NIST National Institute of Standards and Technology

INFORMATION TECHNOLOGY LABORATORY

Trustworthy Networking Program

Border Gateway Protocol (BGP): Investigation of Vulnerabilities and Simulation Studies of Attack Impacts

Realistic AS Topology, RFD Exploitation, Sensitivity to Policy, and Measurement of Routing Performance Degradation

K. Sriram, D. Montgomery, O. Borchert, O. Kim, and R. Kuhn

National Institute of Standards and Technology

Gaithersburg, MD 20878

Contacts: dougm@nist.gov, ksriram@nist.gov

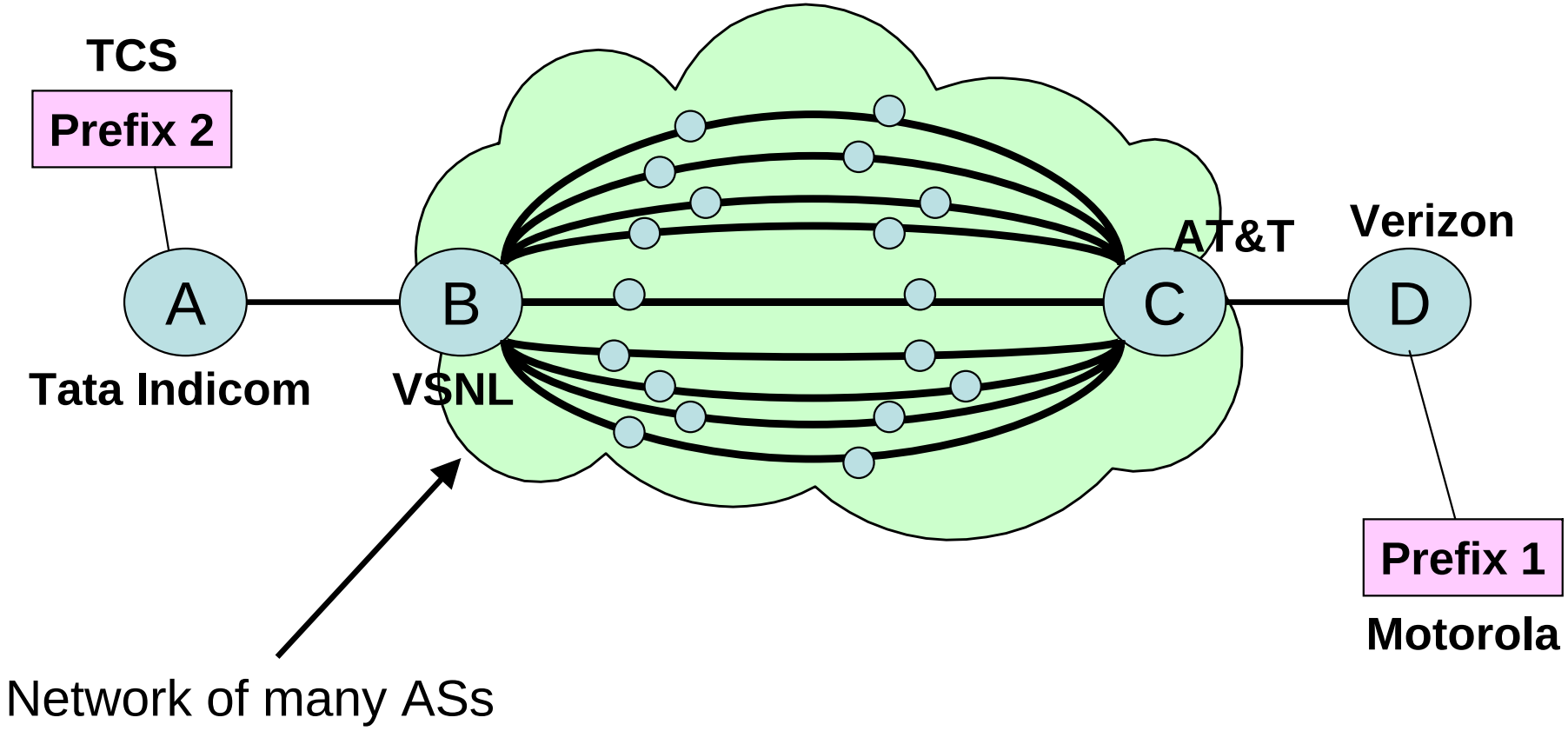
October 27, 2006

This research was supported by the Department of Homeland Security under the Secure Protocols for the Routing Infrastructure (SPRI) program and the NIST Information Technology Laboratory Trustworthy Networking Program.

Outline of the Talk

- Brief BGP tutorial
- BGP vulnerabilities
- RFD exploitation
- Analytical model
- Realistic simulation topology generation
- Simulation methodology
- Simulation results:
  - Grid and realistic topologies
  - Effects of routing policy
  - Metrics and Measurements:
    - Routing performance degradation
    - Denial of Service (DOS) effects

BGP Basics



- Border Gateway Protocol (BGP) -- Inter-domain Routing
- Autonomous System (AS) consists of a provider's network of routers
- ASs originate prefixes and propagate path updates to peers

BGP Vulnerabilities and Risks

- Much speculation..
  - Potential vulnerabilities and consequences.
  - Most threatening might be “bugs” – can cripple a router with a single packet.

WIRED NEWS

Flaw Could Cripple Entire Net Associated Press

Story location: http://www.wired.com/news/technology/0,1282,63143,00.html

11:23 AM Apr. 20, 2004 PT

Researchers found a serious security flaw that left core Internet technology vulnerable to hackers, prompting a secretive effort by international governments and industry experts in recent weeks to prevent global disruptions of Web surfing, e-mails and instant messages.

Experts said the flaw, disclosed Tuesday by the British government, affects the underlying technology for nearly all Internet traffic. Left unaddressed, they said, it could allow hackers to knock computers offline and broadly disrupt vital traffic-directing devices, called routers, that coordinate the flow of data among distant groups of computers.

"Exploitation of this vulnerability could have affected the glue that holds the Internet together," said Roger Cumming, director for England's National Infrastructure Security Coordination Centre.

- Little public analysis or data ….
  - Empirical analysis of vulnerabilities and their potential consequences.
  - Trace data of actual attacks on the routing infrastructure.

Efforts to Understand the Risks and Possible Solutions

Long term solutions in a state of flux.

- S-BGP, SO-BGP, MD5/IPsec, GTSM, Route Verification, Filtering, Listen & Whisper, etc.
- Range of technologies that may, or may not, be viable.
- It depends on what **your** view of the **risks and benefits vs. costs**.

Lack of shared understanding of both the problem & solution space.

- Need to raise community awareness of potential threats, **risks**, mitigation techniques and their **cost**.
- Need to take "systems view" of improving routing's survivability.
- DHS – "need some way of characterizing benefit vs. cost of various solution techniques."

NIST Objectives:

- **Expedite Research** - Help researchers characterize the design space: risks, mitigation techniques and deployment costs.
- **Expedite Development** - Evaluate the effectiveness and impact of proposed technical solutions.
- **Expedite Adoption** - Help users / decision makers understand threats & mitigations.

NIST Efforts

Near Term Efforts:

- ***DHS - "Focus on the problem / design space."***
- **Large Scale Modeling of BGP Attacks**
  - Most modeling / analysis focused on post-mortem analysis of recent worms/viruses, but **"what if" scenarios** of yet unseen attacks may be more important.
  - **Risk analysis** of the potential impact of successful attacks on BGP.
  - Discover and evaluate **new vulnerabilities**.
  - Look for **emergent behaviors** – e.g., cascading failures, congestion collapse, degraded routing.
  - Framework for **characterization of proposed solutions** & deployment scenarios
- **Modeling and Analysis of Proposed Solutions**
  - Characterizing the effectiveness and cost of the various combinations of countermeasures.
  - Characterize the risk associated with the deployment of proposed solutions.
- **Issue Federal Guidance**
  - FISMA guidance on BGP Security.

BGP Attack Tree Enumeration

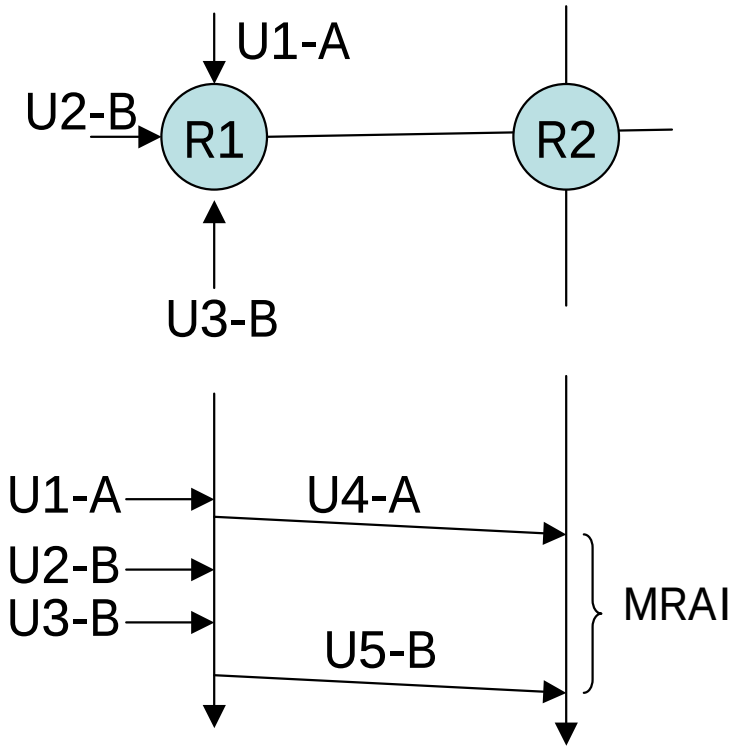
- Broad classification of attacks (IETF drafts):
  - Establish Unauthorized BGP Session with Peer
  - Originate Unauthorized Prefix/Attribute into Peer Route Table
  - Change Path Preference of a Prefix
  - Conduct Denial/Degradation of Service Attack Against BGP Process
  - Reset a BGP Peering Session
  - Send Spoofed BGP Message

BGP Peering Session Attacks

- There are many different attack possibilities on the BGP routing infrastructure (IETF ID: draft-ietf-rpsec-bgpattack-00)
- We focus on attacks that cause BGP peering sessions to be reset
- Common way to reset a BGP peering session is to reset or attack the underlying TCP connection
- Multiple TCP/ICMP vulnerabilities documented - may be exploited to launch TCP connection-reset attacks
  - "Slipping in the window" TCP reset attack (requires correctly guessing a TCP sequence number within a flow control window)
  - ICMP error messages spoofed to cause TCP reset (IETF ID, Dec. 2004)
    - ✓ Does not require guessing the TCP sequence number
    - ✓ Hard ICMP error messages (spoofed)
    - ✓ Soft ICMP error messages (spoofed)

MRAI: Minimum Route Advertisement Interval

- A BGP router sends route advertisements/withdrawals to a peer at intervals no smaller than MRAI
- Jittered MRAI: randomly chosen from a range of 22.5s to 30s (independently at each node)
- MRAI is a sender side discipline for neighbor overload avoidance



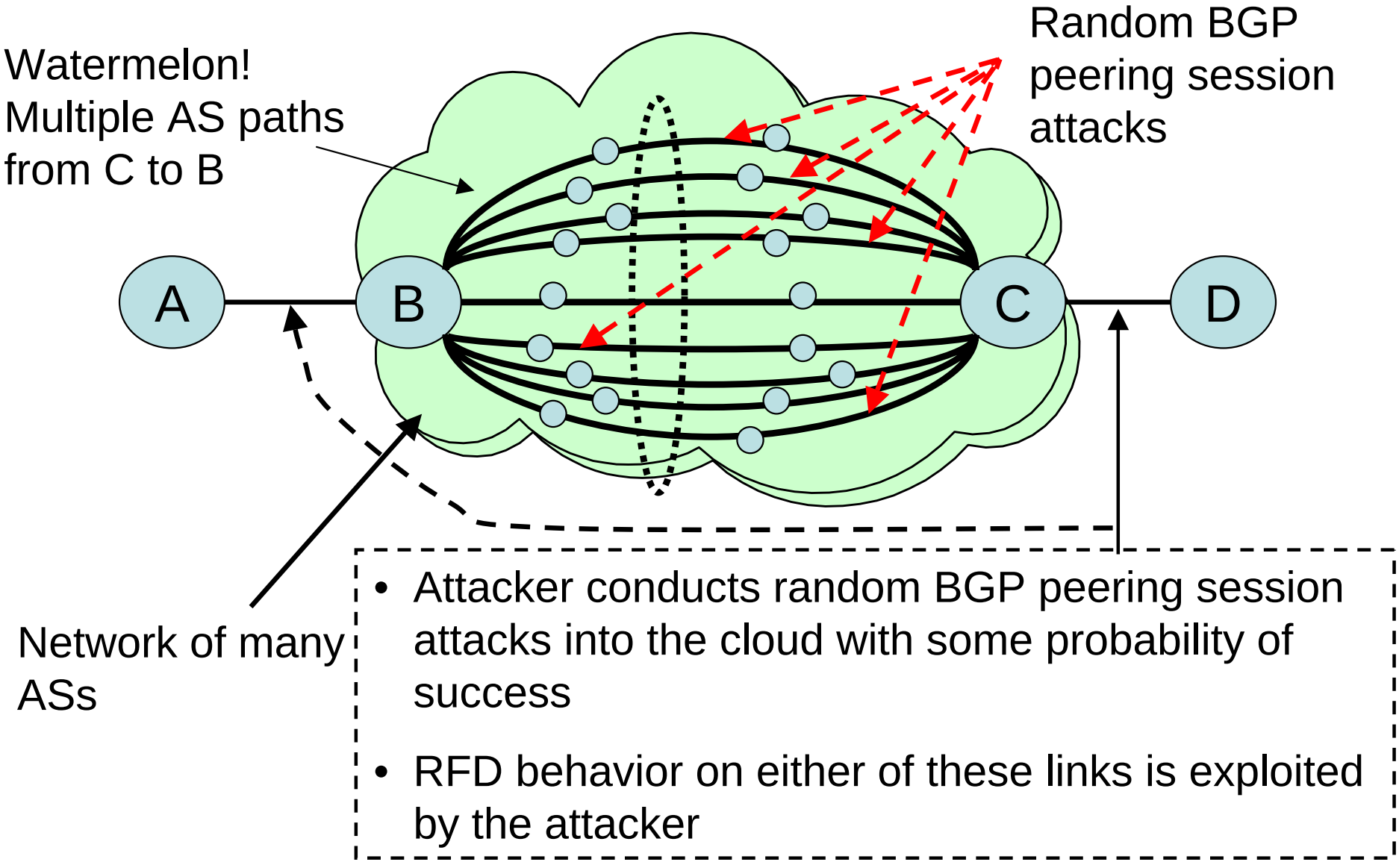
RFD: Route Flap Damping

- An upstream router assigns an incremental RFD penalty to a peer and destination (prefix) combination each time an update is received from that peer for that destination
- If the RFD penalty exceeds a preset cutoff threshold, then the route is suppressed
- RFD is a method for receiver side route monitoring and suppression in the event of frequent updates

RFD Parameter	Vendor A	Vendor B	
Withdrawal penalty	1000	1000	
Re-advertisement penalty	0	1000	
Attribute change penalty	500	500	
Cutoff threshold	2000	3000	
Half-time	900	900	sec
Reuse threshold	750	750	
Max supress time	3600	3600	sec
Max penalty	12000	12000	

- **The two sets of numbers correspond to two commercial implementations**
- **Use the numbers for sensitivity study in our numerical examples**

Exploitation of Route Flap Damping



- Attacker conducts random BGP peering session attacks into the cloud with some probability of success
- RFD behavior on either of these links is exploited by the attacker

Illustration: How It Works (MRAI = 30 s)

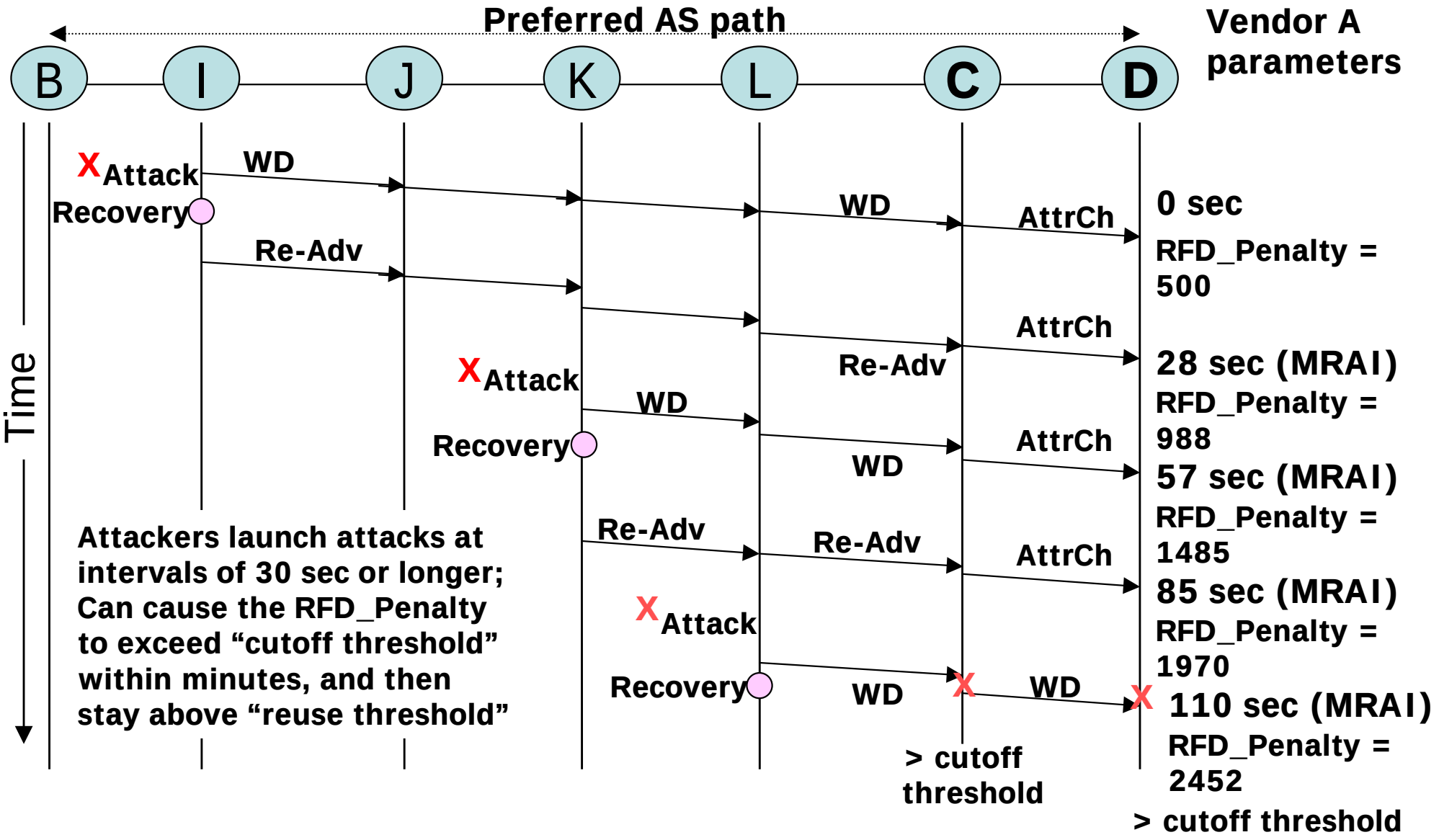
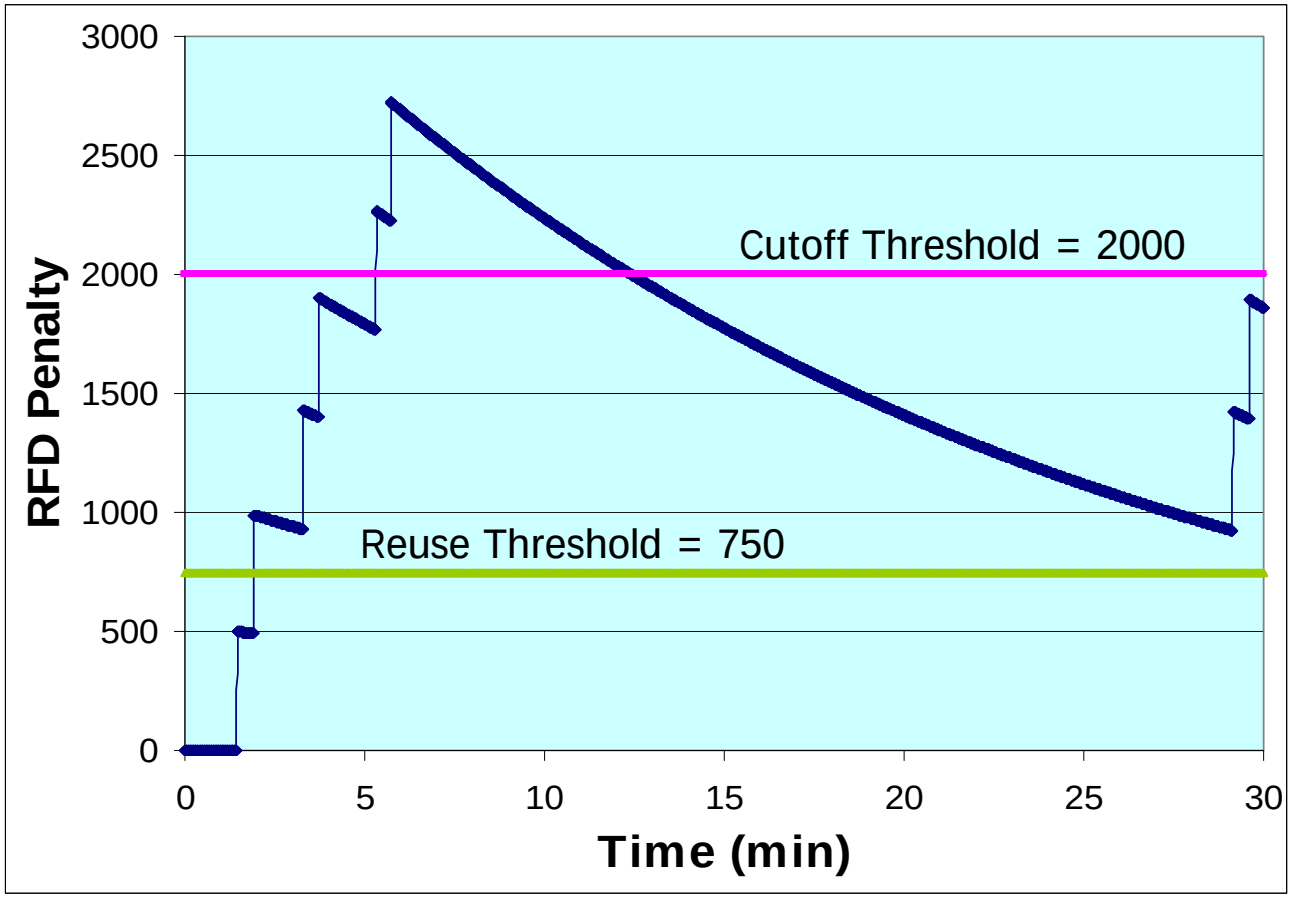
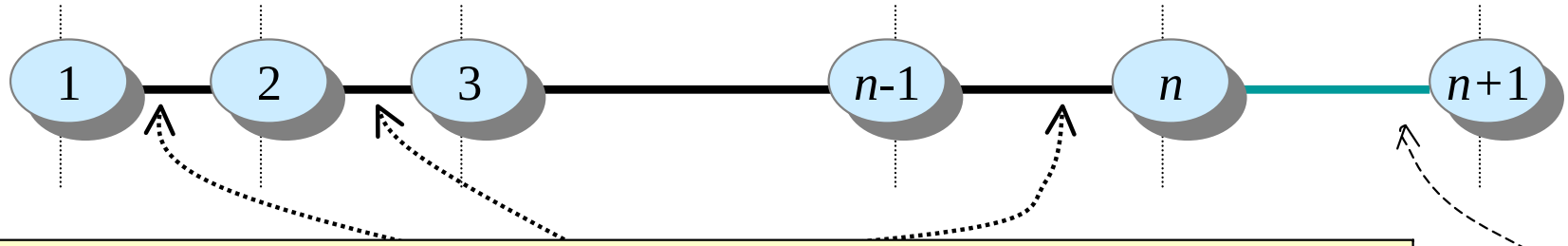


Illustration: How It Works (MRAI = 30 s)



- The update interval is effected by MRAI
- Attackers need to successfully attack one of the BGP peering sessions on the preferred path for the penalty to go higher
- 30 sec MRAI allows enough time for the damaged BGP session to recover within the MRAI
- The waves of attacks would be spaced at intervals equaling approximately MRAI
- To achieve prolonged AS isolation, it is enough if only some of the attacks succeed
- Once RFD penalty is exceeded, the attack interval can be made larger (although attackers don't know when they have succeeded)

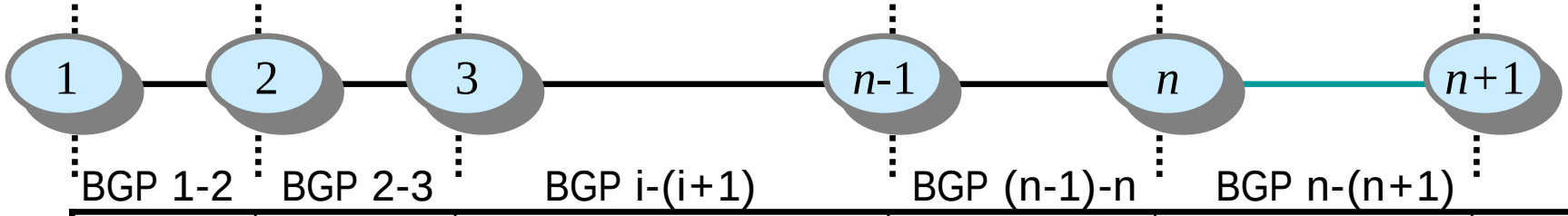
Analytical Model for AS Isolation Probability



- ***n*-1 BGP peering sessions**
- **Attacks are assumed to be spaced at roughly MRAI intervals**
- **Each router is subjected to an attack with probability *p* in each interval**
- **Each BGP peering session can be attacked with probability *q* if there is a router at either end that is subjected to attack**

- **Model predicts the probability that update rejections due to Route Flap Damping are imposed at router *n*+1 for peer *n* and destination 1**
- **Model also predicts the sustenance probability that the attackers can sustain the RFD in update rejection state and thus cause prolonged isolation between router *n*+1 and destination 1 (also all subsequent destinations reachable via router 1).**

Attacks and RFD Penalty Accumulation Model



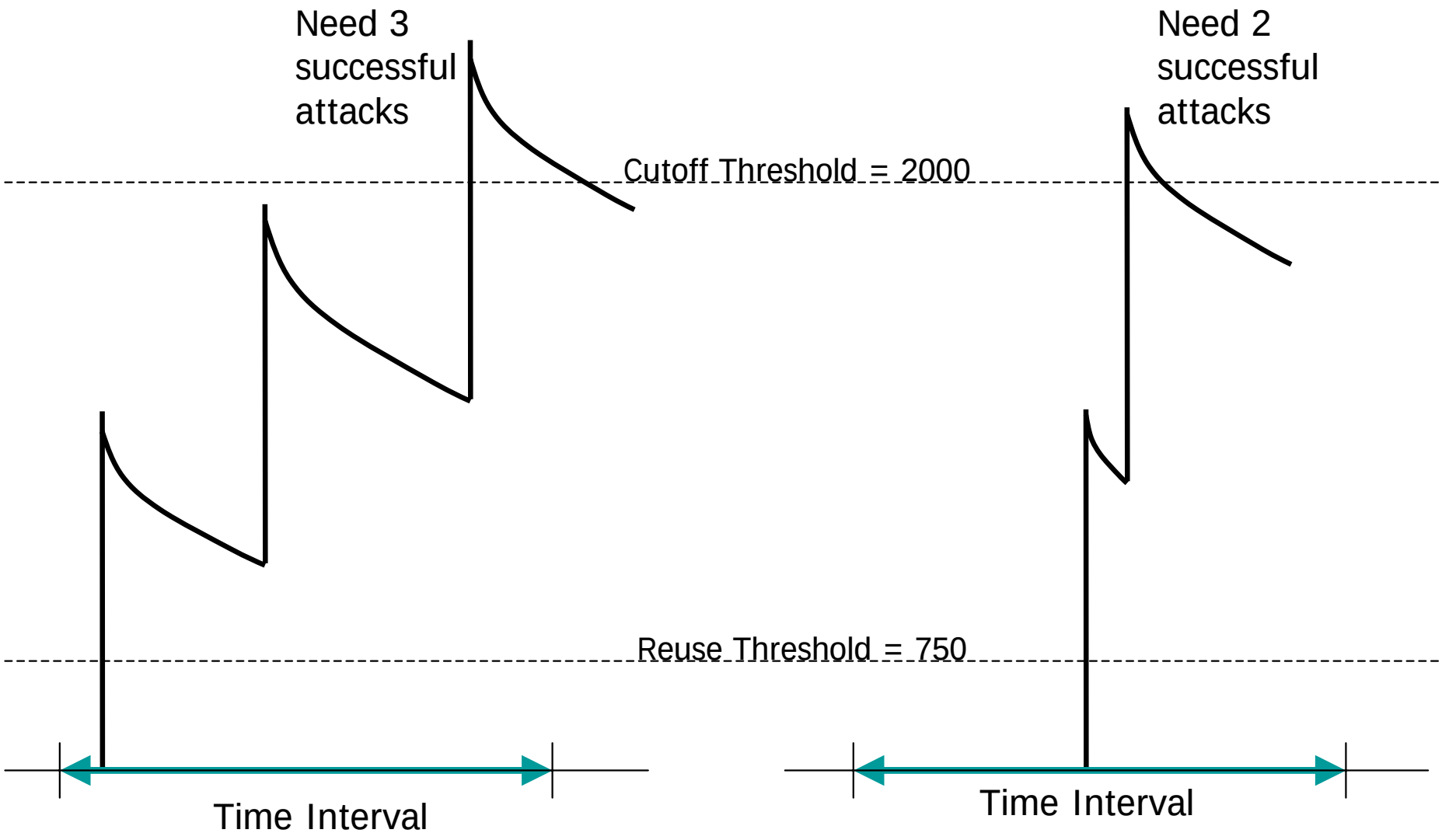
	X		**Withdrawal** **Re-Adv**	**AttrCh** **AttrCh**	**MRAI i**
					MRAI i+1
			X	**AttrCh** **AttrCh**	**MRAI i+2**
X		X	**Withdrawal** **Re-Adv**	**AttrCh** **AttrCh**	**MRAI i+3**
		X		⋮	**MRAI i+4**
				RFD cutoff state	
				▼	

Time (n x MRAI)

X = Successful BGP peering session attack

Note: Router *n* has alternate routes to Router 1

Estimation of Attacks Needed to Push Penalty Above Cutoff



Attacks and RFD Penalty Accumulation Model

$C = cutoff\ threshold$,

$R = reuse\ threshold$,

$H = half\ time$ (decay parameter),

T = MRAI time (≈ 30 sec),

P = incremental penalty incurred per successful attack event,

n = number of BGP nodes in the AS path subject to attacks,

$Q = \Pr\{$a BGP peering session attack is successful$\}$,

$\theta = \Pr\{$AS path of n ASes is successfully attacked at one or more BGP peering sessions$\}$,

E = Elapsed time from the time of beginning of BGP session attacks (in multiples of MRAI)

$R_P(n+1; n, 1; iT)$ = RFD penalty at router $n+1$ for peer n and destination 1 at time iT

$\alpha(n,k) = \Pr\{R_P(n+1; n, 1; iT) > C$ for some $i \in (0,k) \mid E = kT\ )$

Attacks and RFD Penalty Accumulation Model

$$\theta = 1 - (1 - Q)^{n-1}$$

RFD cutoff threshold check (for j attacks in k MRAI intervals):

$$P \sum_{i=0}^{j-1} 2^{\left\{ -\frac{ikT}{(j-1)H} \right\}} > C$$

Let $j_{\min}(k)$ be the smallest j that satisfies the above inequality.

Then,

$$\alpha(n,k) = \sum_{j_{\min}(k)}^{k} \beta_i(n,k)$$

where,

$$\beta_i(n,k) = \frac{k!}{i!(k-i)!} \theta^i (1-\theta)^{k-i}$$

==========

AS/Peer Isolation Sustenance Probability:

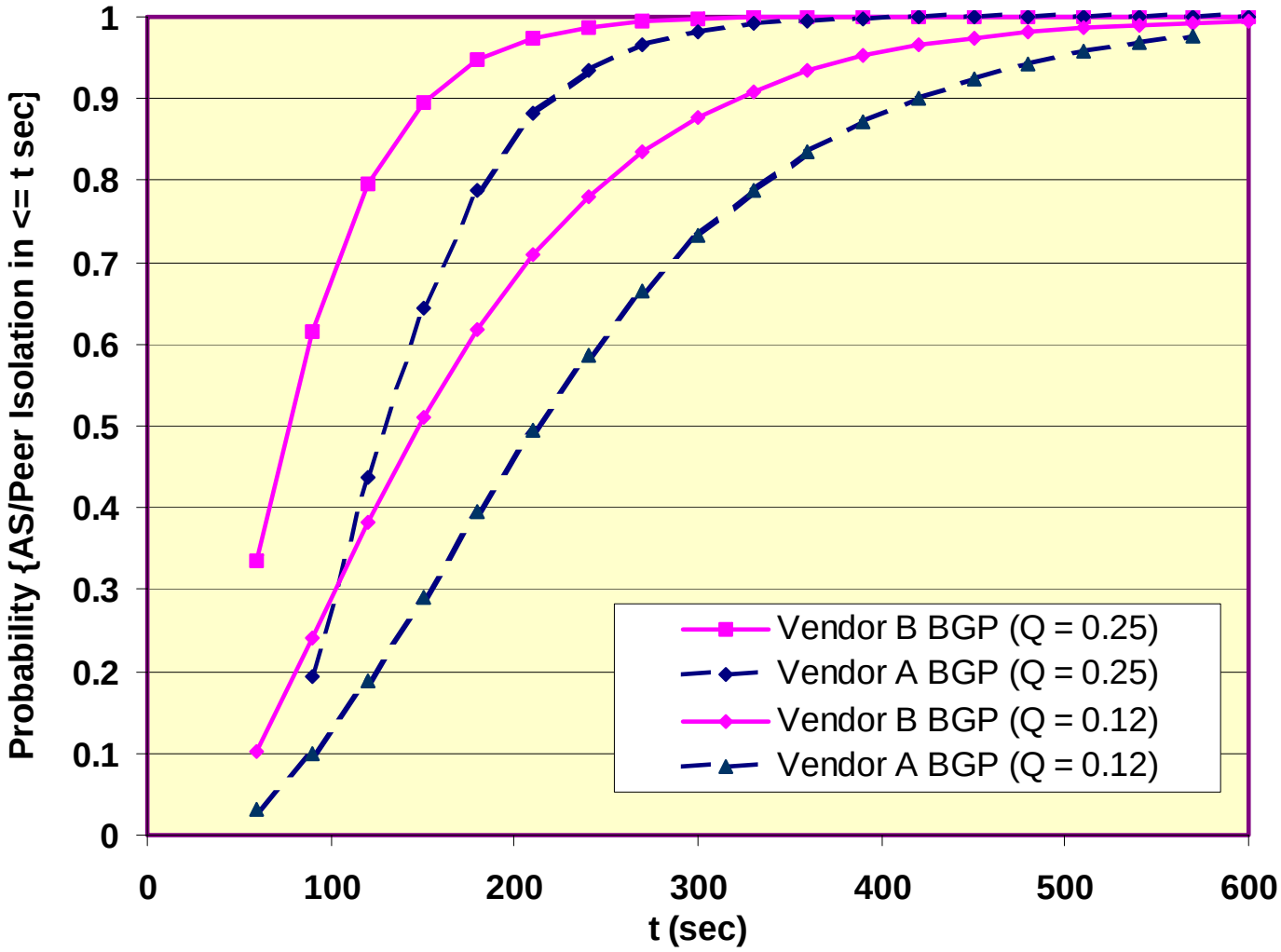
$$P_{sus} = 1 - (1-\theta)^{\left\lceil H\left(\log_2 \frac{C}{R}\right)/t_M \right\rceil}$$

Probability of AS-Prefix Isolation

Probability that AS-Prefix isolation occurs within t sec from start of attacks:

- Sensitivity to vendor settings of RFD parameter values is quite significant
- $n = 4$

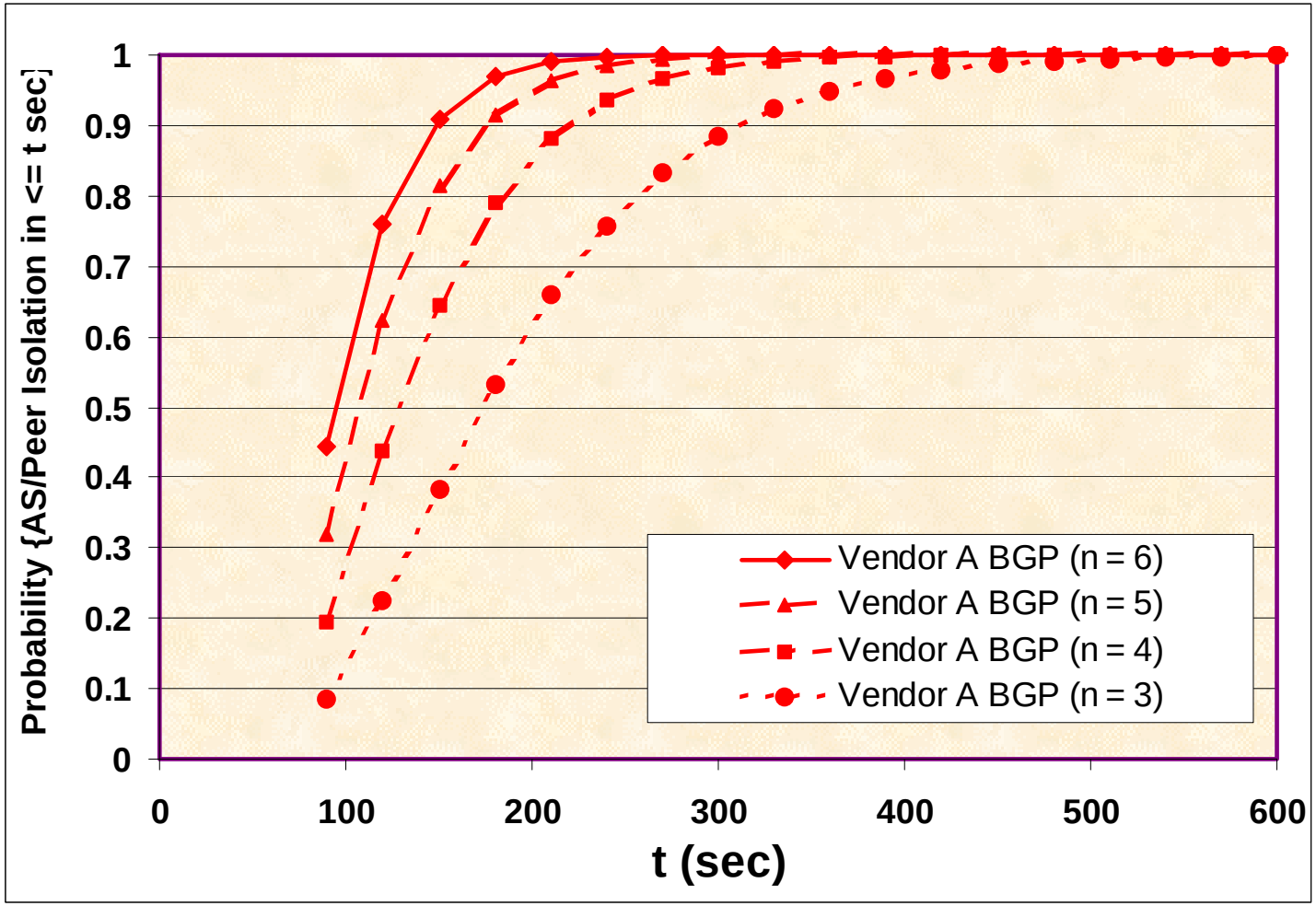
(#ASes in AS path)



Probability of AS-Prefix Isolation

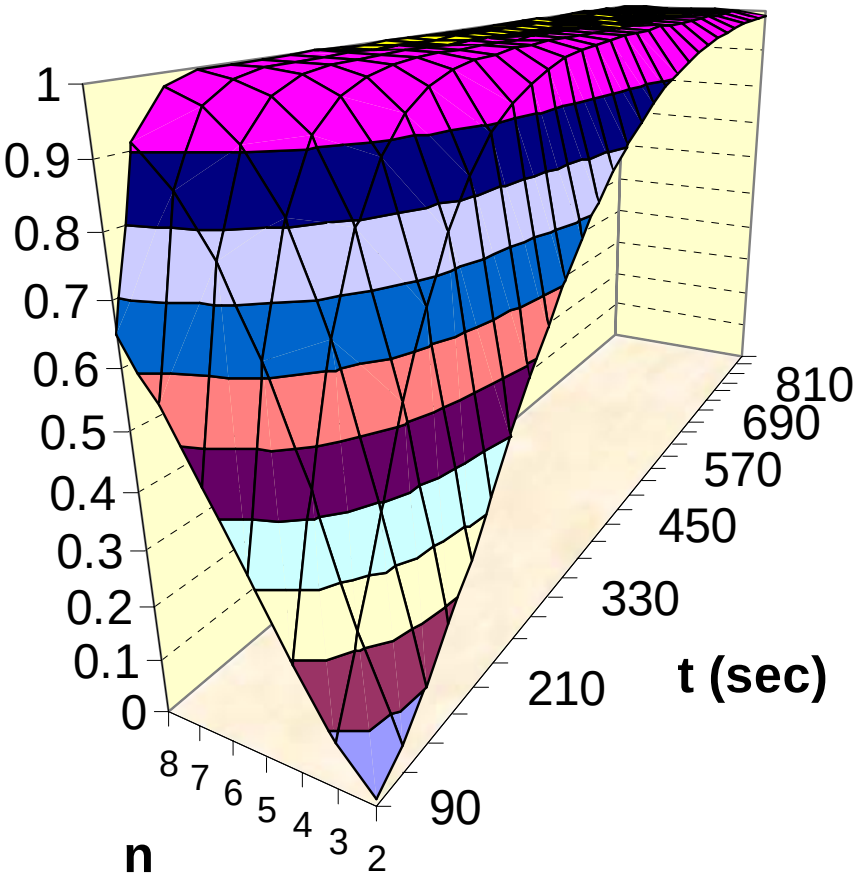
- Vulnerability is higher if AS path-lengths within the attack area are higher
- $Q = 0.25$

Probability that AS-Prefix isolation occurs within t sec from start of attacks:



Probability of AS-Prefix Isolation

Probability that AS-Prefix isolation occurs within t sec from start of attacks:

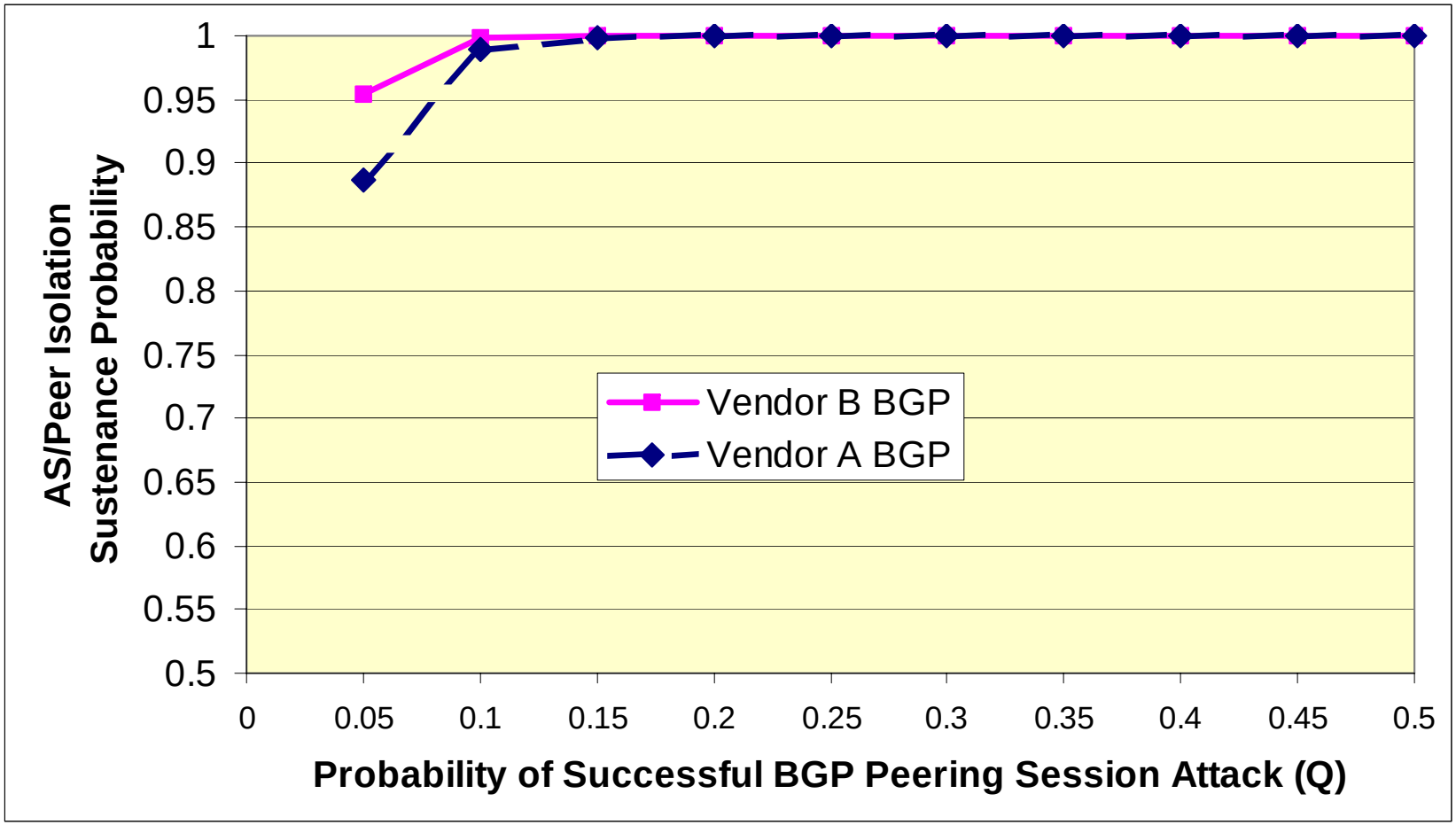


- Attack goal is reached sooner if targeted AS paths have longer lengths

$Q = 0.25$

Probability of Sustenance of AS-Prefix Isolation

Given that an AS-Prefix isolation occurred, what is the probability that it can be sustained for a prolonged period by the attackers:



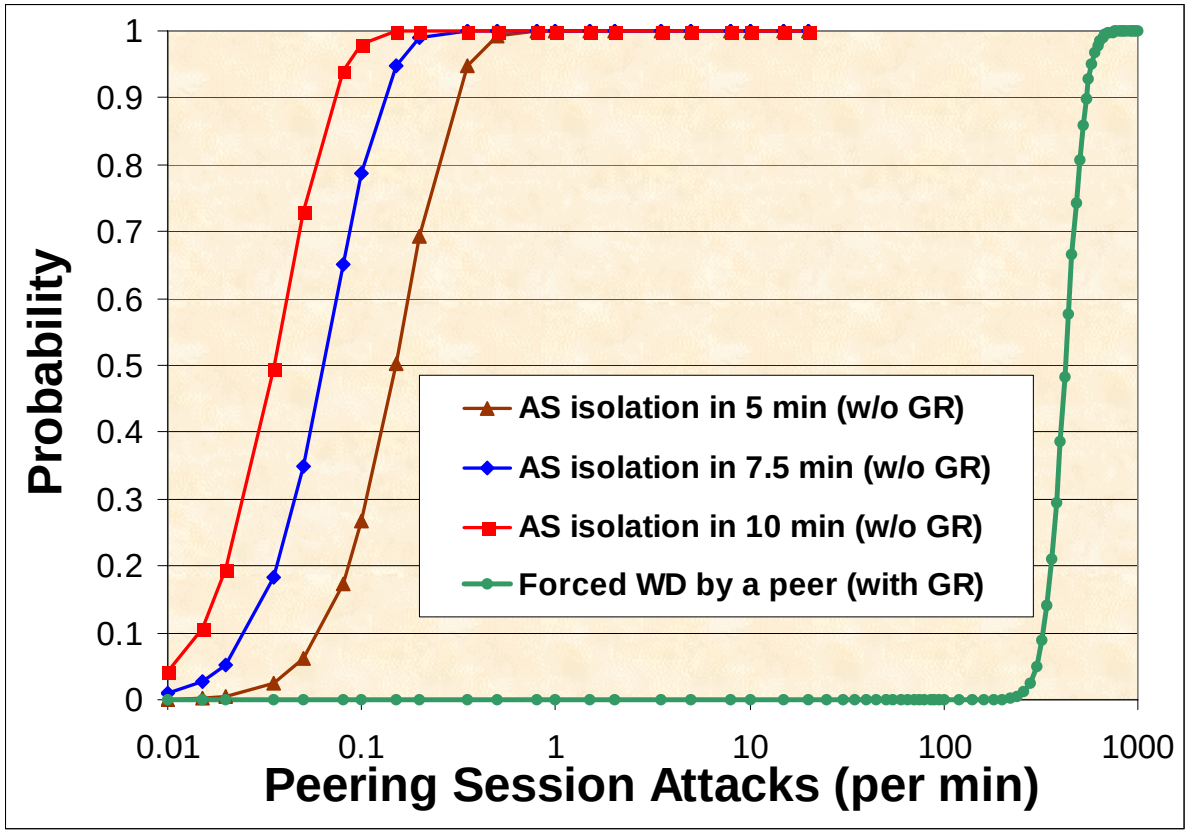
$n = 4$

BGP Graceful Restart: Brief Description

- Gives downed router time to restart without peers withdrawing its routes
- Option negotiated at OPEN
- Two flag bits in capability advertisement
  - Restart bit = router has restarted
  - Forwarding bit = preserved forwarding state
- During restart, peers do not send withdrawals for the restarting router; prevents route flapping
- Restart timer:
  - Restart-time determines how long peer routers will wait to delete stale routes before a BGP open message is received
- If restart-time expired: restart failed, routes deleted, withdrawals sent

BGP Graceful Restart: Mitigation of RFD Exploitation Attacks and Avoidance of AS Isolation

- Without BGP-GR, the RFD exploitation attack resulting in AS isolation is much more feasible
- BGP-GR helps mitigate this type of attack
- With BGP-GR, the attackers need a lot more effort (100 times or more) to even induce route withdrawals at a peer
- BGP-GR restart time = 120 s
- BGP session recovery time = 4 s



$n = 4$

$Q = 0.1$

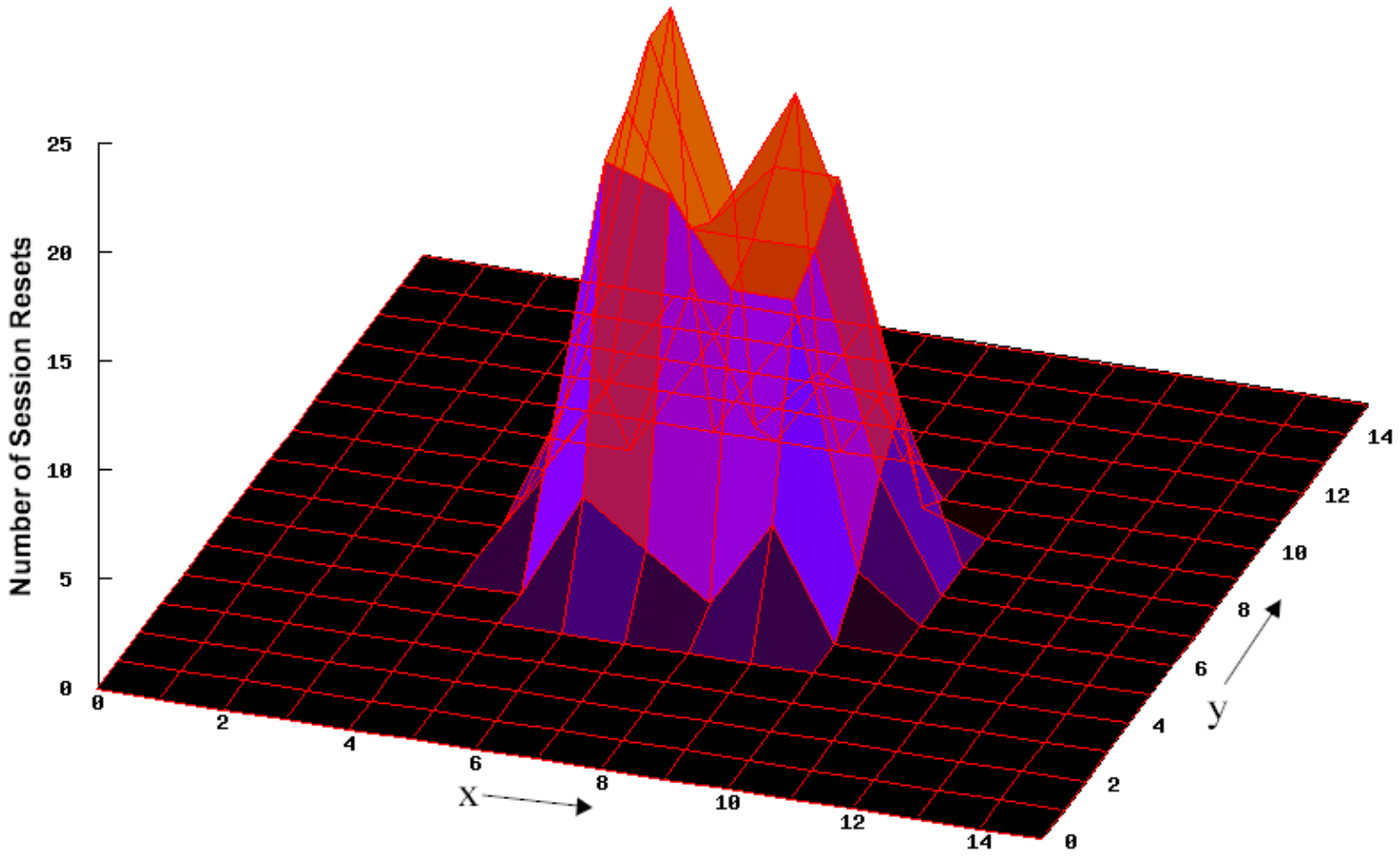
- "Several providers (US) suggest that the cost of implementing this feature outweighs the benefit." – NISCC (UK govt) BGP Best Practices

Simulation Results with Grid Topology

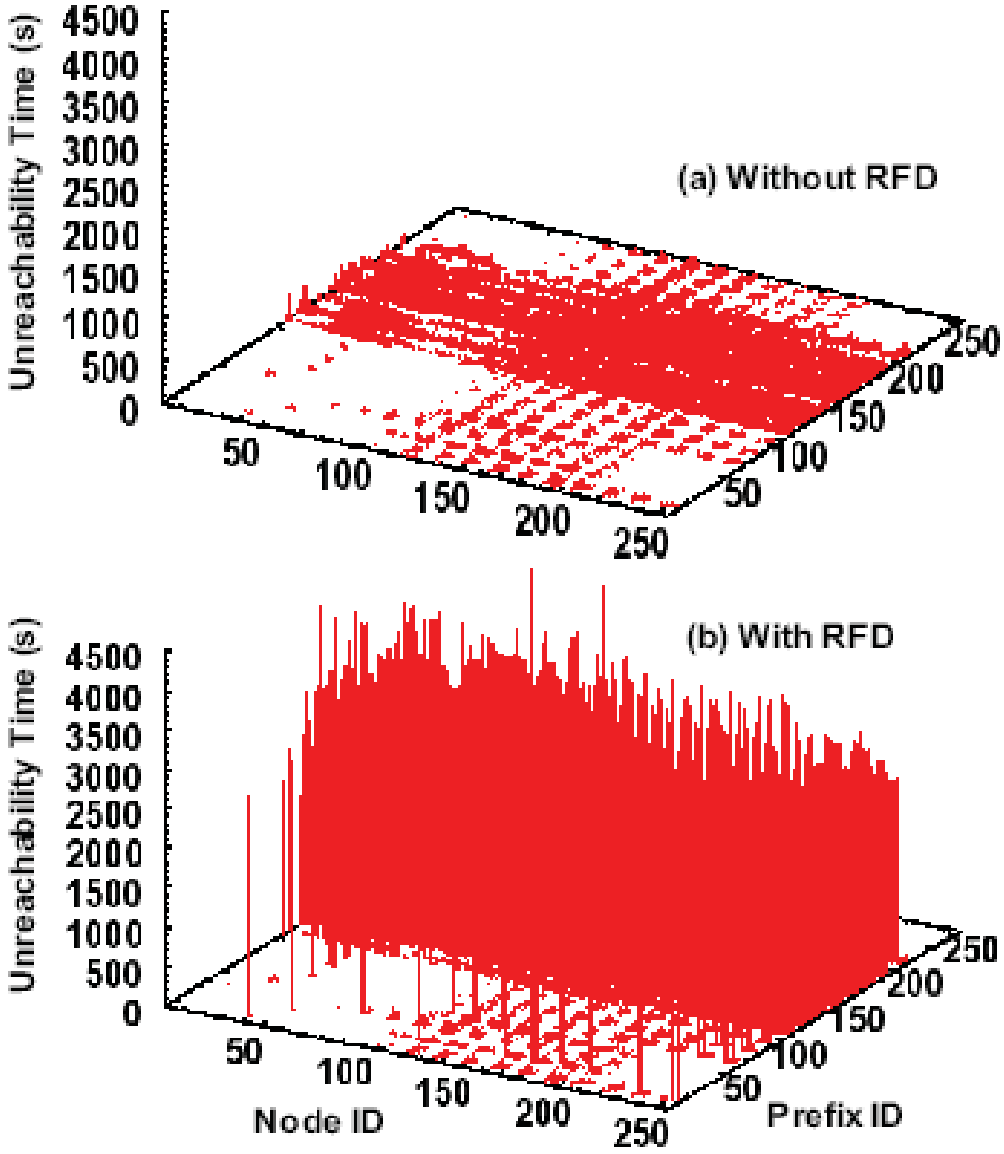
Experiment with Grid Topology

- 256 node grid (16x16)
- Center 8x8 grid attacked
- Total attack duration = 500 sec
- # Attack intervals = 50 (each is 10 sec)
- Prob. of success for each attack = 25%

Measured # BGP Session Resets Plotted over Topology



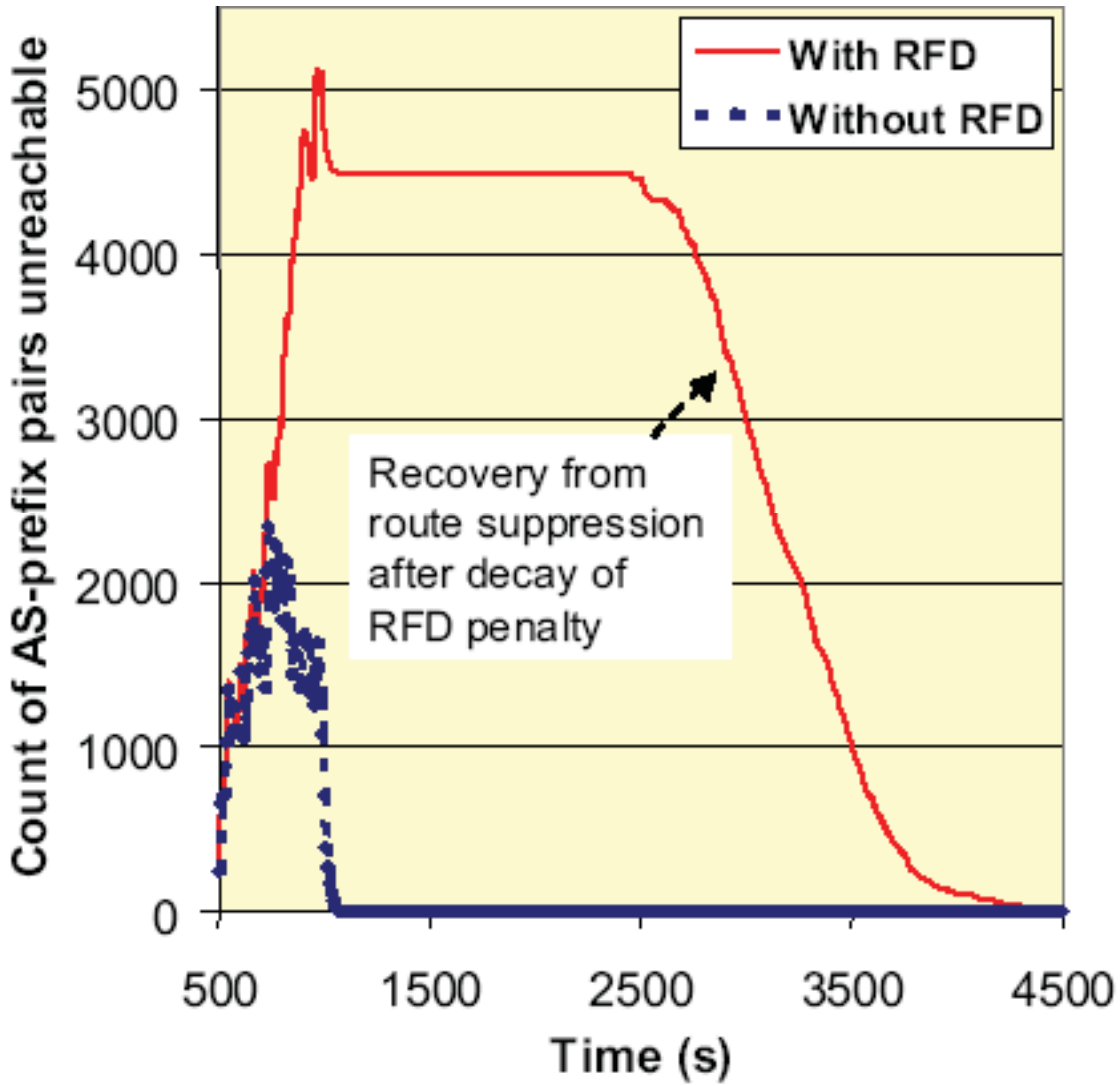
Comparison of Unreachability Time



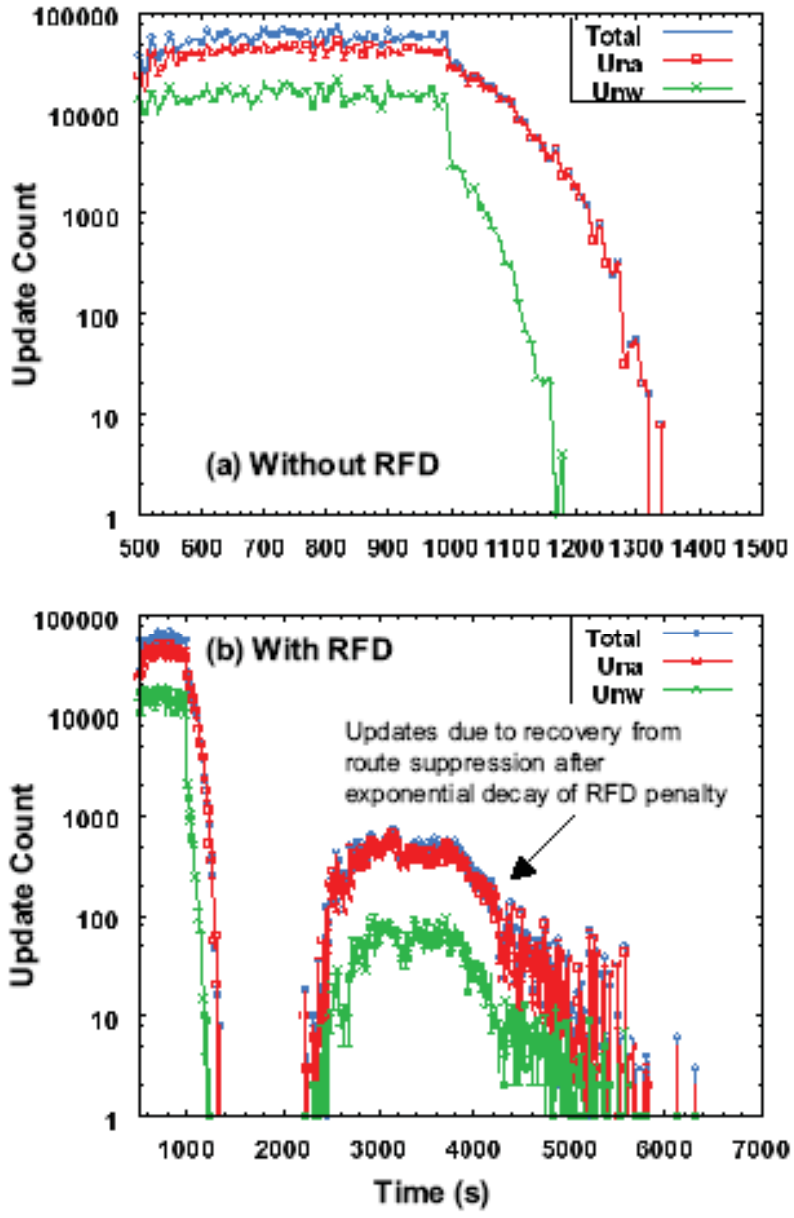
(a) Without RFD

(b) With RFD

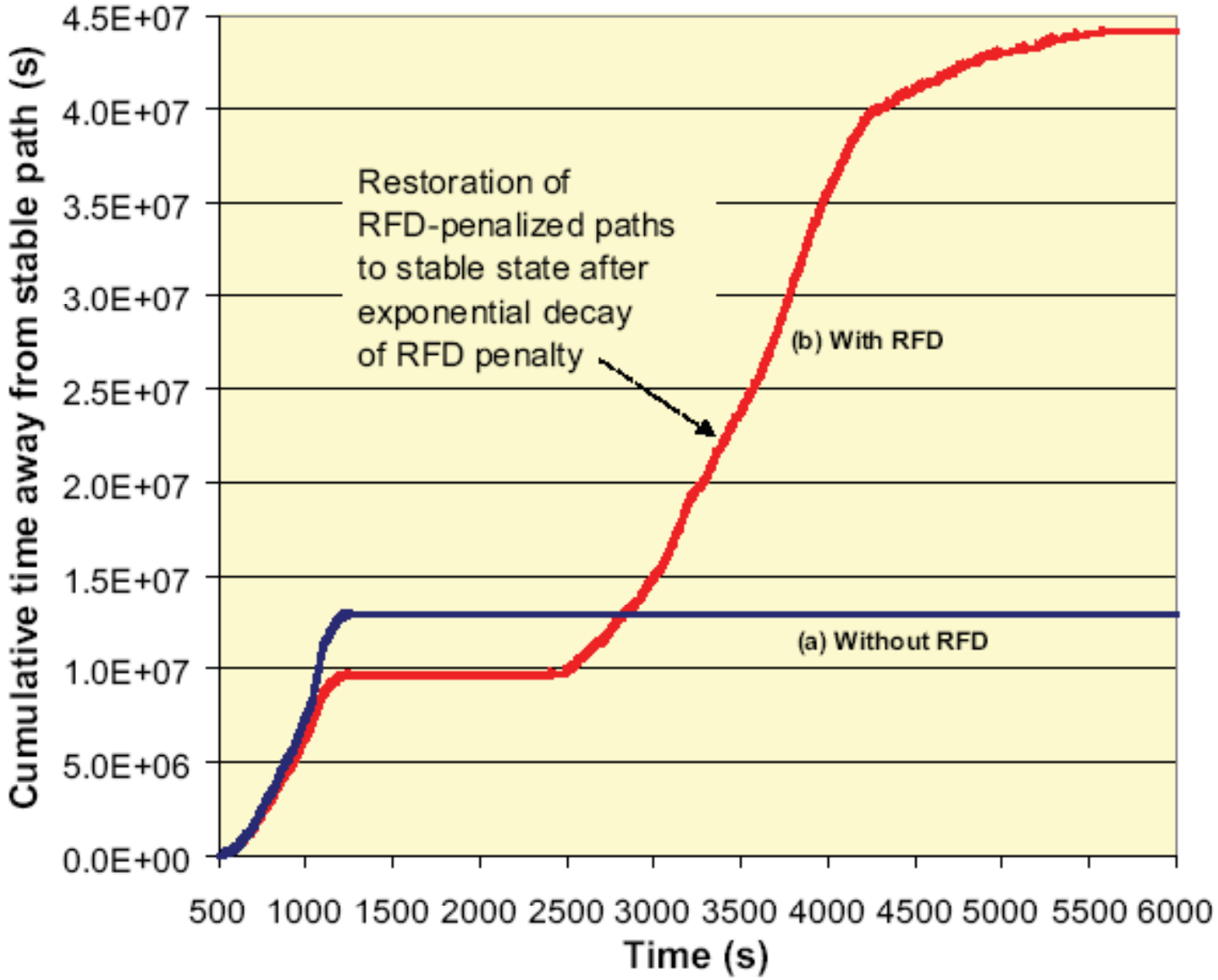
Count of AS-Prefix Pairs Unreachable



Comparison of Update Count



Route Quality: Time Away From Stable Path



Generation of Down-Sampled Realistic AS Topology

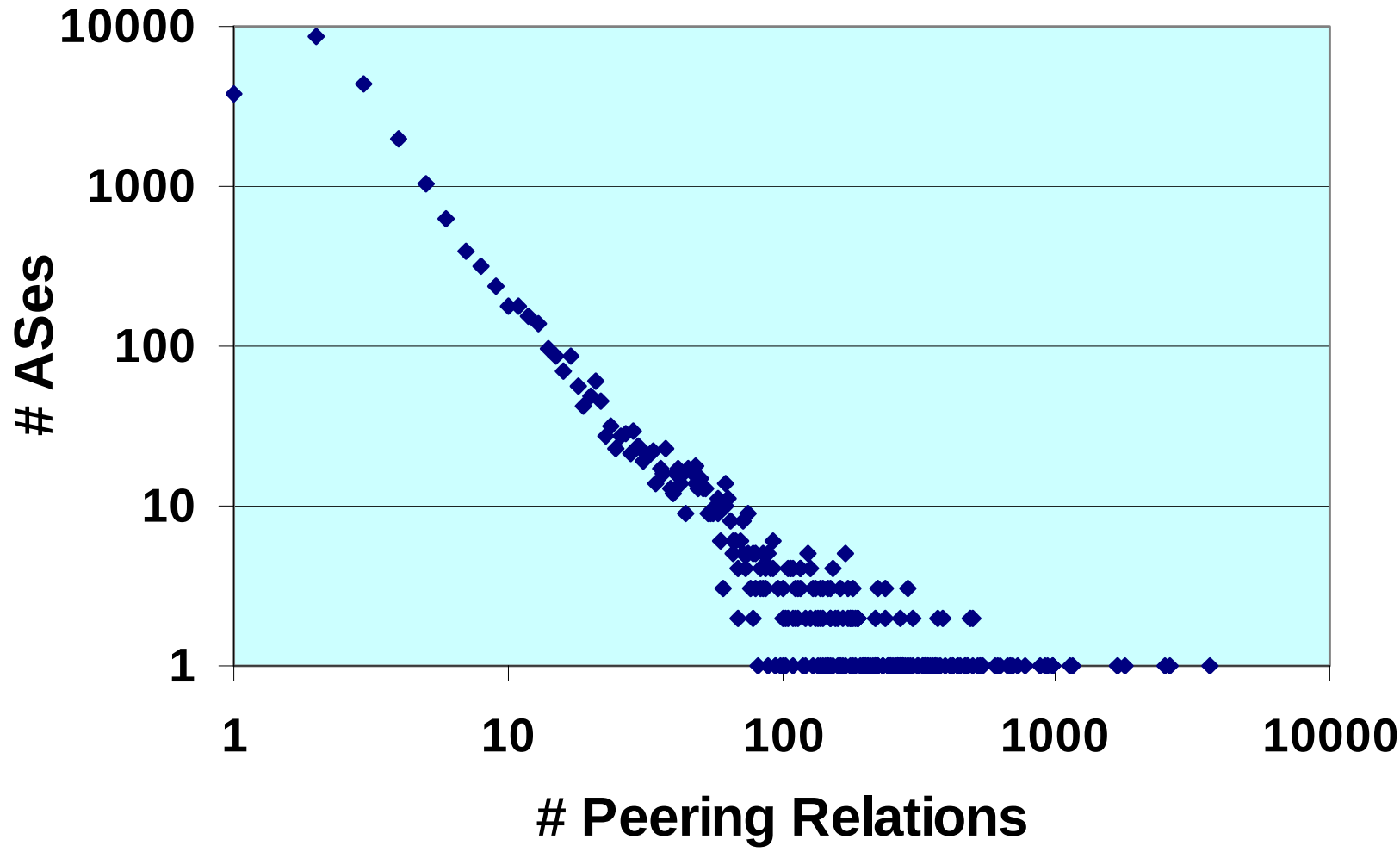
Example AS Topology Data From UCLA

# neighbors	# ASes observed
# neighbors ≤ 100	23380
100 < # neighbors ≤ 1000	266
# neighbors ≥ 1000	8

Degree of Connectivity	# ASes	Percentage
10 or more	2330	10%
9 or more	2570	11%
8 or more	2890	12%
7 or more	3285	14%
6 or more	3917	17%
5 or more	4953	21%
4 or more	6957	29%
3 or more	11278	48%
2 or more	19934	84%
1 or more	23654	100%

- Total # ASes = 23654
- Total number of links = 96445
- Average number of neighbors per AS = 8.15
- The ASes with large numbers of neighbors are large ISPs
- AT&T (AS# 7018) has 2602 neighbors
- UUNET (AS# 701) has 3622 neighbors
- Date data downloaded: December 2005

Peering Statistics (Internet)



Peering Statistics

Ordered (Descending) List of # Neighbors Vs. # ASes

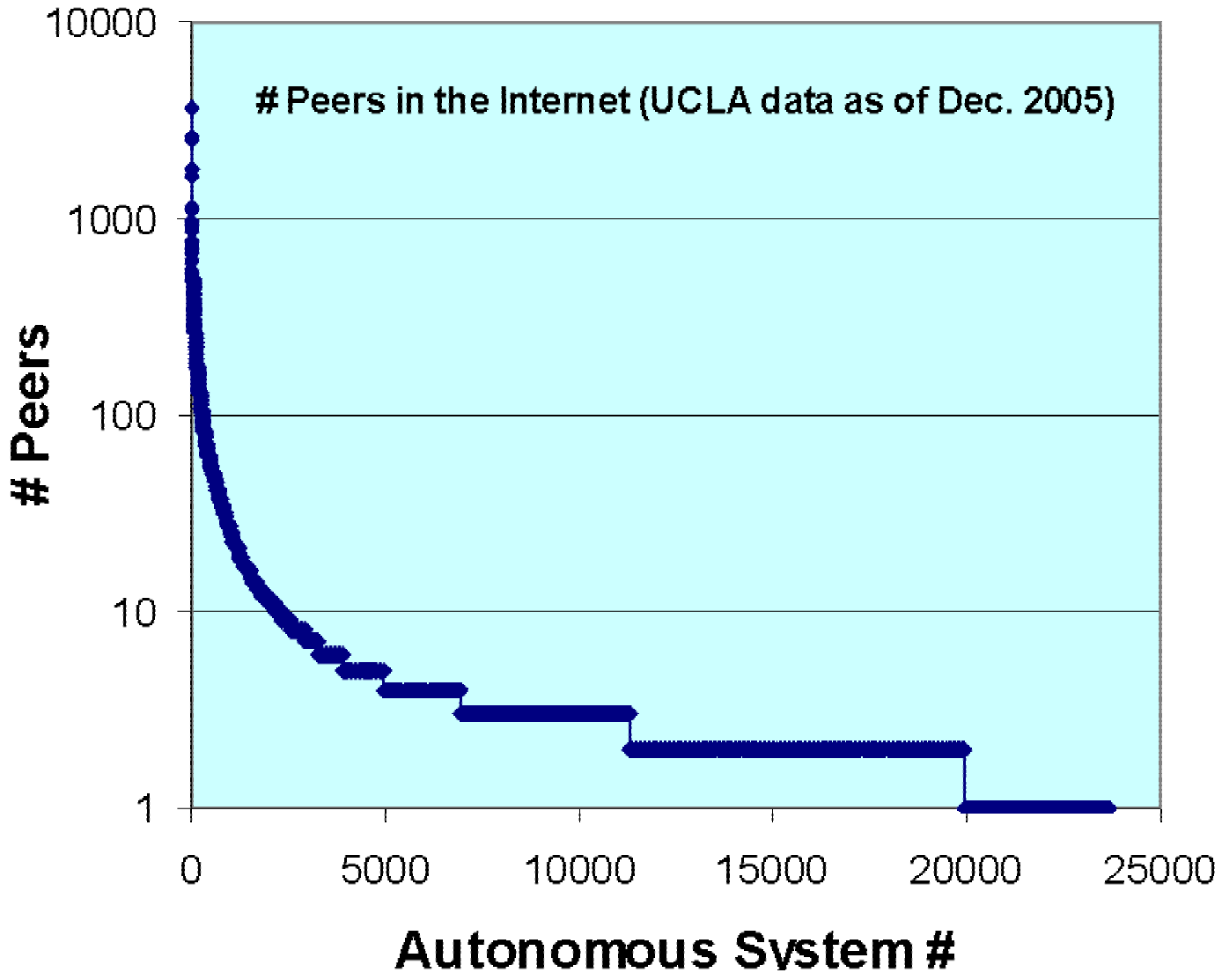
most connected ASes

# neighbors	# ASes
3622	1
2602	1
2510	1
1808	1
1670	1
1664	1
1145	1
1120	1
976	1
973	1
938	1
910	1
870	1
772	1
733	1
691	1
689	1
676	1
667	1
628	1

# neighbors	# ASes
20	49
19	42
18	57
17	87
16	70
15	87
14	98
13	138
12	152
11	177
10	177
9	240
8	320
7	395
6	632
5	1036
4	2004
3	4321
2	8656
1	3720

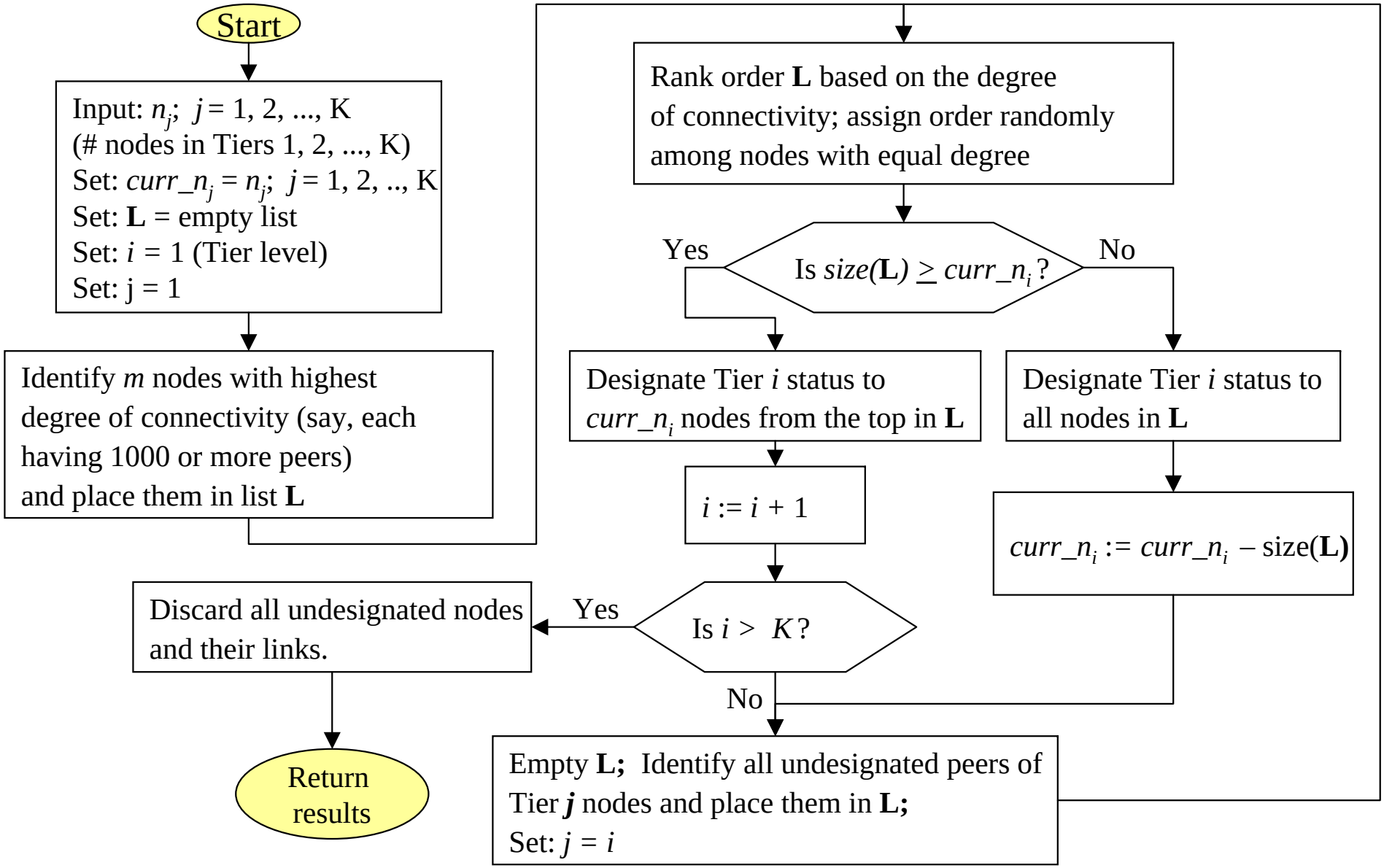
least connected ASes

Peering Statistics (Internet)



(Renumbered to generate the descending plot)

Algorithm for Down-Sampling AS-Level Topology and Tier Assignment to ASes



Some Modifications to the Algorithm

- There are variations of the algorithm that we have considered such as forcing Tier 3 to be mostly stub nodes.
- We proceed to consider an algorithm to further prune the number of peering links. We still see 1300 to 1600 peering links after applying the down sampling algorithms (256 ASes). The next two slides describe an algorithm for further pruning just the peering links (intra-Tier and inter-Tier).

Algorithm for Further Pruning of Links

1. Leave intra Tier 1 peering links as they are (nearly full mesh).
2. For Tier *i* (*i* ≥2) link pruning (Intra-Tier):
   a. Rank order nodes in Tier *i* in accordance with descending degree of connectivity considering peering links only within Tier *i* (ignore peering links to higher or lower Tiers in this step);
   b. Mark those links (intra Tier *i*) as non-removable which have connectivity of only 1 for the node at the other end;
   c. Randomly pick and remove one unmarked link at a time in a round-robin fashion for the rank-ordered nodes in step a;
   d. Stop the round-robin link removal process when the target number of links in Tier *i* has been achieved (or when the remaining intra Tier *i* links are all marked).
3. For Tier *i*-*j* (*j* > *i*) link pruning (Inter-Tier):
   a. Rank order nodes in Tier *j* in accordance with descending degree of connectivity considering peering links only with Tier *i*;
   b. Modify the rank-ordered list of nodes by removing from the list any nodes which have a degree of connectivity equal to one to Tier *j*;
   c. Remove one Tier *j* to Tier *i* peering link at a time in a round-robin fashion for the rank ordered nodes in step b;
   d. Stop the round-robin removal process when the target number of peering links for Tier *j* to Tier *i* has been achieved.
4. Stop and return results when Step 2 and Step 3 have been repeated adequately to consider all intra- and inter-Tier peering link targets.

Comments on Algorithm for Further Pruning of Links

- The pruning algorithms is designed such that the AS-level topology would not get partitioned into unconnected networks. All nodes in the pruned topology remain connected and reachable (under normal operation).
- This property holds because:
  - The down-sampling algorithm guarantees that each AS in Tier *j* is connected to at least one node in Tier *i* (*j* > *i*)
  - The pruning algorithm further guarantees that if a node in Tier *j* has only one link to reach nodes in Tier *i* (*j* > *i*), then that link will not be removed in the pruning process.

Example Inputs to Algorithm for Further Pruning of Links

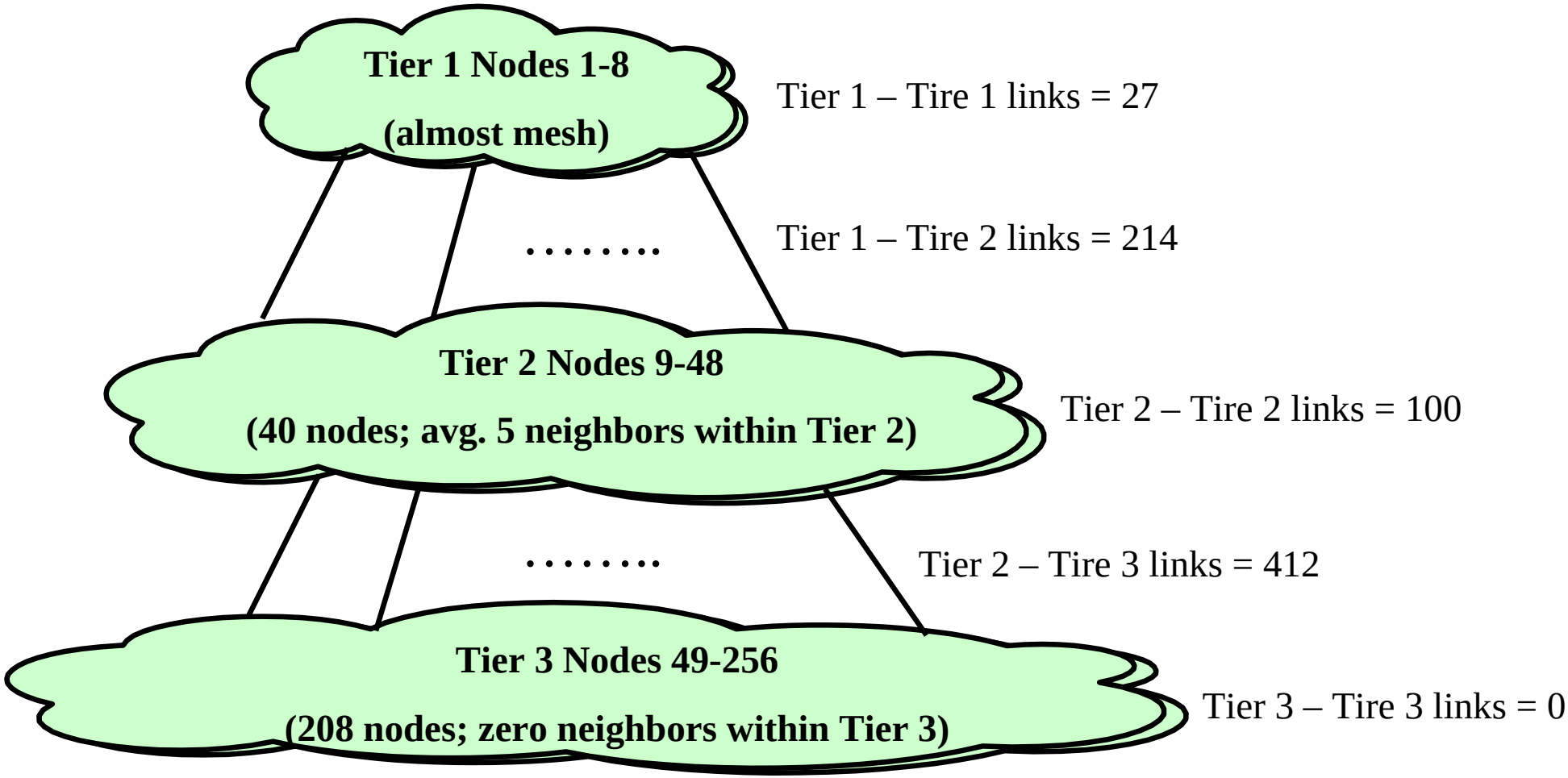
	# Nodes	# Links (after down sampling)	Target peer connectivity (average)	Multiplier to determine # links (average)	# Links (after pruning)
Tier 1	8				
Tier 2	40				
Tier 3	208				
Tier 1-1		27	Almost full mesh		27
Tier 1-2		214	Each Tier 2 to four Tear 1	4	160
Tier 2-2		603	Each Tier 2 to six other Tear 2	3	120
Tier 2-3		469	Each Tier 3 to 1.5 Tear 2	1.5	312
Tier 3-3		4	Almost all stub nodes	0	0
Total	256	1317			619

Parameterized choices we make for the degree of pruning

Down-Sampled/Pruned Topology Data

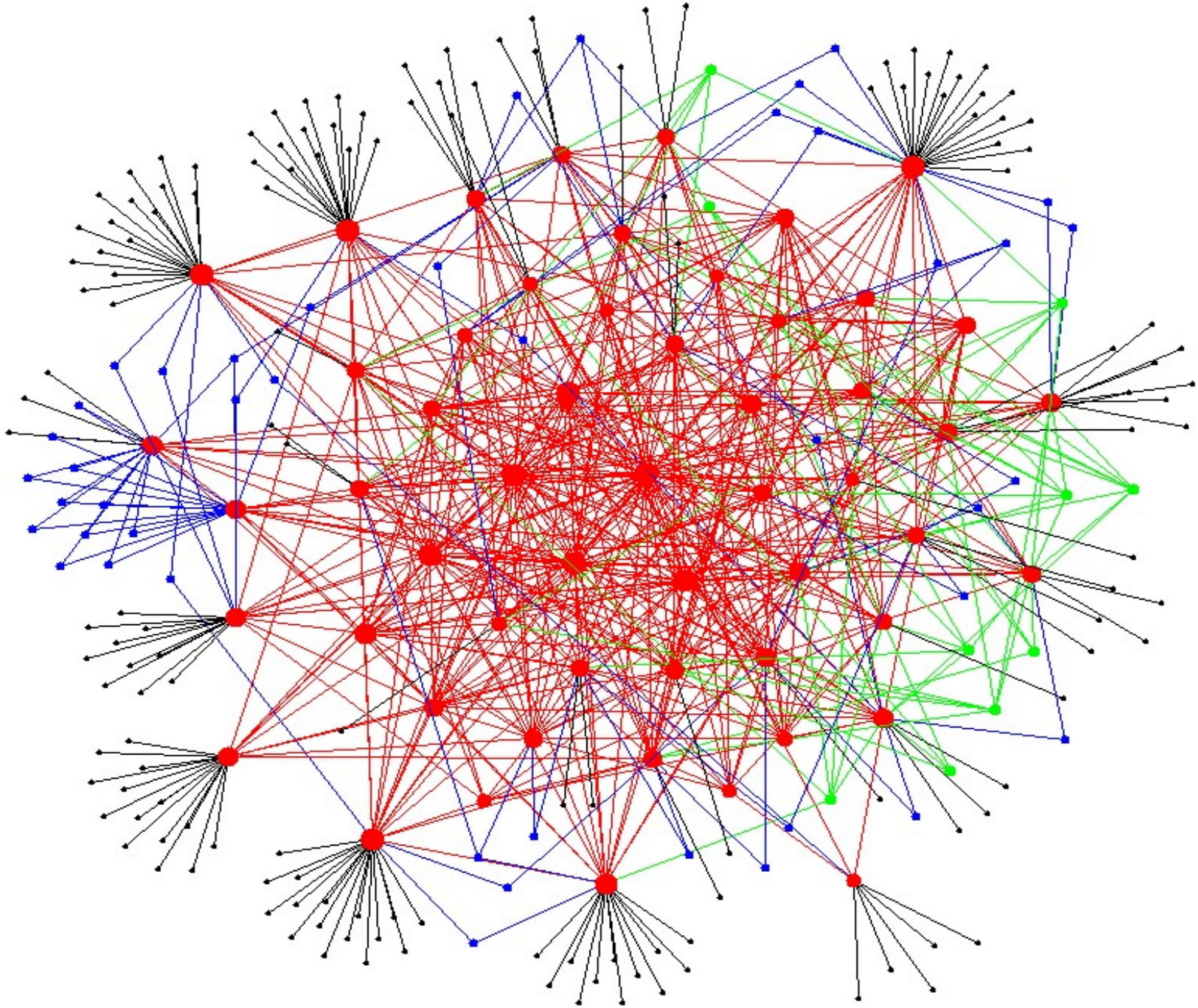
	# Nodes	# Links (after down sampling)	# Links (after pruning)
Tier 1	8		
Tier 2	40		
Tier 3	208		
Tier 1-1		27	27
Tier 1-2		214	214
Tier 2-2		603	100
Tier 2-3		412	412
Tier 3-3		4	0
Total	256	1260	753

Down-Sampled/Pruned Topology



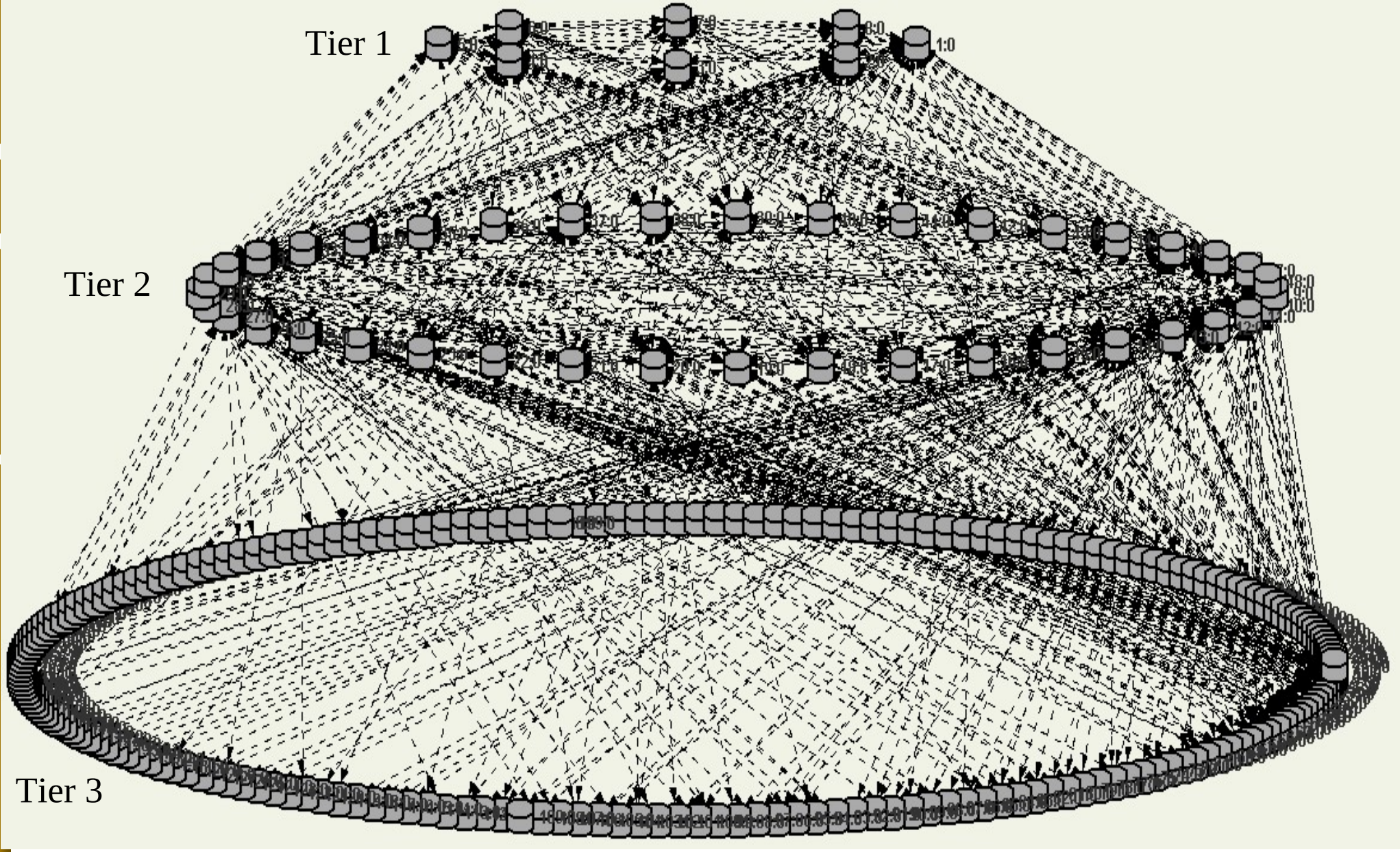
Many Tier 3 nodes are stub nodes

Down-Sampled/Pruned Topology Graph

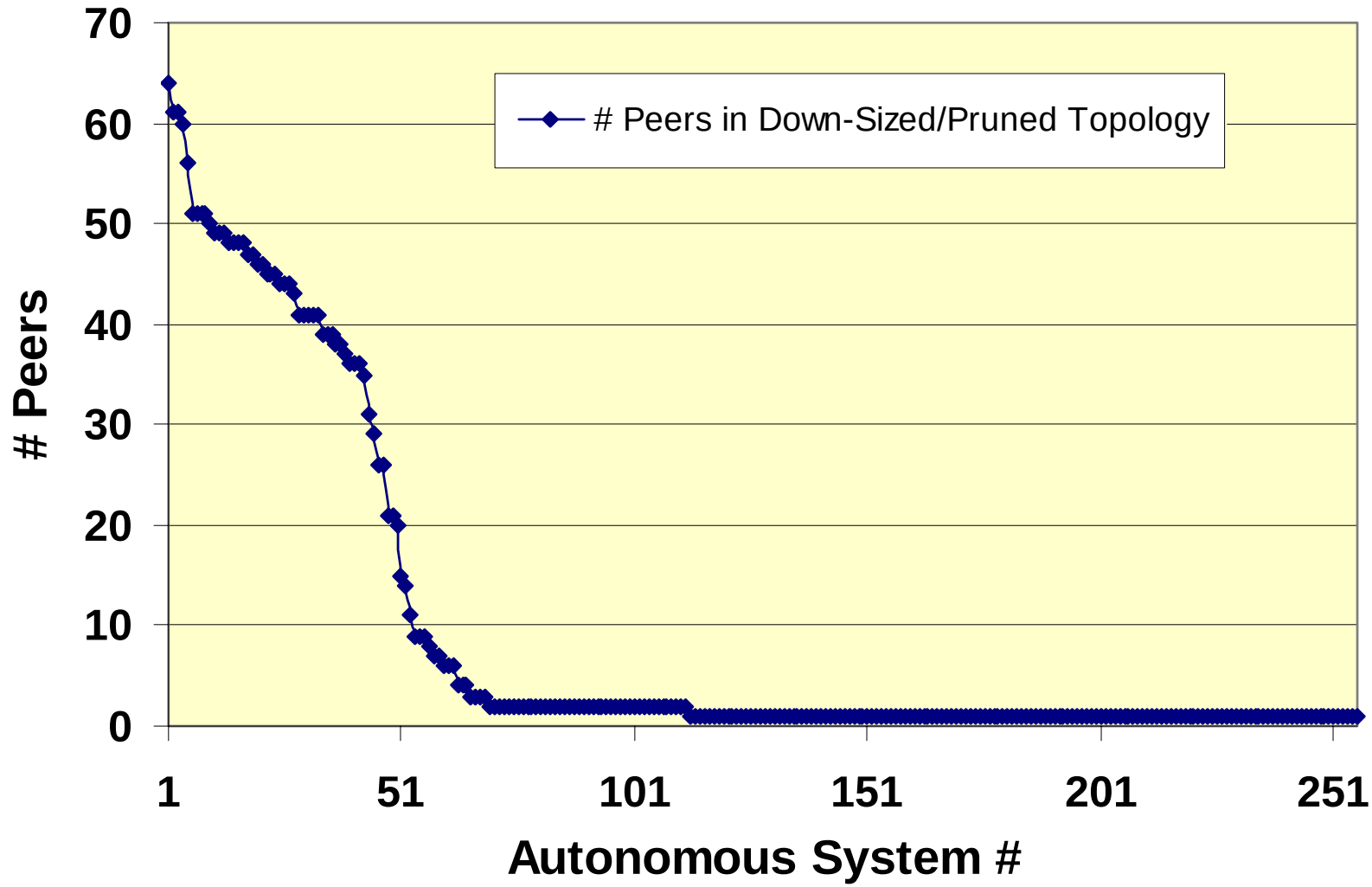


Down-Sampled/Pruned Topology Graph

(Nodes in Each Tier Arranged in Oval Shape)

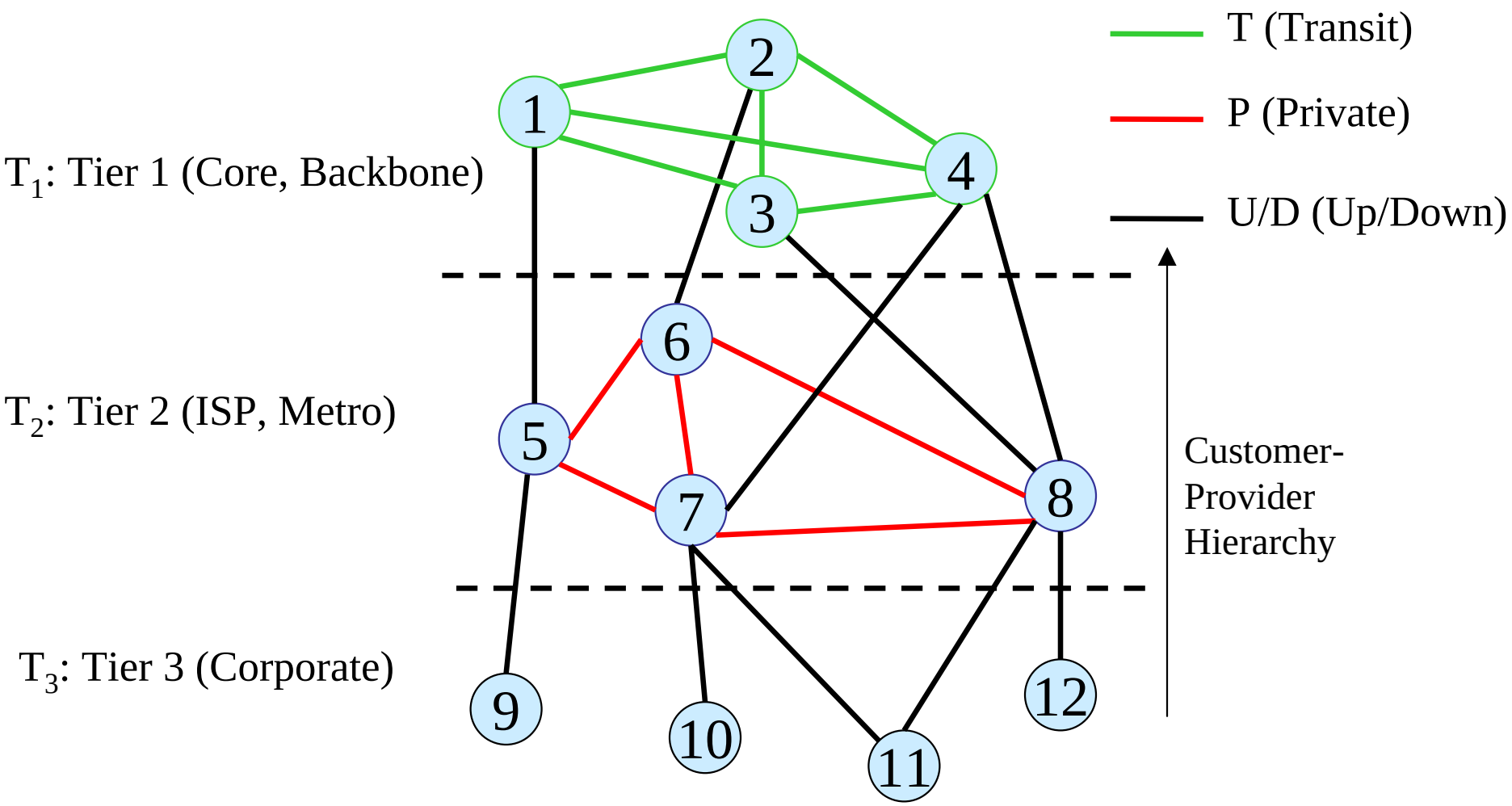


Peering Statistics (Down-Sized & Pruned Topology)



Path Selection Policies

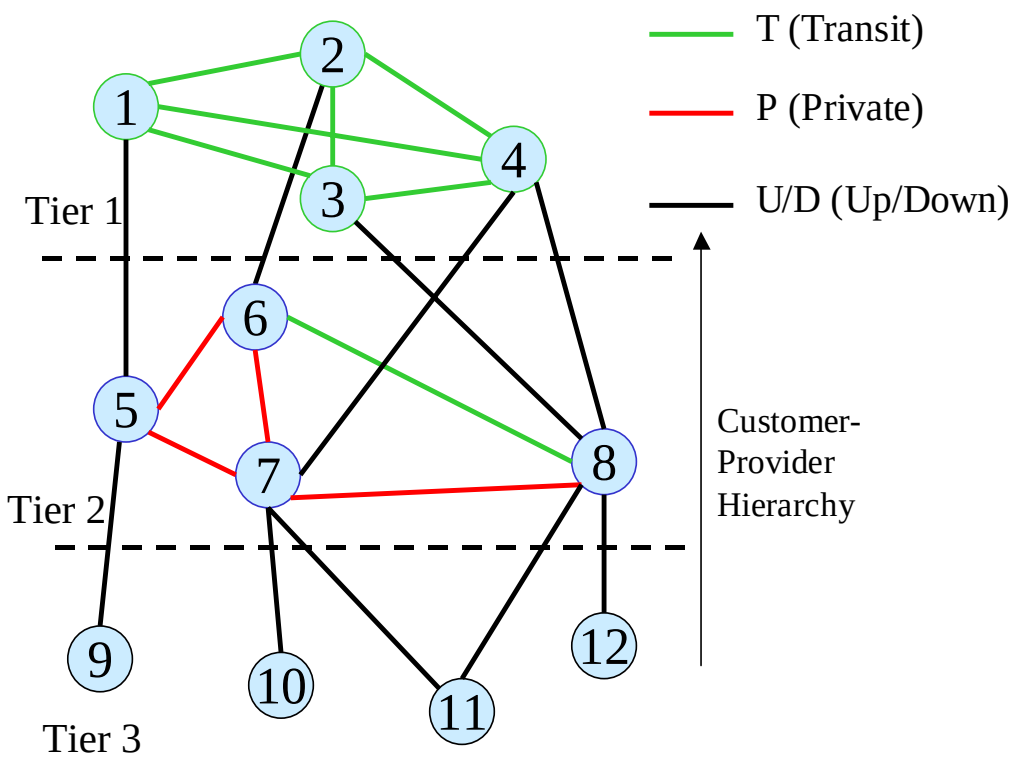
Network Hierarchy and Peering Relations



- T (Transit) link permits any source-destination traffic
- U/D (Up/Down) link also permits any source-destination traffic
- P (Private) link permit only traffic between customers of the two ASes which it connects

Routing Policies

	Rule	Comments
Policy 1	[U|T]*[D|T]*	All links within a Tier are T (none are P)
Policy 2	{[U|T]*[D|T]*} OR {[U]*[P]?[D]*}	All links within Tier 1 are T, but all most all links in Tier 2 are P's (may be with a few exceptions)

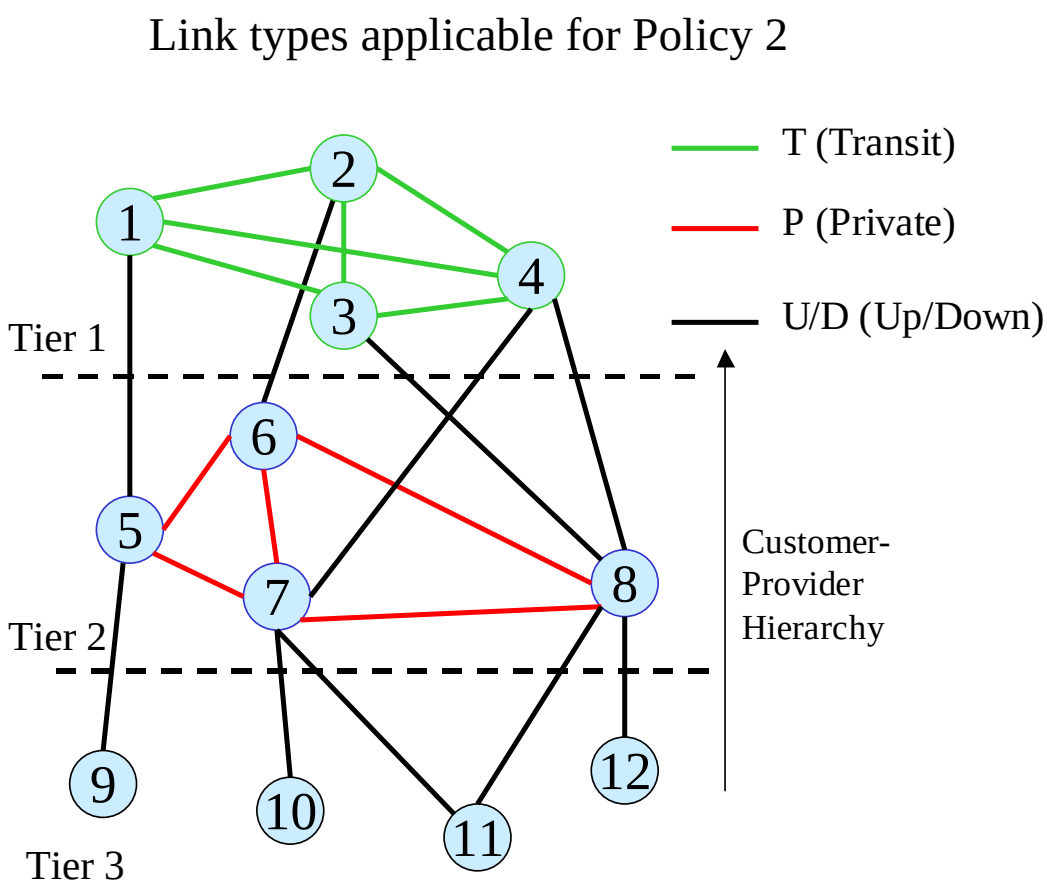


- Policy 1: Once you have gone on a D-link, you can not go on a U-link anymore (currently implemented in prst19 and prst20)
- Policy 2: (1) Rule of Policy 1 still applies, additionally (2) A P-link can be preceded by only U-links, and (3) A P-link can be followed by only D-links.

Assumption: Every lower Tier node is connected to at least one node in the Tier immediately above.

Routing Policies

Example Route	Policy 1 (all internal Tier 2 links are T)	Policy 2 (all internal Tier 2 links are P)
9-5-6-8-12	Allowed	Not Allowed
9-5-1-3-8-12	Allowed	Allowed
9-5-7-10	Allowed	Allowed
9-5-6-7-10	Allowed	Not Allowed
9-5-1-4-7-10	Allowed	Allowed



- Policy 1: Once you have gone on a D-link, you can not go on a U-link anymore (currently implemented in prst19 and prst20)
- Policy 2: (1) Rule of Policy 1 still applies, additionally (2) A P-link can be preceded by only U-links, and (3) A P-link can be followed by only D-links.

Experiment Design With Routing Policies

Experiment	Attack region	Policy
E1P0	All links subjected to attacks	no Policy
E1P1	- do -	Policy 1
E1P2	- do -	Policy 2
E2P0	T1-T1 and T1-T2 links subjected to attacks	no Policy
E2P1	- do -	Policy 1
E2P2	- do -	Policy 2
E3P0	Only T2-T3 links subjected to attacks	no Policy
E3P1	- do -	Policy 1
E3P2	- do -	Policy 2

- Policy 2 is realistic considering common ISP business practices
- Attack duration is 500s (50 intervals of 10s each); Attack success probability is 25%

Routing Policy Comparisons: What to Expect?

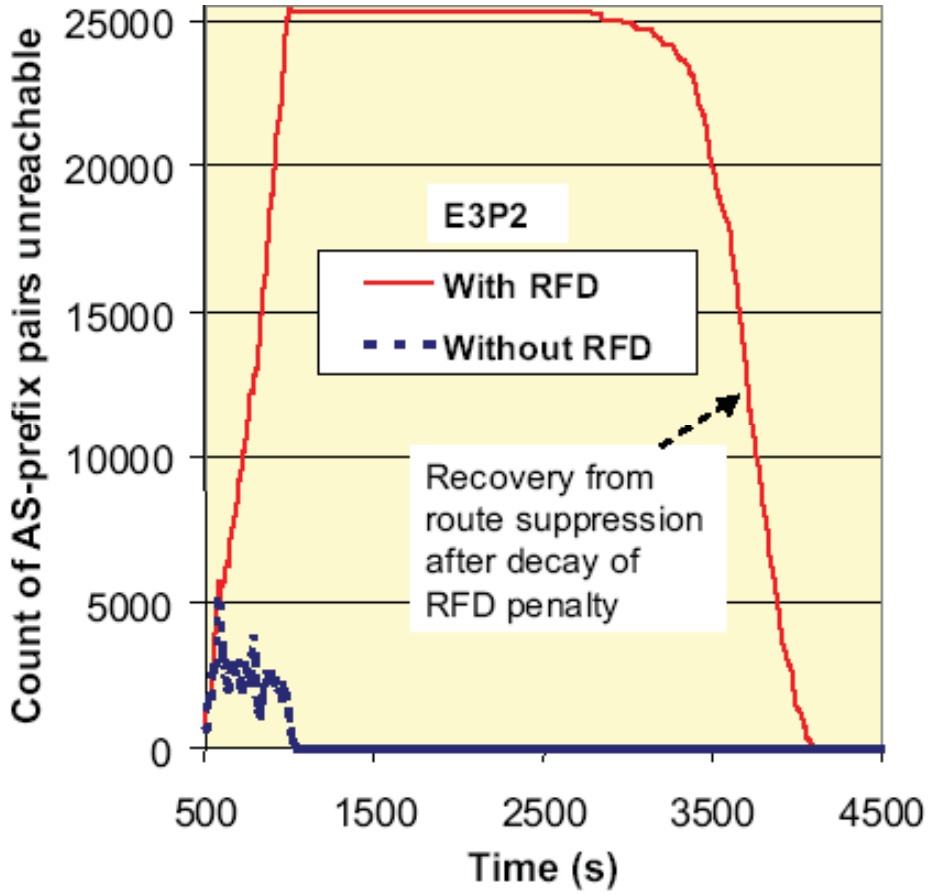
- Policy 2 is more restrictive than Policy 1; allows fewer alternate paths as compared to Policy 1
- Expect unreachability to get worse in this order:

  w/o Policy ⟶ Policy 1 ⟶ Policy 2

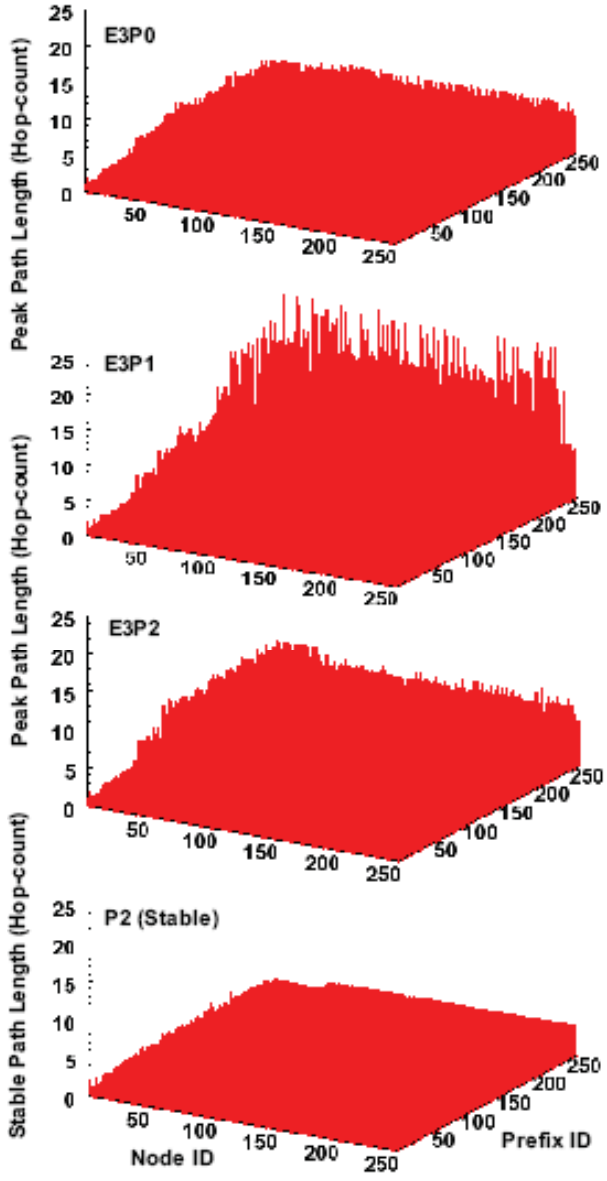
  - The case w/o policy allows the best use of alternate paths followed by Policy 1. Policy 2 allows the least use of alternate paths.
- Also, the differences will be more pronounced when attack region is **at the edges of the network** as compared to the other two attack-topology cases
  - In the former case, RFD suppression vulnerability will be higher.
- Results highlight increased BGP vulnerability to attacks when policy is in effect

Simulation Results with Down-Sampled Realistic AS Topology and Path Selection Policies

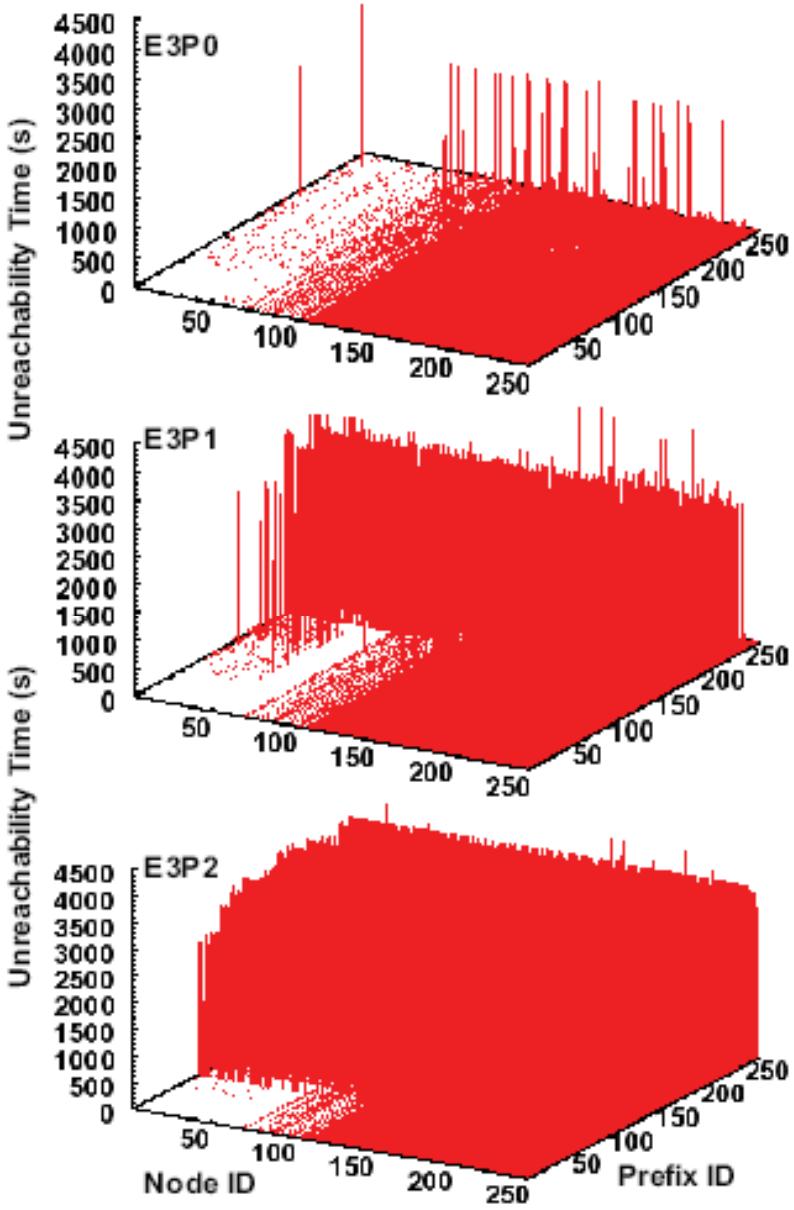
Count of AS-Prefix Pairs Unreachable



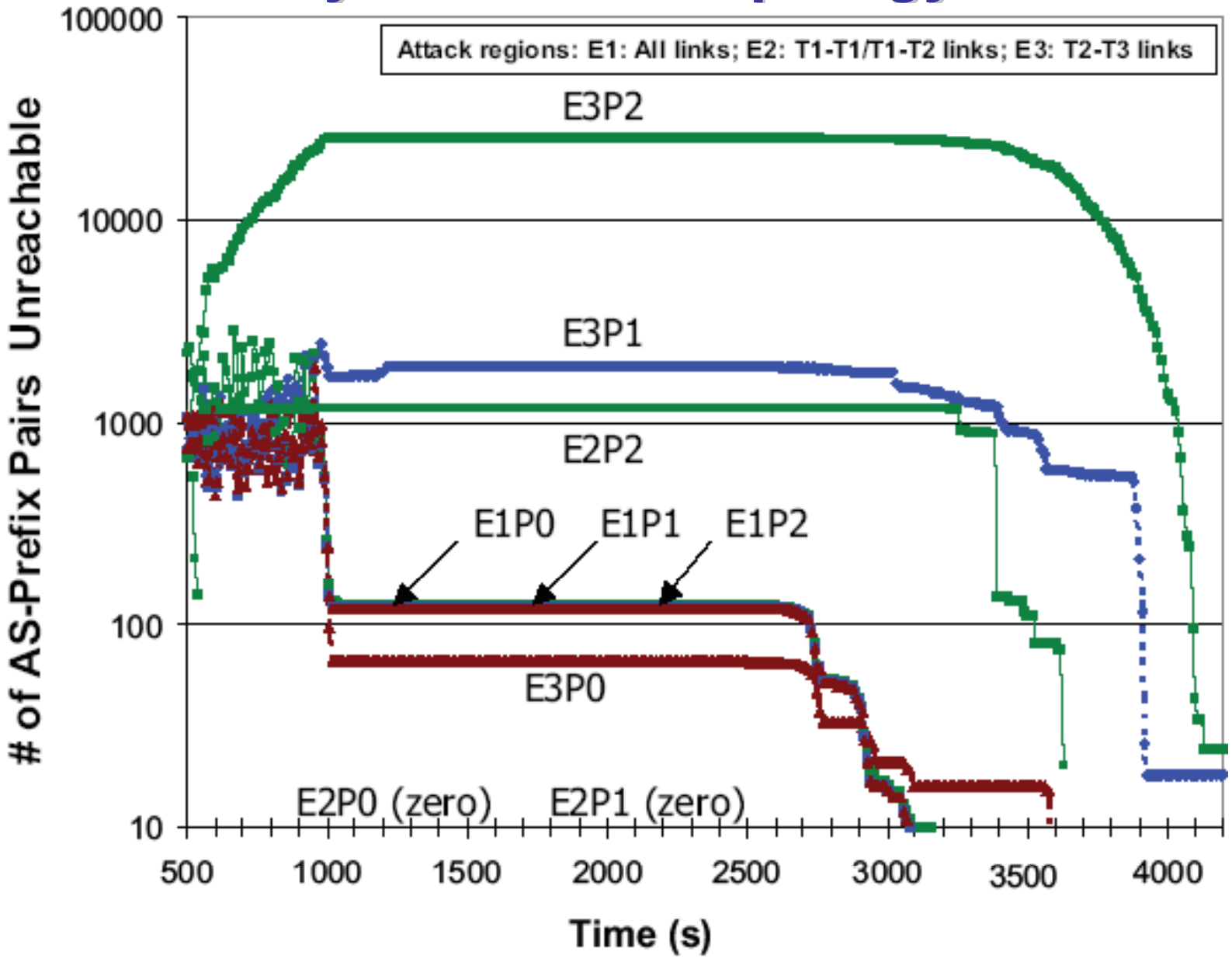
AS Path Length Degradation Vs. Policy



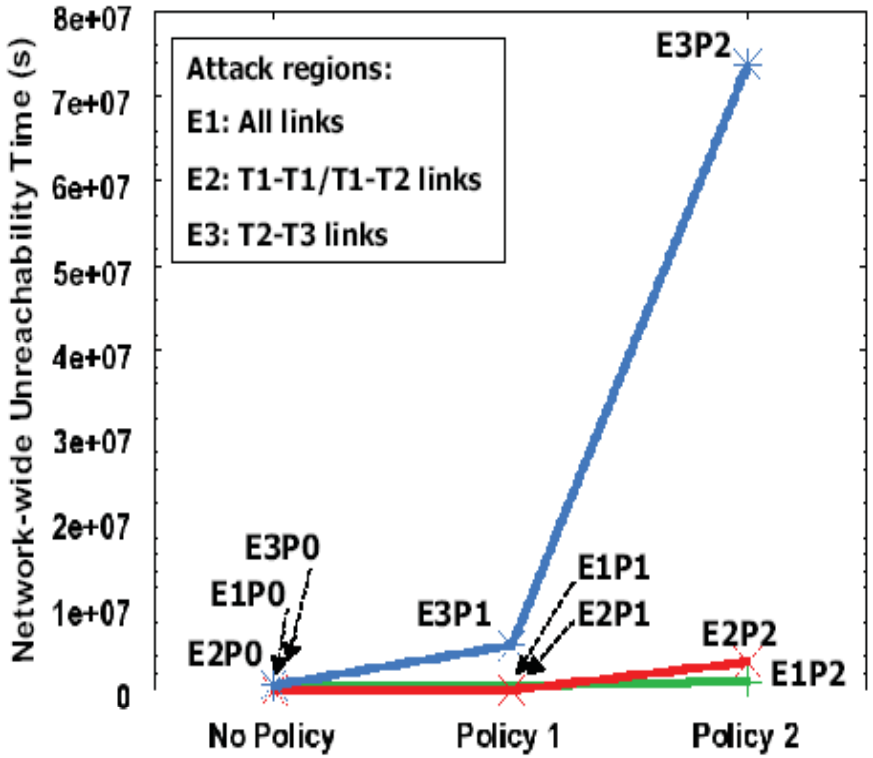
AS-Prefix Unreachability Time Vs. Policy

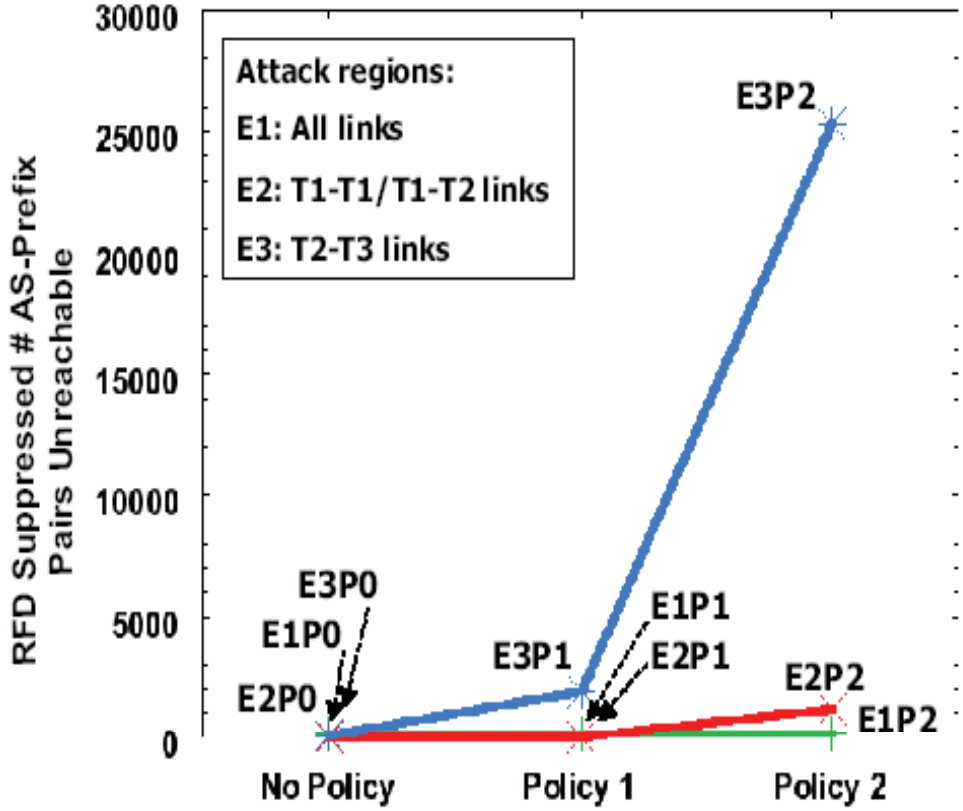


Unreachable Vs. Time: Sensitivity to Attack-Topology and Policy



Performance Degradation: Sensitivity to Attack-Topology and Policy





Conclusions

- Attackers can exploit RFD behavior to cause extended AS isolation
- The attack rate need be no more than about one successful attack every few MRAI intervals
- With BGP Graceful Restart (BGP-GR), the effort involved goes several orders of magnitude higher; so use of BGP-GR can add significant resiliency
- ISP's reluctant to enable BGP-GR?
  - "Several providers (US) suggest that the cost of implementing this feature outweighs the benefit." – NISCC (UK govt) BGP Best Practices
  - "Customers prefer to use an alternate route rather than BGP-GR because staleness of FIB issue with use of BGP-GR" – one source from an ISP says

Conclusions (contd.)

- Simulated semi-realistic topologies obtained by down-sampling from measured AS data (UCLA)
  - 256 nodes, 3 tiers, 753 links
- Studied the impacts of policy on the service disruptions due to attacks
  - Real-life service provider routing policies shown to result in significant amplification of service disruption following BGP session attacks
- Study being extended to encompass Spoofed Message Update Attacks (false prefix announcements and other types of spoofed updates)

Detailed paper in *IEEE JSAC*

K. Sriram, D. Montgomery, O. Borchert, O. Kim, and R. Kuhn, "Study of BGP Peering Session Attacks and Their Impacts on Routing Performance," *IEEE JSAC*: *Special Issue on High-speed Network Security*, Vol. 24, No. 10, October 2006, pp. 1901-1915.

http://www.antd.nist.gov/~ksriram/BGP_Security_Sriram_IEEE_JSAC.pdf

BGP Security Recommendations Pub

- D.R. Kuhn, K. Sriram, and D. Montgomery, "Border Gateway Protocol Security," NIST Special Publication 800-54 (Guidance Document for the Telecom Industry and US Government agencies), Draft circulated for comments, September 2006).

http://csrc.nist.gov/publications/drafts/800-54/Draft-SP800-54.pdf