Survivable Network Design

David Tipper
Graduate Telecommunications and Networking Program
University of Pittsburgh
Telcom 2110 Slides 12

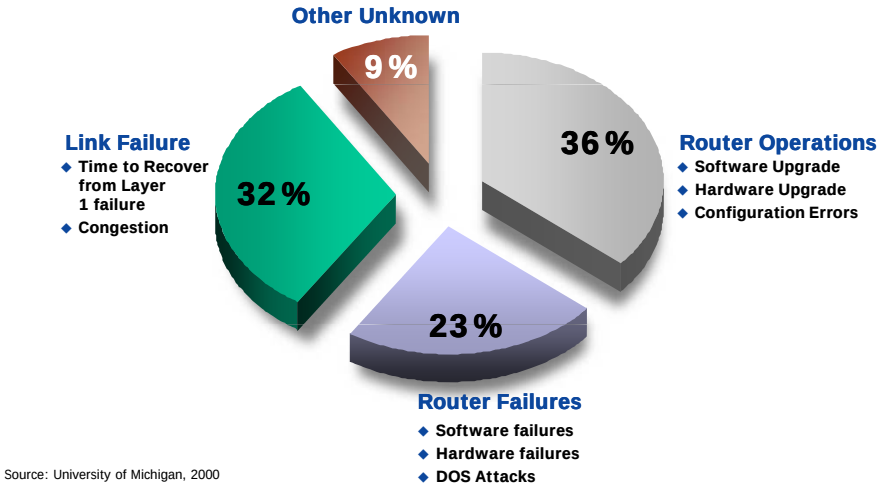
Motivation

- **Communications networks need to be survivable?**
- *Communication Networks are Critical Infrastructure (CI)* (PCCIP 1996) the systems, assets and services upon which society and the economy depend
- Communication infrastructure often considered *most important* CI due to reliance on it by other infrastructures
  - banking and finance, government services
  - power grid SCADA, etc.
- **Increasing Impact and Rate of Failures**
  - Increased bandwidth of links (WDM technology in fiber optic network)
  - Increased societal dependence
  - Multiple network operators and vendor equipment

Causes of Network Outages

- According to Sprint a link outage in IP backbone every 30 min on average
- Accidents
  - cable cuts, car wreck, etc.
  - According to AT&T 4.39 Cable cuts / year / 1000 km
- Human errors
  - incorrect maintenance, installation
- Environmental hazards
  - fire, flood, etc.
- Sabotage
  - physical, electronic
- Operational disruptions
  - schedule upgrades, maintenance, power outage
- Hardware/Software failures
  - Line card failure, faulty laser, software crash, etc.

Backbone Failures



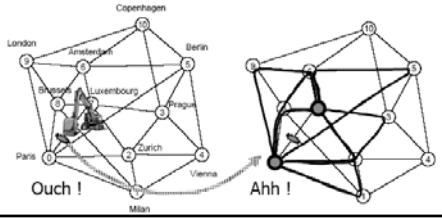
Source: University of Michigan, 2000

Network Survivability

- Definition
  - *Ability of the network to support the committed Quality of Services (QoS) continuously in the presence of various failure scenarios*
  - Includes performance as well as availability
- Survivability Components
  - Analysis: understand failures and system functionality after failures
  - Design: adopt network procedures and architecture to prevent and minimize the impact of failures/attacks on network services.
  - Goal: maintain service for certain scenarios at reasonable cost
- Self – Healing network

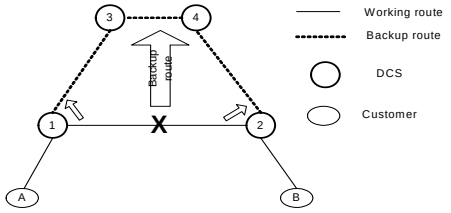
Survivable Network Design

- Three steps towards a survivable network
  1. Prevention:
     - Robust equipment and architecture (e.g., backup power supplies)
     - Security (physical, electronic), Intrusion detection, etc.
  2. Topology Design and Capacity Allocation
     - Design network with enough resources in appropriate topology
     - Spare capacity allocation – to recover from failure
  3. Network Management and traffic restoration procedures
     - Detect the failure, and reroute traffic around failure using the redundant capacity



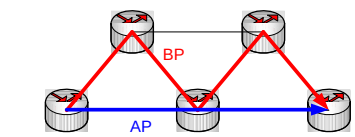
Survivability – Basic Concepts

- Working path and Backup path (recovery path):
- Working path: carry traffic under normal operation
- Backup path: an alternate path to carry the traffic in case of failures



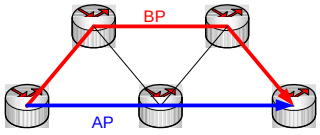
Survivability – Basic Concepts

- To survive against a network failure
  - working path and backup path must be disjoint
  - So that both paths are not lost at the same time
- Disjoint = ? (depending on a failure scenario)
  - Link disjoint
  - Node disjoint
  - (Shared Risk Link Group) SRLG disjoint



Source Destination

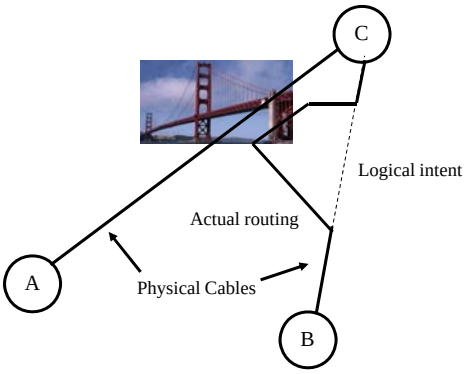
Link-disjoint



Source Destination

Node-disjoint

Shared Risk Link Group (SRLG)



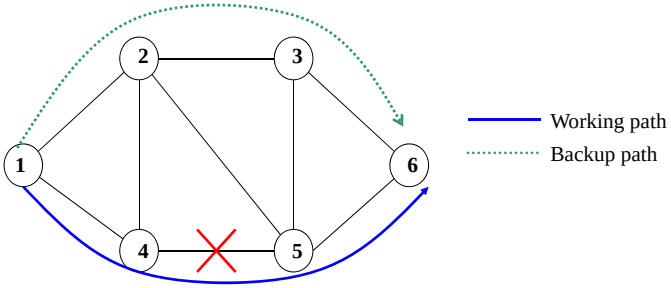
- Two fiber cables share the same duct or other common physical structure (such as a bridge crossing).
- Two cables can fail simultaneously

Classification of Survivability Techniques

- Path-based (Global) versus Link-based (Local)
- Failure Dependent vs. Failure Independent
- Protection versus Restoration
- Dedicated-Backup versus Shared- Backup Capacity
- Ring versus Mesh topology
- Dual homing
- *P* cycle

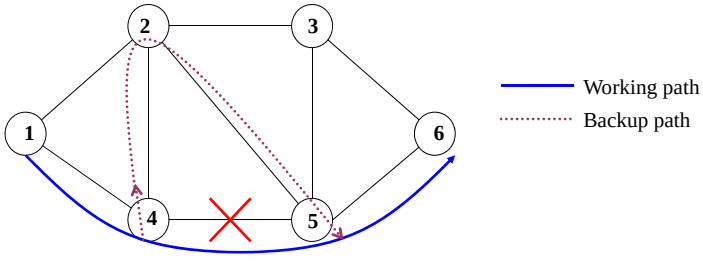
Path-based versus Link-based

- Path-based Scheme (Global)
  - Disjoint alternate routes are provided between source and destination node



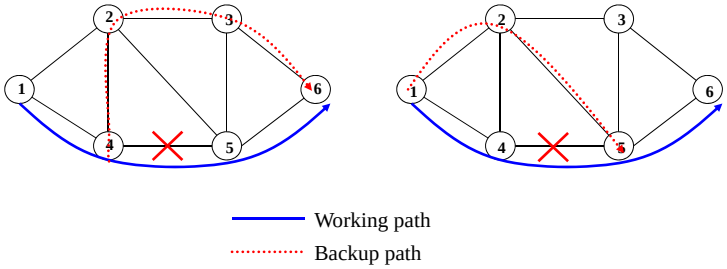
Path-based versus Link-based

- Link-based Scheme (Local)
  - Alternate routes are provided between end nodes of the failed link
  - Can have backhaul situation which wastes bandwidth



Partial Path Scheme

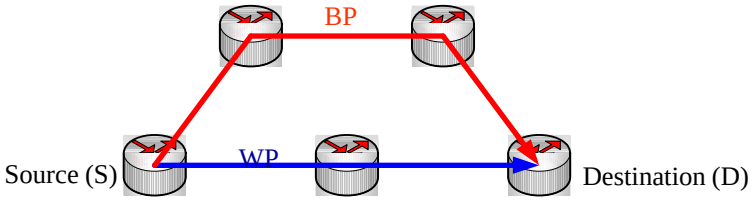
- Partial Path Scheme
  - Alternate routes are from the upstream node to destination node or from the downstream node to source node



Path-based versus Link-based

	Bandwidth efficient	Simpler	Faster recovery speed
Path-based	✓		
Link-based		✓	✓

What Does Survivability Get You?



- A_i is an availability of link i
- Availability of a connection between S-D:

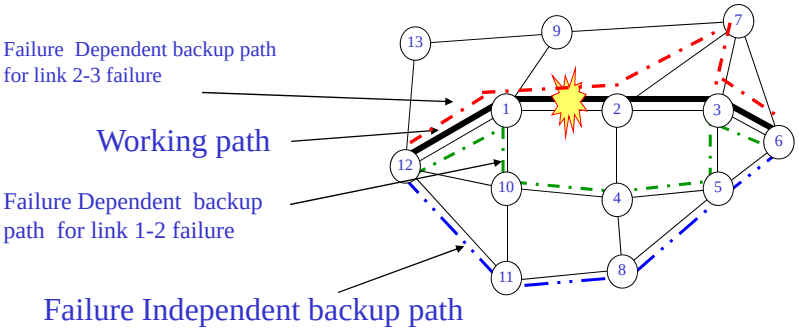
$$A_{no-protection} = \prod_{i \in WP} A_i$$

$$A_{protection} = \prod_{i \in WP} A_i + \prod_{i \in BP} A_i - \prod_{i \in WP \cup BP} A_i$$

  - Given A_i = 0.998297,
    - $A_{no\text{-}protection}$ = 0.996597, $A_{protection}$ = 0.999983

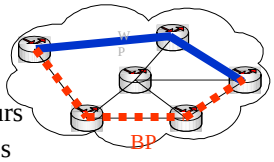
Failure Dependent vs. Failure Independent

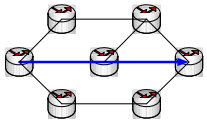
- Failure Dependent – the backup path depends on which device fails – need a set of paths one for each failure case
- Failure Independent – backup path link and node disjoint with working path - one backup path per working path
- Example:



Protection versus Restoration

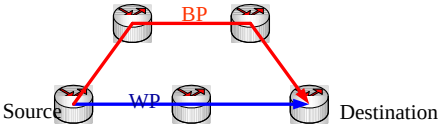
- When to establish the backup paths?
- Protection
  - Backup paths are fully setup before a failure occurs.
  - When failure occurs, no additional signaling is needed to establish the backup path
  - Faster recovery time
- Restoration
  - Backup paths are established after a failure occurs
  - More flexible with regard to the failure scenarios
    - backup paths are setup after the location of failure is known
  - More capacity efficient
    - due to its shared-backup nature,
    - Utilize any spare capacity available in the network
  - But cannot guarantee 100% restorability after failures





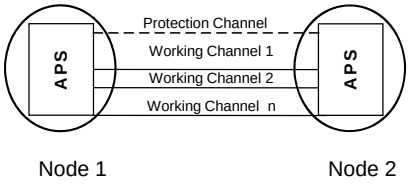
Protection

- Protection Variants
  - 1+1 Protection (dedicated protection)
    - Traffic is duplicated and transmitted over both working and backup paths
      - Fastest recovery speed, but not bandwidth efficient
  - 1:1 Protection (dedicated protection with extra traffic)
    - During normal operation (failure free), traffic is transmitted only over working path; backup path can be used to transmit extra traffic (low priority traffic) → better bandwidth utilization
    - When the working path fails, extra traffic is preempted, and traffic is switched to the backup path



Protection

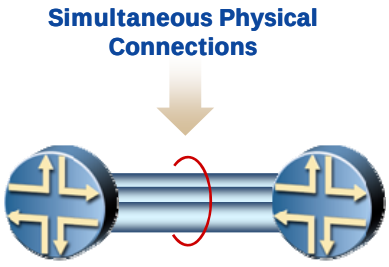
- 1:N Protection (shared recovery with extra traffic)
  - One protection entity for N working entities



- M:N Protection (M ≤ N)
  - M protection entities for N working entities
- Self Healing Rings are a form of Protection

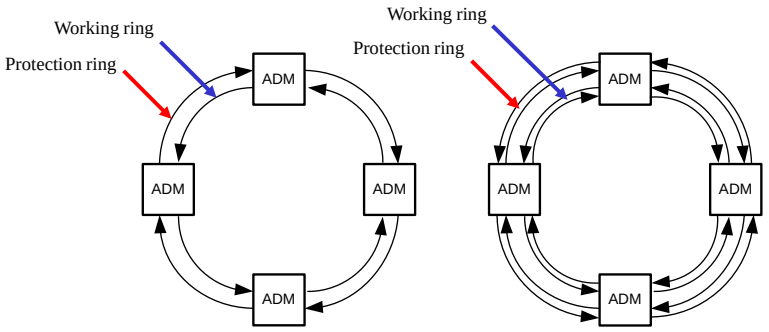
Link Redundancy

- Link Bundling
  - Link failure does not affect forwarding
  - Load redistributed among other members of bundle



- Parallel Link Technologies
  - MLPPP – T1/E1 Link aggregation
  - 802.3ad – Ethernet aggregation
  - SONET/SDH aggregation
  - Multi-Link Frame Relay

Types of Self-healing Rings

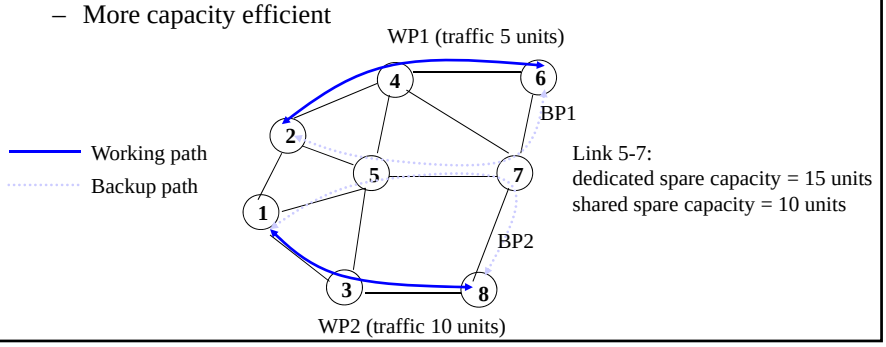


1:1 Uni-directional self-healing ring (USHR)

1:1 Bi-directional self-healing ring (BSHR)

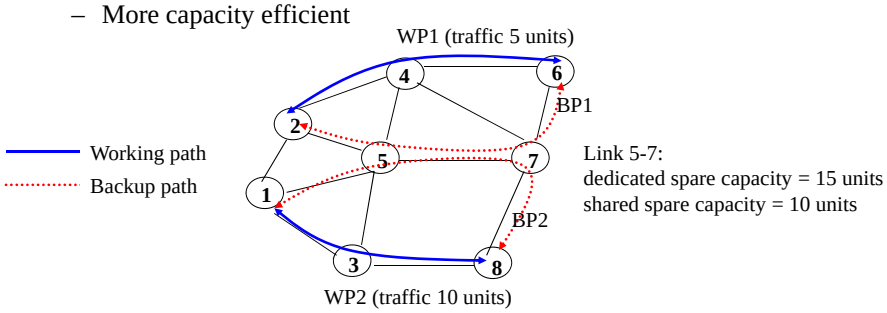
Dedicated versus Shared - Backup

- Dedicated-Backup Capacity
  - Backup resource can be used only by a particular working path
- Shared-Backup Capacity
  - Backup resource between several working paths can be shared
  - Rule: backup resource can be shared only when corresponding working paths are not expected to fail at the same time
  - More capacity efficient



Dedicated versus Shared - Backup

- Dedicated-Backup Capacity
  - Backup resource can be used only by a particular working path
- Shared-Backup Capacity
  - Backup resource between several working paths can be shared
  - Rule: backup resource can be shared only when corresponding working paths are not expected to fail at the same time
  - More capacity efficient



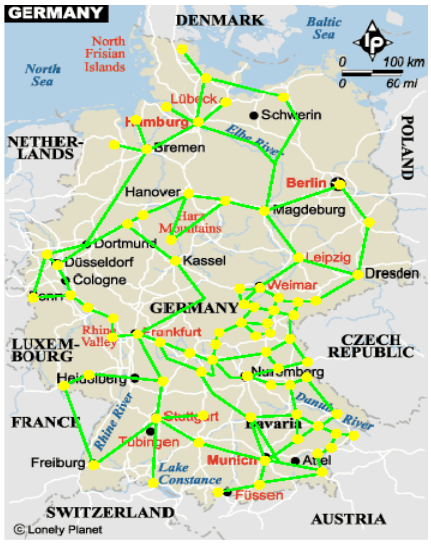
Ring vs Mesh Architectures

Advantages of Rings:

- More cost efficient at low traffic volumes
- Fast protection switching, some capacity sharing

Advantages of Mesh:

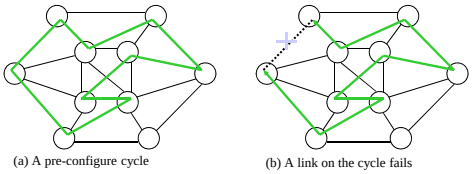
- More cost efficient at high traffic volumes
- Facilitates capacity and cost efficient mesh restoration
- More flexible channel re-configuration



P Cycles

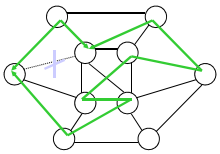
Protection (P) Cycle

- Closed cycles are formulated in the mesh network.
- Affected traffic is rerouted along these cycles.
- For a large network will have a number of p-cycles

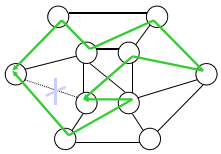


(a) A pre-configure cycle

(b) A link on the cycle fails



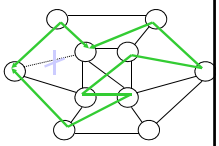
(c) A link not on the cycle fails



(d) Another link not on the cycle fails

P-Cycles: Basics

- For meshed networks
- Pre-reserved protection paths (before failure)
- Based on cycles, like rings
- Also protects *straddling* failures, unlike rings
- Local protection action, adjacent to failure (in the order of some 10 milliseconds)
- Shared capacity

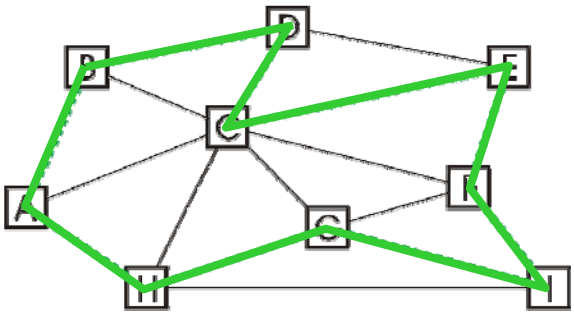


(c) A link not on the cycle fails

- *"**p**re-configured **p**rotection **cycles**"* → *p*-cycles
- Developed at TRLabs

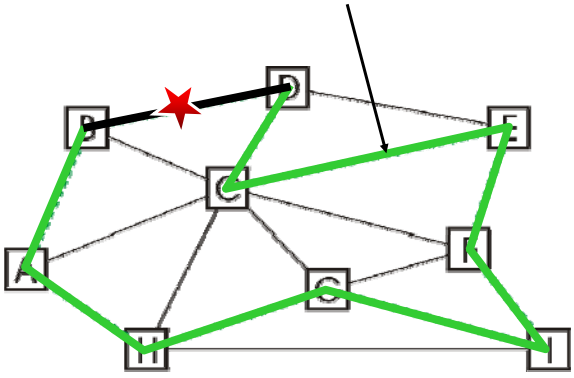
P-Cycles: Basics

- A single *p*-cycle in a network:



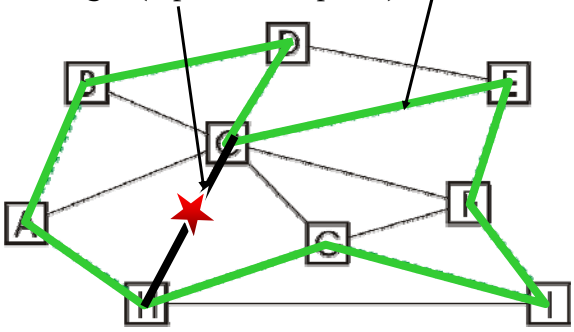
p-Cycles: Basics

- Protected spans:
- *9 „on-cycle“* (1 protection path)

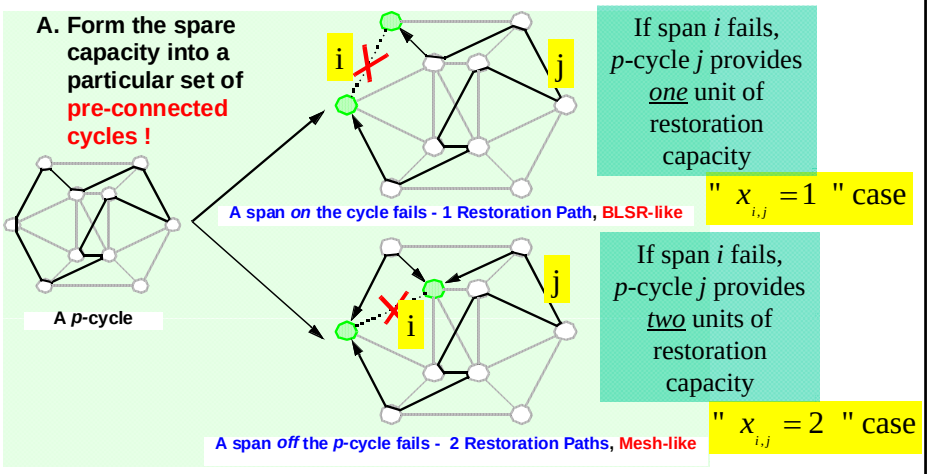


p-Cycles: Basics

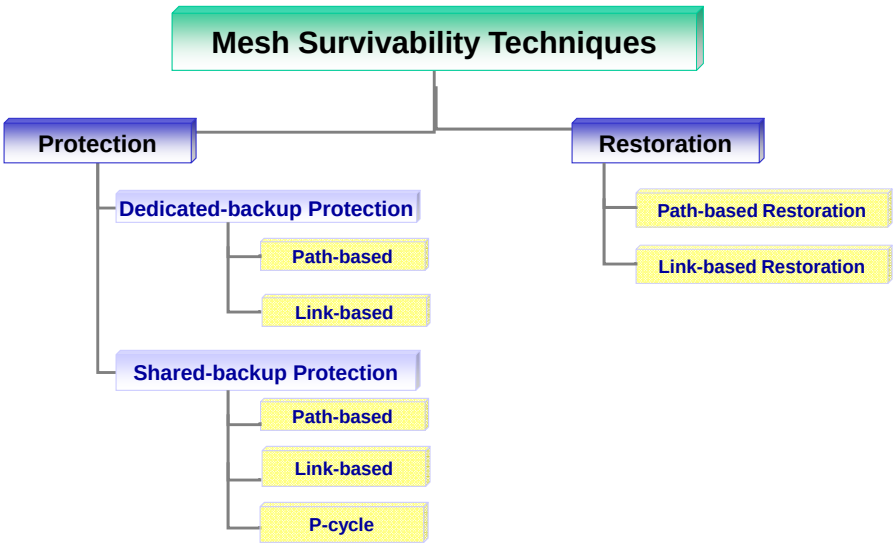
- Protected spans:
- *9 ``on-cycle''* (1 protection path)
- *8 ``straddling''* (2 protection paths)



Restoration using p-cycles



Mesh Survivability Techniques

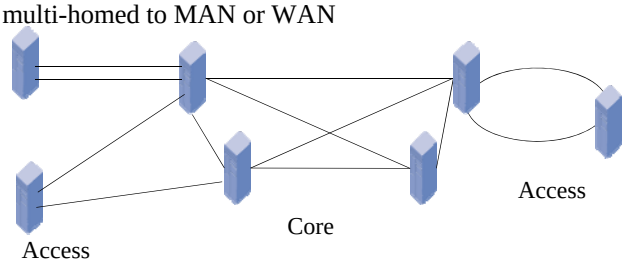


Survivability Technique Metrics

- Scope of failure coverage
  - single link failure, single node/link failure, multiple failures, etc.
- Recovery time
  - 50ms in SONET Ring
- Backup capacity requirement (redundancy, $R_r = \frac{amount\ of\ spare\ capacity}{amount\ of\ working\ capacity}$)
- Guaranteed bandwidth vs. non guaranteed
- Reordering and duplication
  - switching between WP and BP
- Additive latency and jitter
  - quality of backup path, backup path length, congestion on backup path
- State overhead
- Scalability
- Signaling requirements
- Notion of resilence classes (QoR)
  - Different level of connection availability, restorability and recovery time

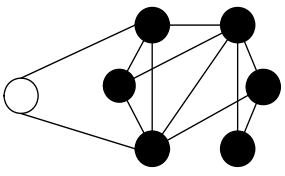
Transport Survivability

- Number of techniques exist
  - APS
  - Multi-homing (with or without trunk diversity)
  - Link restoration
  - Path restoration
  - Self healing rings
  - p-cycles
- See a mixture of techniques in real networks
- Usually little or no survivability at the far edge (CPE – last mile)
- Edges are multi-homed to MAN or WAN

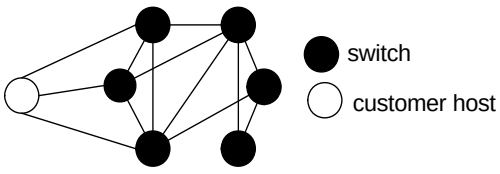


Dual/Multi-homing Topologies

- Dual-homing
  - Customer host is connected to two switched-hubs.
  - Traffic may be split between primary and secondary paths connecting to the hubs.
  - Each path serves as a backup for another.
- Multi-homing
  - Customer host is connected to more than two switched hubs.
  - Greater protection against a failure.

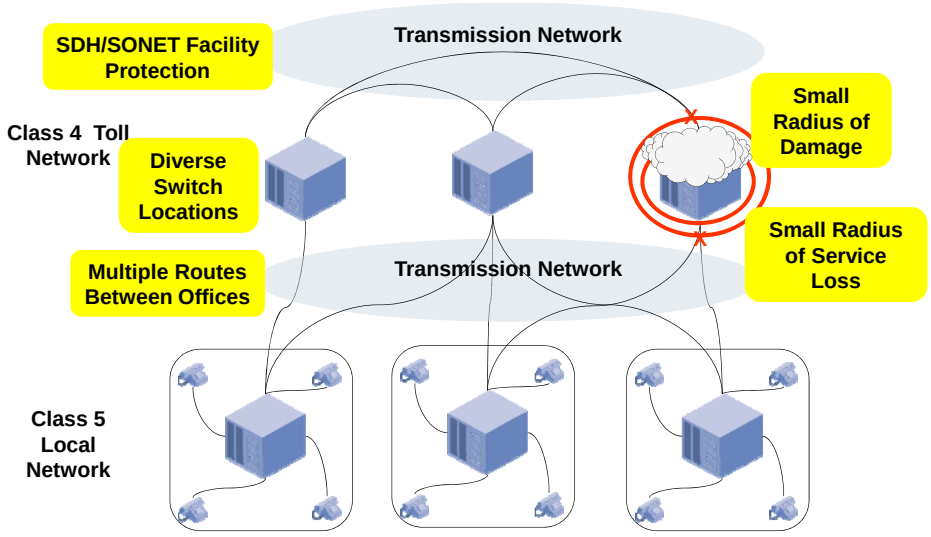


Dual-homing topology



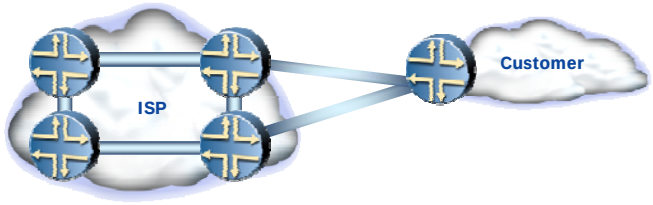
Multi-homing topology

Dual-homing in Telephone Network



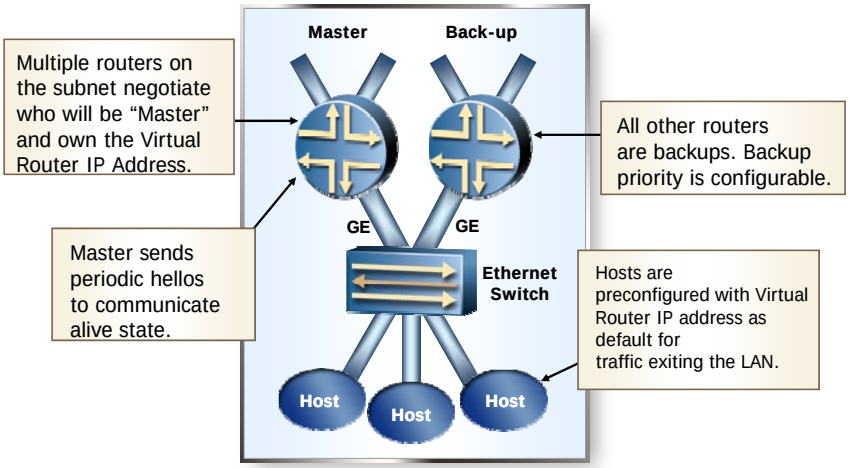
Resilient Edge Connectivity

- Multi-Homing for resilient Internet and IP-VPN connectivity
- Solves link failure and ISP node failure problems
- What about failure of customer edge router?

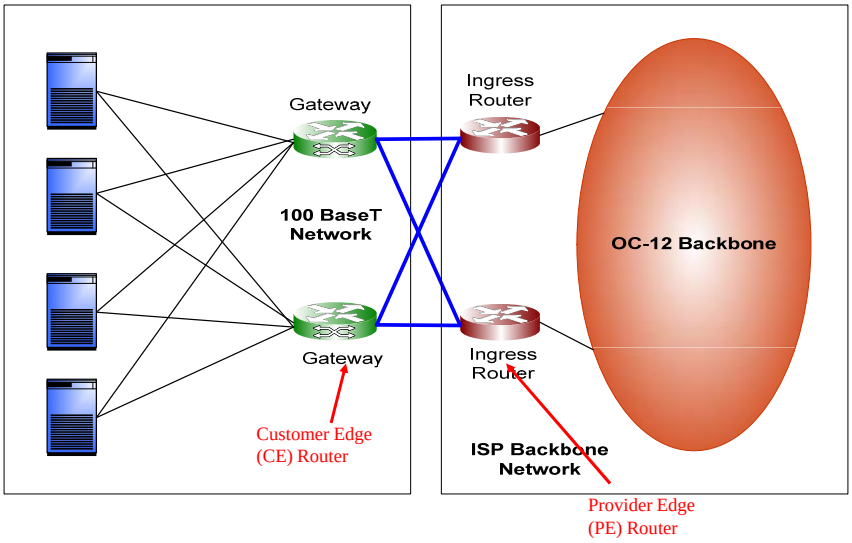


Virtual Router Redundancy Protocol

- Redundant default gateways:VRRP (RFC 2338)

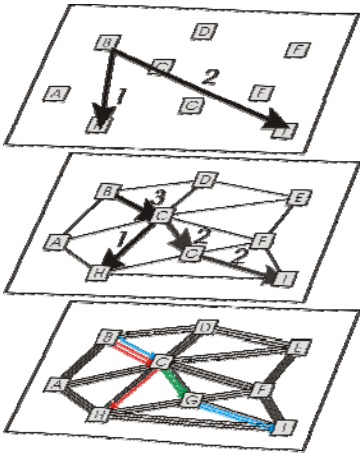


Dual-homing in Data Network



Implementation

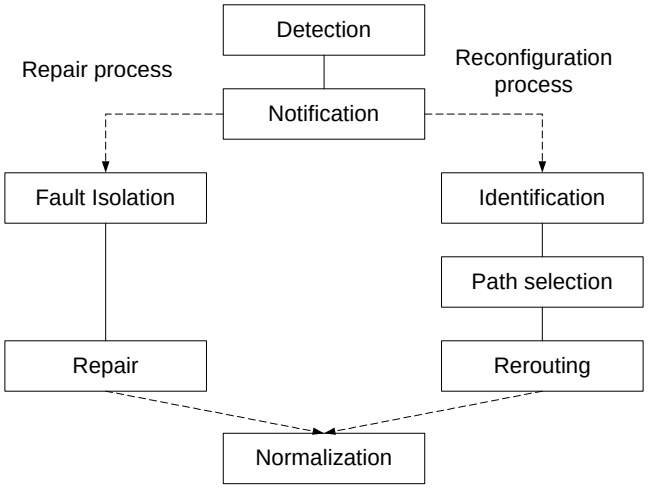
- Multi-layered:
  - Demand Topology
  - Logical Transport Topology
  - Fiber/Optical Topology
- Can implement survivability techniques at each layer
- Need to consider
  - Failure propagation
  - Alarm Setting
  - Speed of recovery
  - Cost
  - Management
  - Traffic Grooming
  - Etc.



Traffic Restoration Capabilities

- A survivability scheme and spare capacity doesn't accomplish restoration by itself, must be used in conjunction with dynamic restoration techniques.
- Need to detect failure and do path rearrangement given that there is enough spare capacity in the networks.
- For example a dual-homing approach guarantees surviving connectivity, but it doesn't restore the circuits/connections in itself.
- Need network management procedures to perform path rearrangement.

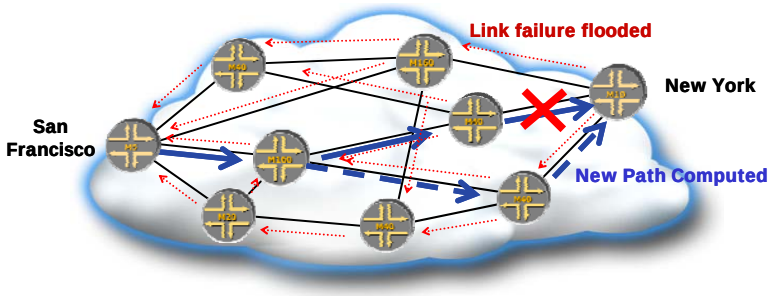
Steps in Traffic Recovery



IP Survivability Options

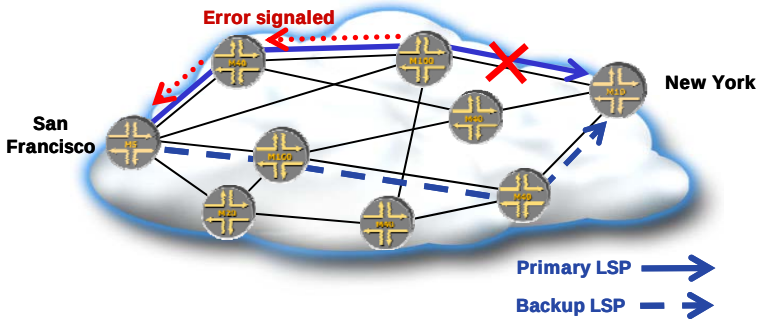
- Several techniques to improve survivability in IP networks
- IP layer –
  - adjust link weights and timers for faster failure recovery
  - prestore second shortest paths, etc,
- Adopt Optical Transport techniques from Telco operators (survivable rings, APS, path restoration, etc.)
- MPLS logical layer restoration

IP Dynamic Routing



- OSPF or IS-IS computes path
- If link or node fails, New path is computed
- Response times: Typically a few seconds
  - Can be tuned to ~1000's milliseconds
  - According to Sprint data – usually ~ 7secs to recover

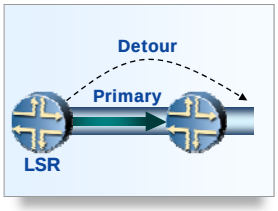
Backup Label Switched Paths



- Primary (working) LSP & backup LSPs established a priori
- If primary fails
  - Signal to head end, Use backup
- Faster response, requires wide area signaling

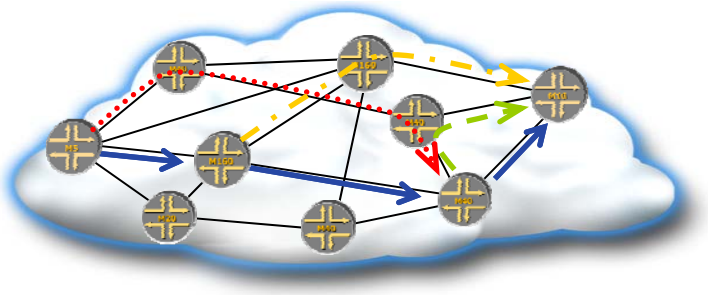
MPLS Fast Reroute

- Increasing demand for "APS-like" redundancy
  - MPLS resilience to link/node failures
  - Control-plane protection required
  - Avoid cost of SONET APS protection
- Solution: MPLS Fast-reroute
  - RSVP Extensions define Fast Reroute
  - LSPs can be set up, a priori, to backup:
    - One LSP across a link and optionally next node, or
    - All LSPs across a particular link

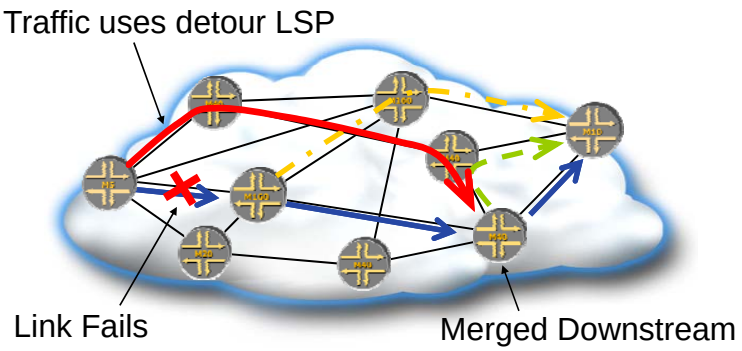


1:1 Protection

- For each LSP, for each node
  - Set up one LSP as backup
  - Merge into primary LSP further downstream
  - Backs up link and downstream node

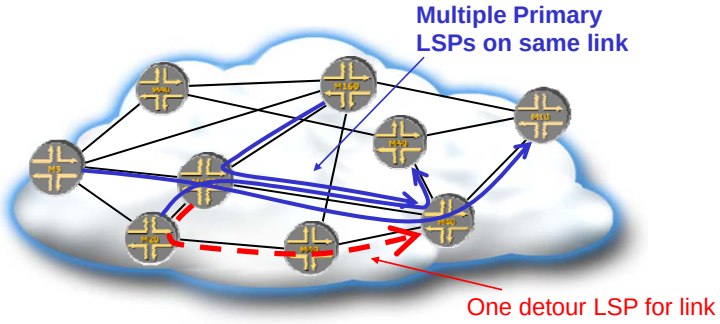


1:1 LSP Protection

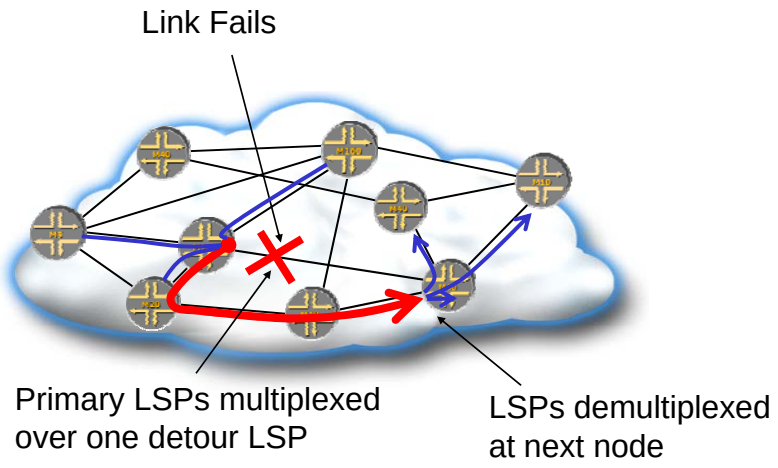


1:N Link Protection

- For each link, for each neighbor
  - Set up one detour LSP to backup the link as a whole
  - Uses LSP Hierarchy to backup all LSPs which were using failed link



1:N Link Protection



1:N Link and Node Protection

- For each link
  - For each node 2 hops away
    - Detour LSP backs up link & intermediate node
    - Uses LSP Hierarchy to backup all LSPs to that node
    - If there are two 2-hop paths to that node, setup two detour LSPs
  - For each node 1 hop away
    - Detour LSP backs up LSPs ending at that node

MPLS Fast Reroute

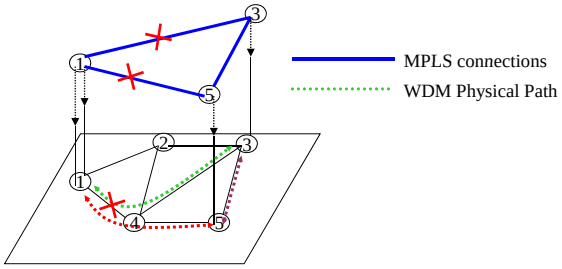
- Provides fast recovery for LSP failure
  - Based on a priori backup of detour LSPs
  - (eg, ~5 millisecond for tens of LSPs with 1:1)
- There are significant tradeoffs between the approaches
  - Number of LSPs required
  - Whether node failures are protected
  - Ability to reserve resources for backup LSPs
  - Optimality of routes

Summary of MPLS Methods

- End-to-End disjoint backup LSP – one per working LSP in the network
- MPLS Fast Re-Route
  - 1:1 LSP link or link + node protection
  - 1:N Link protection
  - 1:N Link plus node protection
- All of these are interoperable based on IETF standards
- Sink Trees are under study
- Does MPLS solve all the problems?
  - Can't recover from IP Layer Failure
  - Doesn't provide protection of layer 1 customers
  - Fault Propagation Issue

Multilayer Networks

- WAN networks have multiple technology layers
  - Converging toward IP/MPLS/WDM
- Multiple Layers present several survivability challenges
  - Coordination of recovery actions at different layers
    - Which layer is responsible for fault recovery?
  - Spare Capacity Allocation **(SCA)**
    - How to prevent over allocation, when each layer provides spare resources?
  - Failure Propagation
    - Lower layer failure can affect multiple higher layer links!

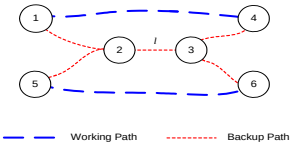


Optimization Based Design

- In implementing the chosen survivability technique (e.g., link protection, p-cycles) at a particular layer (e.g., optical)- optimization techniques are usually adopted.
- First design working network and working/active paths
- Then determine survivability design (often called spare capacity network design)
- Examples in ITU Planning document
- Consider shared backup path protection

Spare Capacity Allocation

- Single Layer Spare Capacity Allocation (SCA) Problem
  - given working paths and network (or virtual network) topology
  - provision spare capacity and find backup routes for fault tolerance
  - Goal: *minimum* spare capacity or cost
- Matrix based formulation*
  - ***P*** path link incident matrix, ***Q*** backup link incident matrix
  - Relate to spare provision matrix ***G***, and spare capacity reservation ***s***
  - Assume path restoration with disjoint backup routes
  - Shared backup path protection for any single link failure



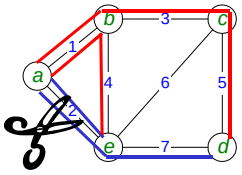
* Y.Liu, D.Tipper, and P. Siripongwutikorn, "Approximating Optimal Spare Capacity Allocation by Successive Survivable Routing," *ACM/IEEE Transactions on Networking*, Vol. 13., No. 1, pp. 198-211, Feb., 2005 .

Matrix model for SCA

Working and backup path matrices related to spare provision matrix $\mathbf{G = Q^T P}$

g_{ij} = spare capacity needed on link *i* when link *j* fails

From **G** find spare capacity allocation **s**=max(**G**)



Link *i*	s	G: 1	2	3	4	5	6	7	Q^T: Backup path link incident matrix: 1	2	3	4	5	6	7	8	9	10
1	2	0	**2**	1	1	0	1		0	0	**1**	**1**	0	1	1	0	0	0
2	2	2	0	2	1	1	0	0	1	1	0	0	0	1	1	0	0	0
3	1	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	1	0
4	1	1	1	1	0	0	1	0	1	0	0	1	1	0	0	0	1	0
5	2	1	1	1	0	0	0	2	0	1	1	0	0	0	0	0	0	1
6	1	0	0	1	0	1	0	1	0	0	0	0	1	0	0	1	0	1
7	2	1	0	2	0	2	0	0	0	1	0	0	0	1	0	1	0	0
	11																	

P: Working path link incident matrix	Flows	1	2	3	4	5	6	7	8	9	10	src	dst
	1	0	0	0	0	0	0	1	a	b			
	2	1	0	1	0	0	0	0	a	c			
	3	0	**1**	0	0	0	0	1	a	d			
	4	0	**1**	0	0	0	0	0	a	e			
	5	0	0	1	0	0	0	0	b	c			
	6	0	0	1	0	1	0	0	b	d			
	7	0	0	0	1	0	0	0	b	e			
	8	0	0	0	0	1	0	0	c	d			
	9	0	0	0	0	0	1	0	c	e			
	10	0	0	0	0	0	0	1	d	e			

An Example:
1. Link 2 fails
2. Flow 3,4 affected
3. Backup paths up
4. Spare BW=2 on Li

Optimization model for link failures

$$\min_{Q,s} \quad S = \boldsymbol{e}^T \boldsymbol{s}$$

Total spare capacity

$$\text{s.t.} \quad \boldsymbol{s} \geq \mathbf{G}$$

Enough spare capacity on each link

$$\mathbf{G} = \mathbf{Q}^T \mathbf{M} \mathbf{P}$$

Calculation of spare provision matrix

$$\mathbf{P} + \mathbf{Q} \leq 1$$

Link-disjointed backup paths

$$\mathbf{Q}\,\mathbf{B}^T = \mathbf{D} \pmod 2$$

Flow conservation of backup

$\mathbf{Q}$ is a binary matrix

Integer programming

Decision variable: $\mathbf{Q}$, $\boldsymbol{s}$

Given: **M** – traffic demand matrix
P – working path link incidence matrix
B and **D** – node-link & flow-node incidence matrices
Mixed Integer Programming problem NP Hard

Heuristic Solution Algorithm

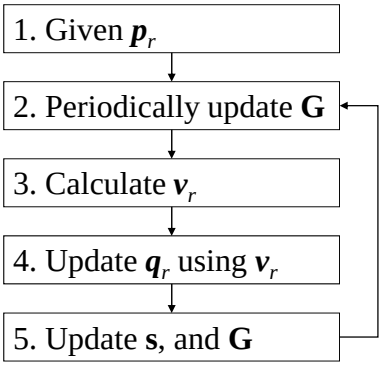
- Successive survivable routing algorithm*
  - Decompose multi-commodity flow → multiple single flows
  - *Goal*: Each flow seeks a backup path with minimal *additional spare capacity*
  - Using shortest path algorithm for each flow to
    - route link-disjointed backup paths
    - using spare provision matrix G to calculate **link cost** – incremental spare reservation $\mathbf{v}_r$;
- Flows **successively** update their backup paths

→termed: *successive survivable routing* (SSR)

- Randomly order flows for successively updating.
- Fast computation find near optimal solution

- **Apparatus and Method for Spare Capacity Allocation*, Y. Liu and D. Tipper , U.S. Patent 6,744,727 B2, June 1, 2004
- Presented in *ACM/IEEE Trans. On Networking* Feb.,2005

SSR flowchart of flow *r*

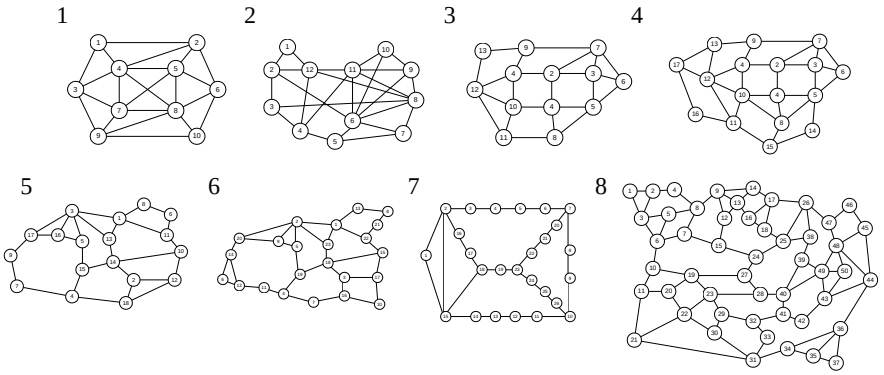


- On source node of flow *r*:
  - p_r, q_r : working and backup path vectors
  - **G**, **s**: spare provision matrix and spare reservation vector
  - v_r : incremental spare reservations, used as link cost
- Stop after no backup path update on the network

Numerical comparison

- Compare different algorithms and bounds
  - **RAFT**: Resource aggregation fault tolerance
  - **SPI**: Sharing with partial information
  - **SR**: Survivable routing (SSR without iteration)
  - **SSR** : Successive survivable routing
  - **SA**: Simulated annealing
  - **BB**: Branch and bound on a path-flow model – optimal
  - **LP**: Linear programming lower bound
- Metrics:
  - % Redundancy = spare capacity/working capacity,
  - execution time

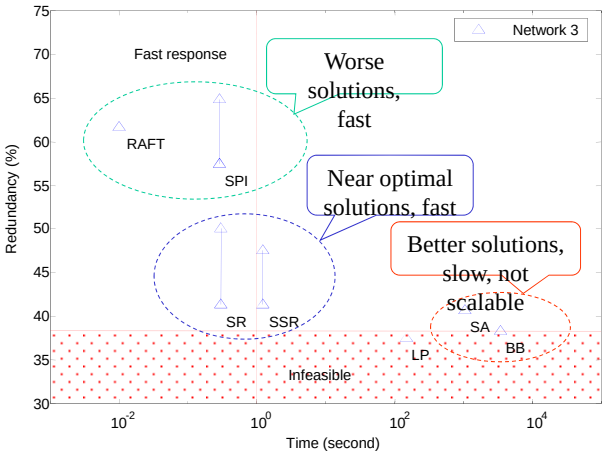
Experiment networks



Network node degree ranges from 2.31 to 4.4
Consider balanced mesh load case

Redundancy versus Time on Network 3

- SSR, SR, SPI have 64 random cases with different flow orders
- Range of solutions
- Time is the sum of time to compute all 64 cases



State of the Art

- Survivable Network Design
  - Important in WAN Backbones
- Basic approach
  - Given particular technology (e.g., WDM, MPLS, etc) assume
    - Traffic restoration scheme (e.g., failure independent path restoration)
    - Failure scenario (any single link failure)
  - Determine least cost survivable network design using optimization formulations with heuristic solutions
- Many tradeoffs identified and studied
  - Protection vs. Restoration
  - Reactive vs. Proactive
  - Shared vs. Dedicated
  - Link vs. Path vs. Rings, etc.
  - Failure Dependent vs. FID
  - Etc.,