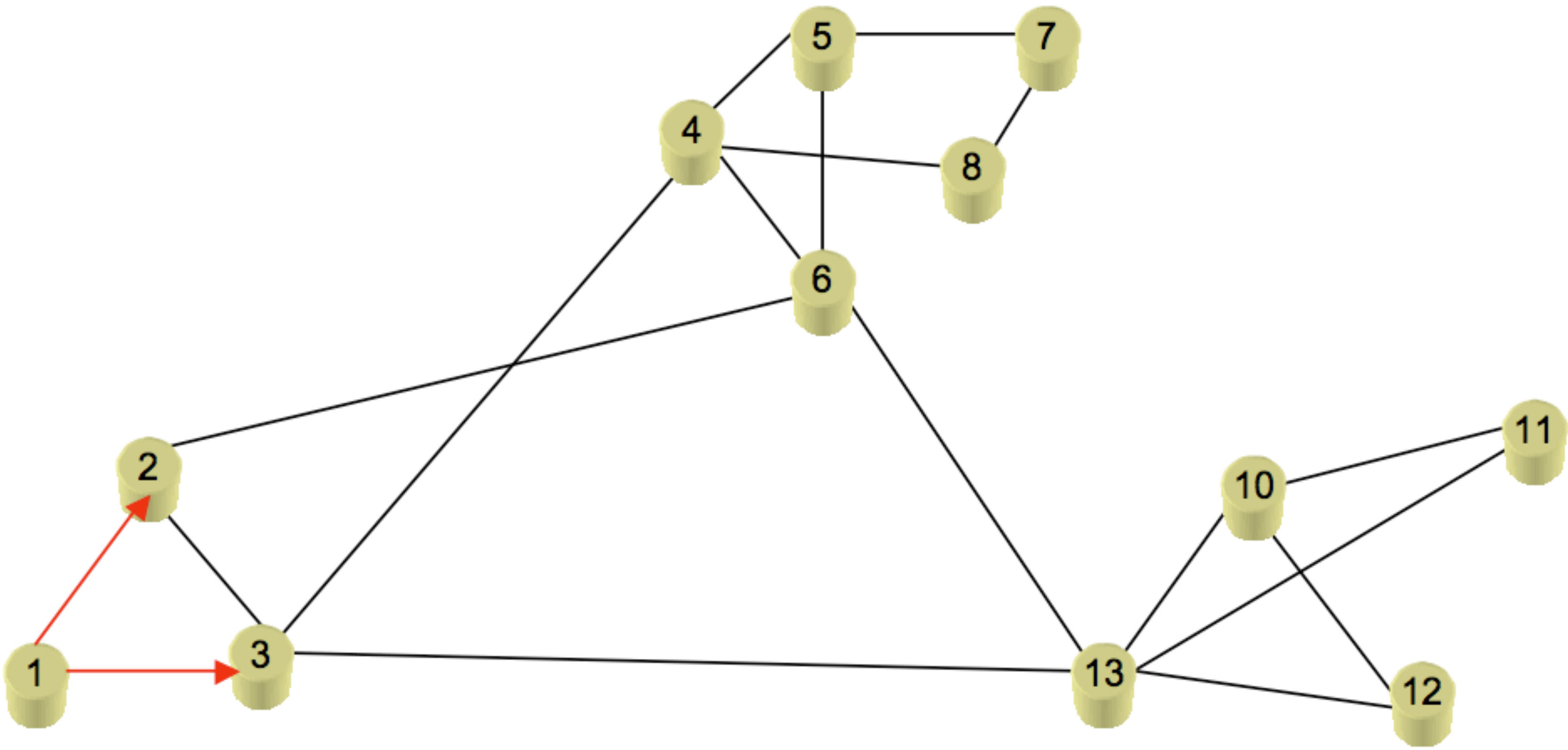
CSC 525:
Principles of Computer Networks

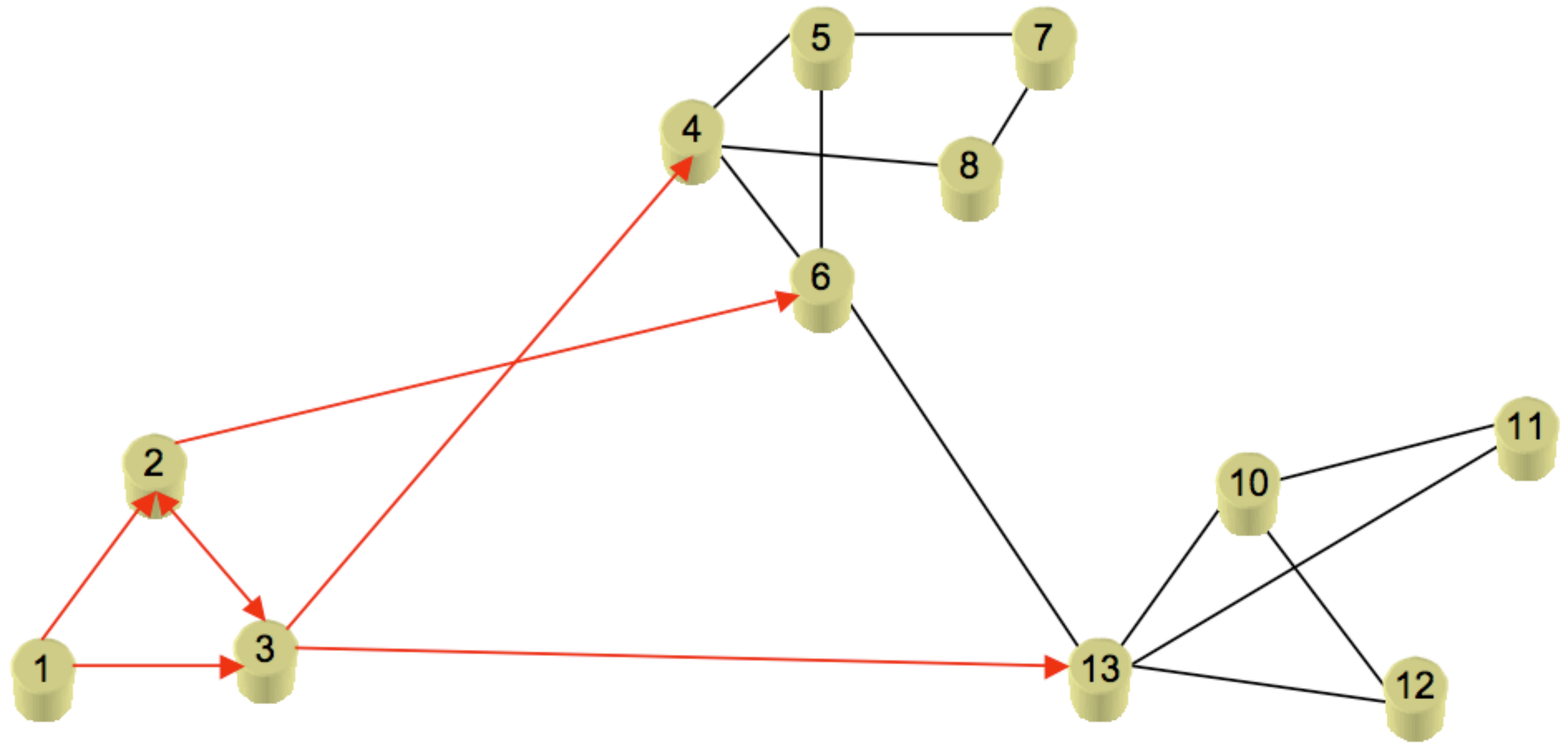
Link State Routing

- The idea:
  - Each node floods its local topology information (i.e., link state announcements, LSA) to the entire network.
  - Each node reassembles the complete network topology (i.e., link state database)
  - Run Dijkstra's algorithm to calculate the shortest paths from the node to every destination.
- The challenge:
  - ensure every node has the same topology, i.e., link state database

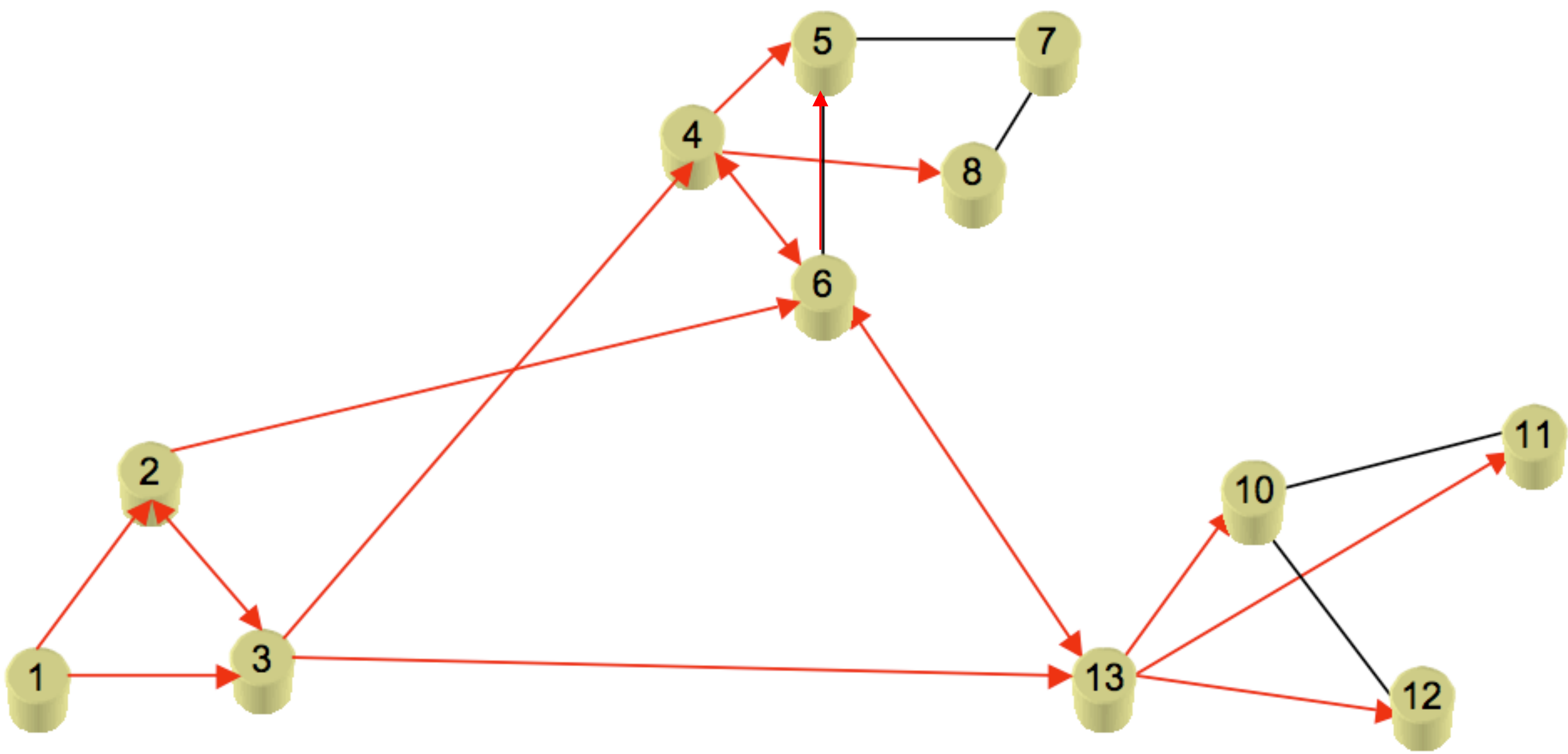
Flooding Example



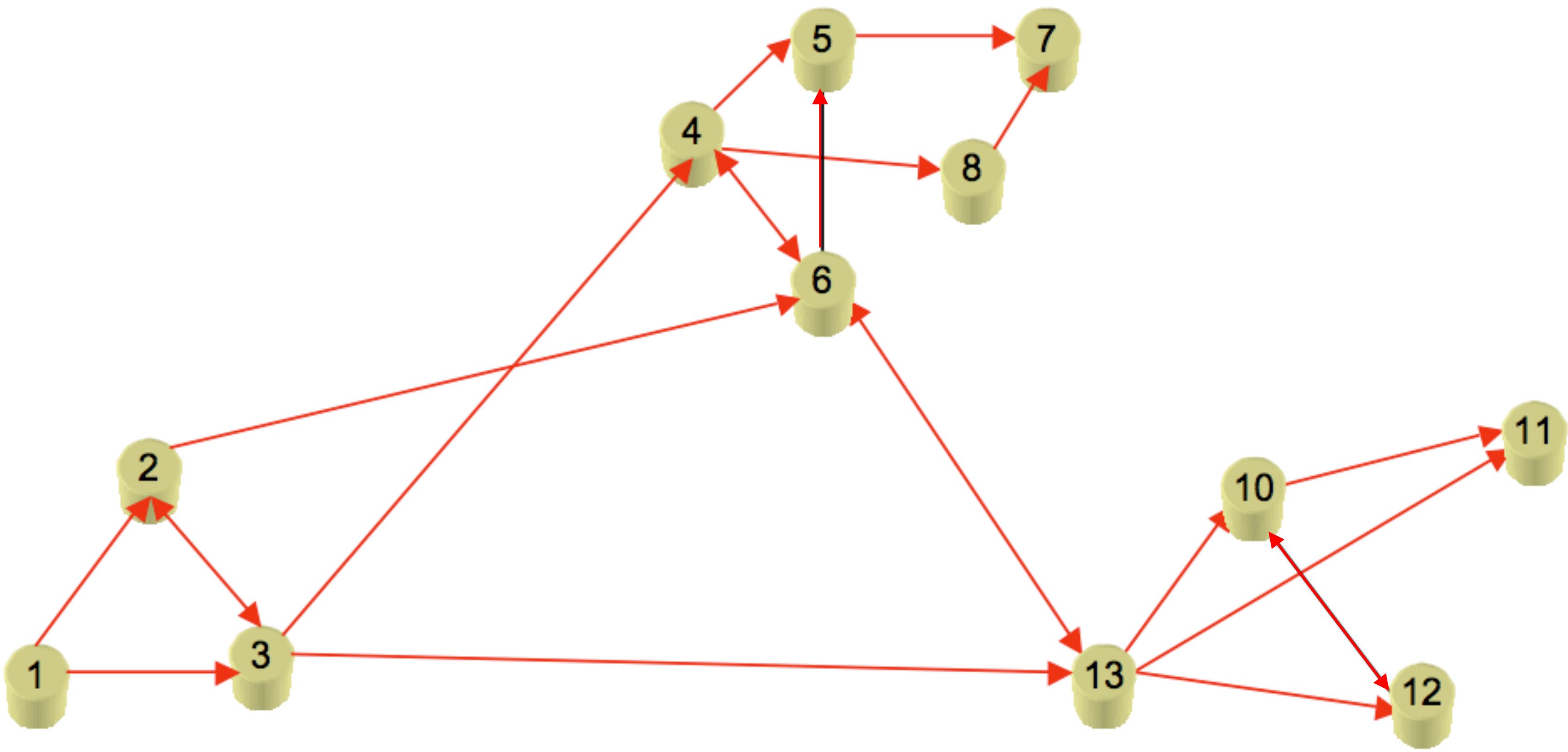
Flooding Example



Flooding Example



Flooding Example



Pitfalls in flooding packets

- Flooding
  - Sending incoming packets to all other outgoing links.
- Problems
  - Duplicate packets
  - Loops
  - Because packet forwarding is memoryless.
- Flooding is used in different network protocols, and each protocol has a mechanism to fix duplicates and loops.
  - It often involves some network states.

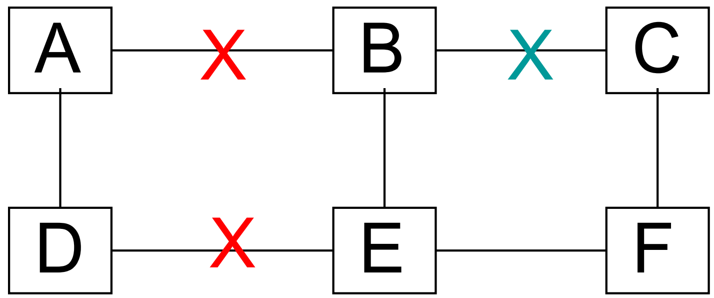
Link State Flooding

- A Link state update contains:
  - Originator's ID (usually an IP address)
  - List of originator's links and their costs
  - List of subnets that the originator connects to (i.e., prefixes).
  - Sequence number
  - Age
- Upon receiving a link state update
  - If this is a **new** update (based on originator ID and sequence number), forward it to all neighbors except the one from which it was received.

Reliable Flooding

- Ensure the update will get to every router and avoid duplicate and loop.
- Sequence number
  - The originator increases seq number by one for each new update
  - A router only forwards updates with newer sequence numbers
  - may wrap around
  - rebooted routers don’t send update immediately
- Per-hop acknowledgement
- Aging
  - During flooding increment by one after each hop
  - In database increment by one every second
  - Max age one hour
  - Remove expired entries
- Originator re-floods link states every half an hour

Bringing Up Adjacencies



- A-B and D-E fail, network is partitioned
- Then B-C fails, flooding is done within B,C,E,F.
- Now A-B and D-E are up
  - A, D won't know that B-C is down!

Database Synchronization

- When a link is up, neighbors exchange "database description."
- Find out inconsistency
  - Who has which links/nodes and what are the sequence numbers of those updates.
- Request and reply corresponding records to synchronize the databases

Link State Protocols

- Two protocols that realize the link state routing:
  - OSPF (open shortest path first)
  - IS-IS (intermediate systems to intermediate systems)
- Both are being used by major ISPs today.

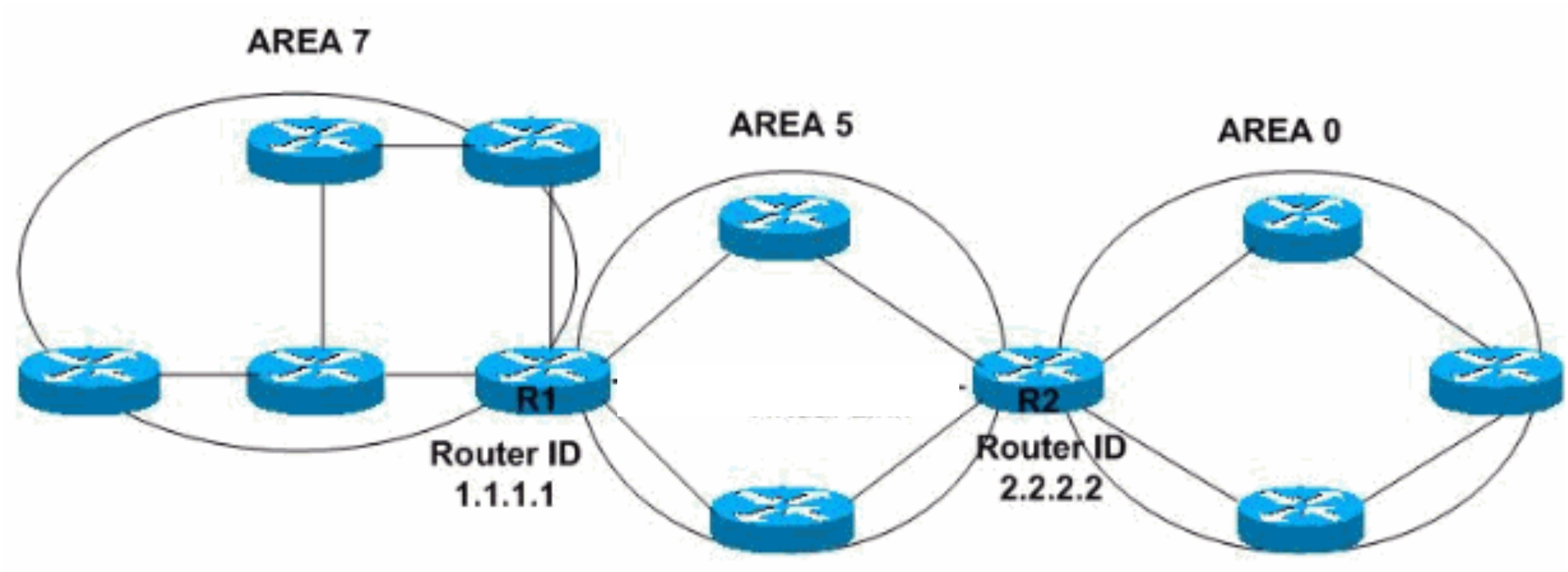
OSPF Operations

- Periodic Hello messages between neighbors
- A newly established link triggers database synchronization between neighbors
- Changes in topology and network are flooded to the entire network
  - re-flood every half an hour
- Run Dijkstra's algorithm locally on the full topology to compute shortest paths.

OSPF Area

- Split a large network into multiple “areas” to increase scalability
- A router within an area knows
  - All links in the same area
  - Reachability information to prefixes in other areas, but don’t know the topology details of other areas
- An area border router
  - Knows the topologies of its bordering areas and reachability information in remote areas
- Reduce routing overhead in communication, memory, and CPU
  - But the routes may not be the “shortest”.

OSPF Area



Distance Vector vs. Link State

- Distance Vector
  - Simple, low overhead
  - Routing loops, count-to-infinity
  - Used by networks with small size or simple topology
- Link State
  - Complex, high overhead
  - Fast convergence, no persistent loops
  - Support multiple metrics and multiple paths of equal cost.
  - Used by major ISPs
- Compare mechanisms:
  - What's in a routing updates?
  - How is a routing update propagated?

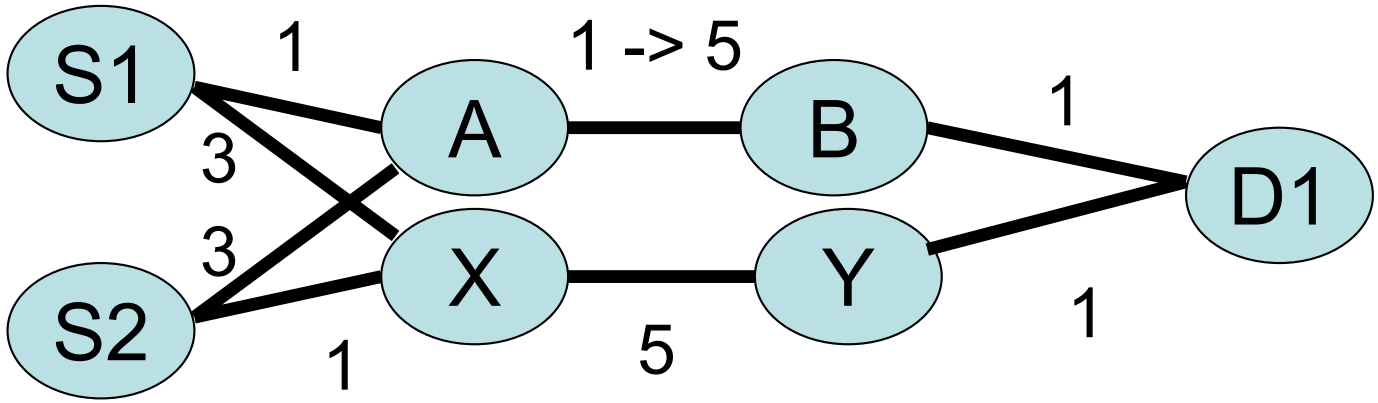
"The Revised ARPANET Routing Metric"

- Objective
  - devise a new *routing metric* that limits the route oscillations under heavy traffic load.
- Approach
  - Limit the range of dynamics of the metric.
  - Smooth the variation of the metric
  - Account for link types
- Illustrate the trade-off between adaptability and stability
- A great example of traffic engineering (TE) by dynamically adjusting link weights in a link-state protocol.

Routing Metric In Use

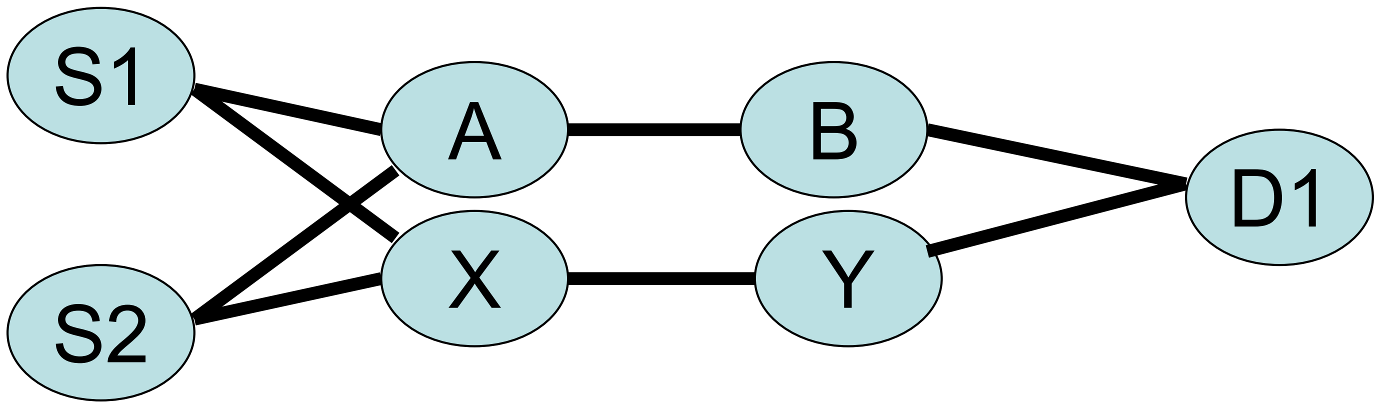
- Distance Vector (1969)
  - Hop count
- Link State (1979)
  - Use link delay as the link cost
- Link delay is the sum of
  - Processing delay, transmission delay, propagation delay
    - independent of traffic load
  - Queuing delay
    - Depend on traffic load
  - The intention was to adapt routing paths to traffic load.
- In today's networks, link metrics are set by operators. They can be either load-sensitive or not.

Example



- Under light traffic
  - Both sources' shortest paths use link A-B.
- With more traffic
  - Higher queueing delay leads to higher link cost at A-B.
  - A change in link cost triggers link state flooding and recomputing shortest paths.
  - Now the shortest path from S2 uses link X-Y, so will the traffic.

Example

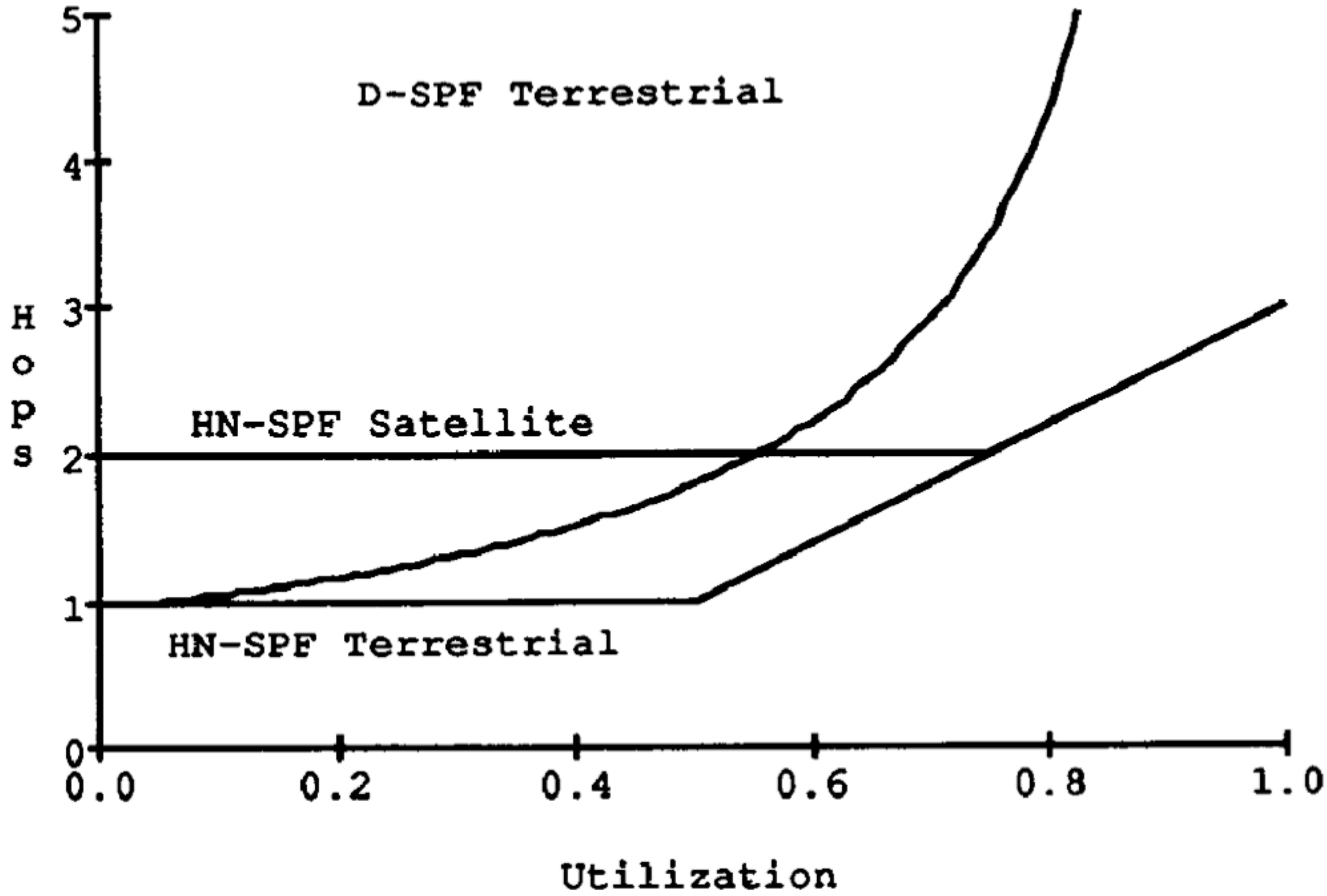


- However, undesired oscillation may happen under heavy traffic
  - Say, cost of AB increases to 10. Both S1 and S2 will switch to link XY. Now the cost of XY will become very high, then all the traffic will switch back to AB, and so on.
  - This will lead to low bandwidth utilization, possible congestion and packet loss, and more routing overhead.

Revised Metric

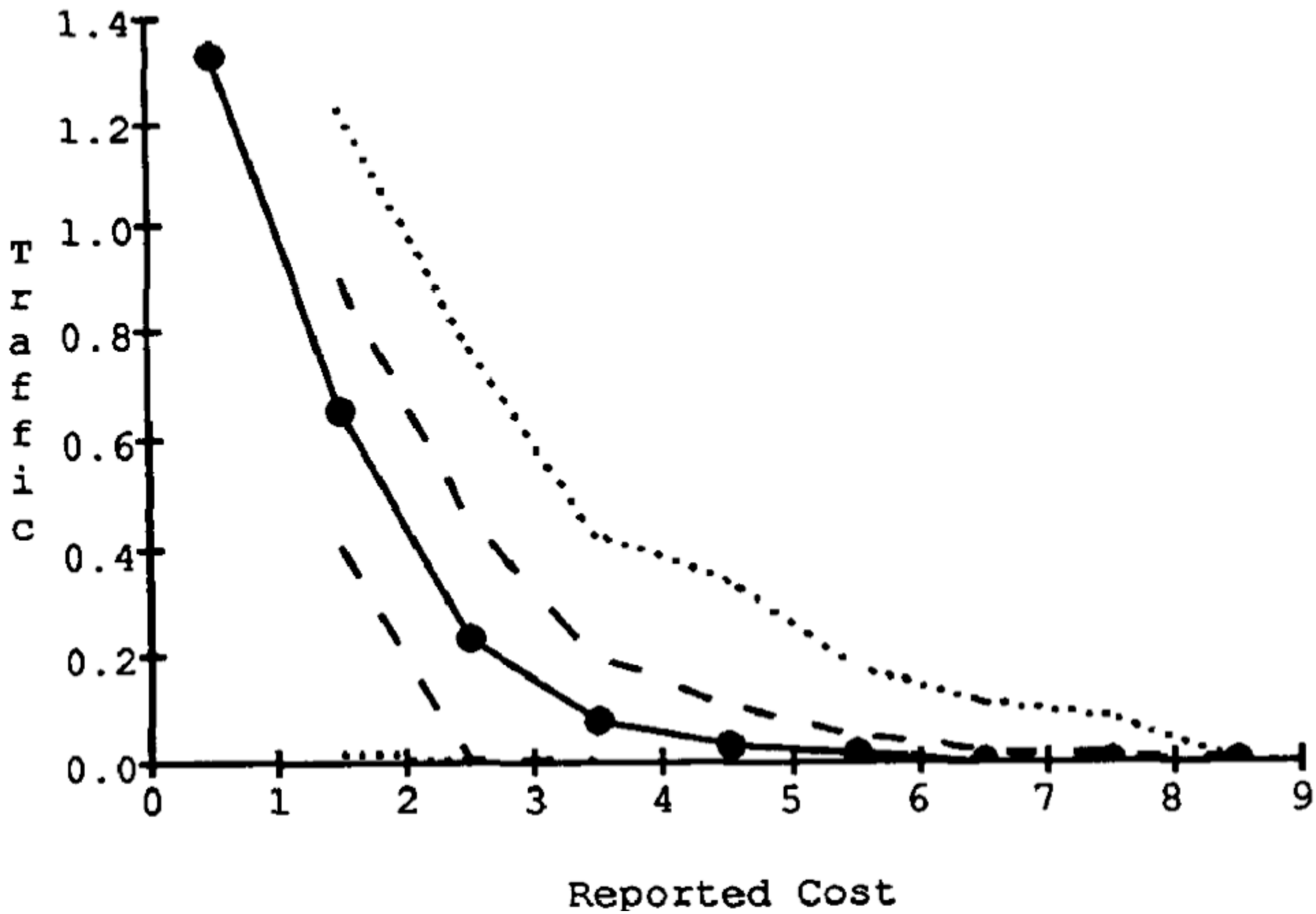
- Normalize link cost to “hops”
- Still take into account traffic load, but
- Limit the range of the metric changes
  - max link cost is 3 hops
  - min link cost depends on link type
  - most expensive link is 7 times of the least expensive
  - consider traffic only when it’s moderate or heavy
- Smooth the variation
  - utilization measurement is averaged with previous results
  - limit the change of link cost in each round to 0.5 hop

Metric Map



Changing traffic load leads to different link costs.

Network Response Map

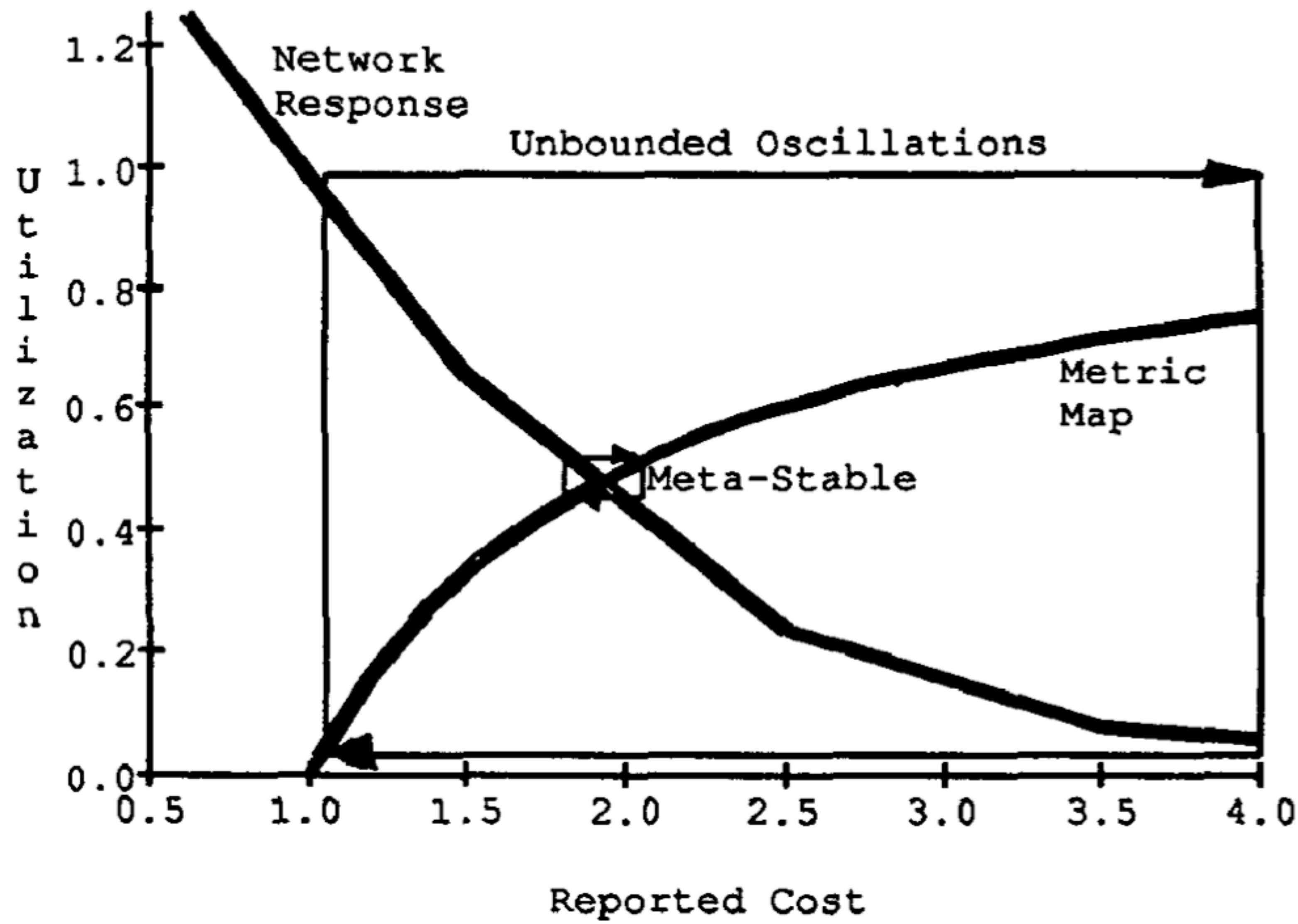


Changing link costs leads to different traffic load.

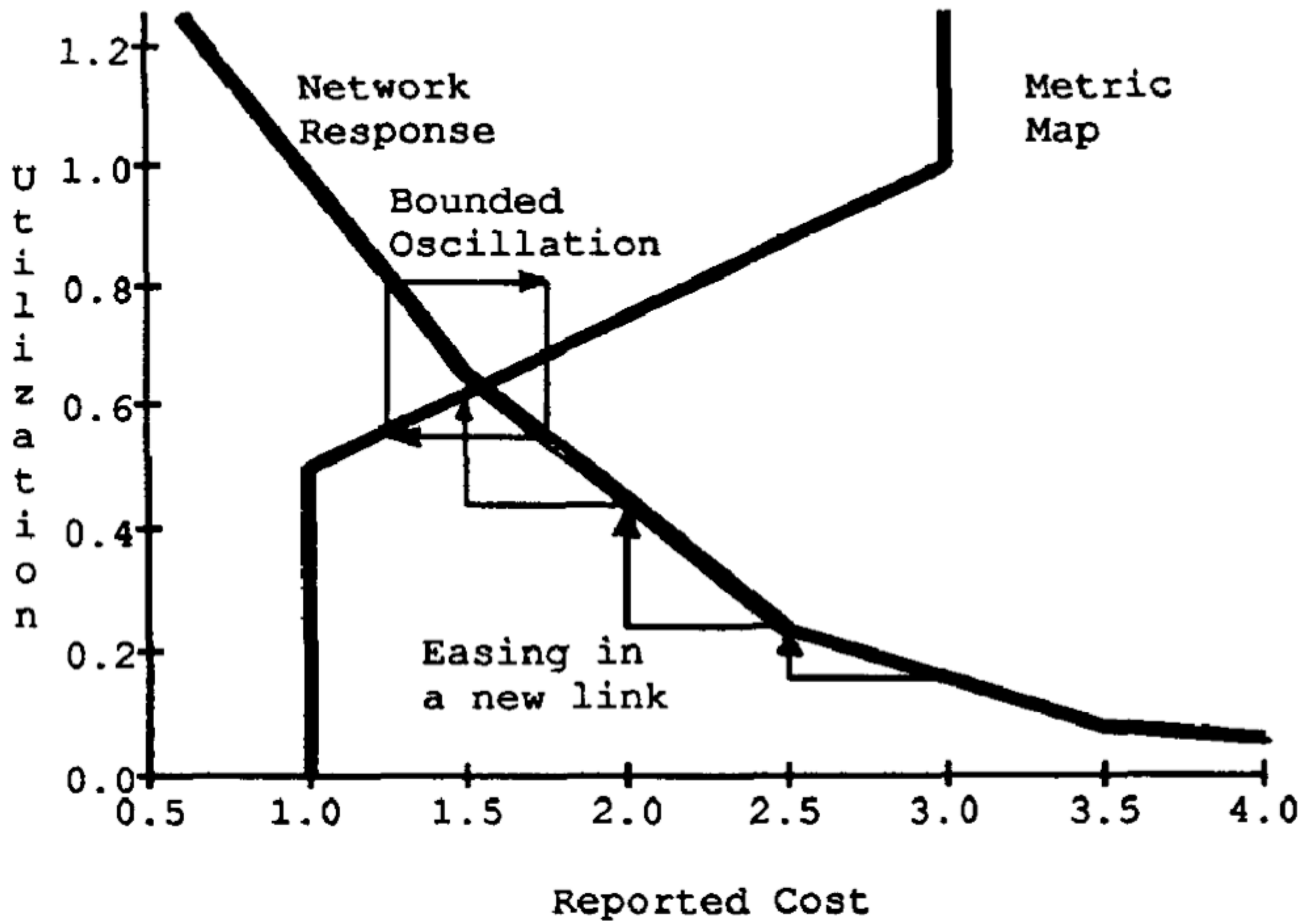
Equilibrium Calculation

- Utilization measurement => Reported Cost => LSA flooding => Path Calculation => Traffic Switch => Utilization measurement
- Equilibrium is reached when there's no more traffic switching
  - Different from routing convergence. Traffic engineering operates at a larger time scale than routing.
- Combine metric map and network response map to get the equilibrium point

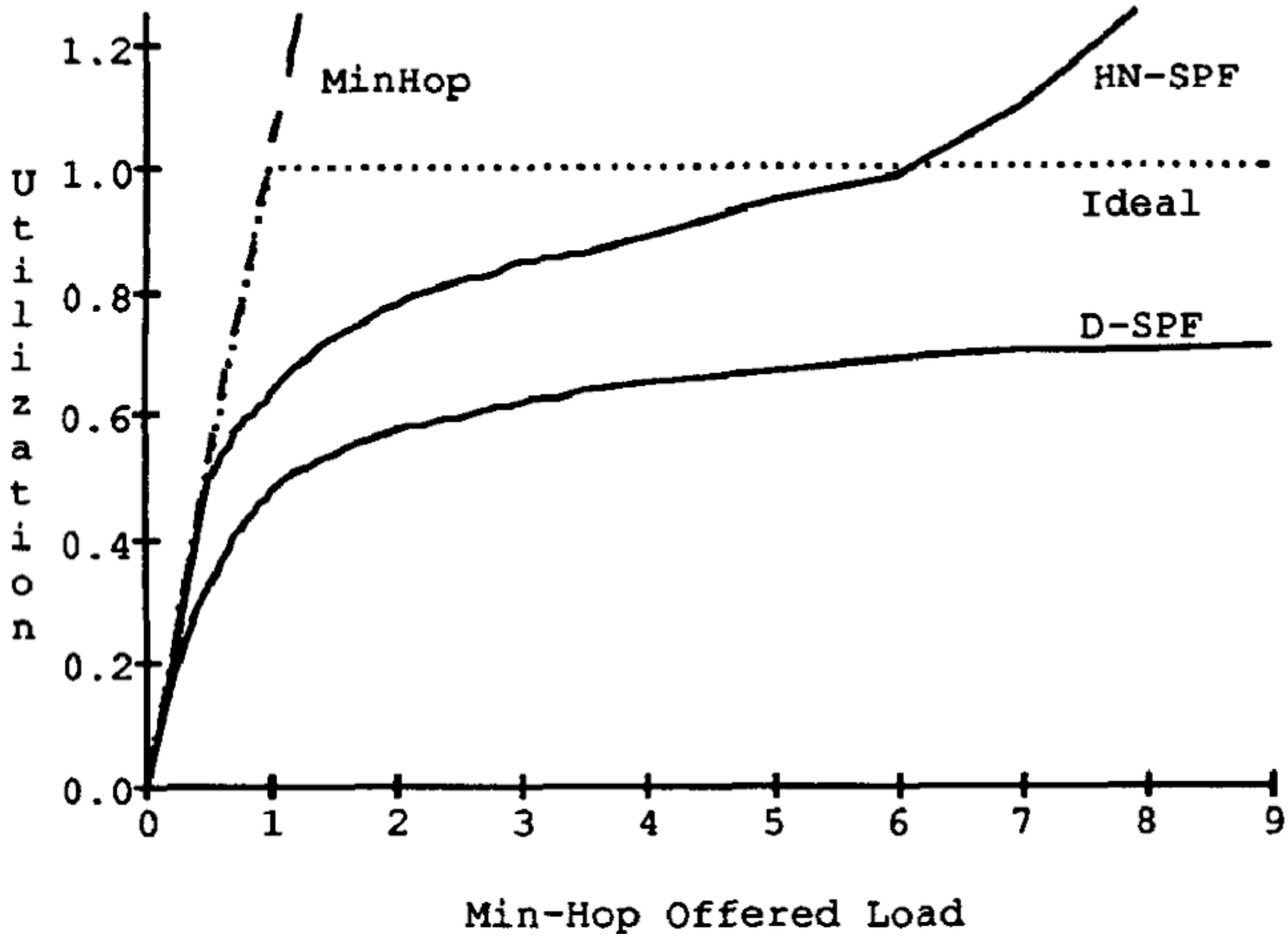
Dynamics of D-SPF



Dynamics of HN-SPF



Comparison of Metrics



Performance Gain of HN-SPF