[simula . research laboratory]

A Resilient Internet for Mission Critical Applications

Stein Gjessing

Simula Research Laboratory
and University of Oslo

Norway

With colleagues at Simula:
Olav Lysne, Tarik Cicic,
Audun Hansen and Amund Kvalbein

ICQNM & ICDS 2007
Gosier, Guadeloupe
January 3., 2007

Contents

- The Internet and the Communication needs of Mission Critical Applications
- Our focus today:
  The effects of link and router failures
- The Solution:
  - Proactive and local rerouting
- Multiple Routing Configurations
- Conclusion

The Internet and Mission Critical Applications

- The Internet was designed for latency tolerant applications.
- **Mission Critical Applications need**

1. Congestion and Admission control with
    - Quality of Service support
    - Class of Service support
2. Forward Error Correction
3. Fast reroute in case of link or router failure

The Internet and Mission Critical Applications

There are solutions for Mission Critical Applications for

1. Congestion and Admission control with Quality of Service - Class of Service
2. Error detection and error correction

However, solutions for the Communication needs of Mission Critical Applications have not been found for link and router failures

The Internet and Link and Router failures

- Frequent faults
- A large number of link failures are transient
- 70% of unplanned failures in an IP backbone are single link failures

- Component (eg. a link) failure starts a system-wide routing table update
- Loops and **packet losses for more than one second**
  - More like 10 seconds
- Unacceptable for real time, Mission Critical Applications

Faulty components: Reactive vs. proactive solutions

- IGP rerouting is **reactive** and **global**
  - Link state information is sent to all routers in the network when a fault is encountered, and new routing tables are built
  - Optimizations does not completely solve the problem (frequent pings, local updates, fast shortest paths calc.)
- A **proactive** and **local** mechanism is needed for fast response and less (no?) packet loss for mission critical applications

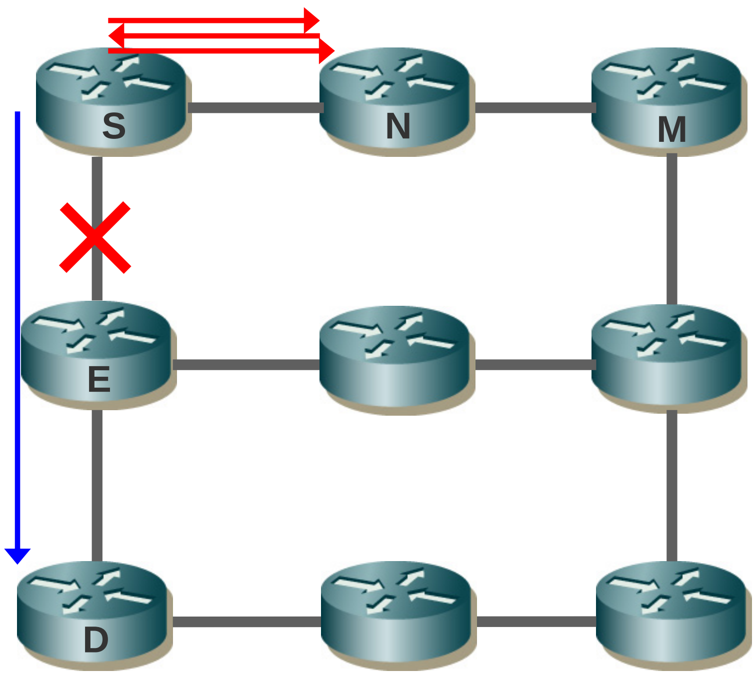
- Possible solution: MPLS: Multi Protocol Label Switching
  - Path oriented. Optional in the Internet.
  - MPLS fast reroute (but requires tons of extra paths)

IGP: Interior Gateway Protocol (OSPF, IS-IS)

Challenges with proactive approaches

- Pre-compute backup next-hops
- Connectionless
  - hop-by-hop decisions
  - Only first router knows the failure and uses backup next hop

→ looping



Decision in Node S:

Dest	Via	Cost
D	**E**	**2**
D	N	6

Decision in Node N:

Dest	Via	Cost
D	**S**	**3**
D	M	5

Proactive and local methods

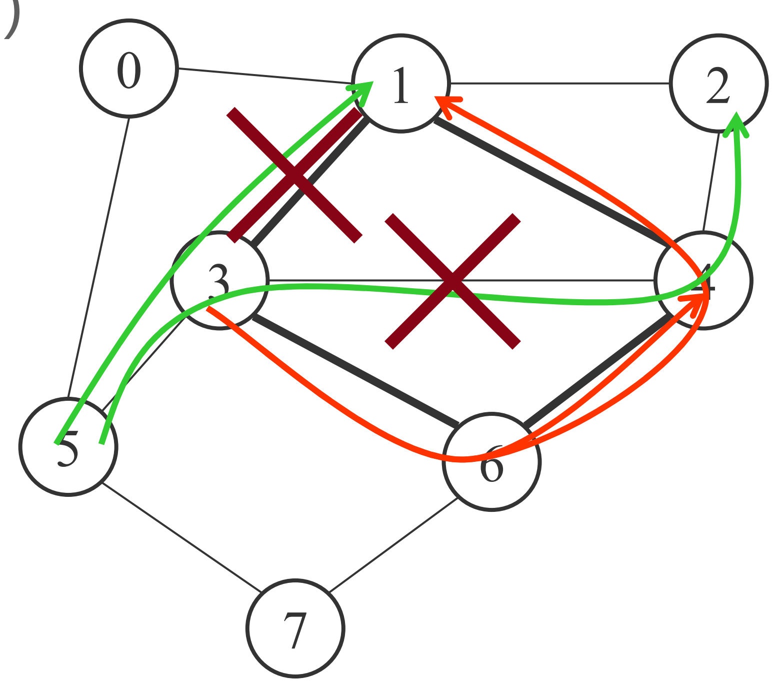
- Pre-calculate backup paths
  - NB! Connectionless
    - Routing table lookup based on destination address
- Store alternatives as
  - Special routing tables or
  - Special addresses
    (extra routing table entries)
- When an error occurs
  - Immediately revert to alternative addresses or tables
- Alternatives must guarantee freedom from loops

Proactive and local methods

- P-cycles
- Failure Inferencing-based Fast Re-routing (FIFR)
- Multiple short paths
  - Equal cost multipaths,
  - Multiple short paths,
  - No U-turn
  - **IETF IP fast reroute ("Not via")**
- **Multiple Routing Configuration**
  - Implemented e.g. by IETF Multi-Topology Routing

p-Cycles Concepts

- Cycle 1-4-6-3 (bidirectional)
- Protects links 1-4, 4-6, 6-3, 3-1 and 3-4
- Substantial research done
- Small or large cycles?
- Leave protection cycle as early as possible?
- Not completely connectionless



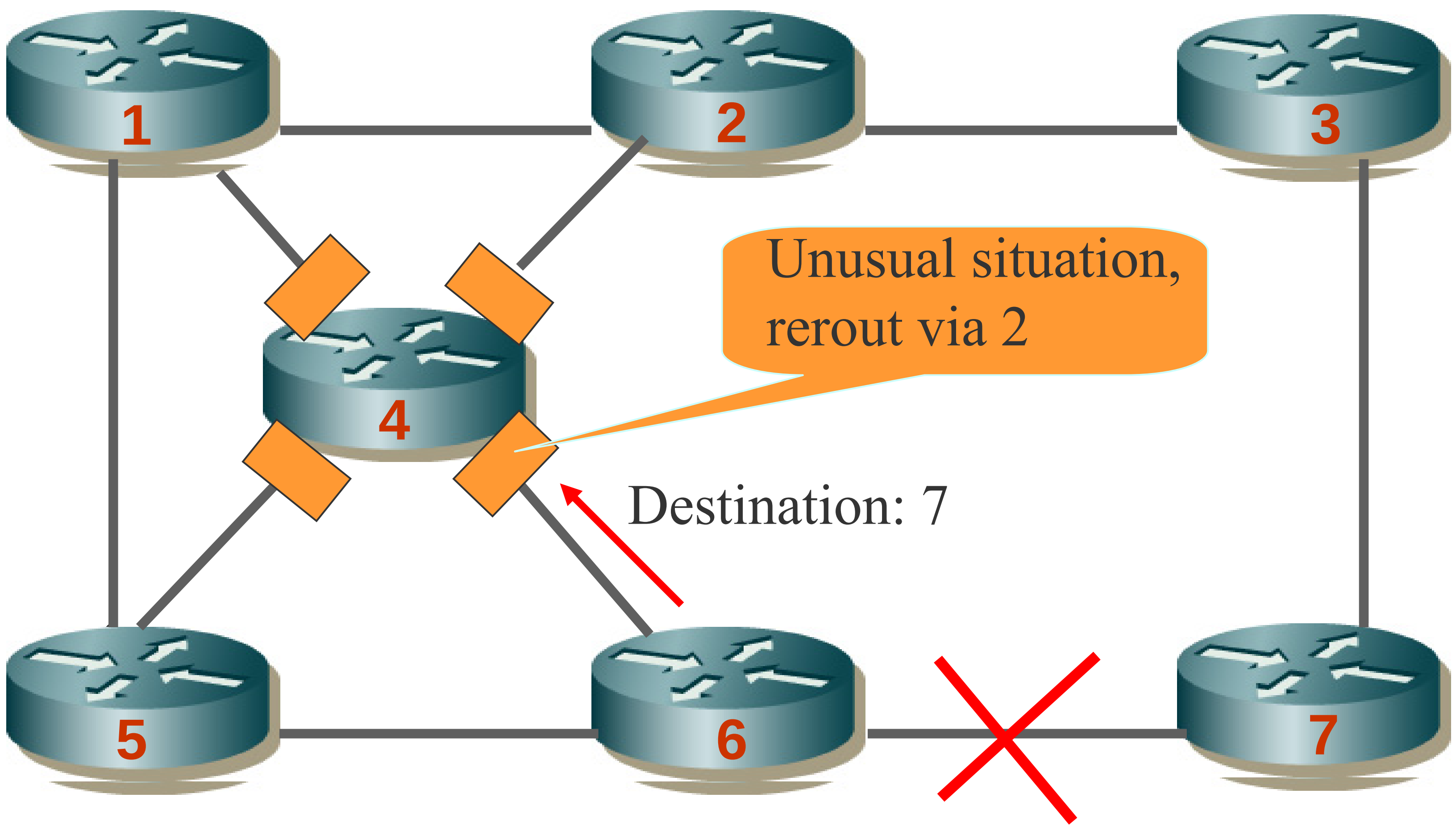
Failure Inferencing-based Fast Rerouting (FIFR)

- One of three main approaches considered in the community
- Main idea:
  - Interface specific forwarding
  - Based on incoming interface, a node know what is a safe next hop

  → Interface specific forwarding tables

FIFR – Line interface cards One routing table per card

Example: From 6 to 7

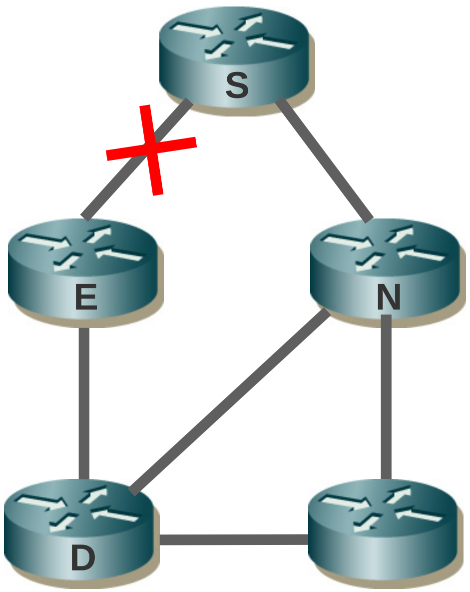


Multiple short paths

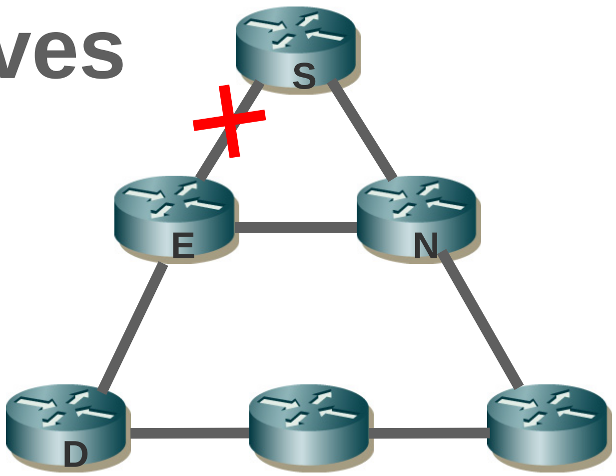
- Equal cost multi-path (ECMP)
- Other short paths (NB! Avoid loops)
- IETF IP Fast Reroute
  - Routing Area Working Group (rtgwg)
  - Called “Not via”
  - M. Shand. IP Fast Reroute Framework. IETF Internet Draft, 2006. draft-ietf-rtgwg-ipfrr-framework-05.txt.

Equal Cost Multi-path (ECMP)

- Obviously loop-free
  - Will not loop back to the failure
- For node failures:
  - make sure that the failed node is not on the ECMP
- Example:
  All same link weights
  S – E – D
  S – N - D



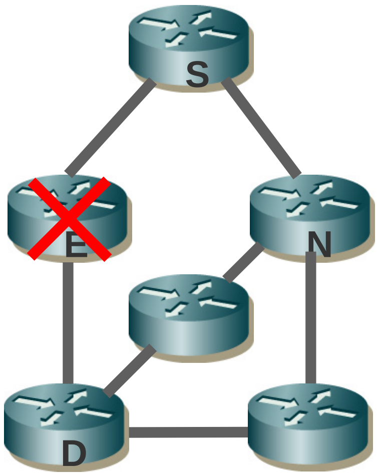
Other loop-free alternatives



- S – D
- A direct neighbor N of the detecting node S has a path to the destination D that does not traverse the failure
- Link failure coverage

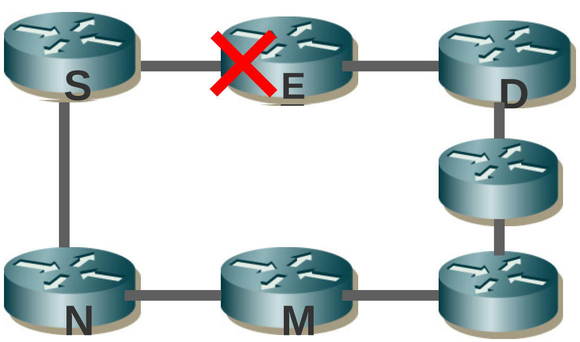
  cost(N,D) < cost(N, S) + cost(S,D)
- Node failure coverage

  cost(N,D) < cost(N,E) + cost(E,D)



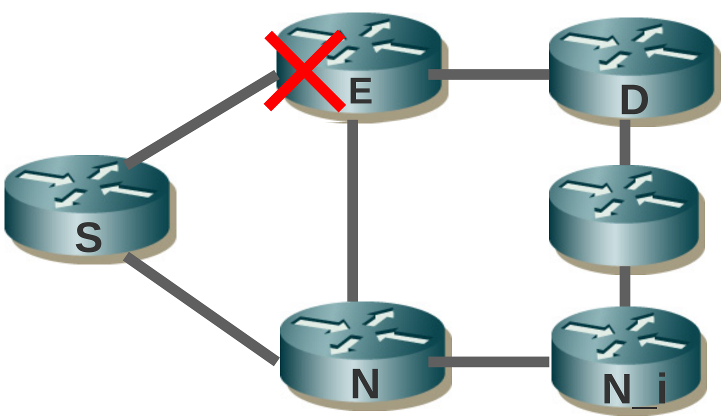
"No U-turn" alternatives

- S – D
- When node S uses node N as backup next hop, node N must not use the primary next hop S towards D, but rather use the loop-free node protecting alternate (node M) towards D
- This means that node N is not allowed to give packets from S a u-turn back to S.



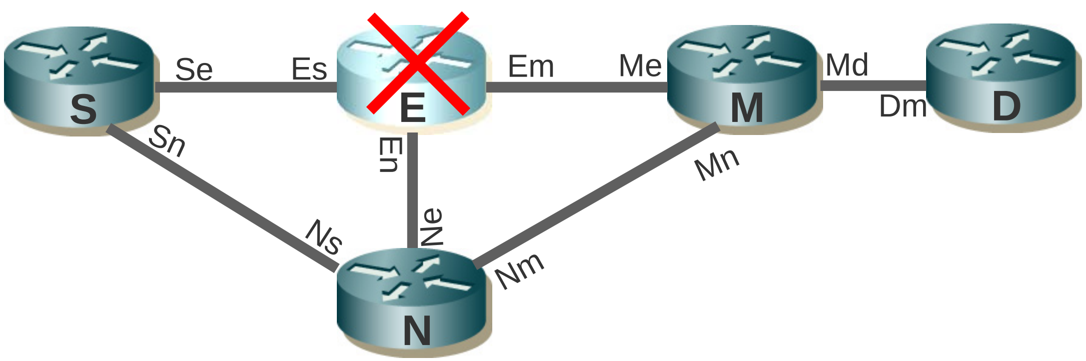
Multi-hop tunneling

- S – D
- Used to steer the packets to a node N_i that is i hops away from S and that has a loop-free path to the destination D without traversing the failure
- Without signaling, using only packet encapsulation.
- Can only be used when the packets tunneled from S to N_i do not traverse the failure



Tunneling using Not-via addresses

- A packet addressed to a Not-via address must be delivered to the router with that address, not via the neighboring router on the interface to which that address is assigned
- In other words, one must ensure that the packets affected by the failure of router E are delivered to router M that according to the primary route to destination D is downstream of E
- Each router in the network must calculate the best path to each Not-via address or group of addresses without the component(s) that the Not-via address is meant to protect



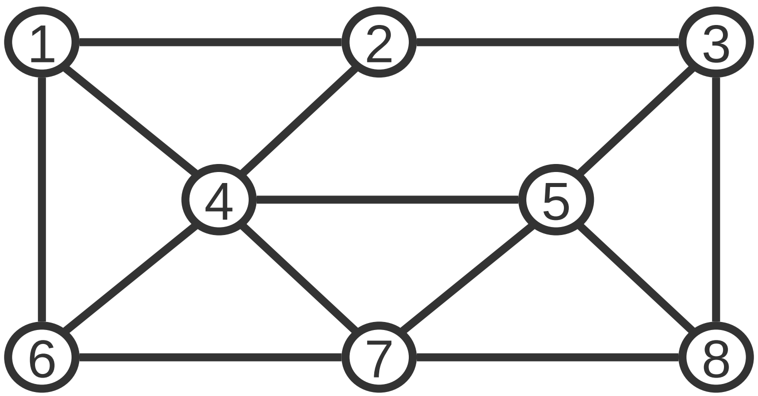
Multiple Routing Configurations - MRC

- Developed at Simula Research Laboratory in Norway

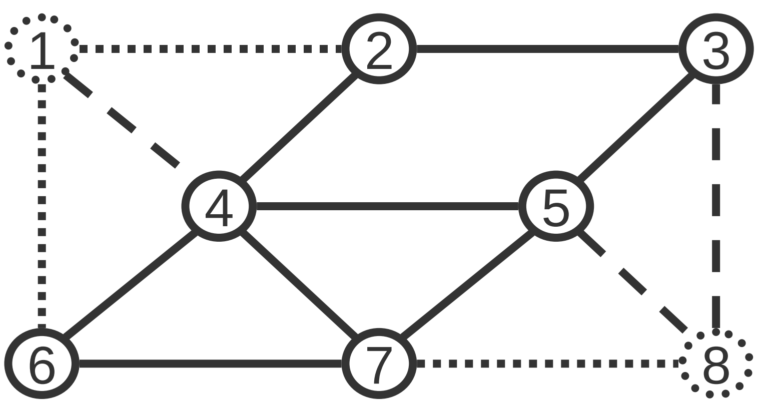
Main idea:

- Use a backup view of the network in case of a failure
- A backup view is called a **backup configuration**
- A backup configuration is like the original, except some links have new weights
- Based on work in interconnects:
  - Ingebørg Theiss and Olav Lysne: FROOTS – Fault handling in Up*/Down* routed networks with multiple roots", In Proceedings of the International Conference on High Performance Computing HiPC. Springer-Verlag, 2003.
  - Ingebørg Theiss and Olav Lysne: "FRoots, A Fault Tolerant and Topology Agnostic Routing technique", IEEE Transactions on Parallel and Distributed Systems 17(10): 1136-1150, 2006.

Multiple Routing Configurations - MRC



Full topology



Example Backup Configuration

Regular node – link:

Restricted link:

Isolated node – link:

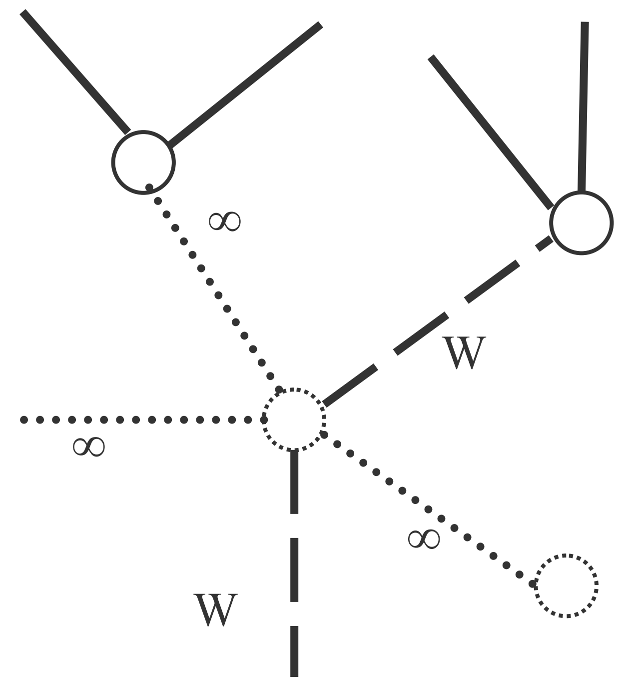
Backup configurations

- A backup configuration is the original topology with a new set of link weights
- One logical routing table per backup configuration
  - or special addresses
- MRC constructs a full set of backup configurations
  - Not used in the failure-free situation
  - Each backup configuration protects a subset of the links and nodes
- Routing in the backup configurations is restricted
  - All nodes must be reachable in every backup configuration
  - All links and nodes are **isolated** in one backup configuration
- A small number of backup configuarations is needed (4 – 6)

Normal link and node

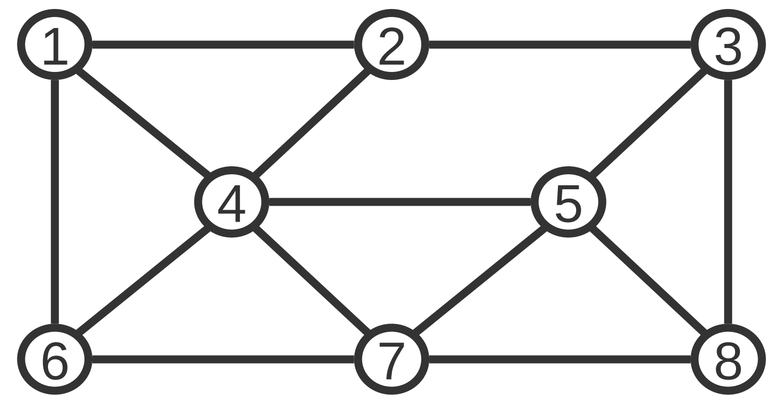
Isolated links and nodes

Restricted links



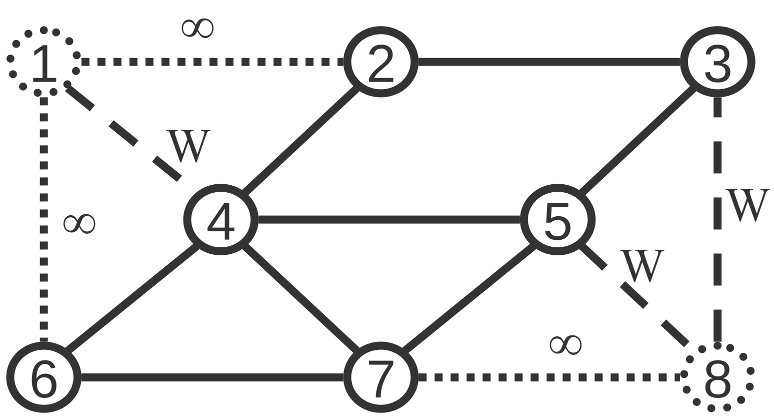
- An isolated link has infinite weight
- A restricted link has a high weight W
  - W is chosen so that the link is used only as a ”last resort”
- A node is isolated when all attached links are either isolated or restricted

Multiple Routing Configurations - MRC



Full topology

Shortest path calculations between all nodes in regular topology

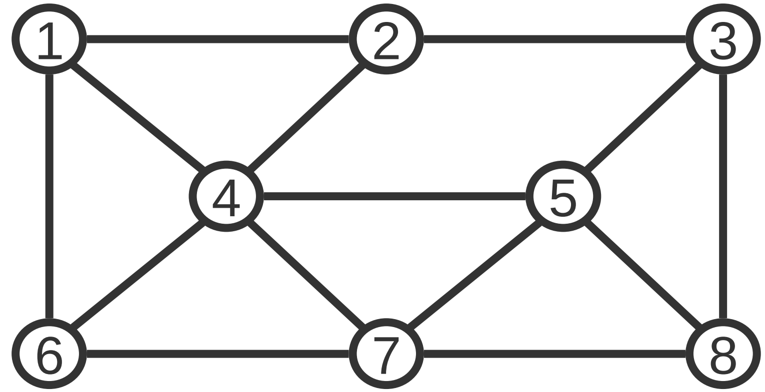


Example Backup Configuration

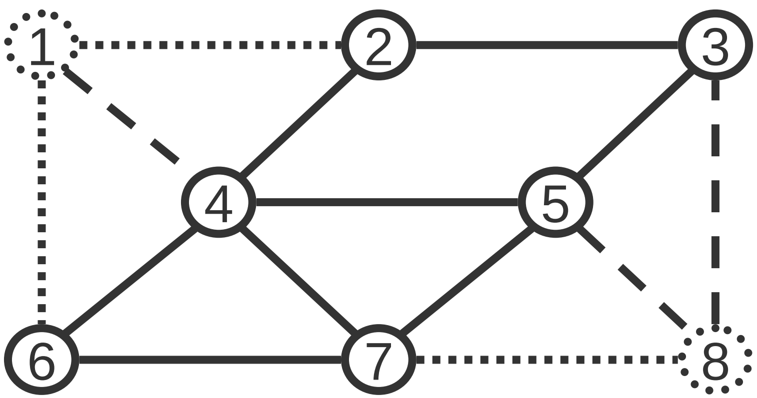
Shortest path calculations between all nodes in every backup configuration

[simula . research laboratory]

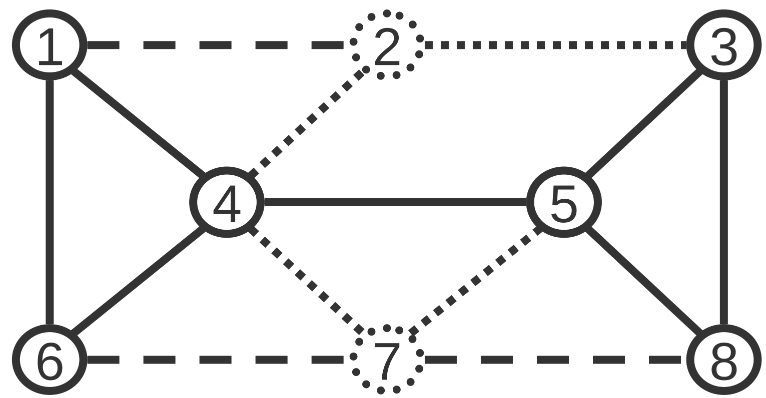
MRC: Example of full set of backup configurations



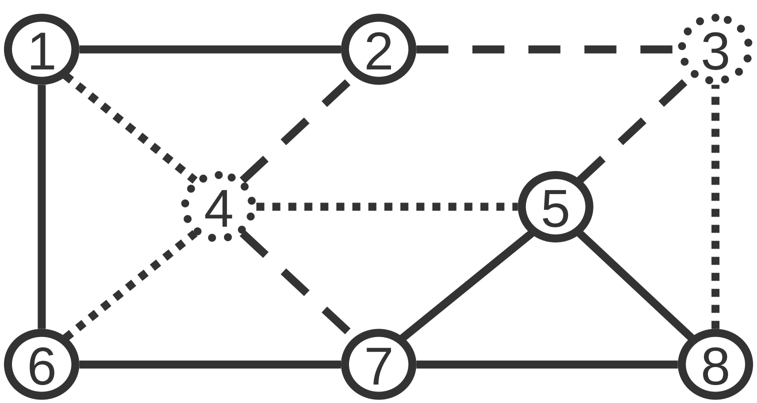
Full topology



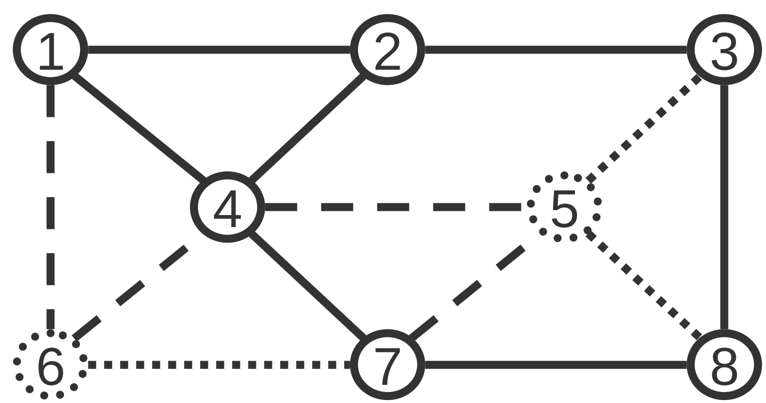
Layer 1



Layer 2

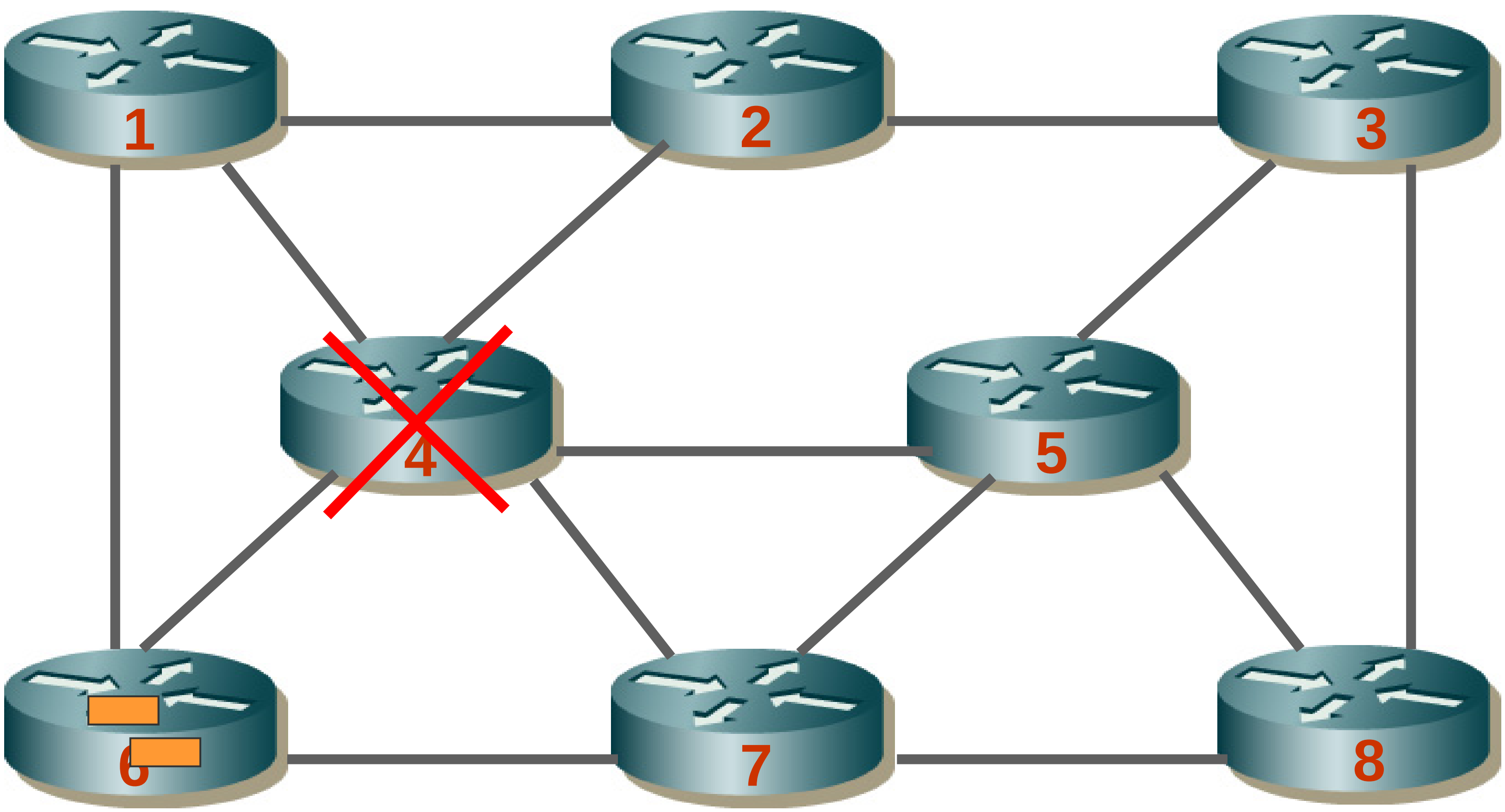


Layer 3

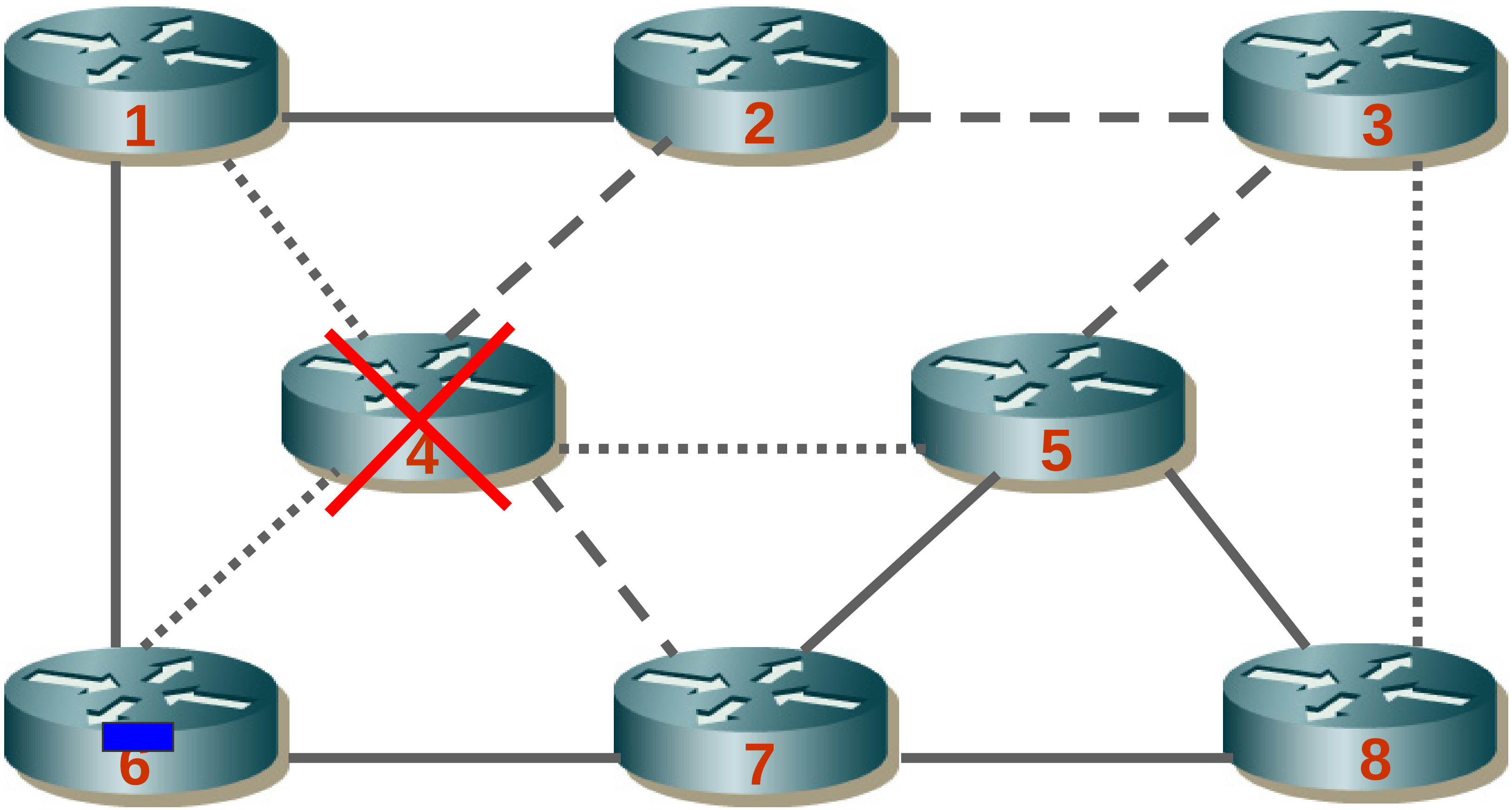


Layer 4

Normal Routing from 6 to 3



Multiple Routing Configurations:
Routing from 6 to 3 when 4 fails, use config. 3



Implementation issues

- Each router maintains one routing table for each backup configurations (or use extra routing entries)
- Identify the current configuration in a packet:
  - Configuration number or special addresses
- Multiple failures:
  - Many nodes/areas can share the same safe layer → shared risk groups (SRG)

Multi-Topology Routing (MTR)

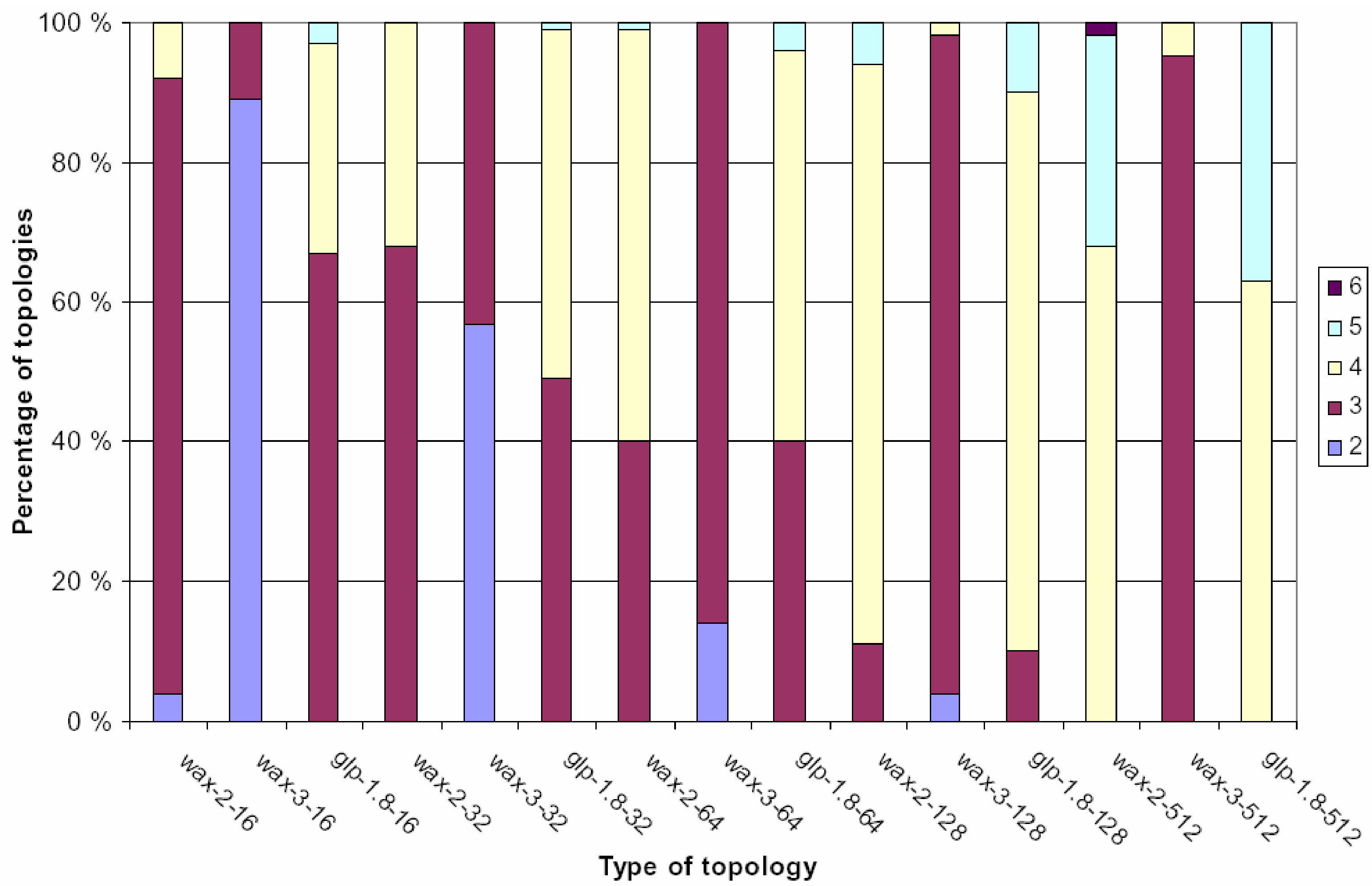
- Standardization within IETF isis and ospf working groups
- Proposed for computing different paths for unicast traffic, multicast traffic, different classes of service, or an in-band network management topology

→ Can be used to implement MRC

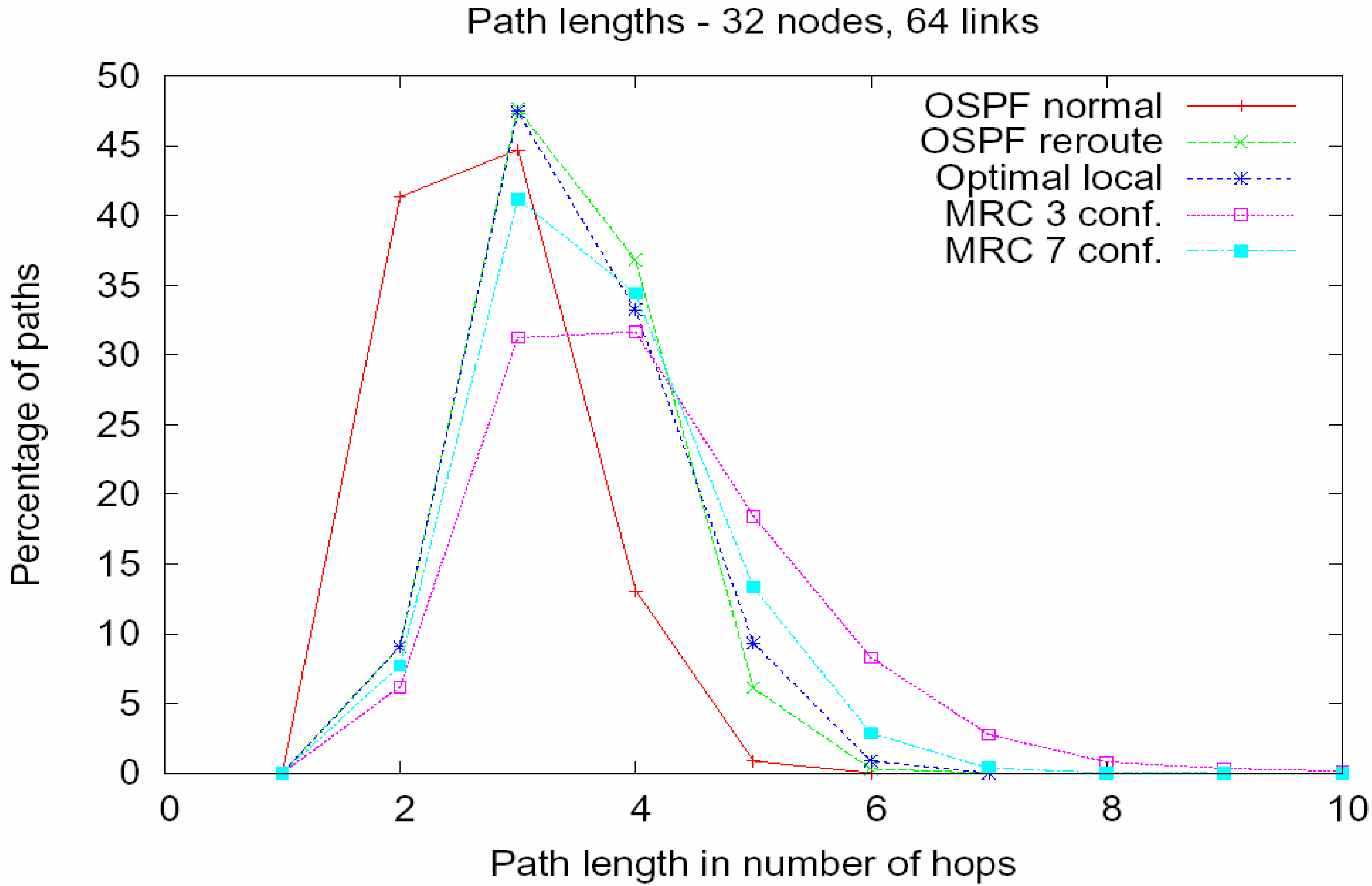
  → One topology is one MRC-configuration

Not-via	MRC	FIFR
Extensive IGP changes	**Less IGP changes (MTR)**	Medium IGP changes
Less simple global view	**Simple global view**	No global view
One extra destination address per Not-via address in the routing table	**One extra routing table per backup topology**	Interface specific routing
One extra SPT calculation per component	**One extra SPT calculation per backup topology**	Additional SPT algorithm
Tunneling	**No tunneling**	No tunneling
Multi-failure??	**Multi-failure OK**	Multi-failure??

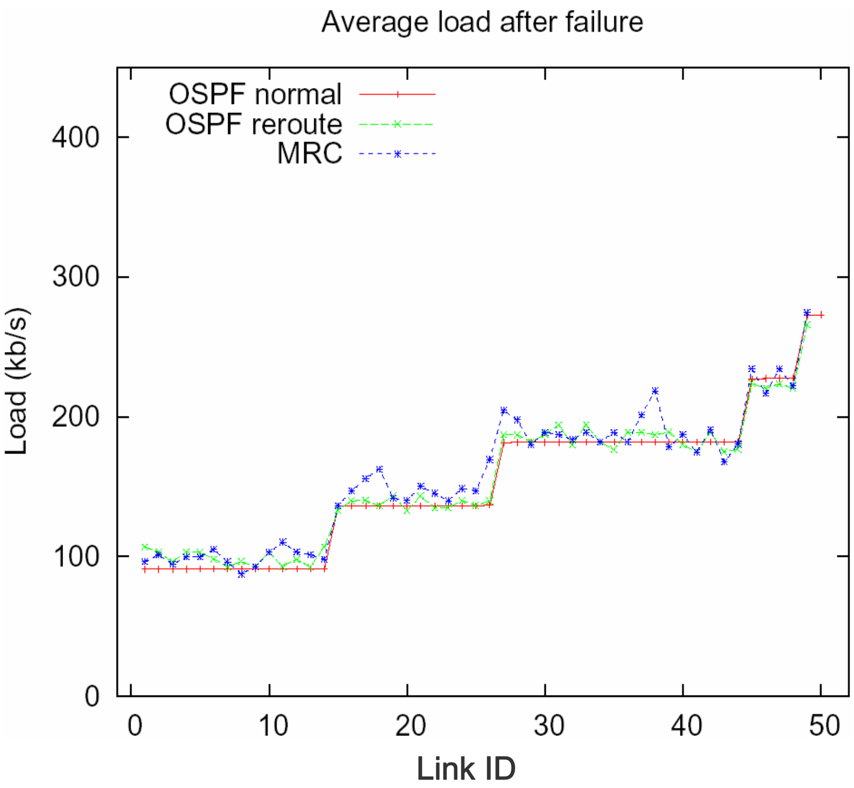
MRC - Number of configurations needed

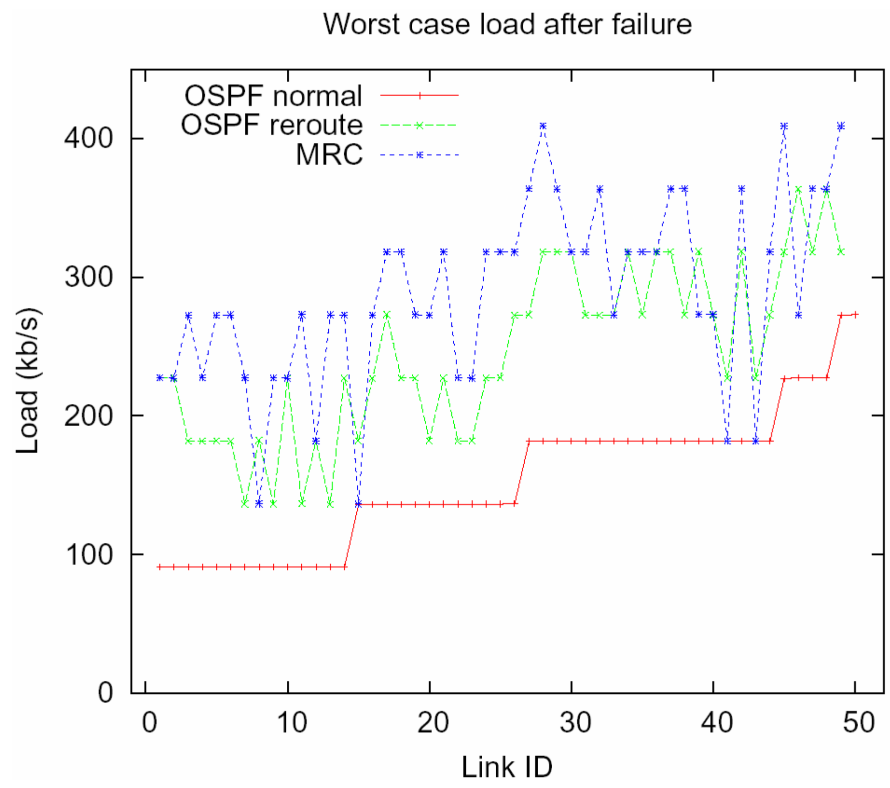


MRC - Path lengths



MRC - Link load after failure





Conclusion 1

Multiple Routing Configurations - MRC

- Gives fast recovery from component failures in IP networks
  - Good for Latency Intolerant, Mission Critical Applications
  - Loop free local reaction to failures, immediately after failure detection
  - 100% coverage against single link and node failures
  - Handles link and node failures with a single mechanism
- Based on maintaining a small set of backup routing configurations – scales well

Conclusion 2

- **A proactive IP-routing solution is needed for Mission Critical Applications**
- IETF IP fast reroute, FIFR and MRC have all pros and cons
- IP fast reroute needs Not-via to obtain full coverage
- MRC provides a very good alternative
- MRC can be implemented with IETF Multi-Topology Routing
- MRC can be extended to guarantee two concurrent failures or all failures in a Shared Risk Group

- A. Kvalbein, A. F. Hansen, T. Cicic, S. Gjessing, and O. Lysne: "**Fast IP network recovery using multiple routing configurations**," INFOCOM, Apr. 2006
- A. Kvalbein, T. Cicic and S. Gjessing: "**Post-Failure Routing Performance with Multiple Routing Configurations**", INFOCOM, May 2007

Some more references

- P-cycles: D. Stamatelakis and W. D. Grover, “IP layer restoration and network planning based on virtual protection cycles,” *IEEE Journal on selected areas in communications*, vol. 18, no. 10, Oct. 2000.
- FIFR: Z. Zhong, S. Nelakuditi, Y. Yu, S. Lee, J. Wang, and C.-N. Chuah,“Failure inferencing based fast rerouting for handling transient link andnode failures,” in *Proceedings of IEEE Global Internet*, Mar. 2005
- Not via: S. Bryant, M. Shand, and S. Previdi, “IP fast reroute using not-via addresses,” Internet Draft (work in progress), Oct. 2005, draft-bryantshand-IPFRR-notvia-addresses-01.txt.
- ECMP: A. Iselt, A. Kirstdter, A. Pardigon, and T. Schwabe, “Resilient routing using ECMP and MPLS,” in *Proceedings of HPSR 2004*, Apr. 2004.
- MRC: A. Kvalbein, A.F. Hansen, T. Cicic, S. Gjessing, O. Lysne, “Fast IP Network Recovery using Multiple Routing Cionfigurations”. In INFOCOM 2006, edited by Arturo Azcorra, Joe Touch, Zhili Zhang. IEEE, Barcelona, Spain, pages 23–29, 2006.