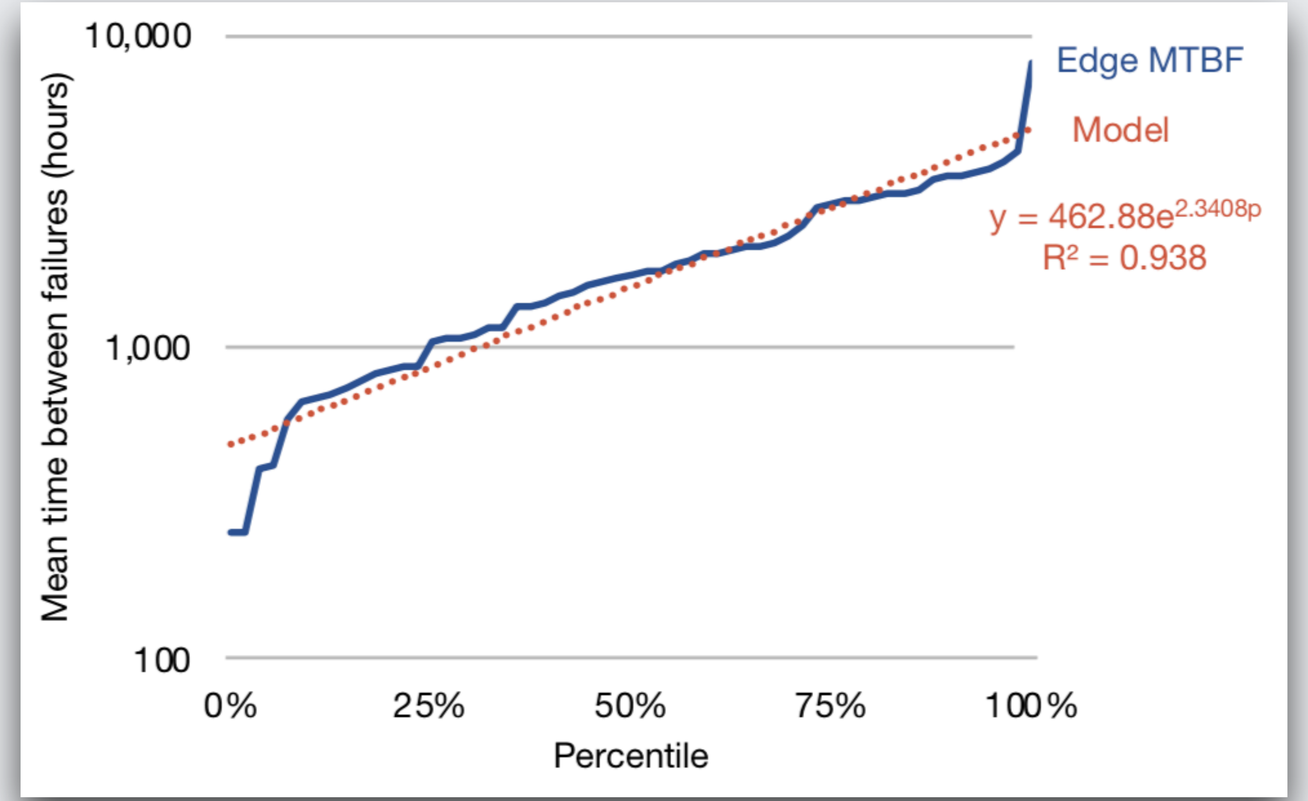
Data center backbones

- Shared resource
- Frequent link failures
- Capacity planning dictates reliability

Measuring backbone reliability

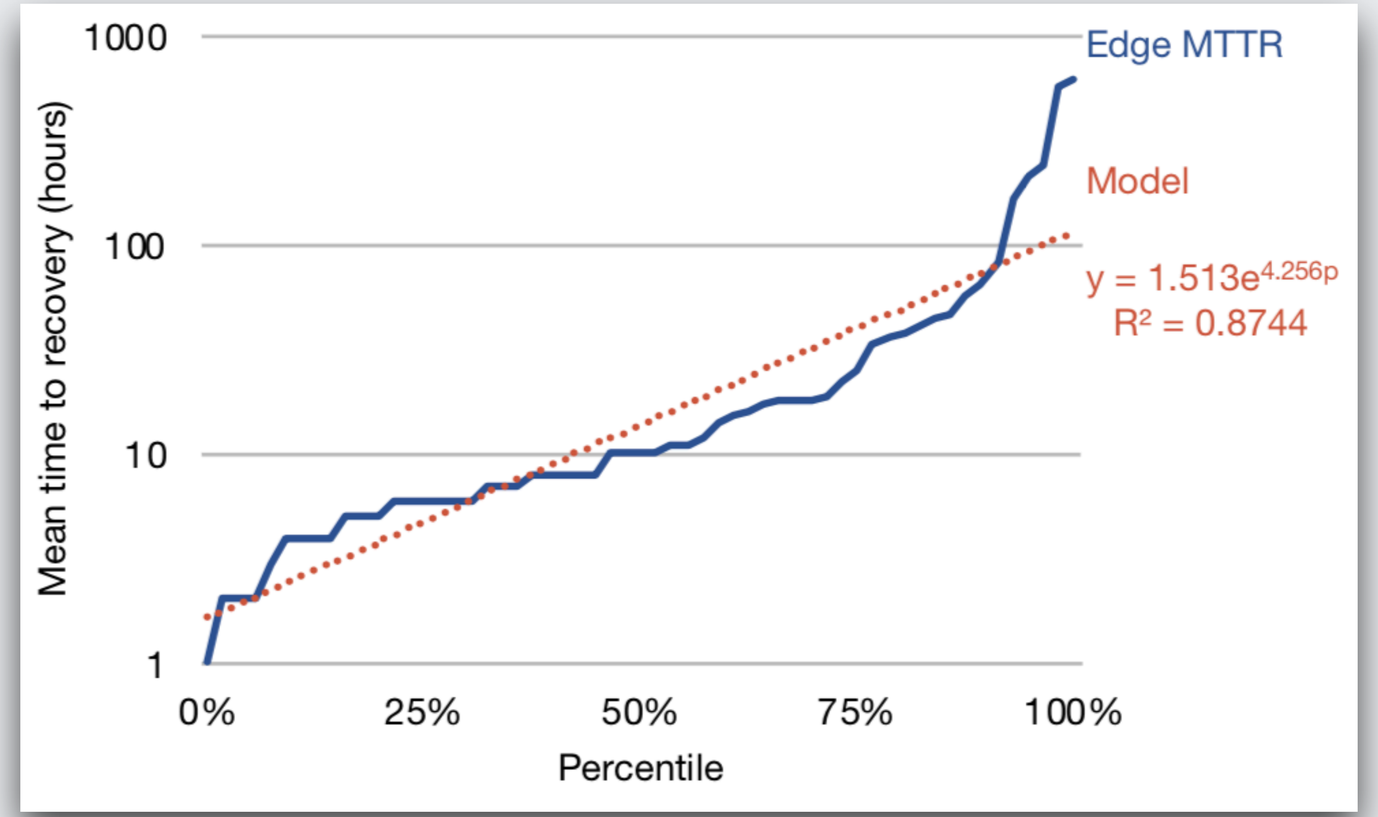
- Sent via email for maintenances and outages
- Parsed and logged in a database
- Used to compute reliability statistics:
  - **Mean time between failures (*MTBF*)**
  - **Mean time to repair (*MTTR*)**

Edge node MTBF distribution



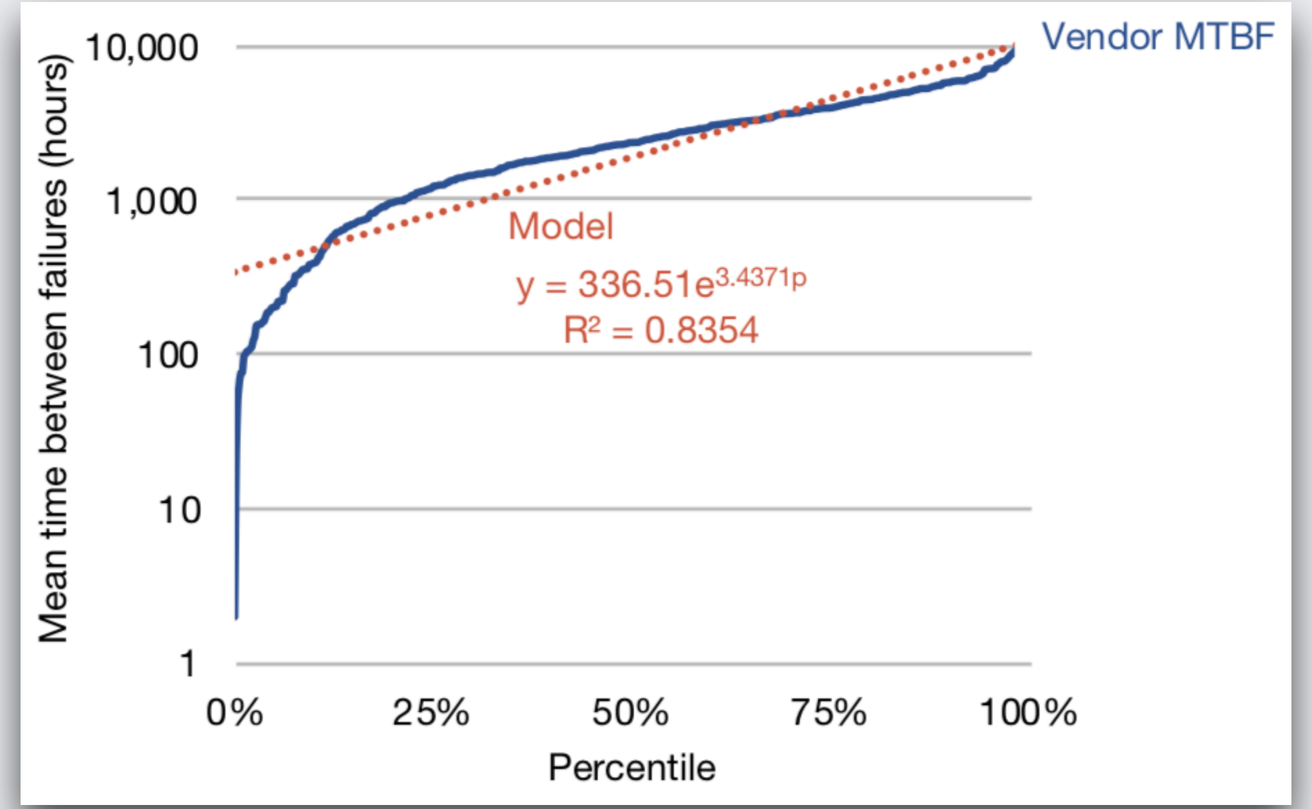
- Typical edge node failure rate is on the order of months

Edge node MTTR distribution



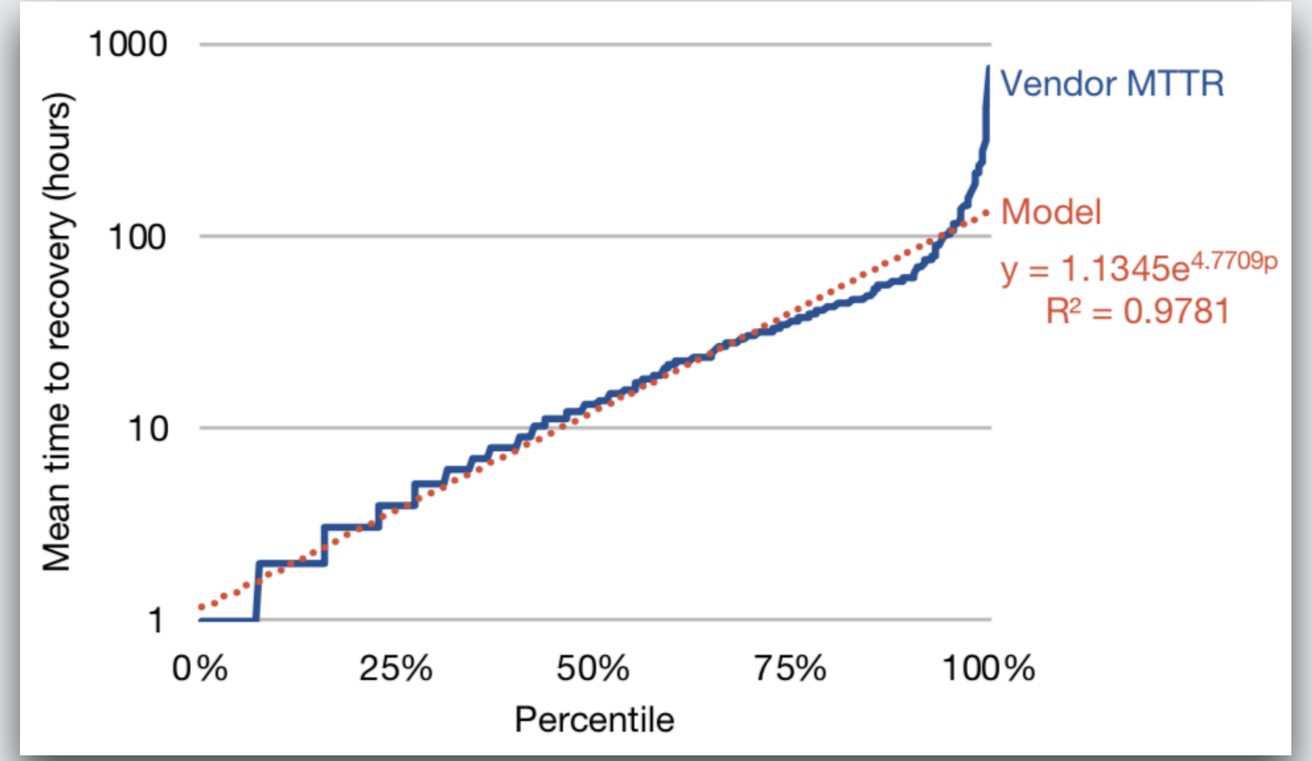
- Edge node mean time to repair is on the order of hours

Fiber vendor MTBF distribution



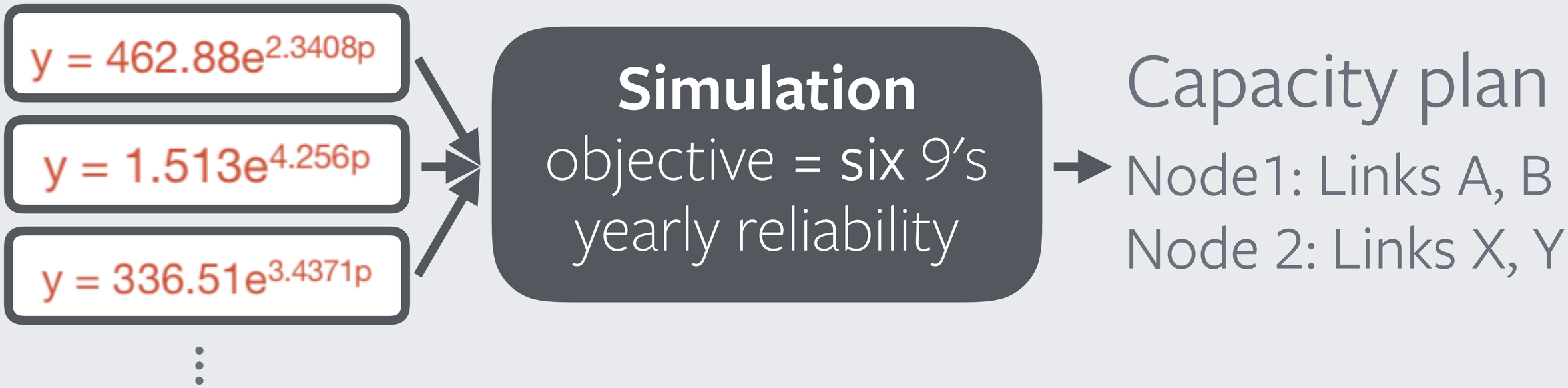
- Typical vendor link failure rate is on the order of months

Fiber vendor MTTR distribution

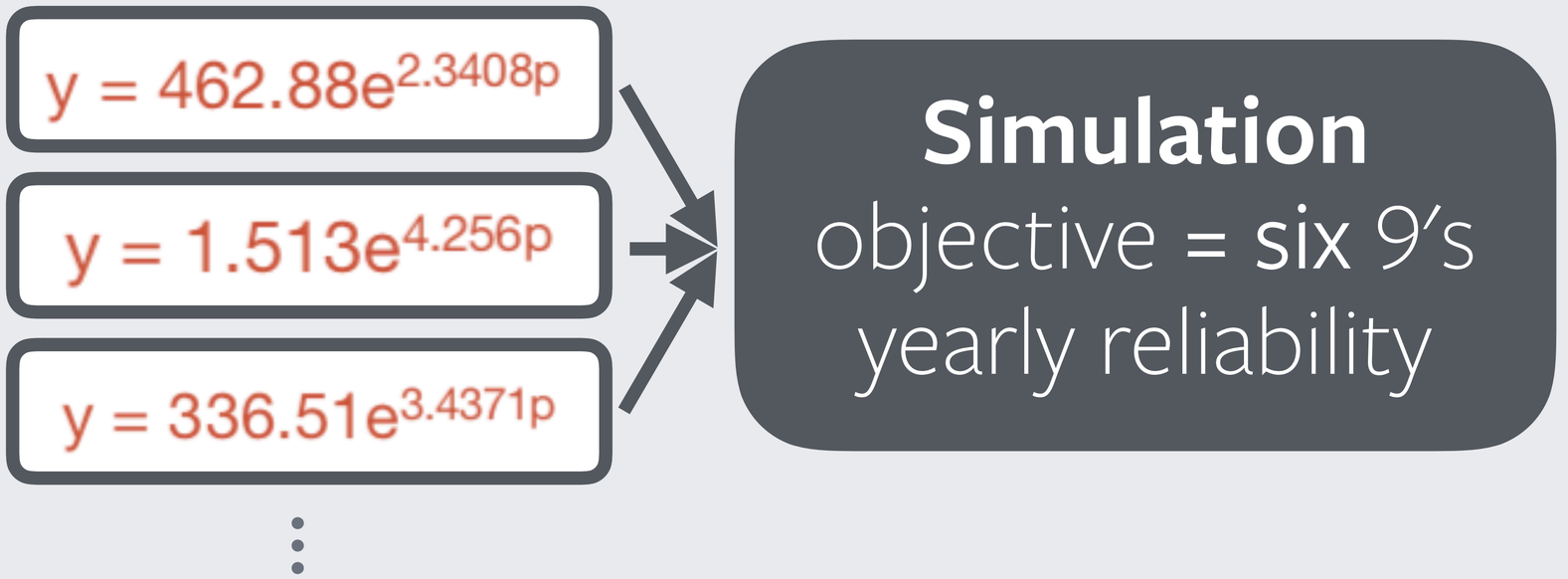


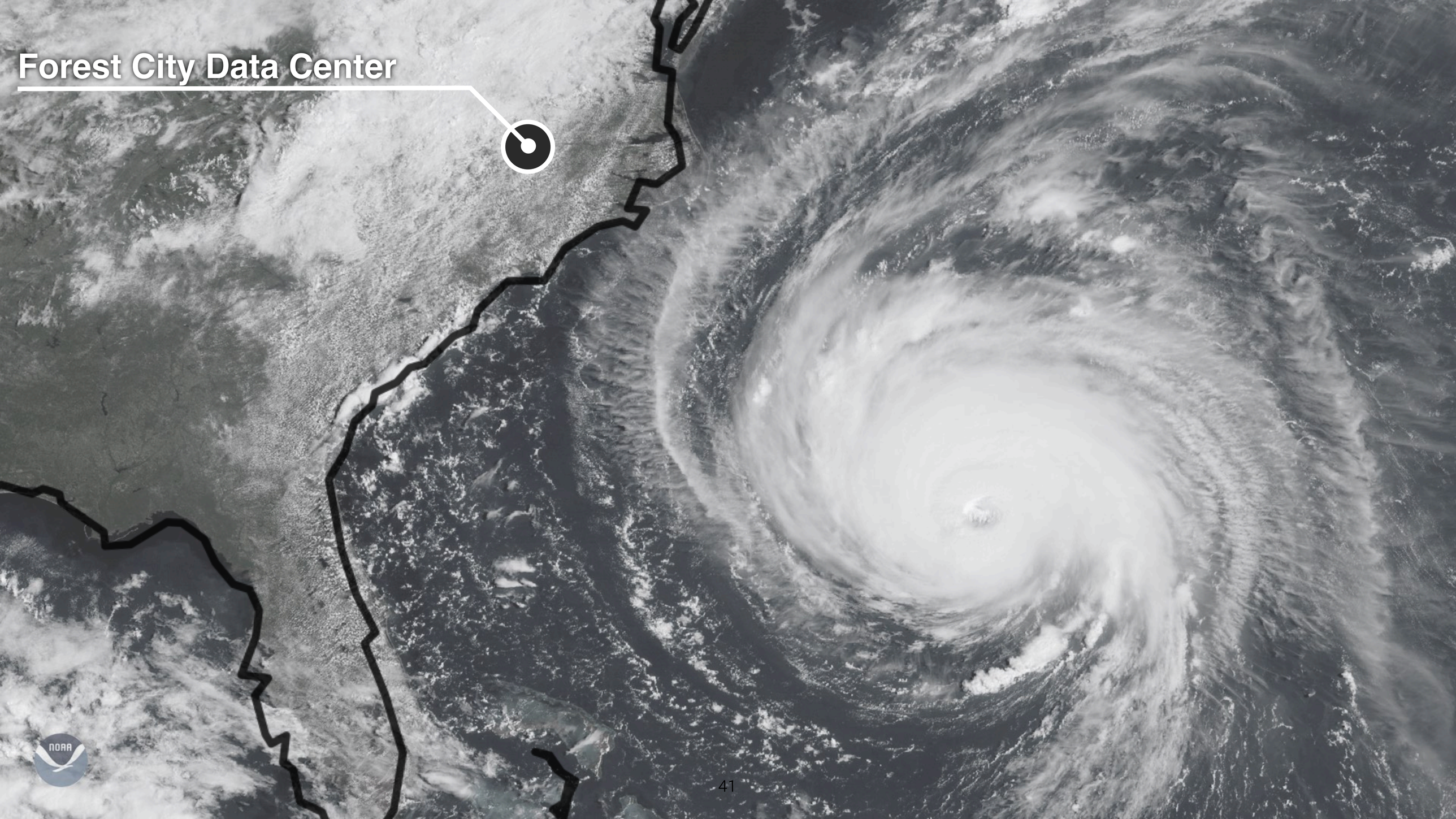
- Vendor MTBF and MTTR span multiple orders of magnitude

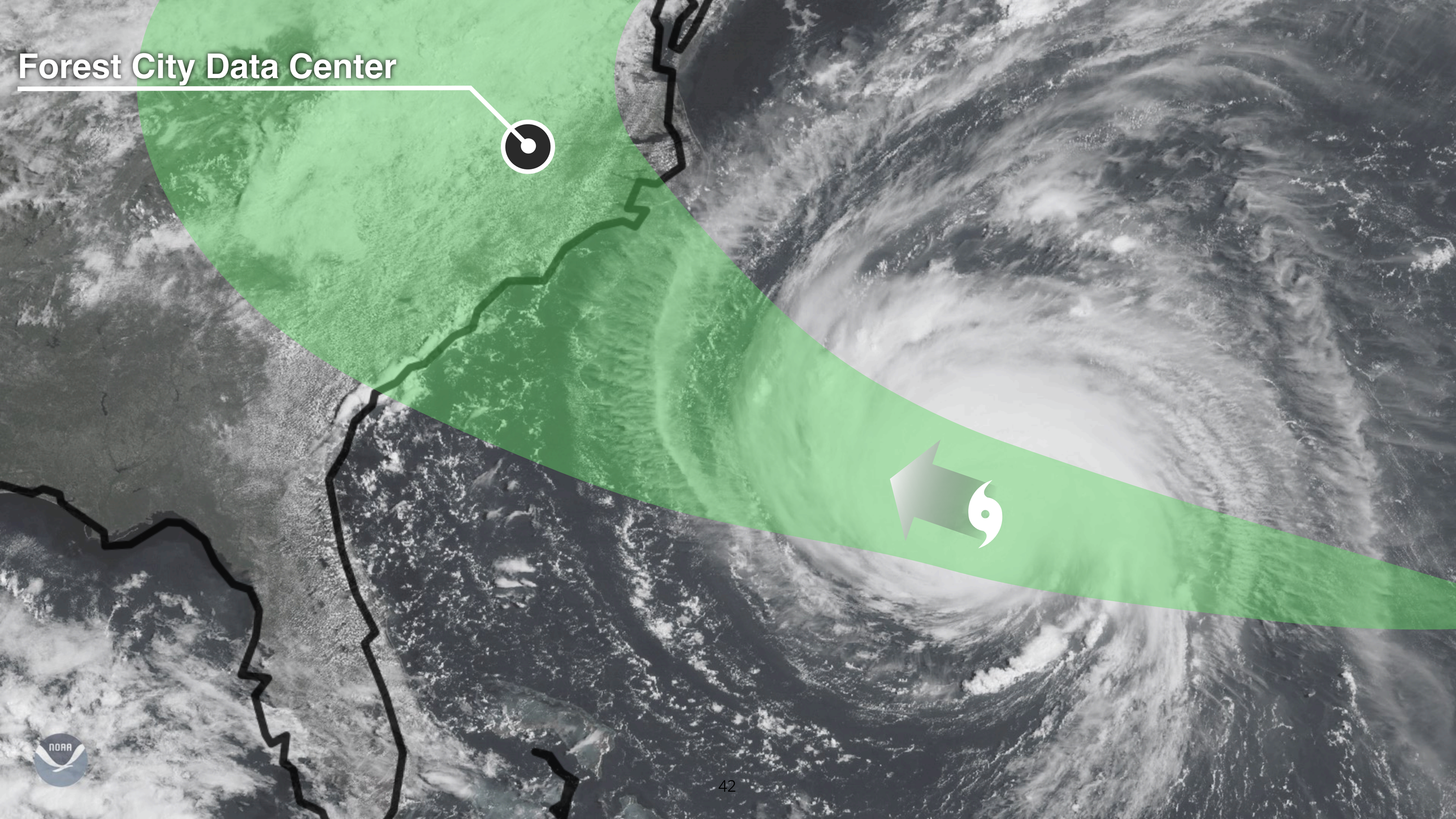
Minimizing backbone outages



Minimizing backbone outages




Forest City Data Center


Forest City Data Center

Roadmap

- Tracking how network failures affect software
- A next challenge for data center network reliability
- Geo replication and backbone capacity planning
- Concluding thoughts

Concluding thoughts

- First and last hop reliability forces us to rethink how network and software share the task of reliability
- Reliable backbone planning is a key enabler for geo replication and software management flexibility


facebook