lecture 09: state management, continued

5590: software defined networking

anduo wang, Temple University

TTLMAN 401B, R 17:30-20:00

about review

about review

what to write (4 parts)

- summary: 3-4 sentences, this is part of the review
- strength
- weakness
- (optional, constructive) comments: suggest what to improve on the technical side, on presentation (writing, organization)

what not to include

- repeat technical details of the paper
  - DO NOT include figures/texts from the paper

about review

about review

language

- be formal
  - e.g., "the authors may want to" instead of "you should …"
- work on grammar

guideline

- keep reviews informative
- an opportunity to start conversation
- write the review for your own understanding
  - remember: your reviews are not graded

statesman: use cases, evaluations …

statesman deployment

10 geographically-distributed datacenter (DC)

- cover switches, links within each DC and across DC (WAN)

three applications

- switch-upgrade
- failure-mitigation
- inter-DC TE

challenges—maintaining globally available and distributed states

- inter-DC
  - due to WAN failures, DCs may be disconnected
- within-DC
  - huge volume of state data: hundreds of thousands of switches and links
  - millions of state variables

challenges—updating DCN states

- heterogeneity: diverse range of network elements expose heterogenous interfaces for updates
- device can fail during an update
- device respond slow, dominating the application control loop

solution—maintaining globally available and distributed states

solution—maintaining globally available and distributed states

partitioning checker's responsibility into impact groups

- one impact group per DC
- one additional impact group with border routers of all DCs and the WAN links

solution—maintaining globally available and distributed states

partitioning checker's responsibility into impact groups

- one impact group per DC
- one additional impact group with border routers of all DCs and the WAN links

partitioning monitor

- split monitor's responsibility into many instances
  - each covers 1k switches

solution—updating DCN states

solution—updating DCN states

heterogeneity

- OpenFlow and command tempaltes

solution—updating DCN states

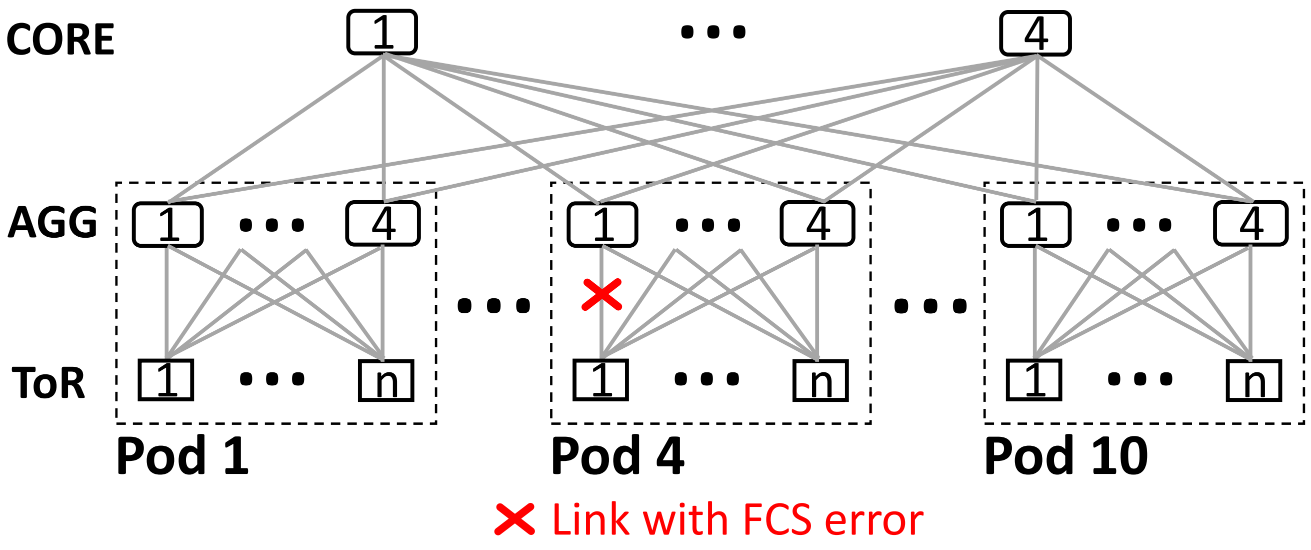
heterogeneity

- OpenFlow and command tempaltes

dynamic failures

- stateless updates
- simply push to the devices the latest OS-TS difference

use case: maintaining invariants

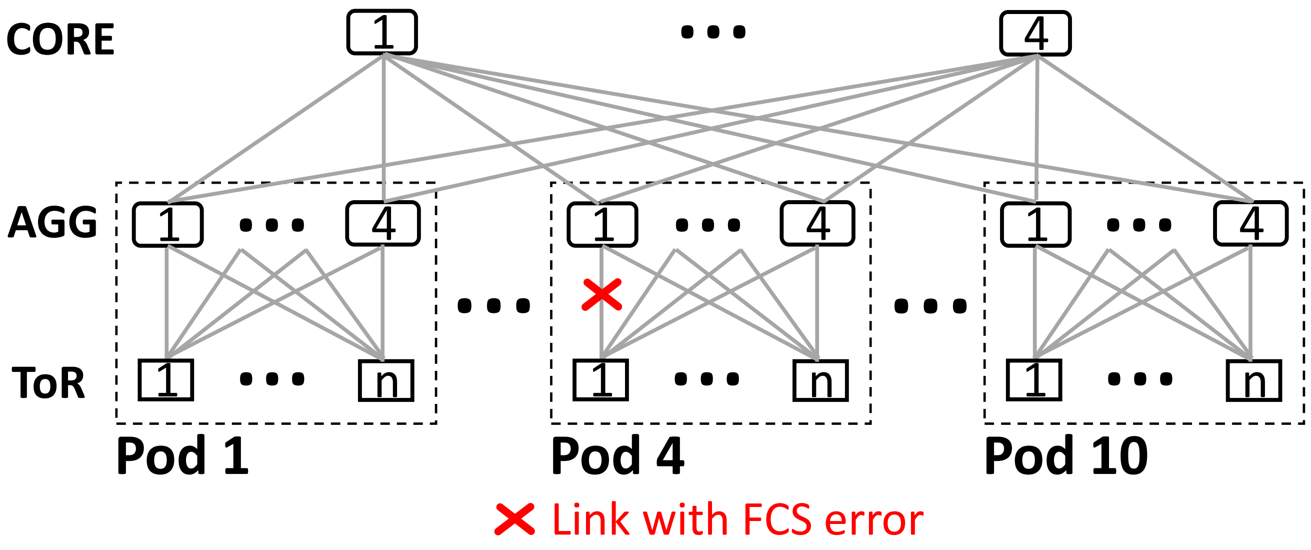


switch_upgrade and failure_mitigation coexist

statesman goal: maintaining capacity invariant

- 99% ToR pairs have at least 50% capacity

use case: maintaining invariants



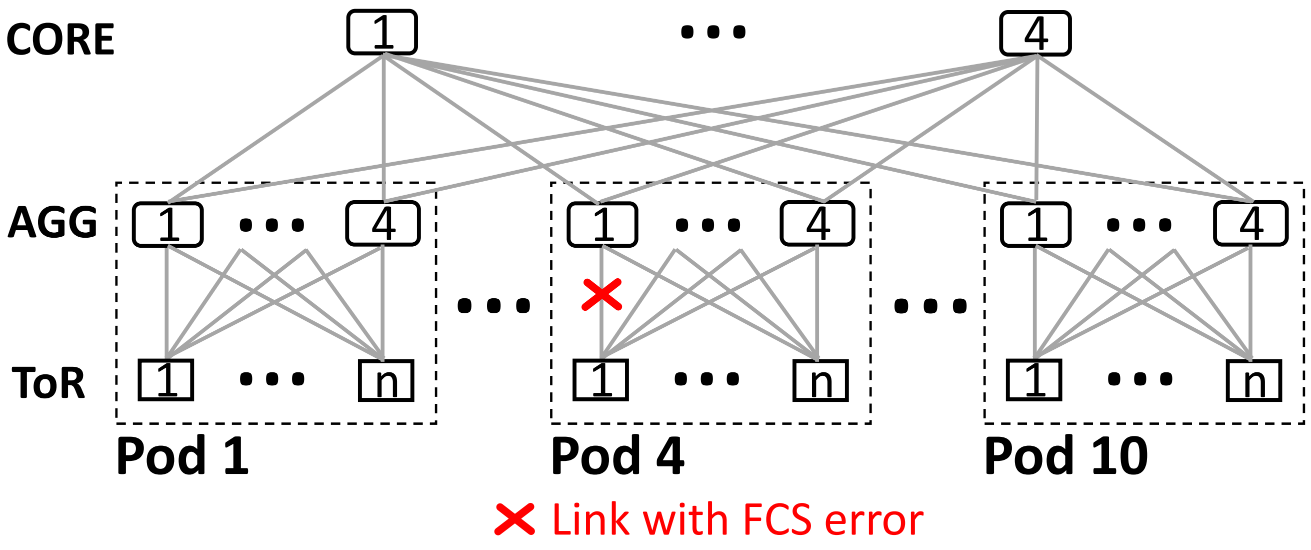
one DC with 10 pods

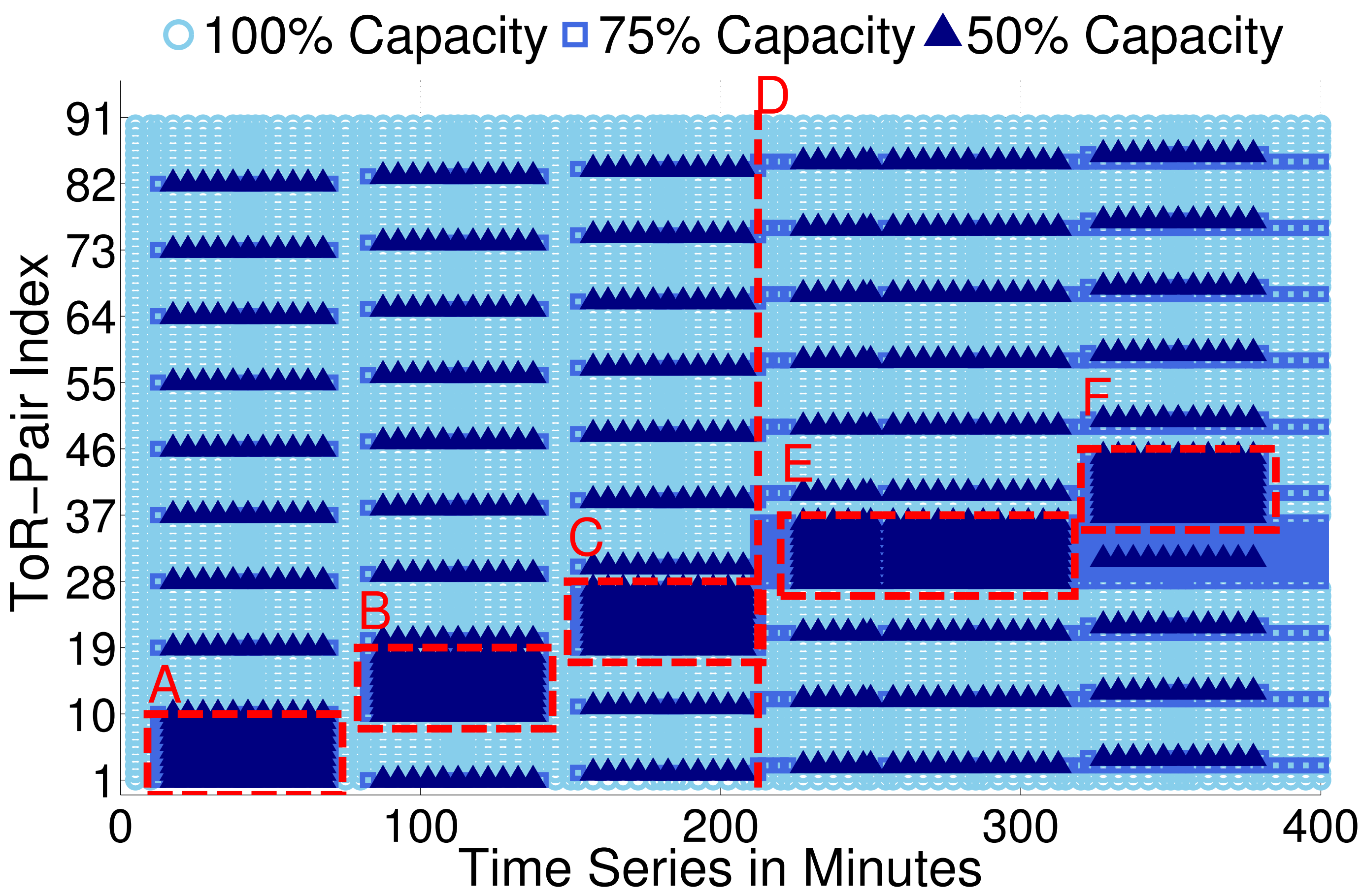
- each pod has 4 AGGs and a number of ToRs

switch_upgrade

- upgrade all 40 AGGs
- (sequentially) pod by pod
- attempt parallel upgrades within each pod

use case: maintaining invariants

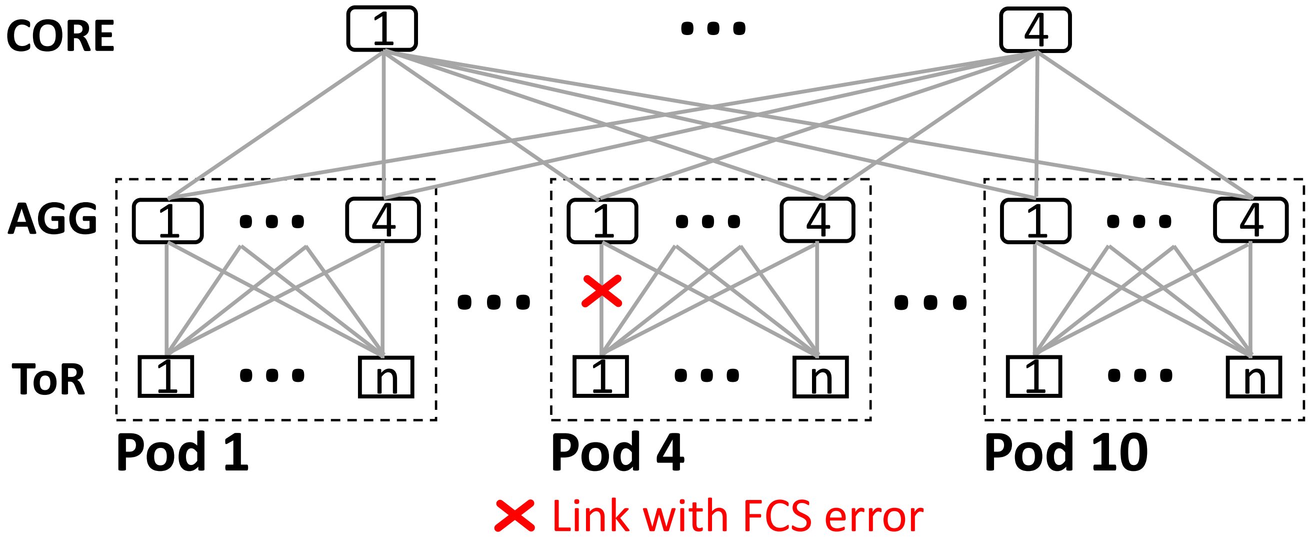


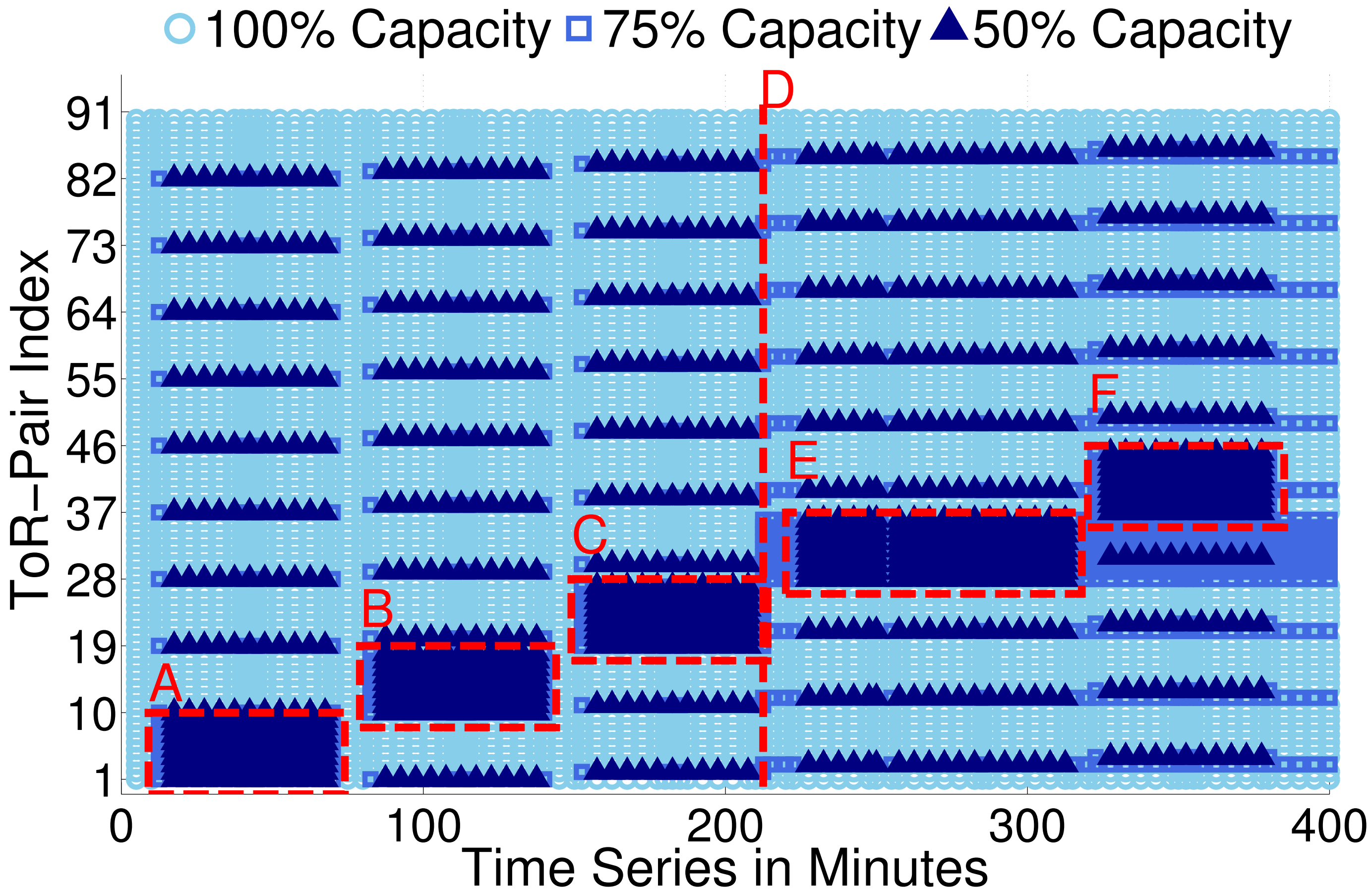


90 ToR pairs

- one ToR from each pod
- put the 9 ToR pairs from the same pods together

use case: maintaining invariants





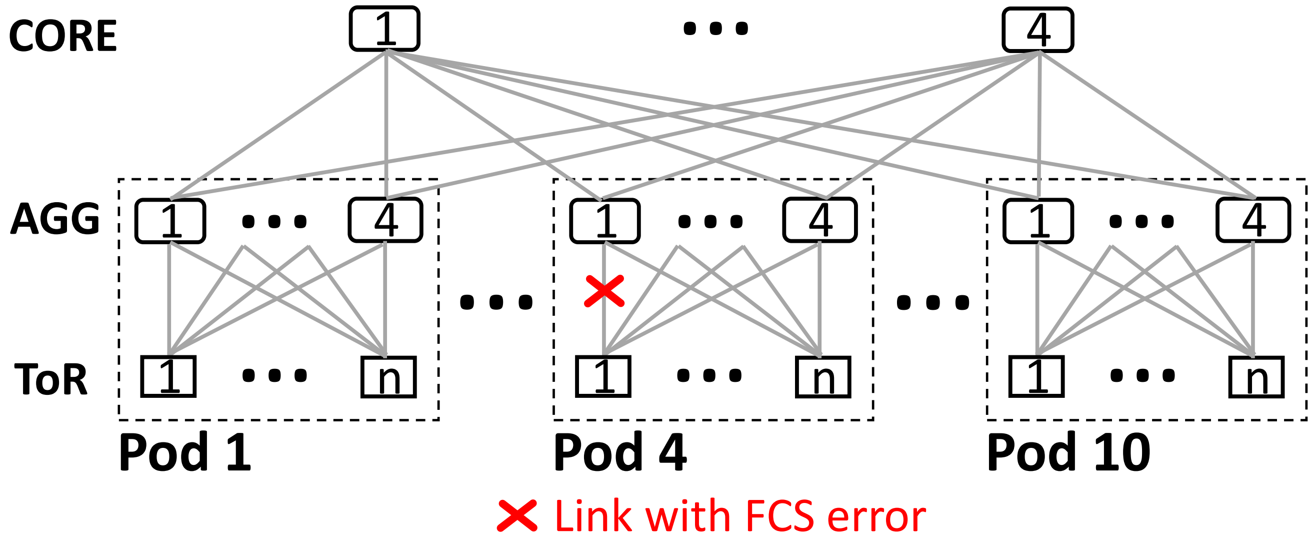
90 ToR pairs

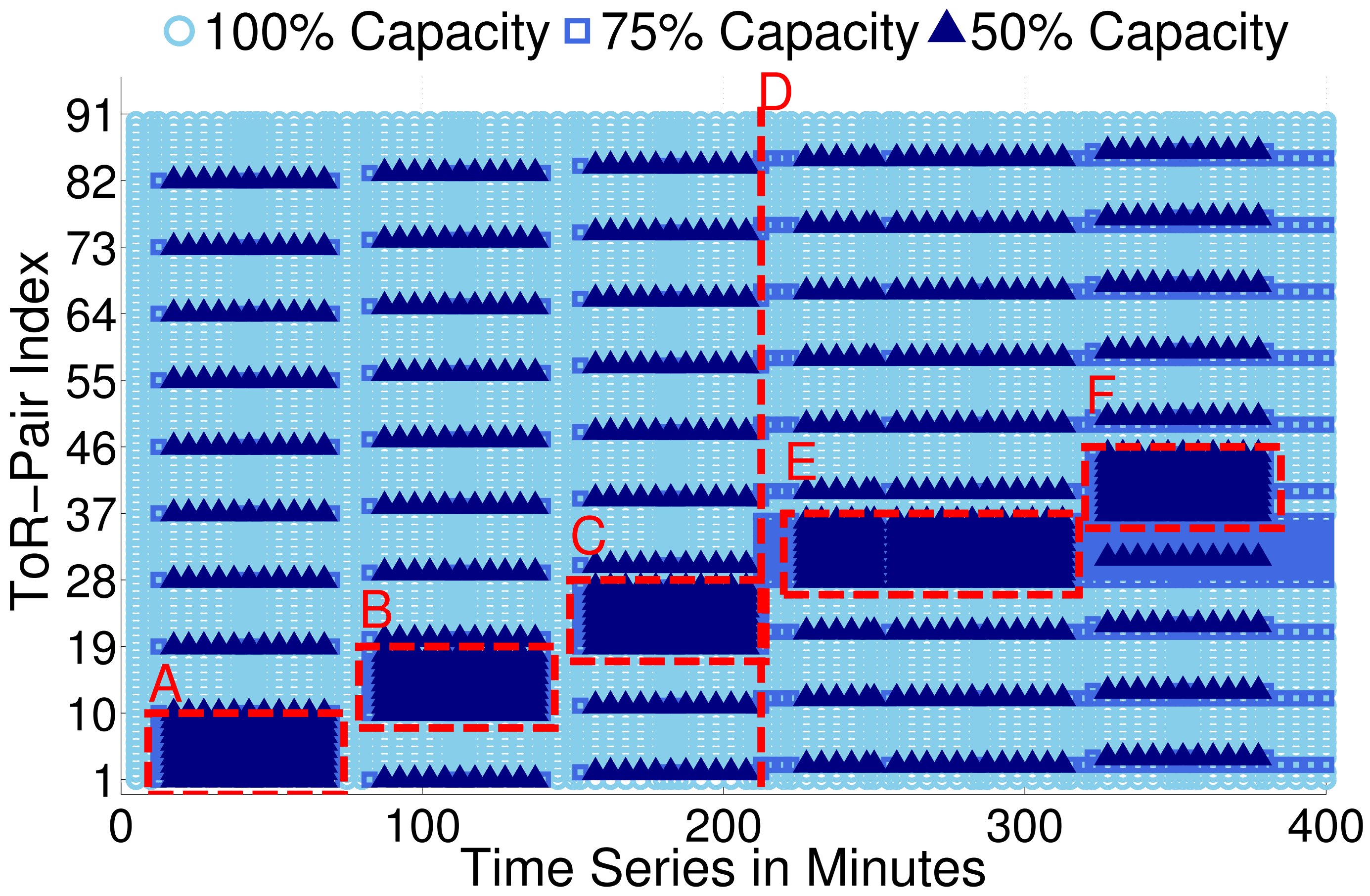
- one ToR from each pod
- put the 9 ToR pairs from the same pods together

A,B,C

- switch-upgrade upgrades pod 1,2,3

use case: maintaining invariants





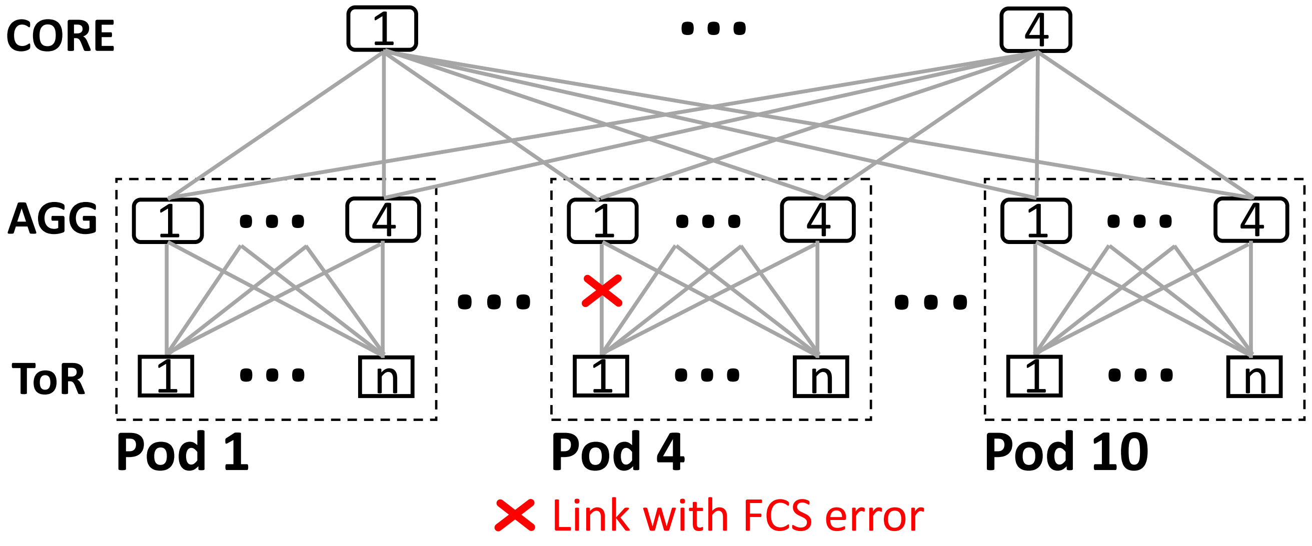
90 ToR pairs

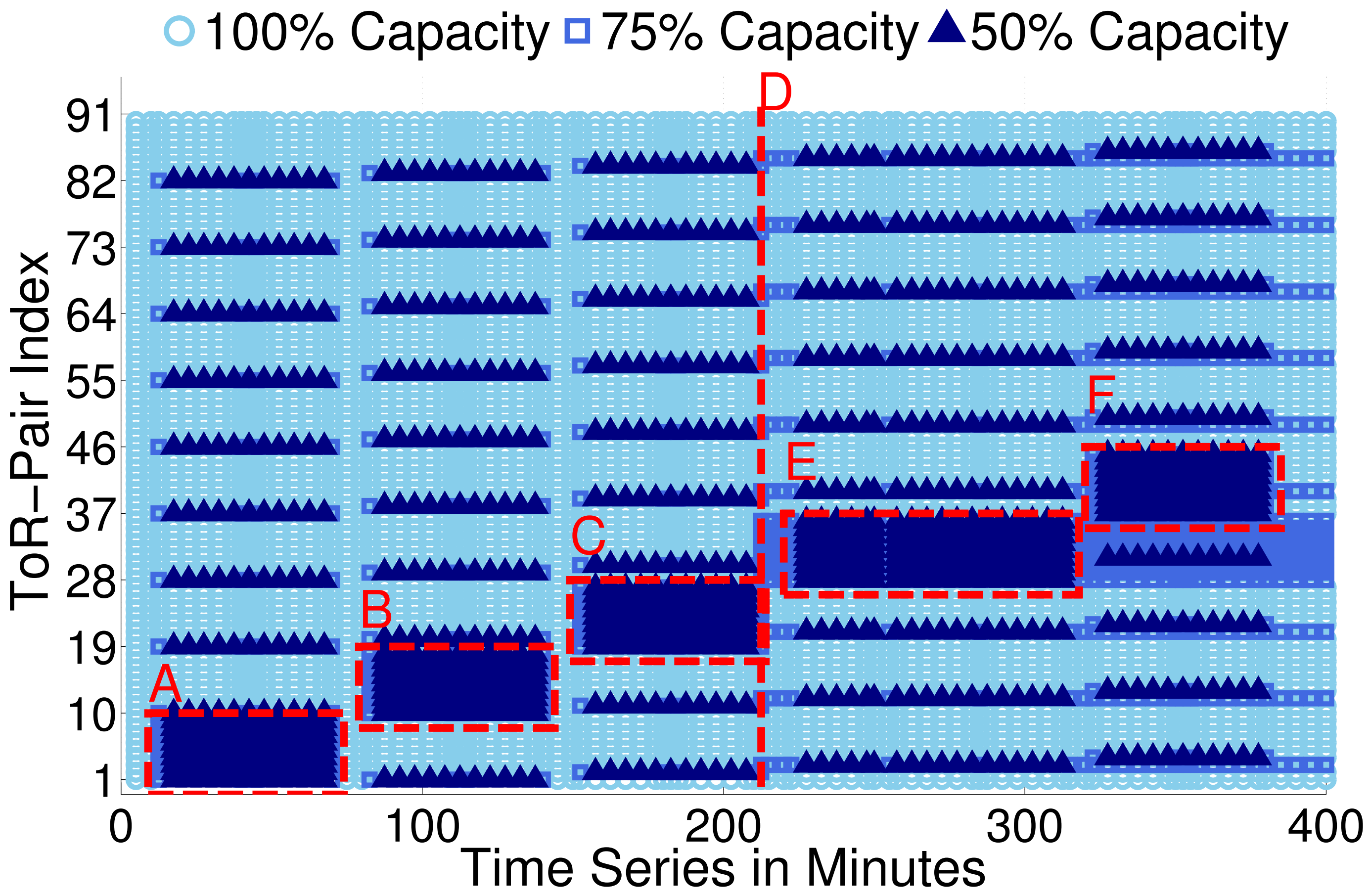
- one ToR from each pod
- put the 9 ToR pairs from the same pods together

D

- failure-mitigation detects problem, shutting down link

use case: maintaining invariants





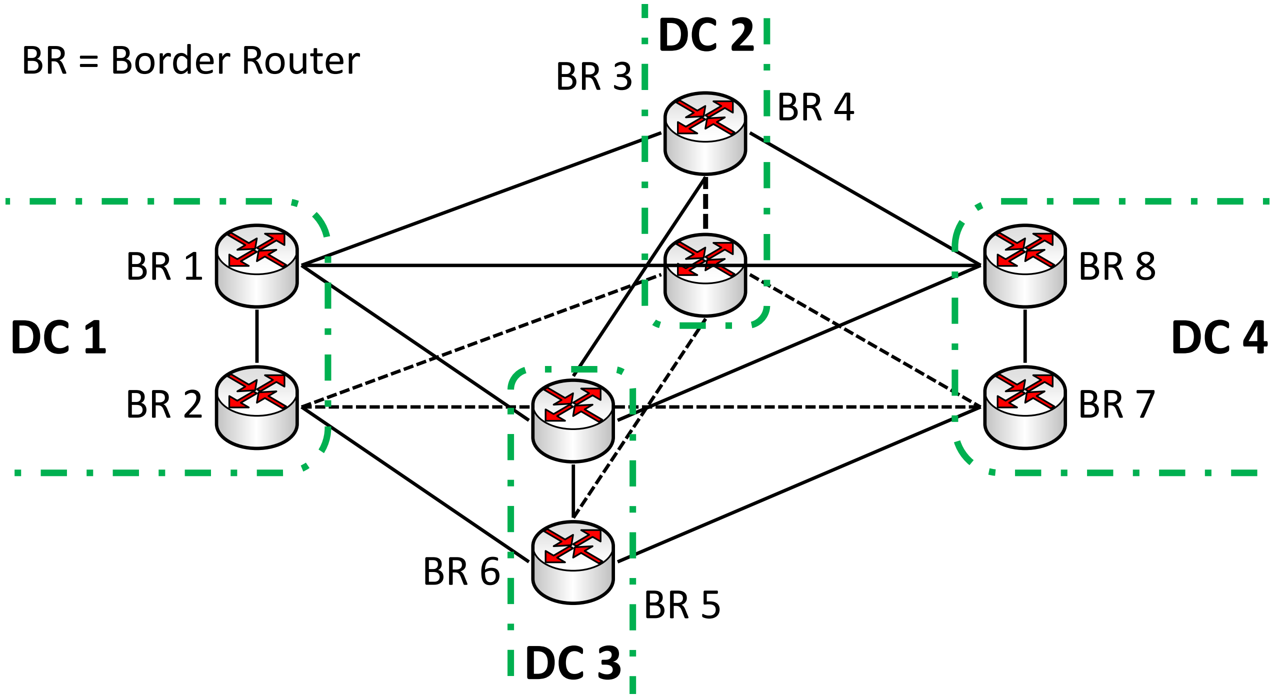
90 ToR pairs

- one ToR from each pod
- put the 9 ToR pairs from the same pods together

E

- switch-upgrade slows down by the checker
  - upgrades Agg 1,2 together; then Agg 3; finally Agg 3

use case: resolving conflicts



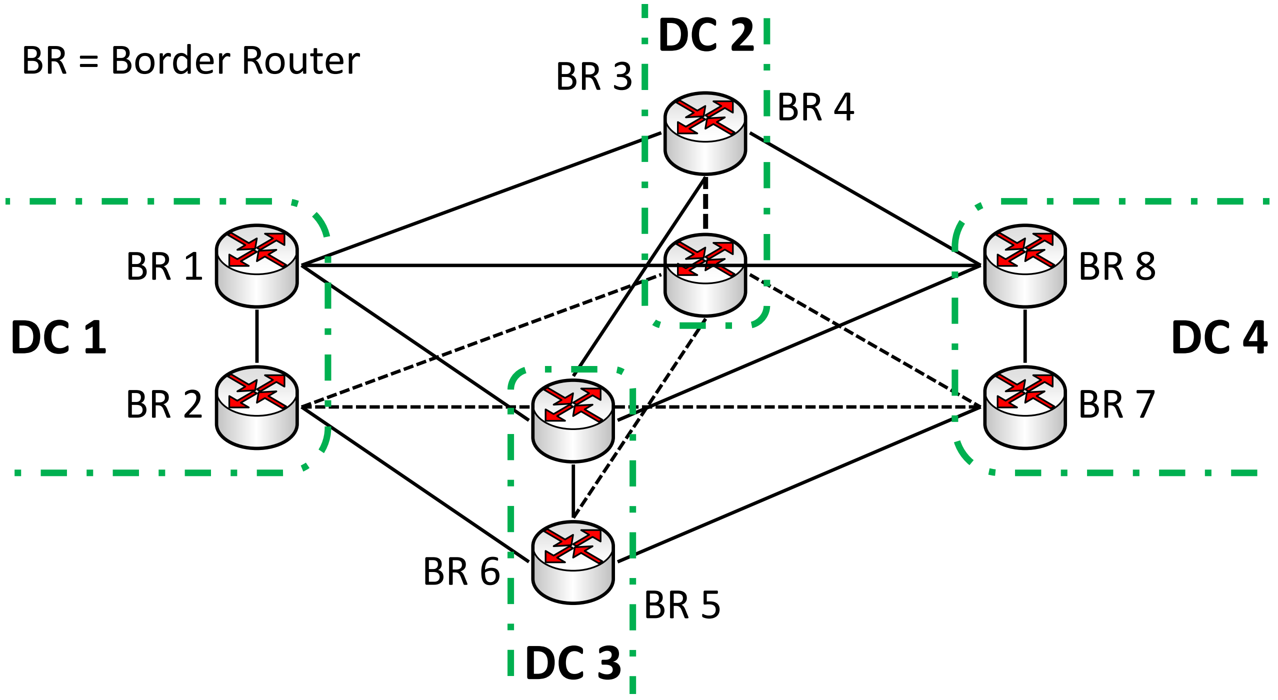
setup

- 8 border routers (BRs)
- 24 (bi-directional WAN) inter-DC links

statesman goal

- upgrade BRs while inter-DC is on

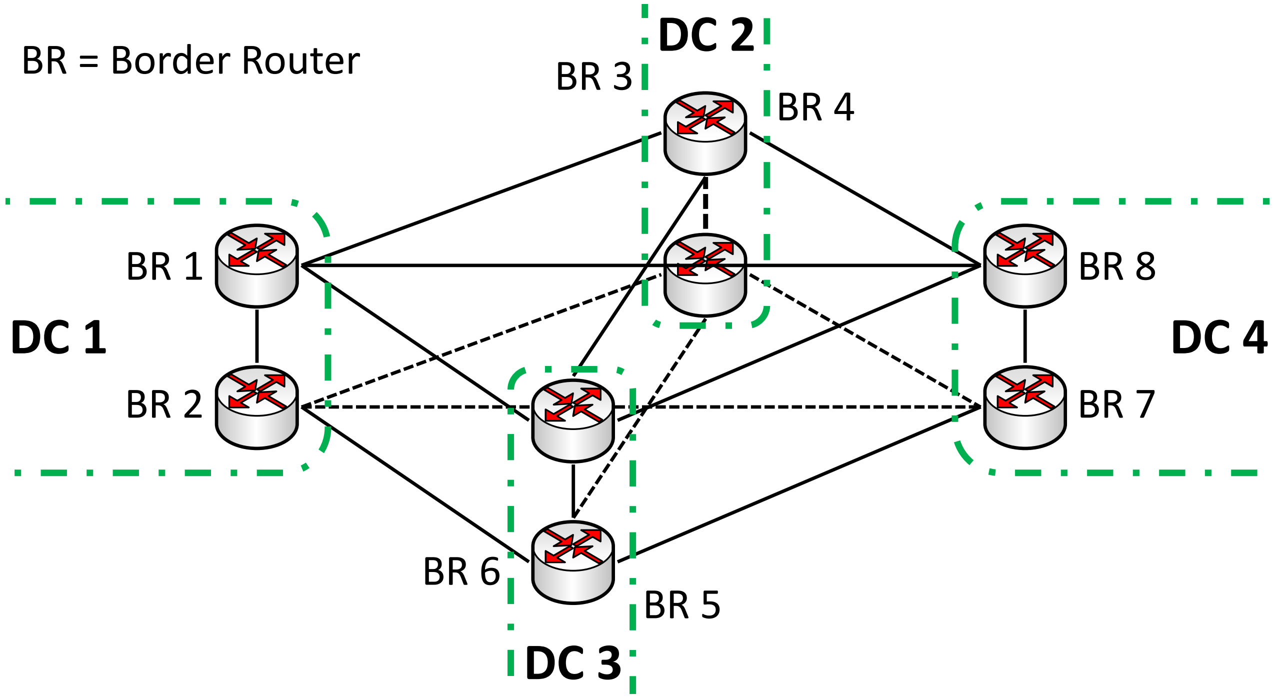
use case: resolving conflicts

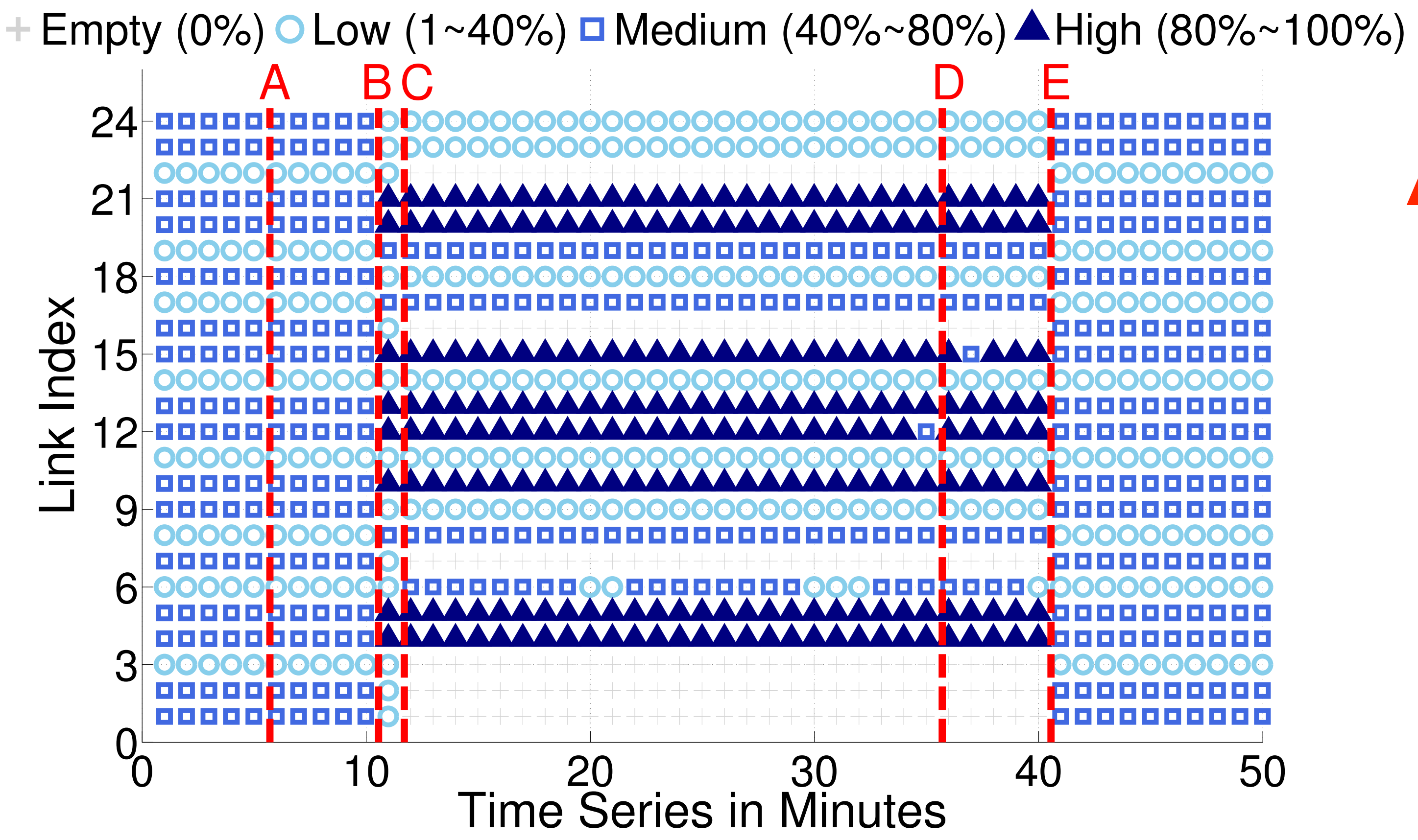


solution: statesman coordinates, by locks, swtich_upgrade, TE

- assign TE low-level lock
- switch_upgrade high-level lock

use case: resolving conflicts





statesman coordinates

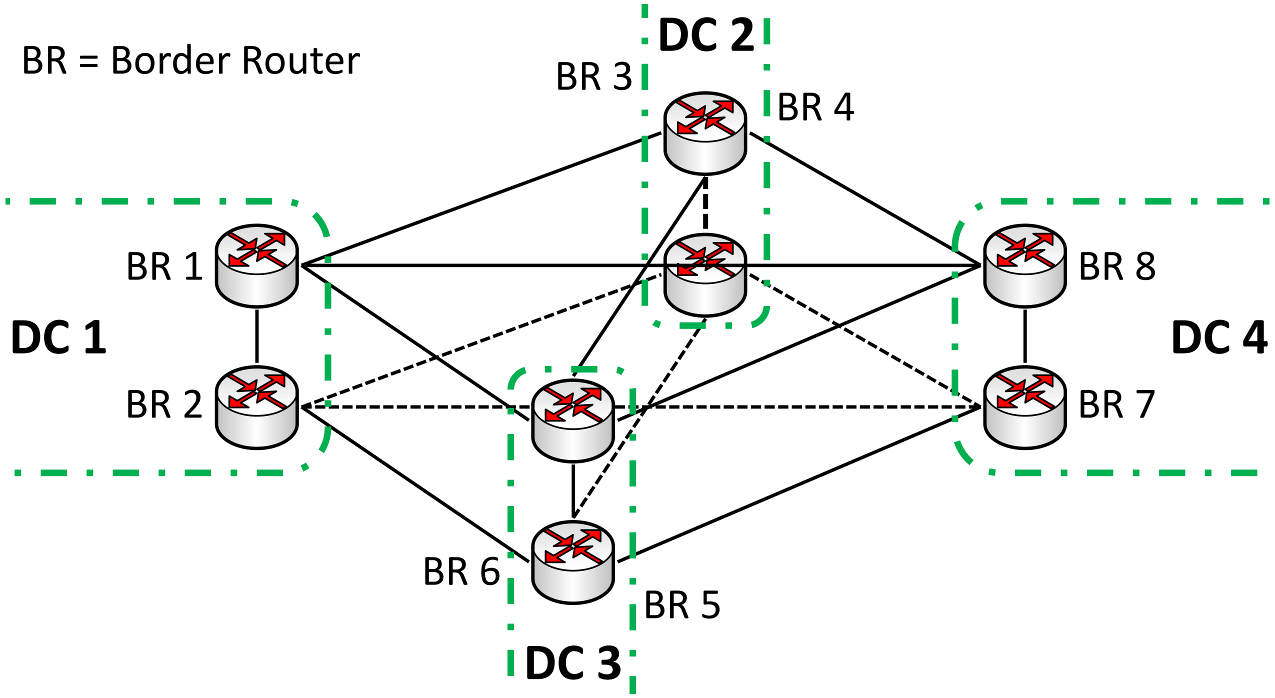
swtich_upgrade, TE

- assign TE low-level lock
- switch_upgrade high-level lock

A

- switch_upgrade acquires high-level lock

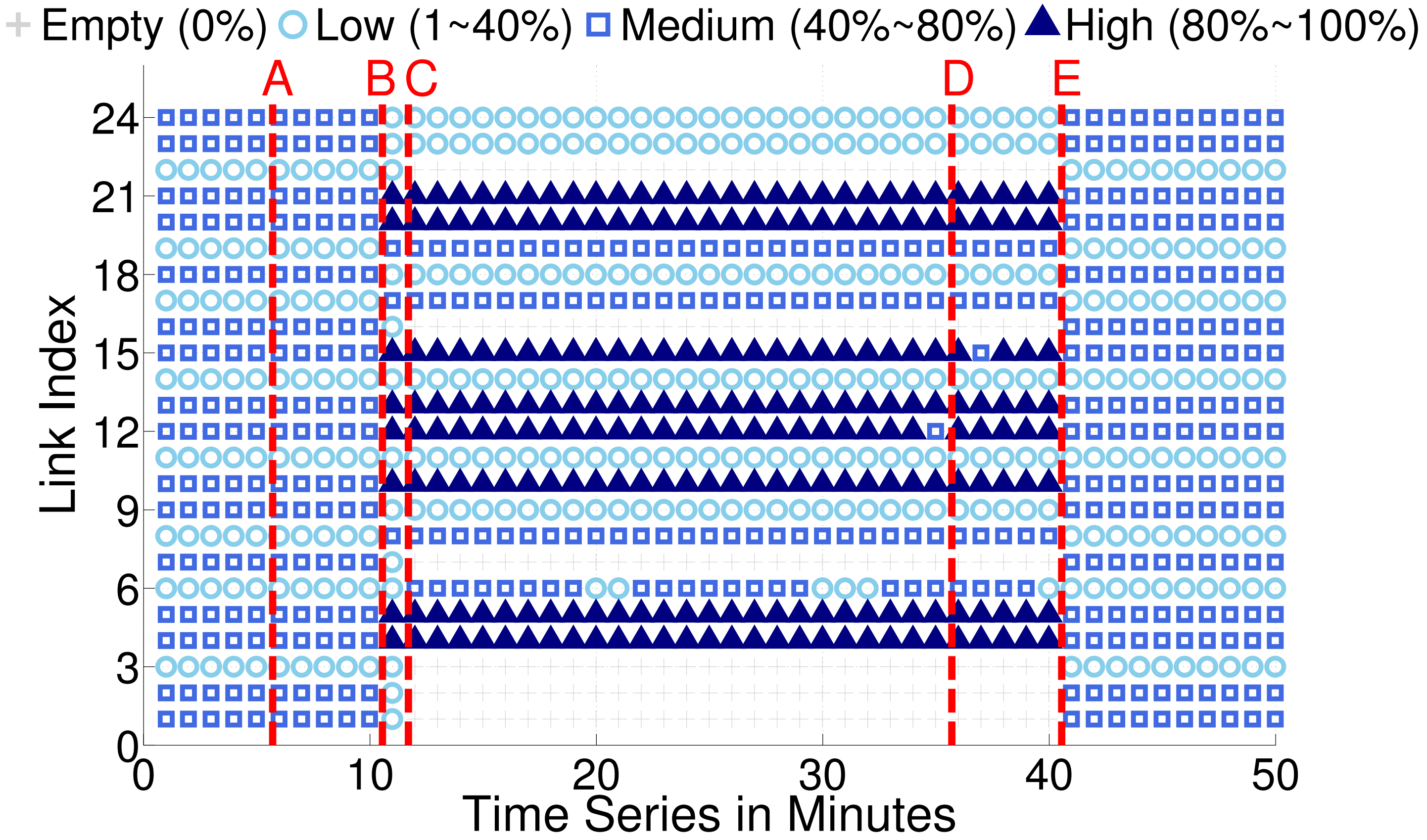
use case: resolving conflicts



statesman coordinates

swtich_upgrade, TE

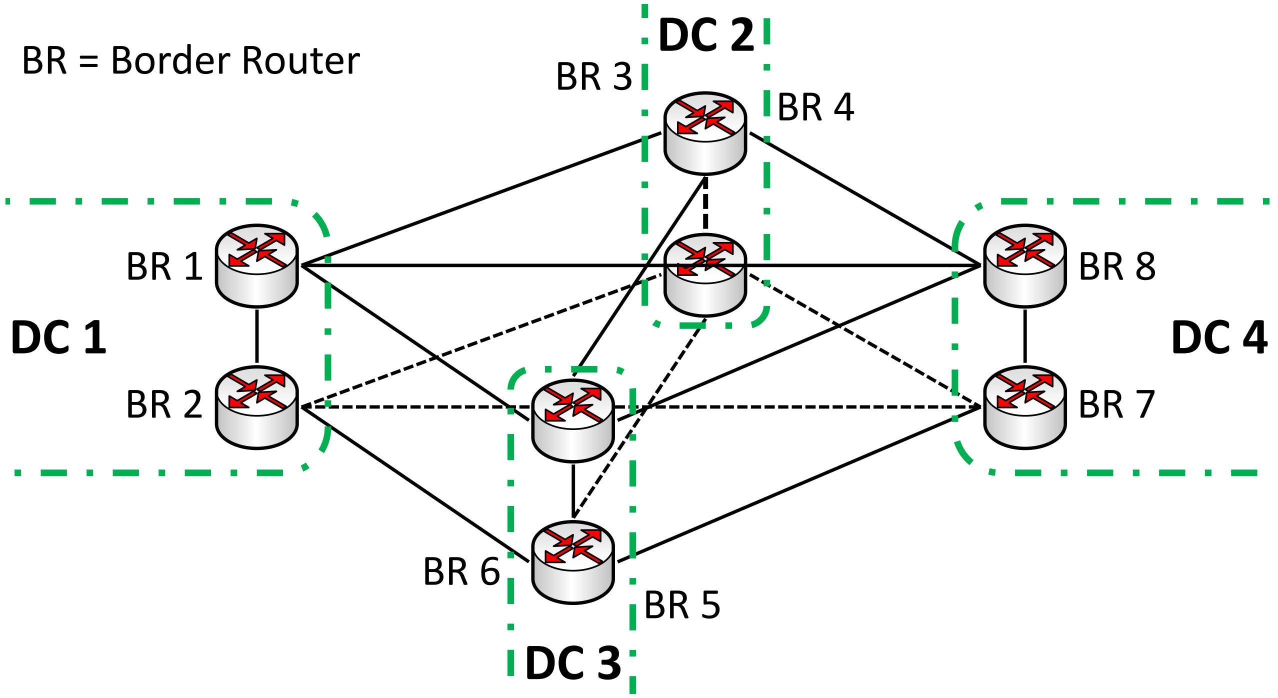
- assign TE low-level lock
- switch_upgrade high-level lock



B

- TE fails to hold low-level lock, moving traffic away

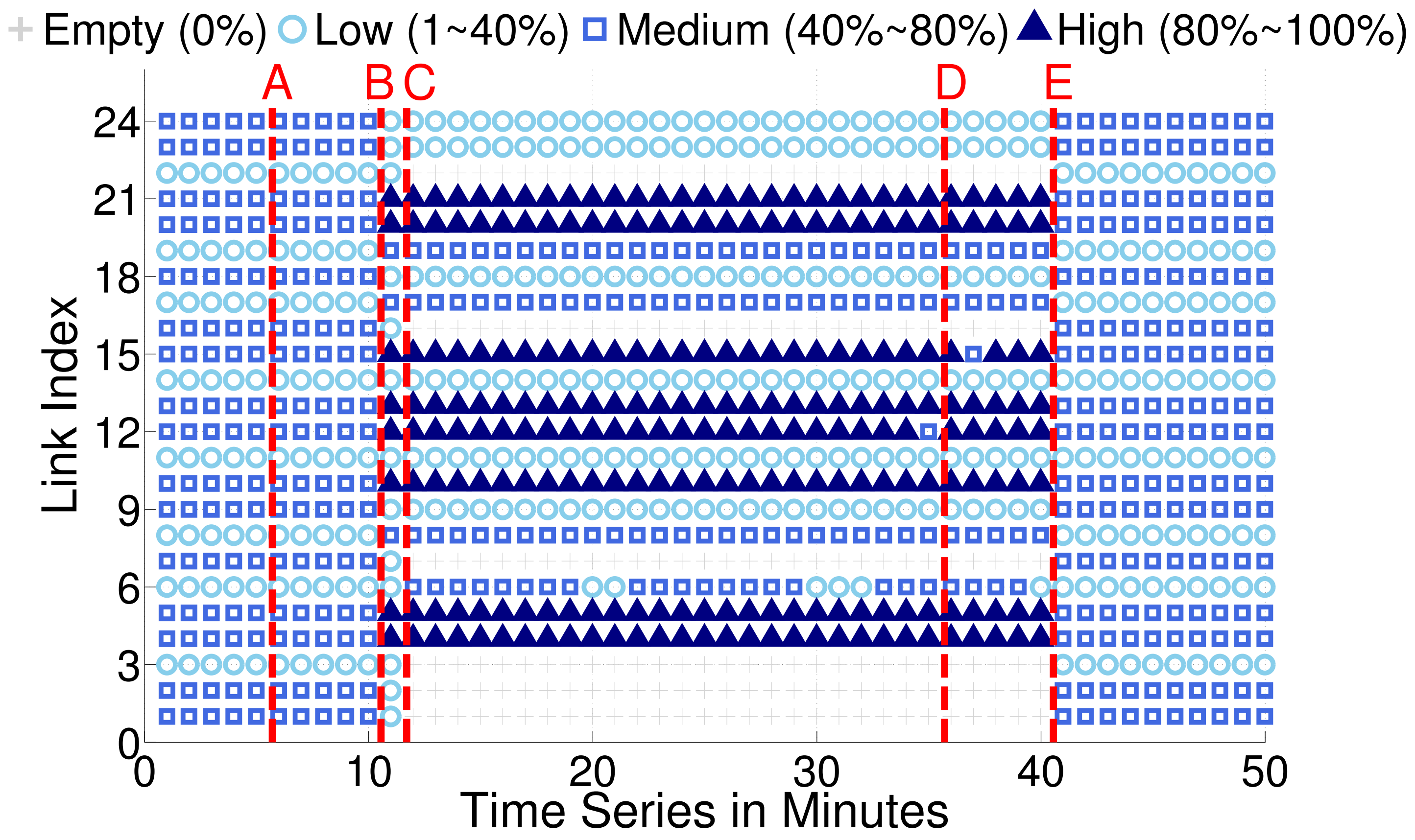
use case: resolving conflicts



statesman coordinates

swtich_upgrade, TE

- assign TE low-level lock
- switch_upgrade high-level lock

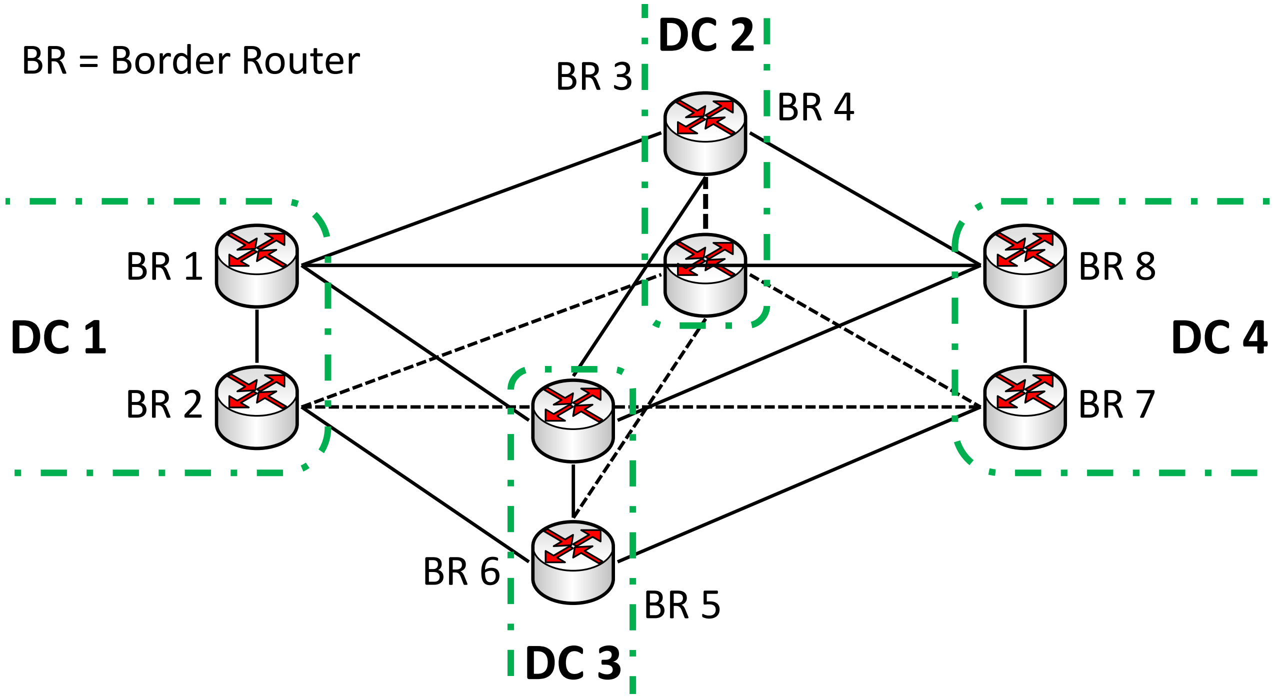


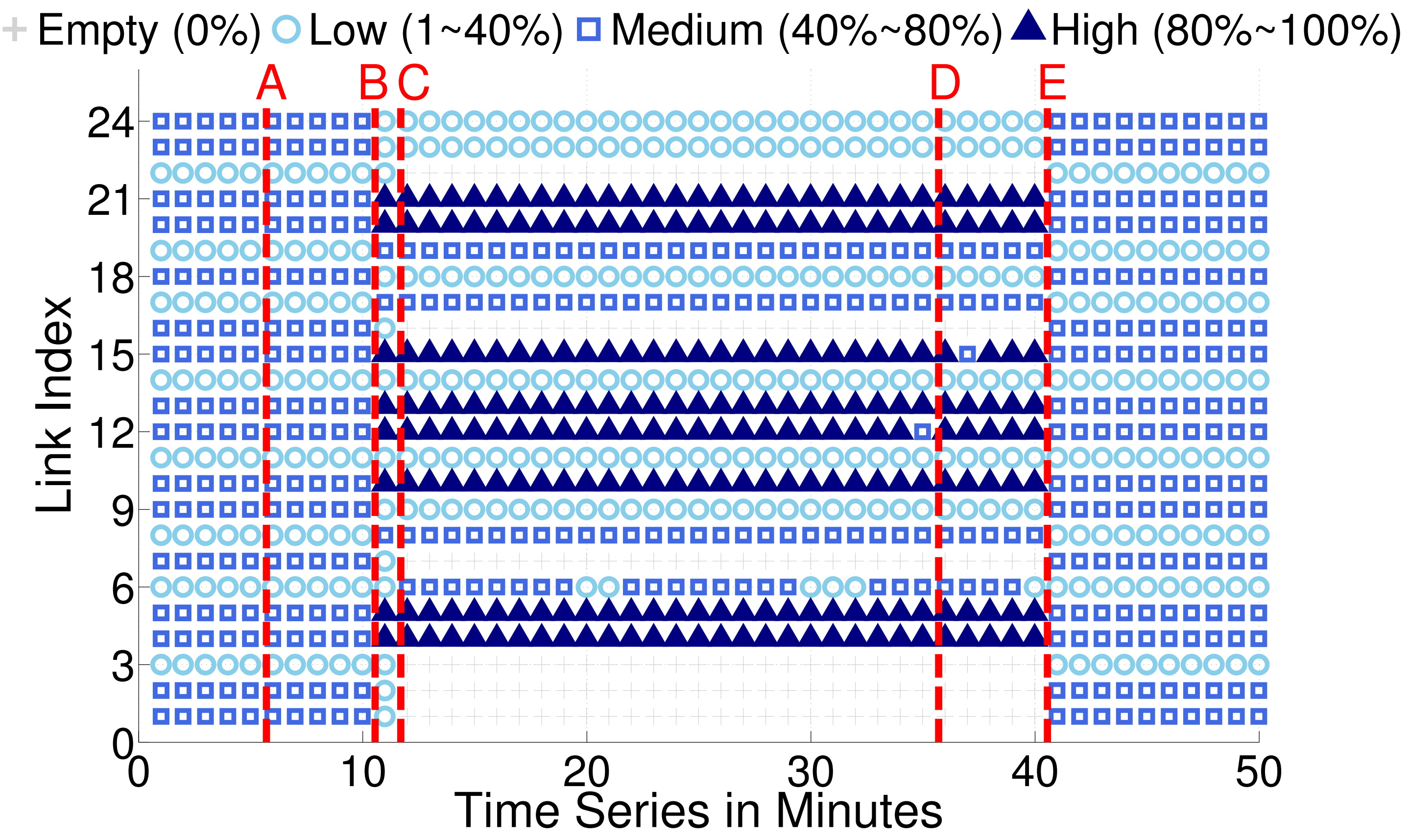
B

- TE fails to hold low-level lock, moving traffic away

how?

use case: resolving conflicts





statesman coordinates

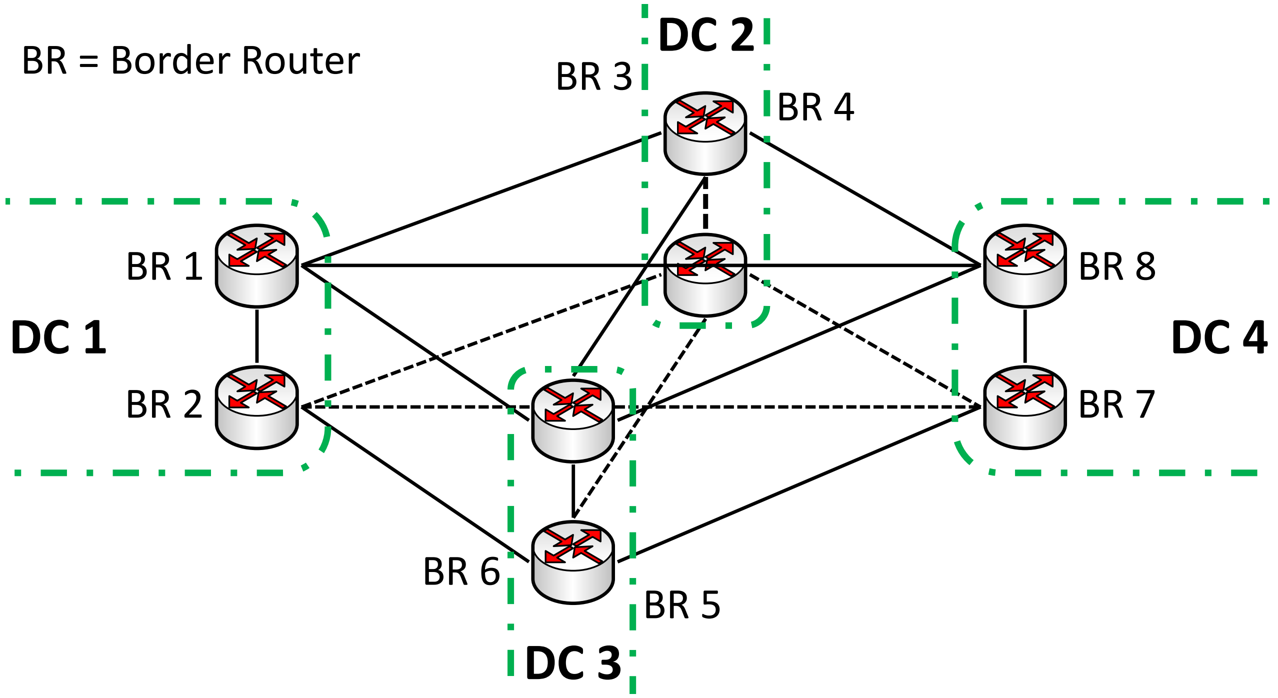
swtich_upgrade, TE

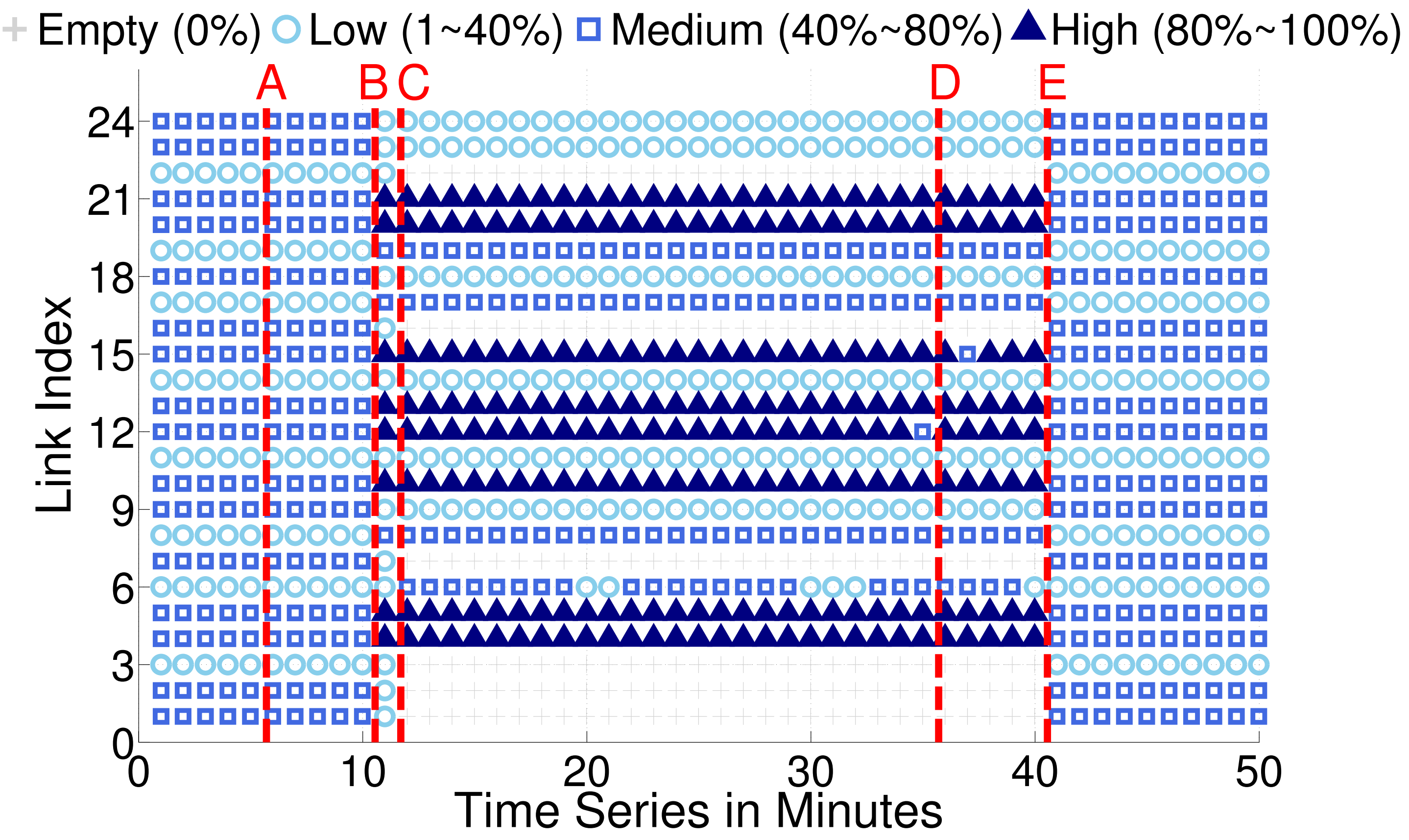
- assign TE low-level lock
- switch_upgrade high-level lock

C, D

- upgrading BRs in progress
- done, releasing high-level lock

use case: resolving conflicts





statesman coordinates

swtich_upgrade, TE

- assign TE low-level lock
- switch_upgrade high-level lock

E

- TE grabs low-level lock, in operation

statesman performance

evaluating latency

- application: (<10ms) negligible
- checker: seconds
- updator: (>50%) dominating