Cross-Layered Multipath Resilient Transport Protocol
for the Future Internet

James P.G. Sterbenz*†‡
司徒傑莫 Джеймс Ф.Г. Штербэнз 송재윤
Yufei Cheng 成宇飞, Anh Nguyễn,

*Department of Electrical Engineering & Computer Science
Information Technology & Telecommunications Research Center
The University of Kansas
†School of Computing and Communications, Infolab 21
Lancaster University
‡The Hong Kong Polytechnic University
jpgs@{ittc.ku.edu|comp.{lancs.ac.uk|polyu.edu.hk}}
https://www.ittc.ku.edu/~jpgs
https://wiki.ittc.ku.edu/resilinets

Where is Kansas?
Geography Lesson

Resilient Transport and Routing
Outline

- ResiliNets initiative
  - multilevel interrealm resilience
  - resilience to attackers
  - resilience to large scale disasters
- ResTP: resilient transport protocol
- GeoDivRP: geodiverse routing protocol

Resilience and Survivability
Motivation and Definition

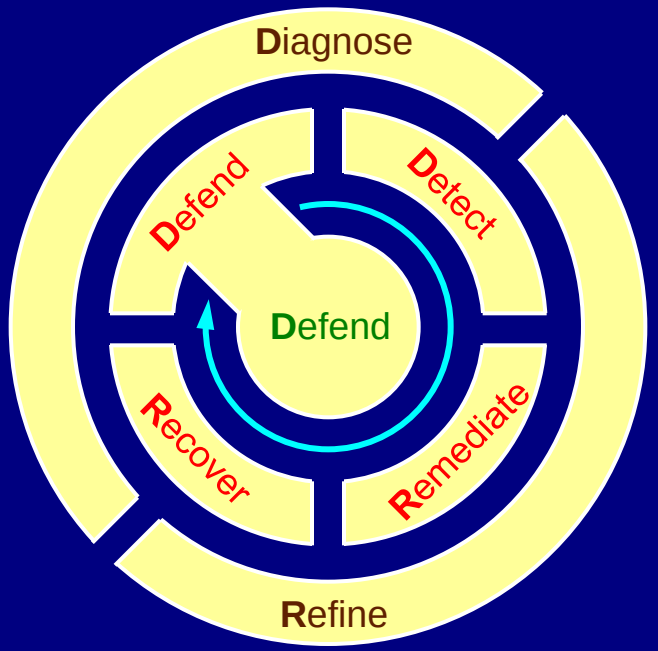
- Increasing reliance on network infrastructure
  - ⇒ Increasingly severe consequences of disruption
  - ⇒ Increasing attractiveness as target from bad guys
- Need *resilience*
  - provide and maintain acceptable service
  - in the face of faults and challenges to normal operation
- Challenges
  - …
  - large scale disasters (natural and human-caused)
  - malicious attacks from intelligent adversaries

ResiliNets Initiative
Goals

- Understand network structure and vulnerabilities
  - develop new models and tools for analysis
- Develop ways to increase network resilience
  - improving existing networks under cost constraints
  - increase cost to attackers
  - Future Internet design
  - validate by analysis, simulation, and experimentation
- Funded primarily by
  - US NSF FIND and GENI programs and NeTS (with Medhi)
  - US DoD
  - EU FP6 and FP7 FIRE programme (with David Hutchison)

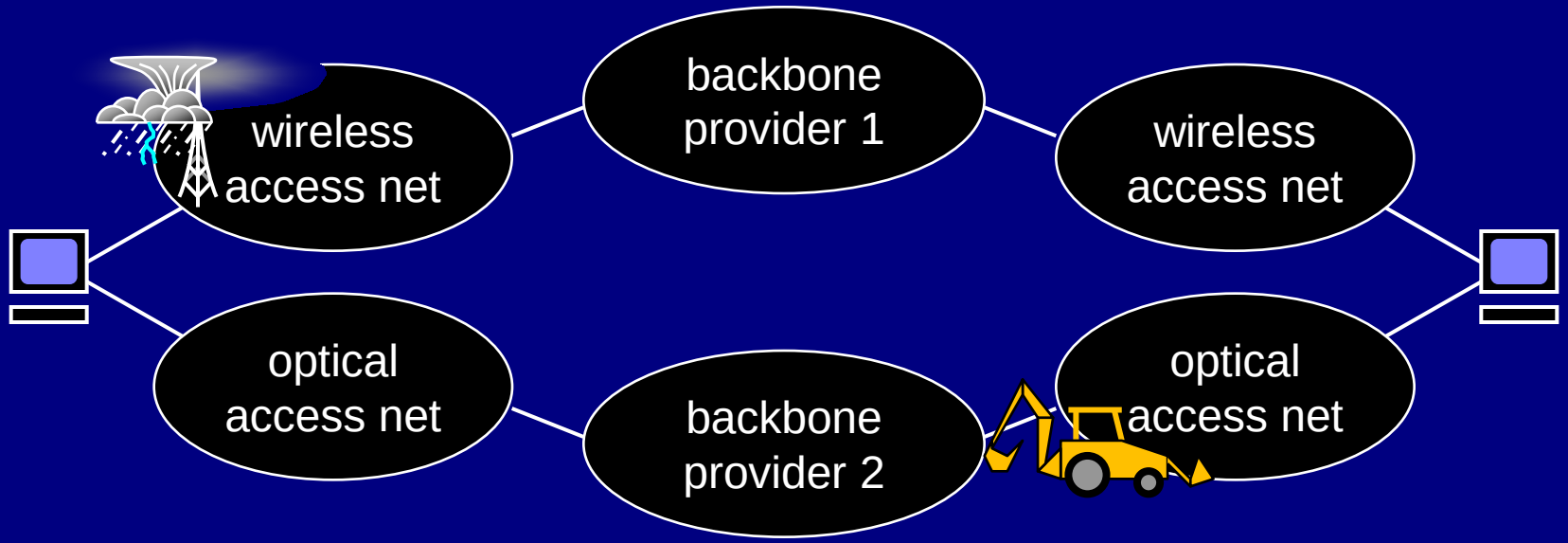
ResiliNets Strategy
D^2R^2 + DR

- Two phase strategy for resilience
- Real time control loop: D^2R^2
  - defend
    - passive
    - active
  - detect
  - remediate
  - recover
- Background loop: DR
  - diagnose
  - refine



[Wiki 2005, *ComNet* 2010]

ResiliNets Principles
Redundancy, Diversity, Heterogeneity



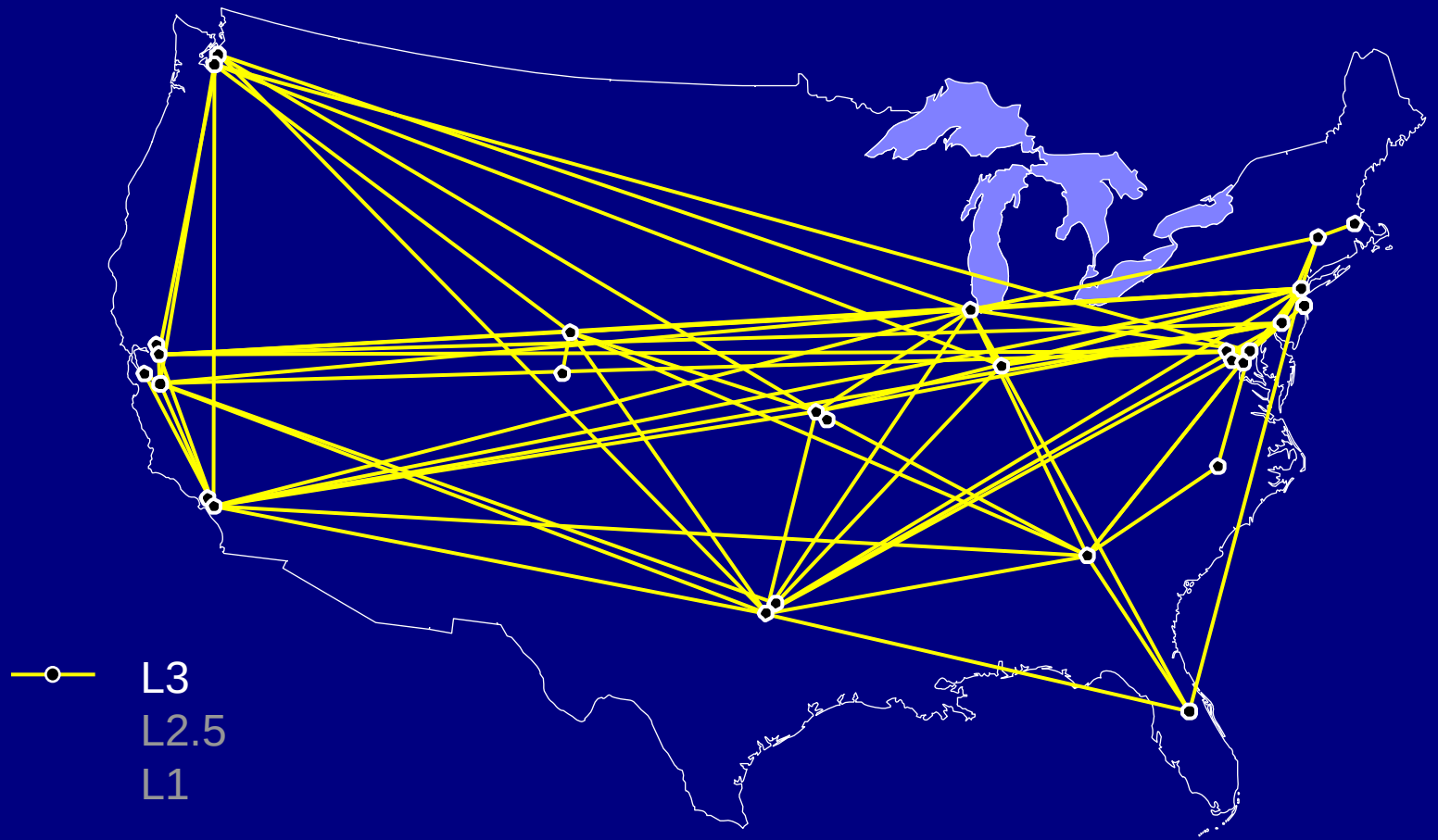
- Diversity
  - mechanism (wired & wireless), provider, *geographic path*
- Multipath transport
  - spreading (erasure code) or as hot-standby

Multilevel Structural Diversity
Multilevel Interrealm Resilience

- ResiliNets review
- Multilevel interrealm resilience
  - resilience to attackers
  - resilience to large scale disasters
- ResTP: resilient transport protocol
- GeoDivRP: geodiverse routing protocol

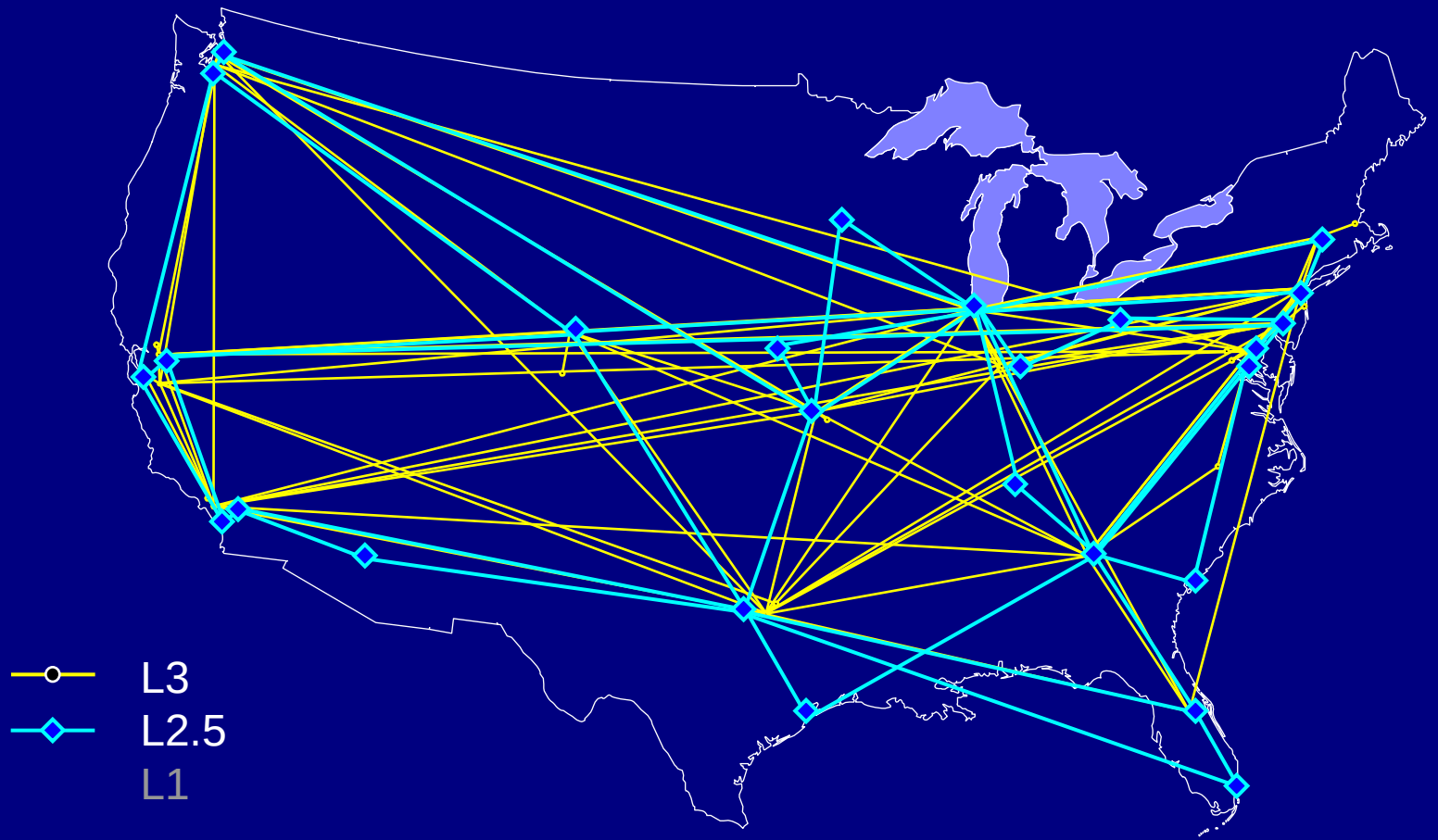
Multilevel Network Topology

Example: Sprint L3 IP PoP Topology



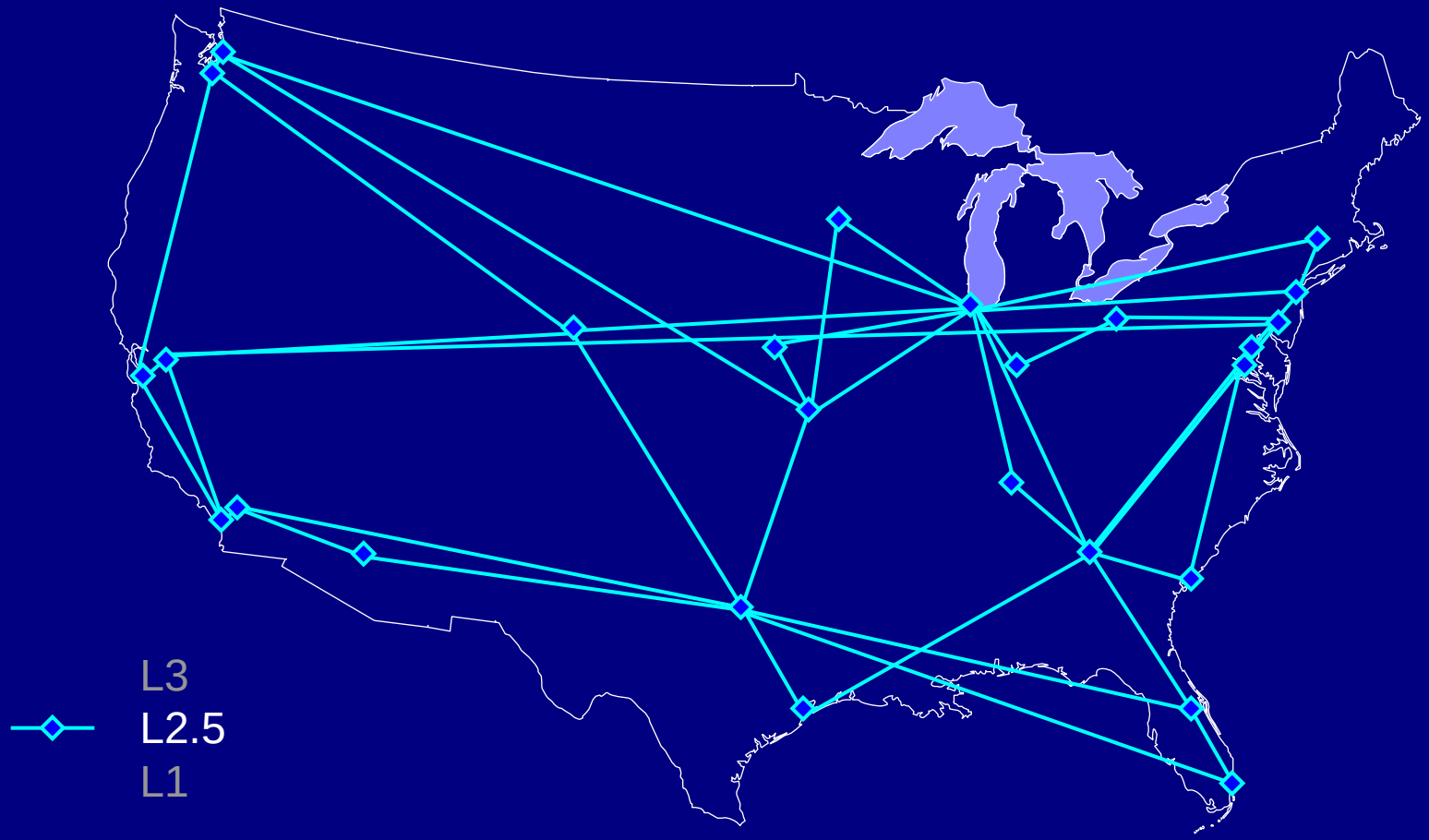
Multilevel Network Topology

Example: Sprint L3 overlay on L2.5



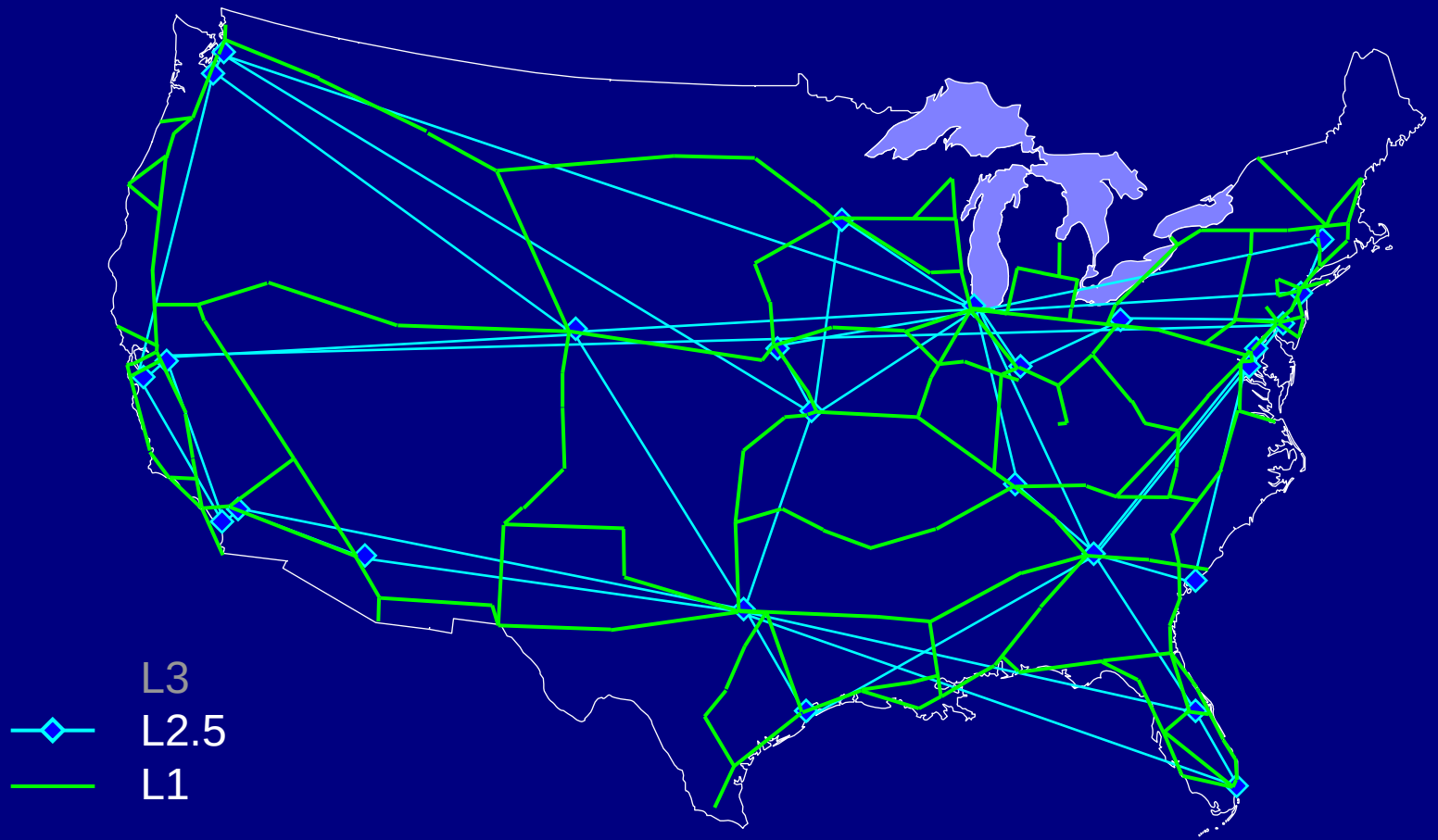
Multilevel Network Topology

Example: Sprint L2.5 MPLS PoP Topology



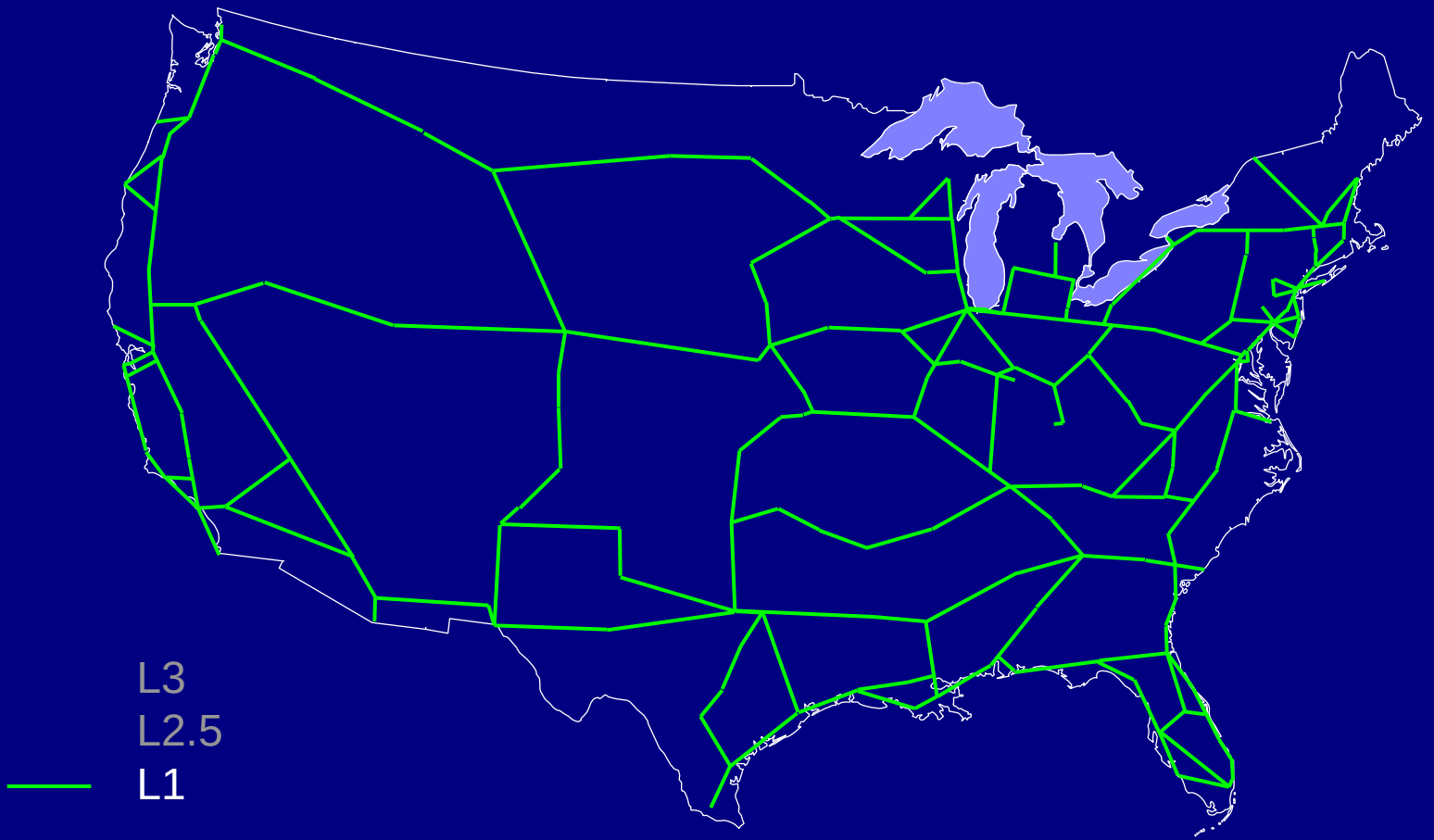
Multilevel Network Topology

Example: Sprint L2.5 overlay on L2/1



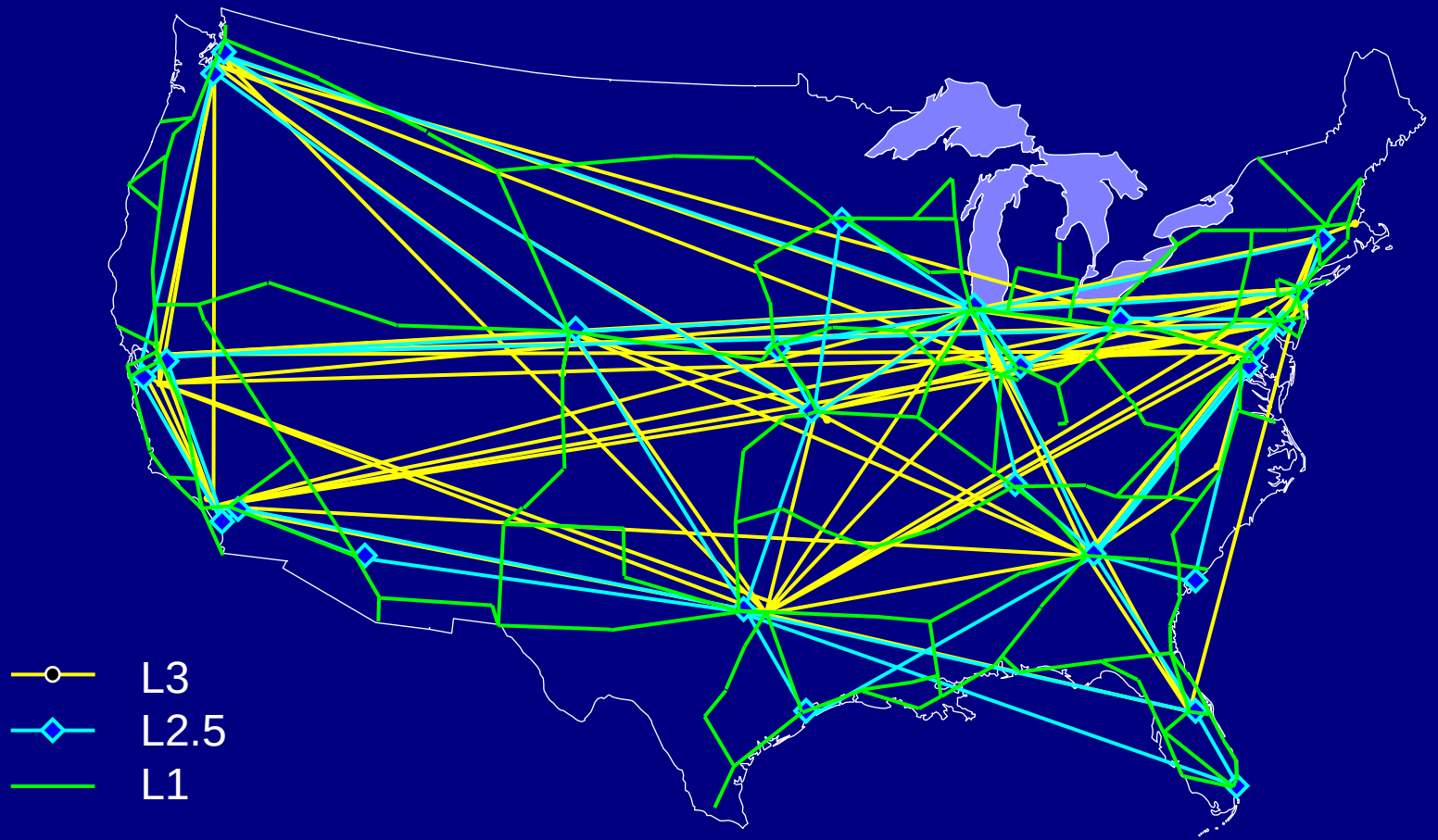
Multilevel Network Topology

Example: Sprint L1 Physical Fiber Topology

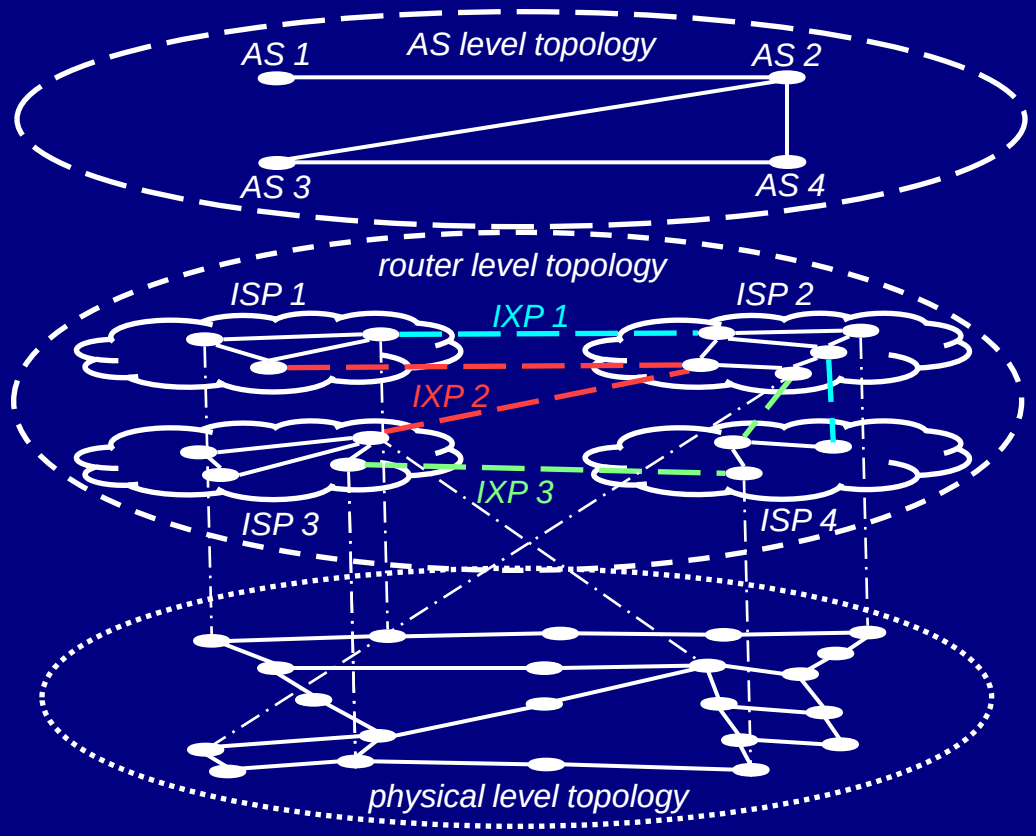


Multilevel Network Topology

Example: Sprint L1–3 Topology



Multilevel Network Analysis
Abstraction of Internet Topology



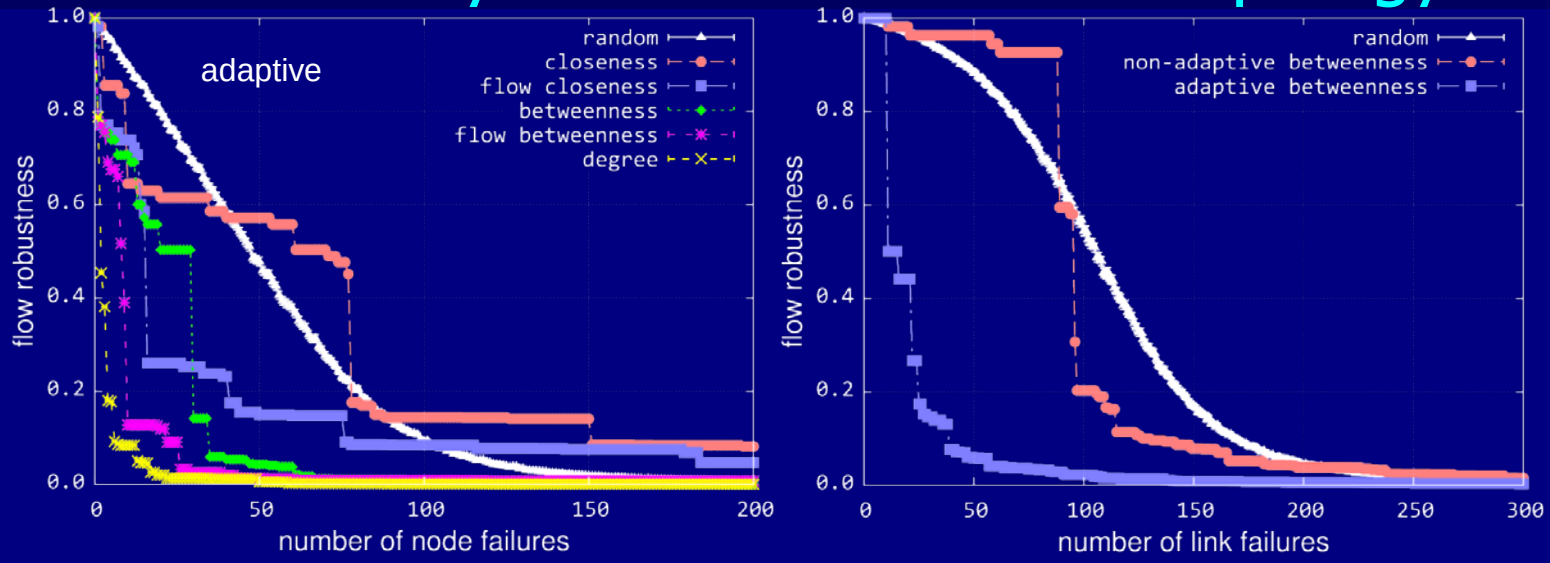
[DRCN 2013]

Multilevel Structural Diversity
Resilience to Large-Scale Disasters

- ResiliNets review
- Challenge Taxonomy
- Multilevel interrealm resilience
  - resilience to attackers
  - resilience to large scale disasters
- ResTP: resilient transport protocol
- GeoDivRP: geodiverse routing protocol

Multilevel Resilience

Effect of Physical Failures on L3 Topology



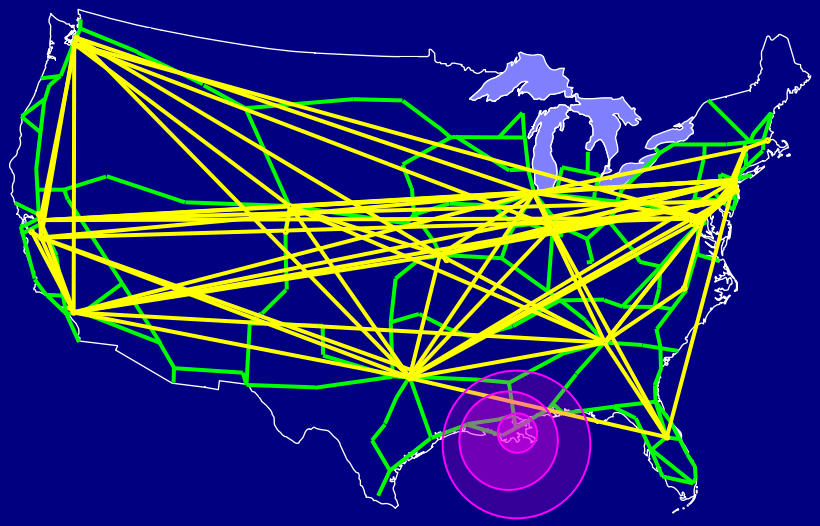
- Attacks against physical infrastructure
  - based on centrality (importance) metrics
  - adaptive recomputes metrics after each node failure)
- Analysis of impact on higher layer flows
  - heuristics to add elements under cost constraints

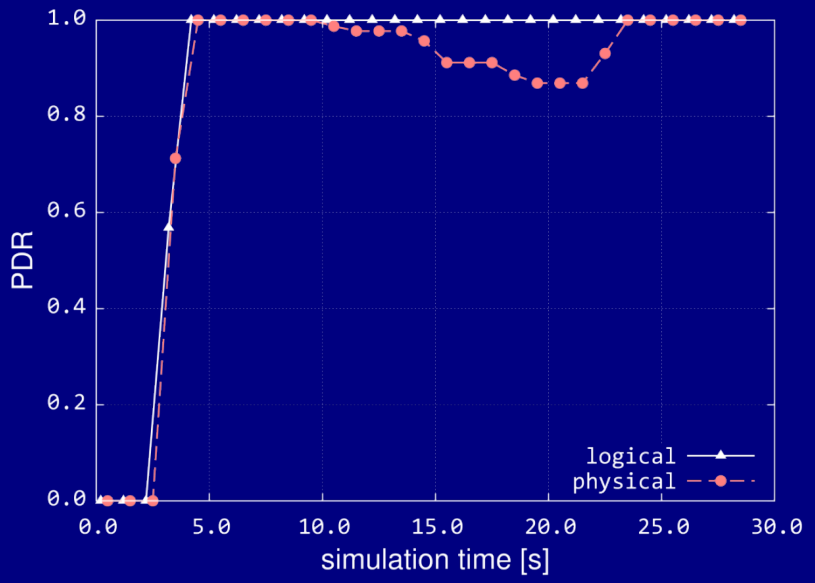
Multilevel Structural Diversity
Resilience to Large-Scale Disasters

- ResiliNets review
- Challenge Taxonomy
- Multilevel interrealm resilience
  - resilience to attackers
  - resilience to large scale disasters
- ResTP: resilient transport protocol
- GeoDivRP: geodiverse routing protocol

Simulation Analysis

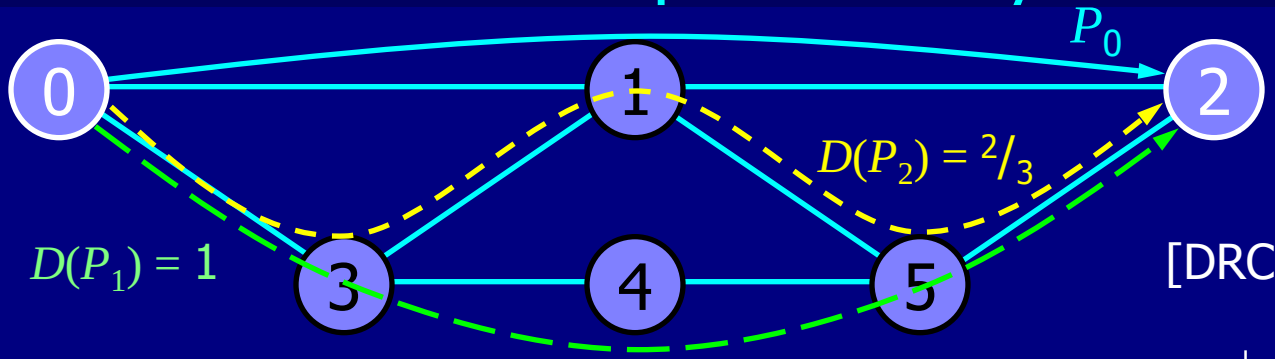
Example: Multilevel Analysis of Disaster





- Hurricane disaster in New Orleans area
- Destruction of physical infrastructure
- Effect on IP-layer network services

Resilience Analysis
Path and Graph Diversity



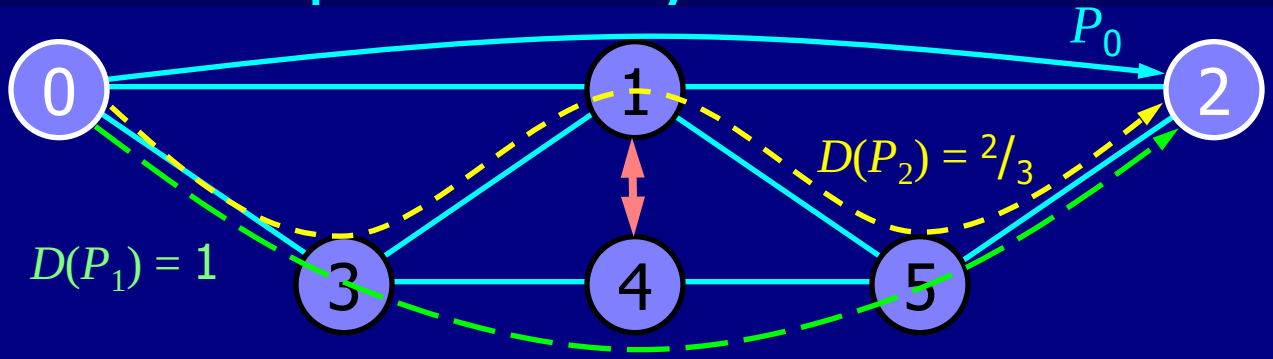
[DRCN 2009]

- Path diversity
  - measure of links and nodes in common

$$D(P_k) = 1 - \frac{|P_k \cap P_0|}{|P_0|}$$

- EPD: effective path diversity [0,1)
  - normalised diversity with respect to a single shortest path
  - measure of E2E flow resilience
- TGD: total graph diversity is average of EPD
  - for all pairs: quantifies available diversity in graph

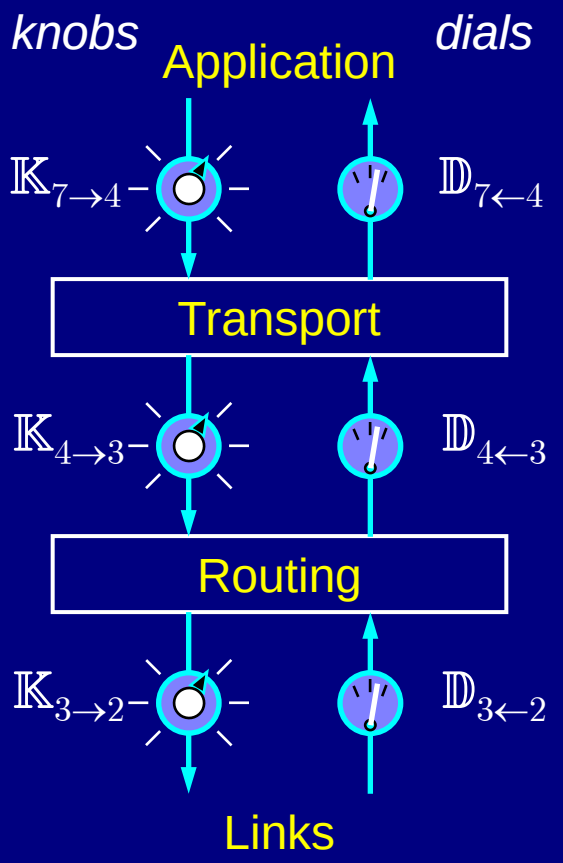
Resilience Analysis
Path and Graph Diversity with Distance Metric



- cTGD: compensated TGD
  - weighted to be predictive of flow robustness [RNDM 2010]
  - algebraic connectivity also fair predictor or flow robustness
- GeoPath diversity
  - distance d between paths beyond source and destination
  - GeoDivRP: $(k, d, [s,t])$ multipath geographic routing
    - number of paths k [RNDM 2013]

ResiliNets Protocols
Cross-Layer Model: Generic

- *Knobs* $\mathbb{K}_{i \rightarrow i-1} = \{k_i\}$ influence behaviour to levels below
- *Dials* $\mathbb{D}_{i+1 \leftarrow i} = \{d_i\}$ expose characteristics to upper levels
- Levels (of significance to ResiliNets)
  - 8: social
  - 7: application
  - 4: end-to-end transport
  - 3i: inter-realm (domain)
  - 3r: routing
  - 3t: logical topology
  - 2: hop-by-hop links
  - 1: physical topology



Multilevel Structural Diversity
Resilience Multipath Transport

- ResiliNets review
- Challenge Taxonomy
- Multilevel interrealm resilience
  - resilience to attackers
  - resilience to large scale disasters
- ResTP: resilient transport protocol
- GeoDivRP: geodiverse routing protocol

Resilient Transport: ResTP
Overview

- ResTP: Resilient Transport Protocol [CFI 2015]
  - flexible and composable (ala TP++ [Feldmeier, MCauley])
- Flexible and composable
  - flow setup and management
    - including multipath support
  - error control
  - transmission (flow and congestion) control
- Cross-layered
  - applications specify service and threat model
  - behaviour based on path characteristics
  - specifies path requirements to GeoDivRP

Resilient Transport: ResTP
Flow Modes

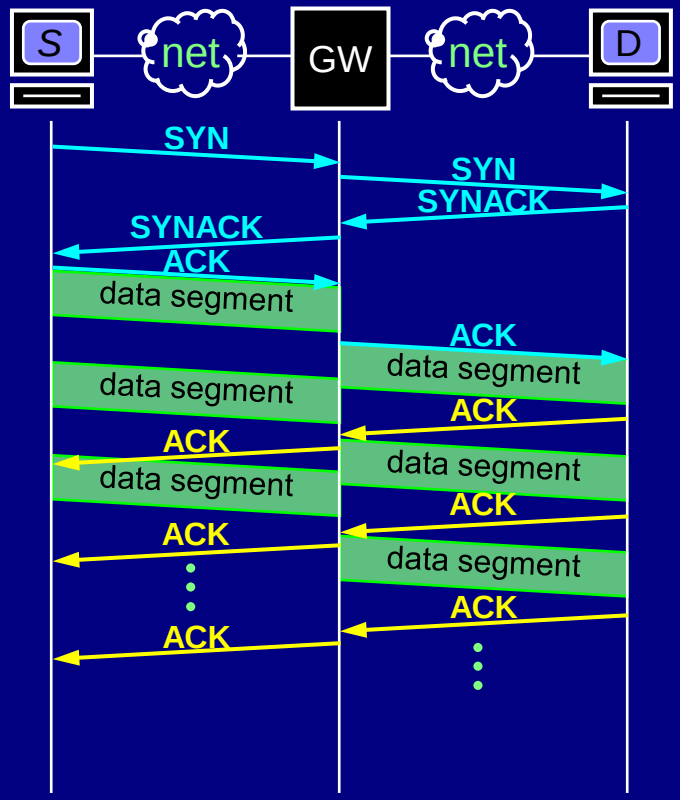
- Multiple flow modes based on `Flow` flags
  - connection oriented `CON=1`
  - opportunistic connections `CON=1 OPT=1`
    - signalling overlaps data
  - custody transfer at realm gateways `CON=1 CXF=1`
    - for DTNs
  - signalled flow with datagrams `CON=0 ARQ=0`
  - individual datagrams `CON=0`

Resilient Transport: ResTP
Transfer Modes

- Multiple transfer modes based on `Flow` & `Error` flags
  - fully reliable (E2E 3-way handshake & ACKs) `CON=1 CXF=0`
    - may be opportunistic flow setup
    - may use HARQ
  - opportunistic connections `OPT=1`
    - signalling overlaps data
  - nearly reliable
    - custody transfer at realm gateways for DTNs `CON=1 CXF=1`
    - AeroTP subset of ResTP uses this
  - quasi-reliable: E2E FEC giving statistical reliability
  - unreliable signalled flow with datagrams `CON=0 ARQ=0`
  - unreliable individual datagrams `CON=0`

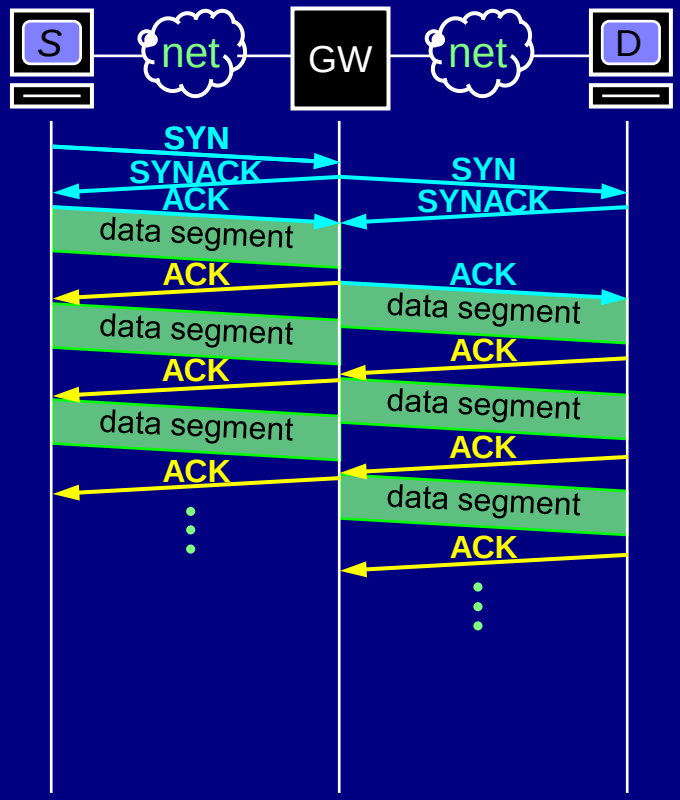
Resilient Transport: ResTP
Fully-Reliable Transfer

- Fully reliable
  - E2E connection management
  - E2E ACKs



Resilient Transport: ResTP
Nearly-Reliable Transfer

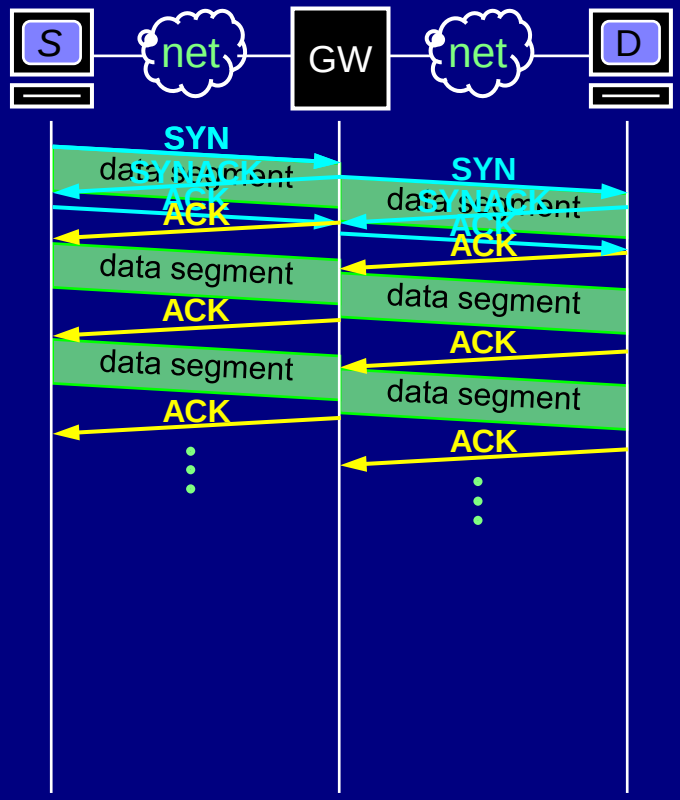
- Nearly reliable
  - e2e connection management
  - e2e ACKs
    - custody transfer at gateways



Resilient Transport: ResTP

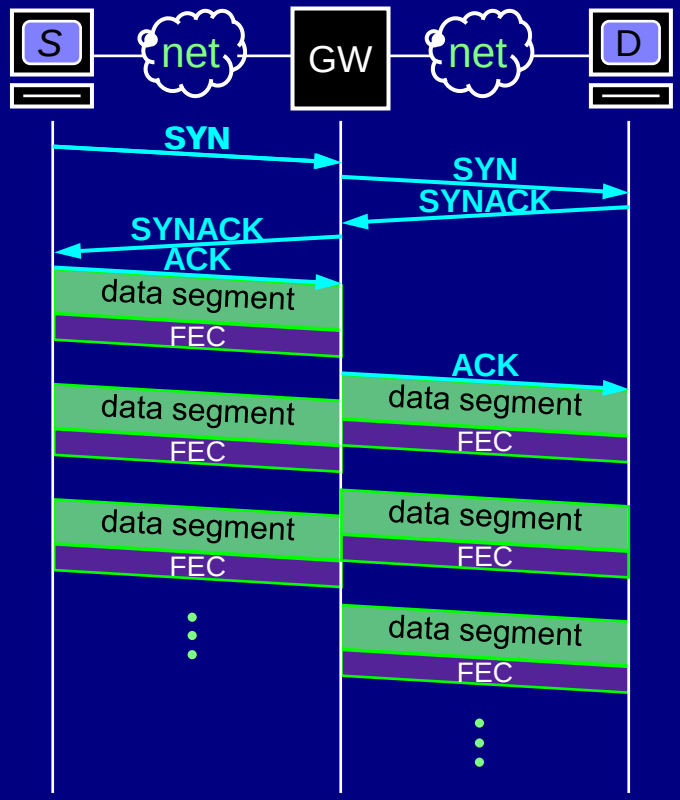
Nearly-Reliable Opportunistic Transfer

- Opportunistic signalling
  - data overlaps
- Nearly reliable
  - e2e connection management
  - e2e ACKs
    - custody transfer at gateways



Resilient Transport: ResTP
Quasi-Reliable Transfer

- Quasi reliable
  - E2E (or e2e) L4 FEC
  - no data ACKs



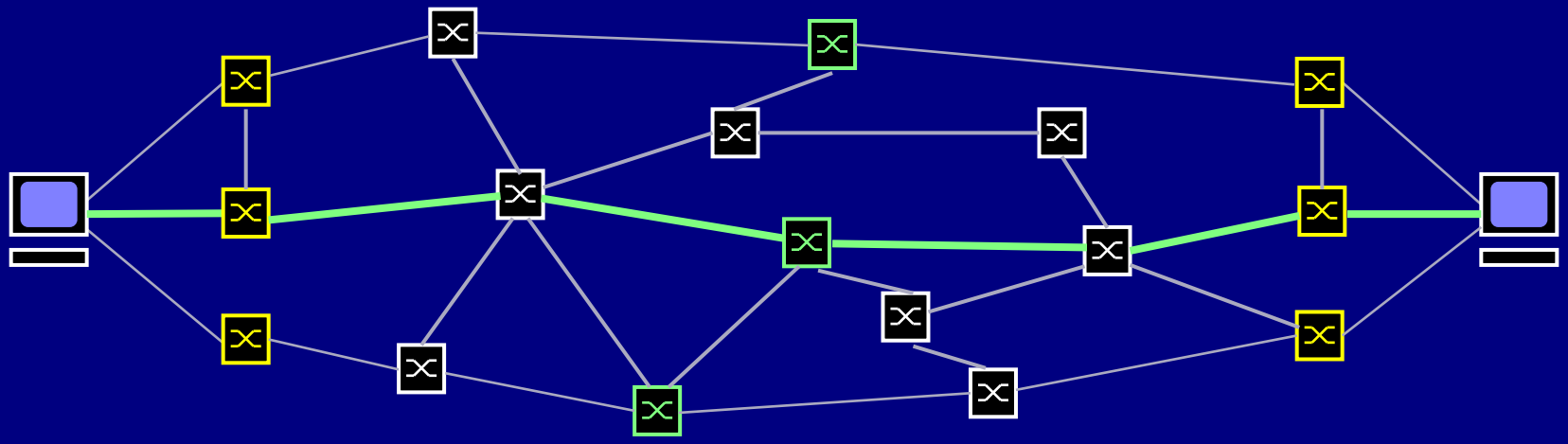
ResTP Operation
Error Control

- Per subflow err control modes based on `Error` flags
  - ARQ for reliable service
    - SACK, MACK, NAK, SNACK (SCPS-style)
  - HARQ for reliable service on lossy path
  - adaptive FEC for quasireliable service
  - none for unreliable service
- Sequence numbers are TPDU numbers
  - not byte sequence numbers as with TCP

Resilient Transport: ResTP
Error Control

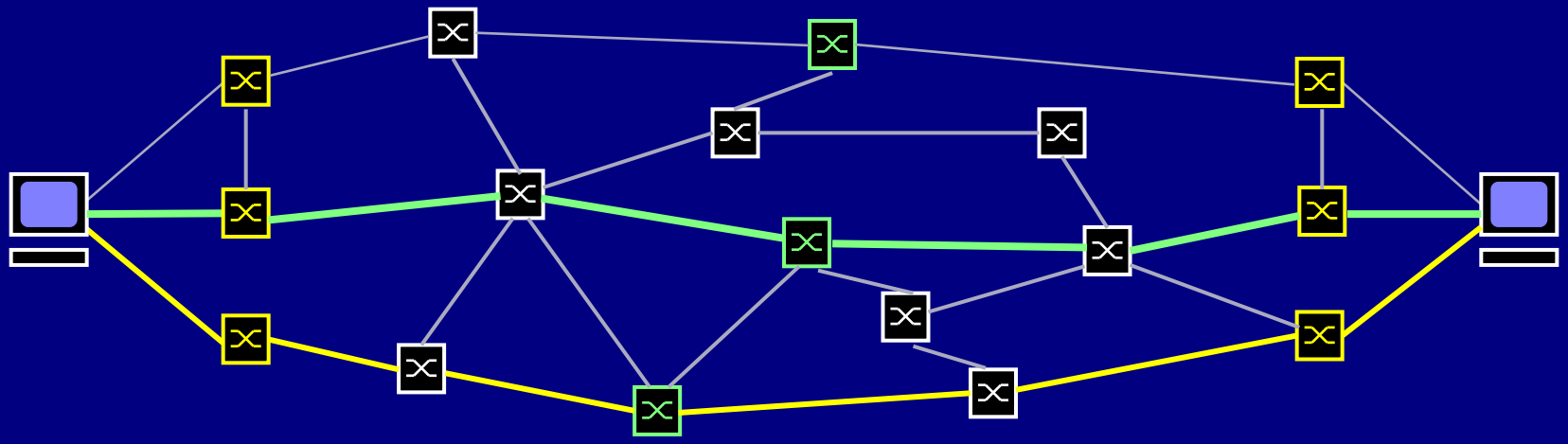
- Multipath modes
  - unipath
    - alternate path may be added on-demand
  - alternate path as hot-standby
  - erasure coding across k paths (typically $k = 3$)
    - best coding for large skew? [UiB?]

ResTP Multipath
Unipath Mode



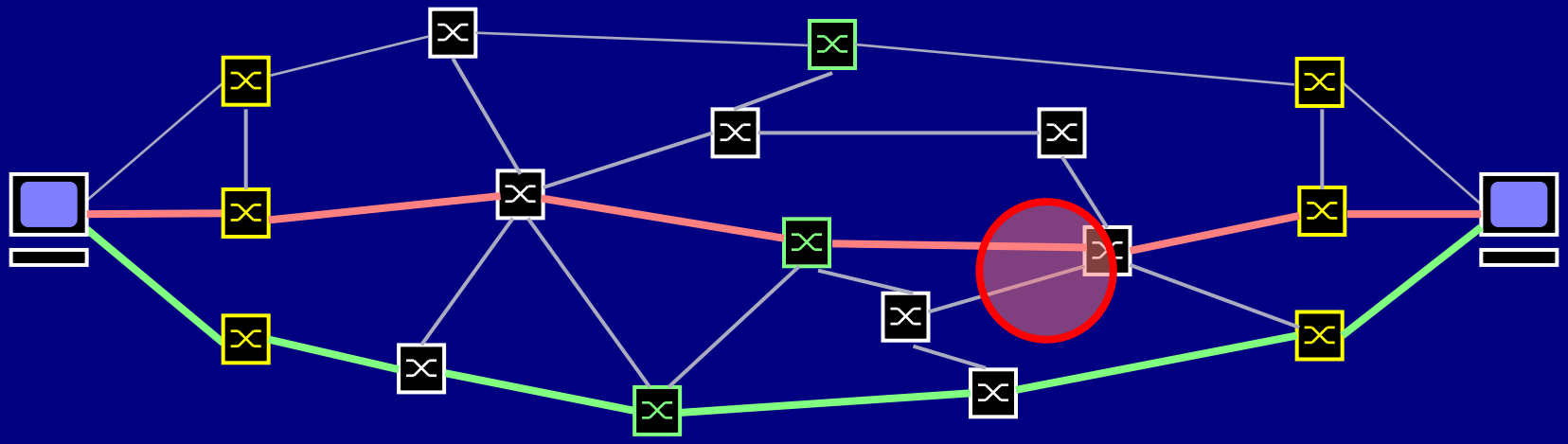
- ResTP unipath
  - MP? = 0 MPM = 0 k = 1
  - conventional unipath E2E communication
  - requires traditional recovery
    - fast restoration before connection timeout
    - E2E connection retry

ResTP Multipath
Hot-Standby Mode



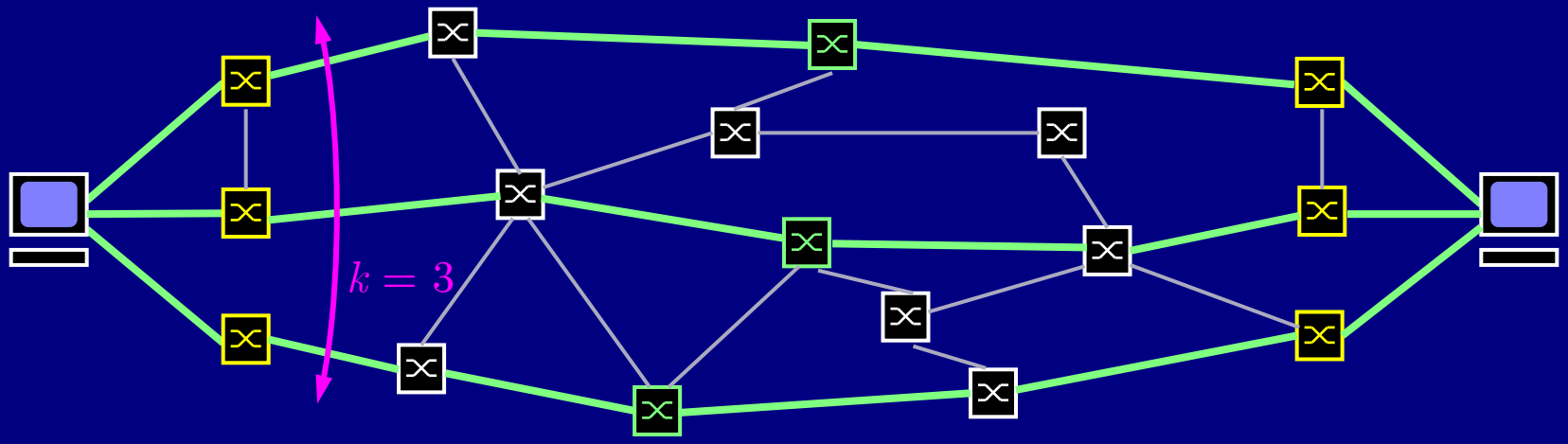
- ResTP hot standby mode
  - MP? = 1 MPM = 0 k = 2
  - information transferred only on primary subflow
  - no information transfer on secondary flow

ResTP Multipath
Hot-Standby Mode



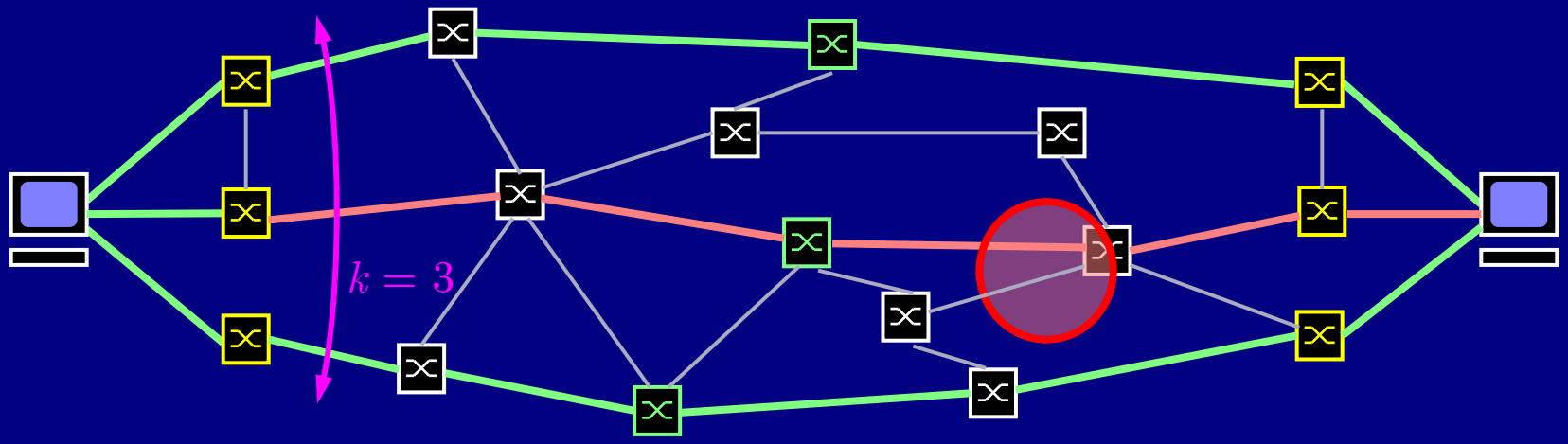
- ResTP hot standby mode
  - MP? = 1 MPM = 0 k = 2
  - information transferred only on primary subflow
  - no information transfer on secondary flow
    - unless primary fails
    - faster restoration but some TPDU loss

ResTP Multipath
Spreading Mode



- ResTP multipath spreading mode
  - `MP? = 1  MPM = 1  k =` k
  - spread information across k subflows (e.g. k =3)
  - coding specified in `E2E coding scheme`

ResTP Multipath
Spreading Mode



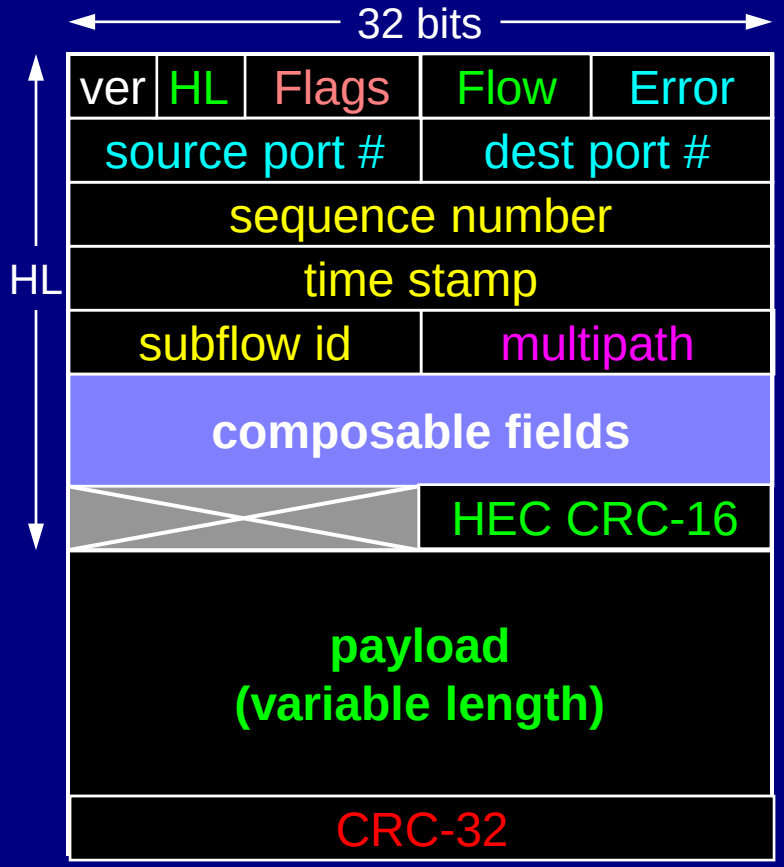
- ResTP multipath spreading mode
  - `MP? = 1  MPM = 1  k =` k
  - spread information across k subflows (e.g. k =3)
  - coding specified in `E2E coding scheme`
  - E2E flow survives path loss with no E2E loss or disruption

Resilient Transport: ResTP
Transmission Control

- Transmission control modes [future work: UiO?]
  - subflow congestion control
  - subflows should generally not share nodes nor links

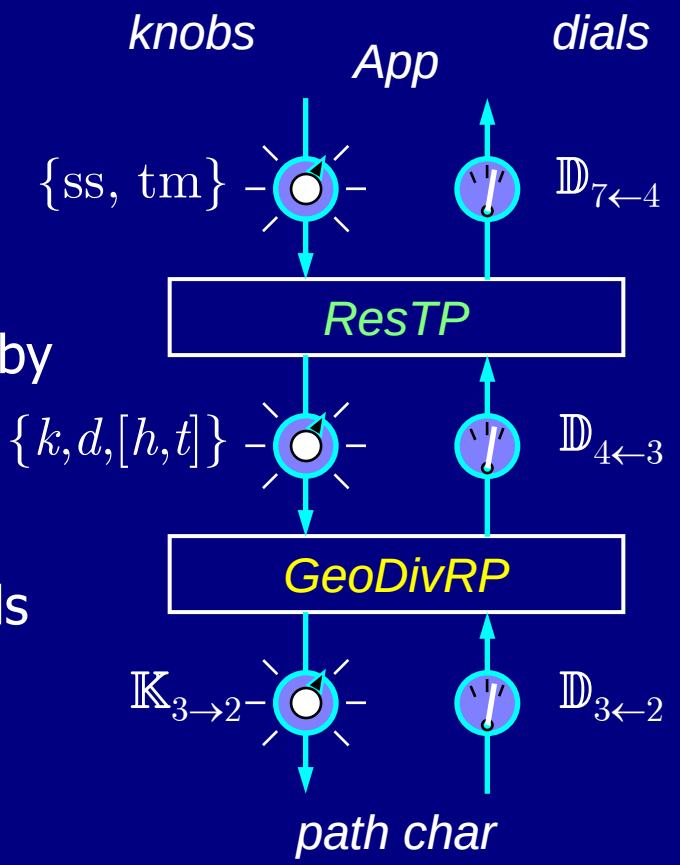
ResTP
TPDU Format

- Header
  - ResTP version
  - header length in words
  - type flags
  - flow type
  - error control
  - src and destination app ports
  - TPDU sequence #
  - timestamp
  - subflow id
  - multipath
  - HEC header check



ResiliNets Protocols
Cross-Layer Model: ResTP/GeoDivRP

- Application
  - $\mathbb{K}_{7\to4} = \{\mathrm{ss}, \mathrm{tm}\}$
    service spec and threat model
- E2E Transport: ResTP
  - erasure spreading vs. hot standby
  - FEC vs. HARQ vs. ARQ
  - $\mathbb{K}_{4\to3} = \{k,\ d,\ [h,\ t]\}$
    k-path diversity over distance d
    opt. stretch h and skew t bounds
- Routing: GeoDivRP
  - construct k d-diverse paths



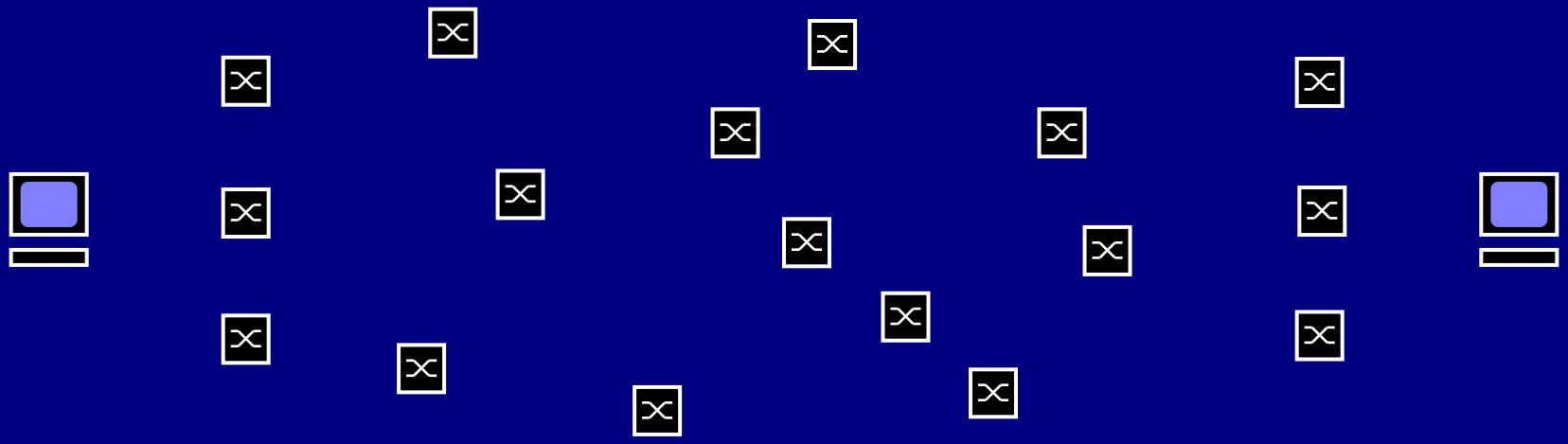
Multilevel Structural Diversity
GeoDiverse Multipath Routing

- ResiliNets review
- Challenge Taxonomy
- Multilevel interrealm resilience
  - resilience to attackers
  - resilience to large scale disasters
- ResTP: resilient transport protocol
- GeoDivRP: geodiverse routing protocol

Geodiverse Routing Protocol
GeoDivRP

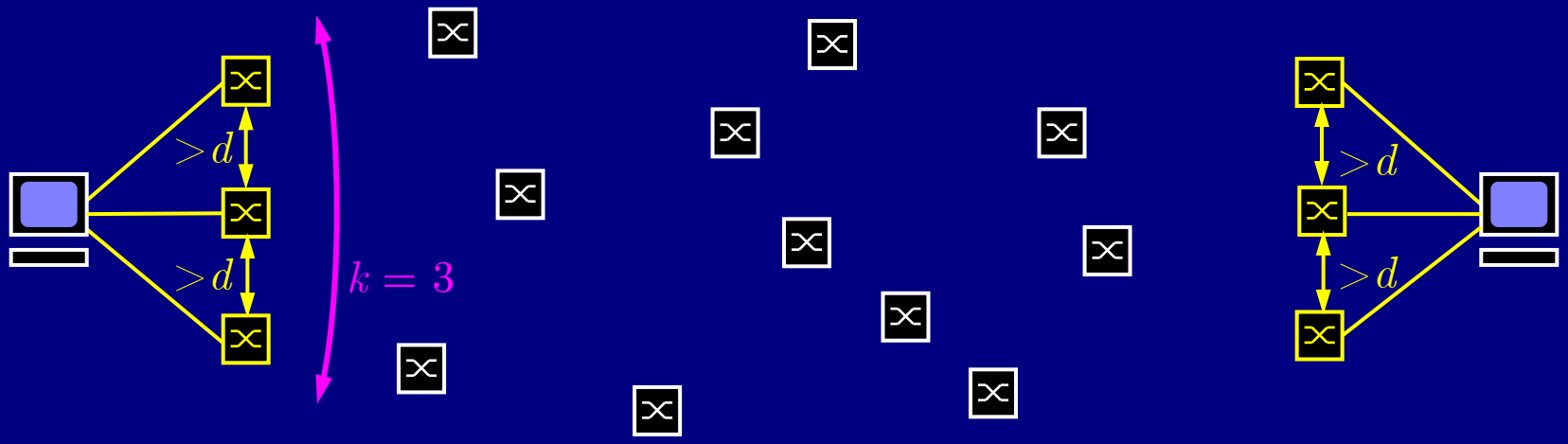
- Two heuristics: iWPSP and MLW
- iWPSP (iterative waypoint shortest path)
  - choose neighbours and waypoints to meet diversity spec
  - splice Dijkstra shortest paths
  - complexity: $2c^2n^2 \log n$ (for average of c neighbours)
  - [Cheng and Sterbenz @ KU: DRCN 2014]
- MLW (modified link weights)
  - modify link weights higher close to primary path
  - forces (weighted) shortest path alternates to be diverse
  - complexity: $2n \log n$
  - [Gardner, May, and Medhi @ UMKC: DRCN 2014]

ResiliNets Protocols
GeoDivRP: iWPSP



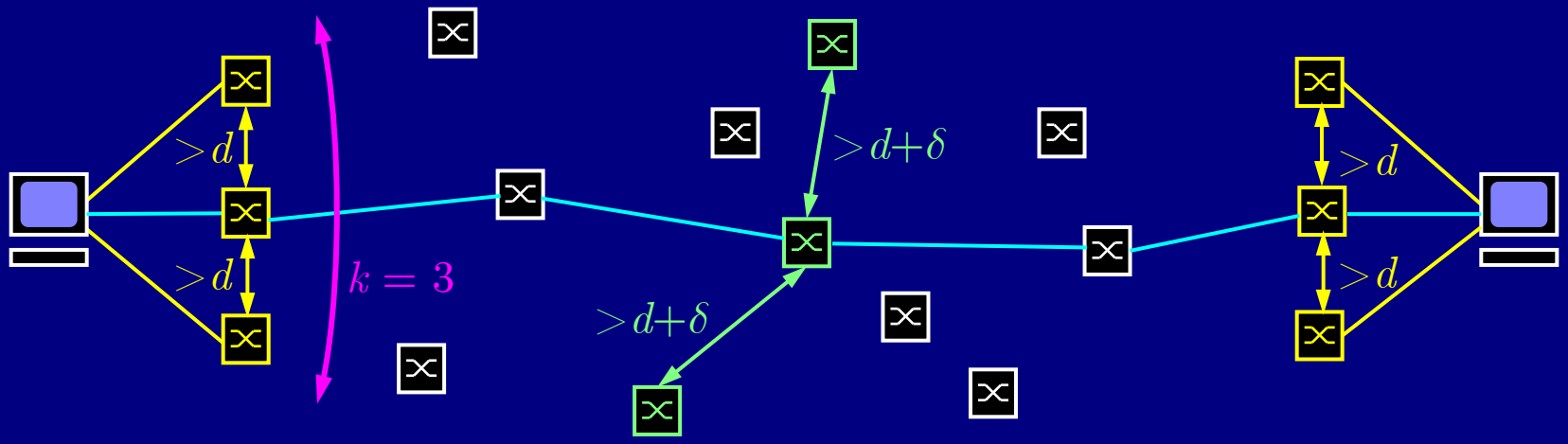
- GeoDivRP: intermediate waypoint algorithm
  - LSAs contain geolocation of routers

ResiliNets Protocols
GeoDivRP: iWPSP



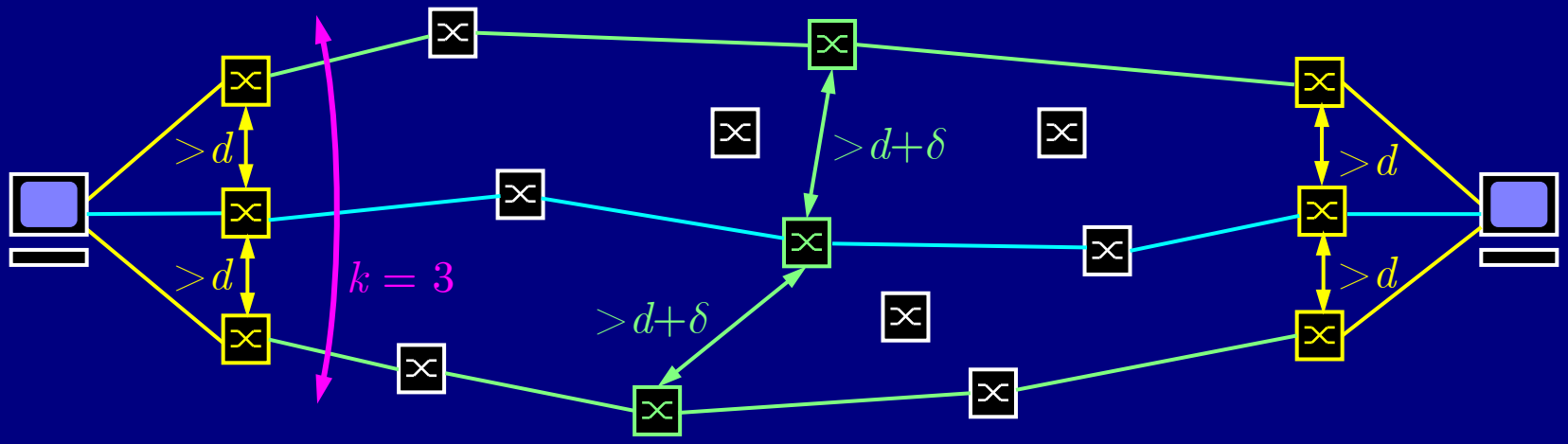
- GeoDivRP: intermediate waypoint algorithm
  - LSAs contain geolocation of routers
  - choose k next hop routers at least d apart if possible

ResiliNets Protocols
GeoDivRP: iWPSP



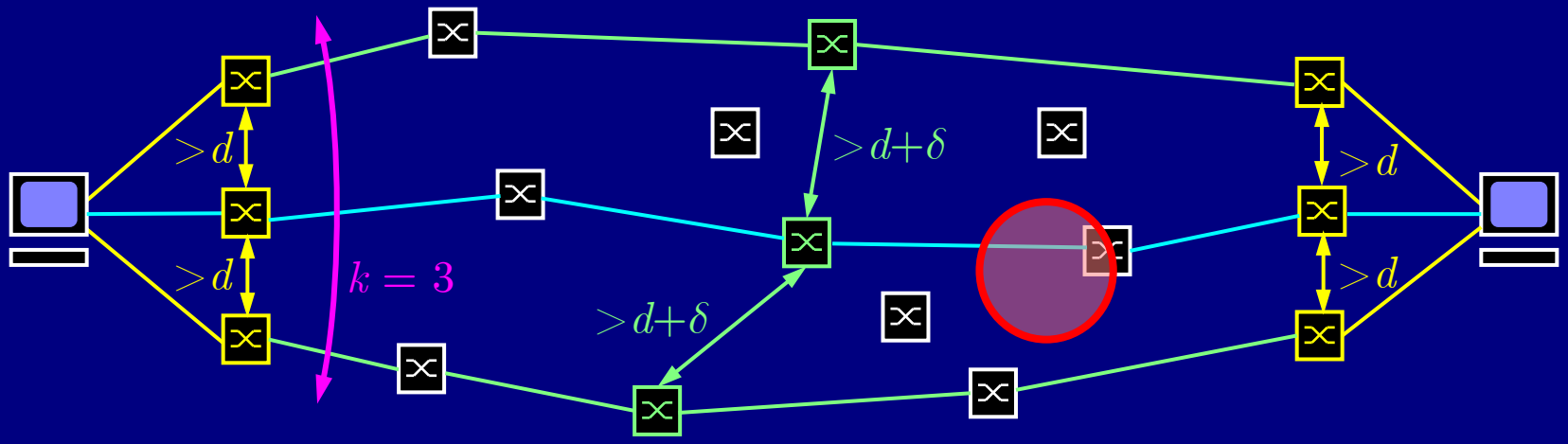
- GeoDivRP: intermediate waypoint algorithm
  - LSAs contain geolocation of routers
  - choose k next hop routers at least d apart if possible
  - choose mid-point waypoints $d+\delta$ wrt to shortest path
    - limit stretch to h and skew to t if specified and possible

ResiliNets Protocols
GeoDivRP: iWPSP



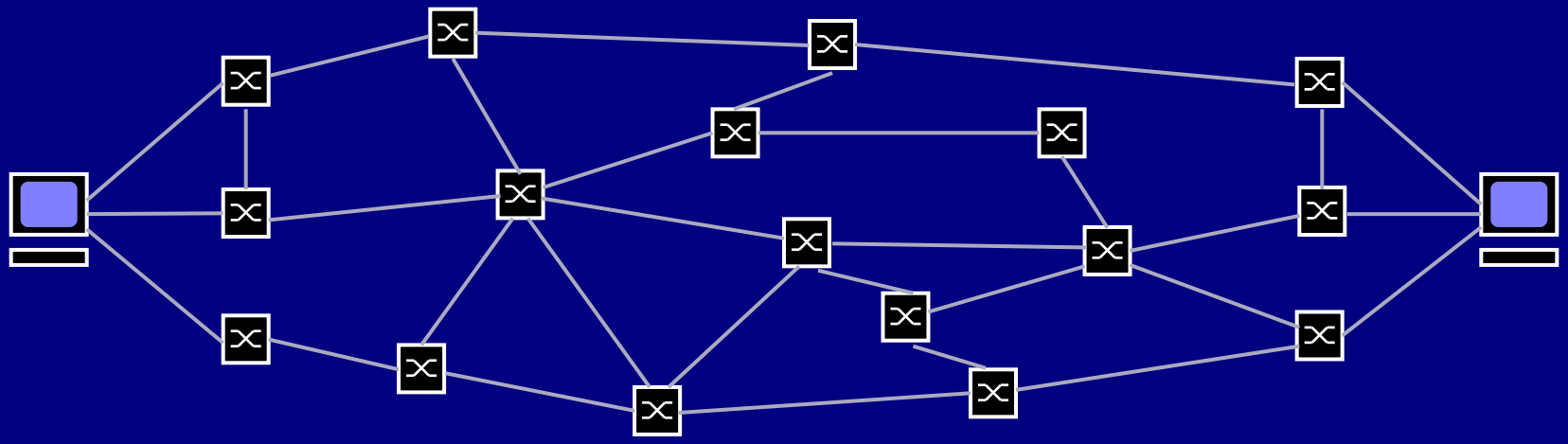
- GeoDivRP: intermediate waypoint algorithm
  - LSAs contain geolocation of routers
  - choose k next hop routers at least d apart if possible
  - choose mid-point waypoints $d+\delta$ wrt to shortest path
    - limit stretch to h and skew to t if specified and possible
  - use conventional SPF (Dijkstra) for paths to waypoints

ResiliNets Protocols
GeoDivRP: iWPSP



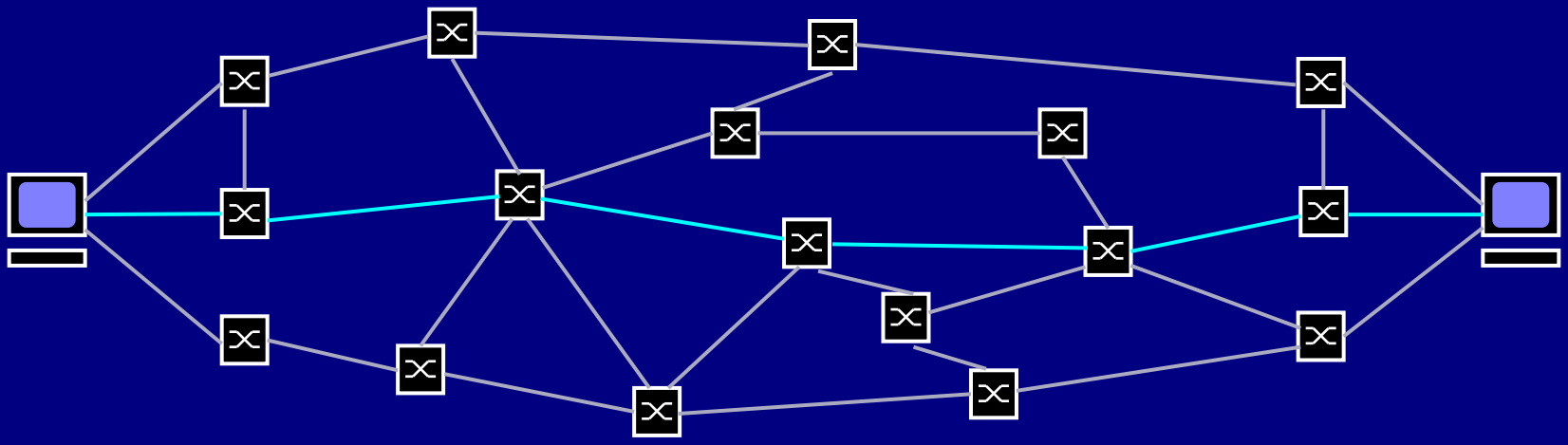
- GeoDivRP: intermediate waypoint algorithm
  - LSAs contain geolocation of routers
  - choose k next hop routers at least d apart if possible
  - choose mid-point waypoints $d+\delta$ wrt to shortest path
    - limit stretch to h and skew to t if specified and possible
  - use conventional SPF (Dijkstra) for paths to waypoints

ResiliNets Protocols
GeoDivRP: MLW



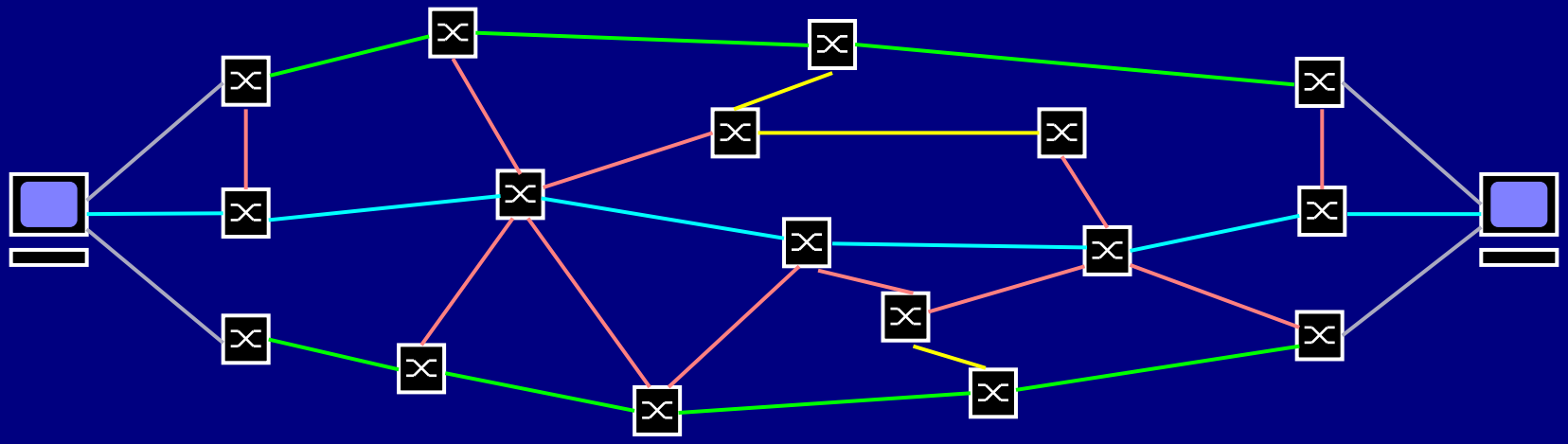
- GeoDivRP: intermediate waypoint algorithm
  - LSAs contain geolocation of routers

ResiliNets Protocols
GeoDivRP: MLW



- GeoDivRP: intermediate waypoint algorithm
  - LSAs contain geolocation of routers
  - choose primary shortest path

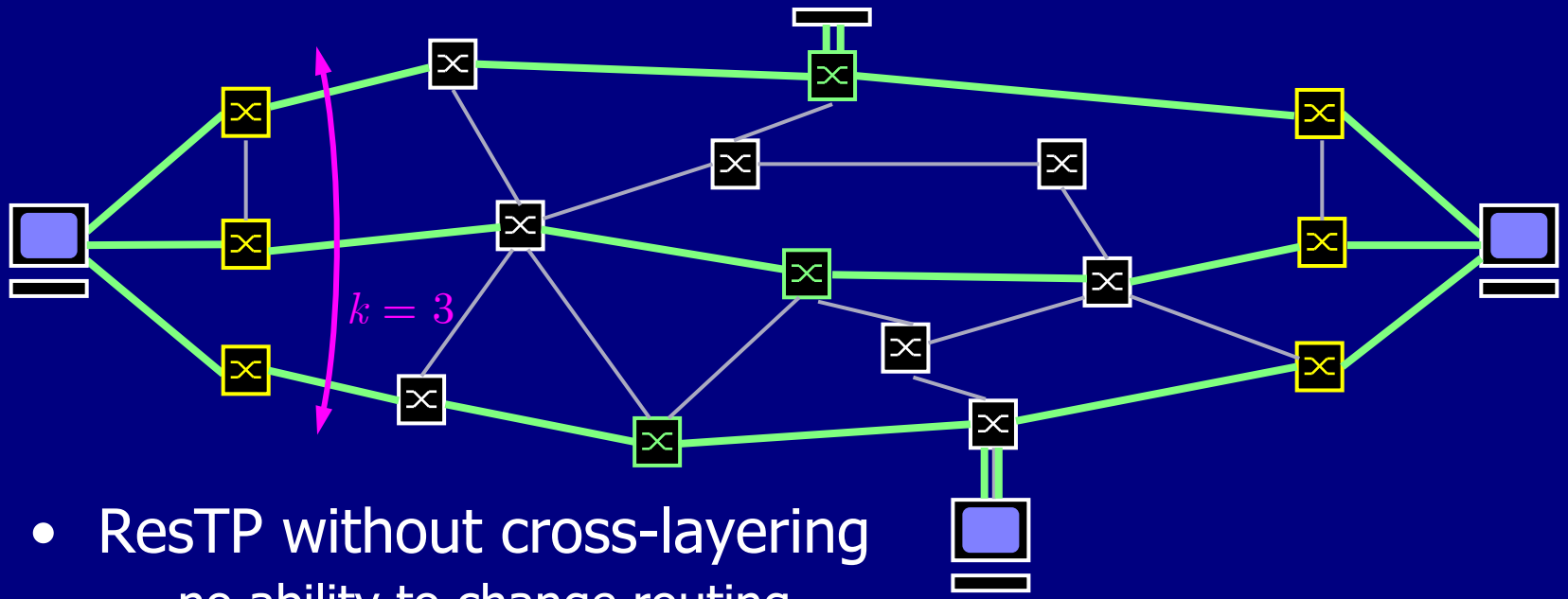
ResiliNets Protocols
GeoDivRP: MLW



- GeoDivRP: intermediate waypoint algorithm
  - LSAs contain geolocation of routers
  - choose primary shortest path
  - modify link weights higher close to primary path
  - forces (weighted) shortest path alternates to be diverse

ResTP Multipath

ResTP without GeoDivRP



- ResTP without cross-layering
  - no ability to change routing
  - no knowledge of network topology
- Assume coöperation from other distributed locations
  - choose geo-diverse relays; multihomed ISPs when possible

References
ResTP and GeoDivRP

- Trúc Anh Ngọc Nguyễn, Justin P. Rohrer, and James P.G. Sterbenz,
  "ResTP – A Transport Protocol for FI Resilience",
  ACM Conference on Future Internet Technologies (CFI),
  June 2015, pp. 9–11.
- Yufei Cheng 成宇飞, M. Todd Gardner, Junyan Li,
  Rebecca May, Deep Medhi, and James P.G. Sterbenz,
  "Optimised Heuristics for a Geodiverse Routing Protocol",
  IEEE/IFIP Design of Reliable Communication Networks (DRCN),
  Ghent, Belgium, April 2014, pp. 1–9.

End