

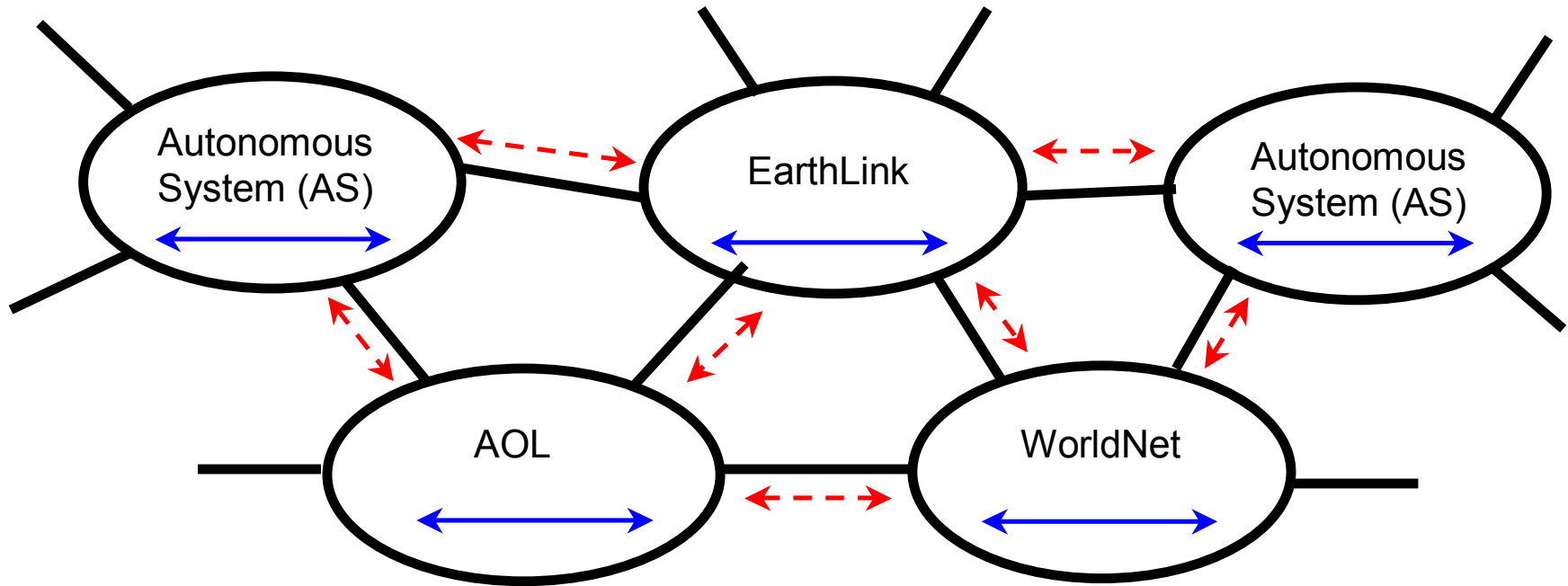
Interdomain Routing

- General Overview
- SPYCE Activity

Joan Feigenbaum, Yale University
<http://www.cs.yale.edu/~jfb>

Most of these slides were provided by
Tim Griffin of AT&T Research Labs.

Connecting Networks



Autonomous System (AS): A collection of IP subnets and routers under the same administrative authority.

———— Interior Routing Protocol (e.g., Open Shortest Path First)

----- Exterior Routing Protocol (e.g., Border Gateway Protocol)

BGP-4

- **BGP** = Border Gateway Protocol
- Is a Policy-Based routing protocol
- Is the de facto EGP of today's global Internet
- Relatively simple protocol, but configuration is complex and the entire world can see, and be impacted by, your mistakes.

- 1989 : BGP-1 [RFC 1105]
 - Replacement for EGP (1984, RFC 904)
- 1990 : BGP-2 [RFC 1163]
- 1991 : BGP-3 [RFC 1267]
- 1995 : BGP-4 [RFC 1771]
 - Support for Classless Interdomain Routing (CIDR)

AS Numbers (ASNs)

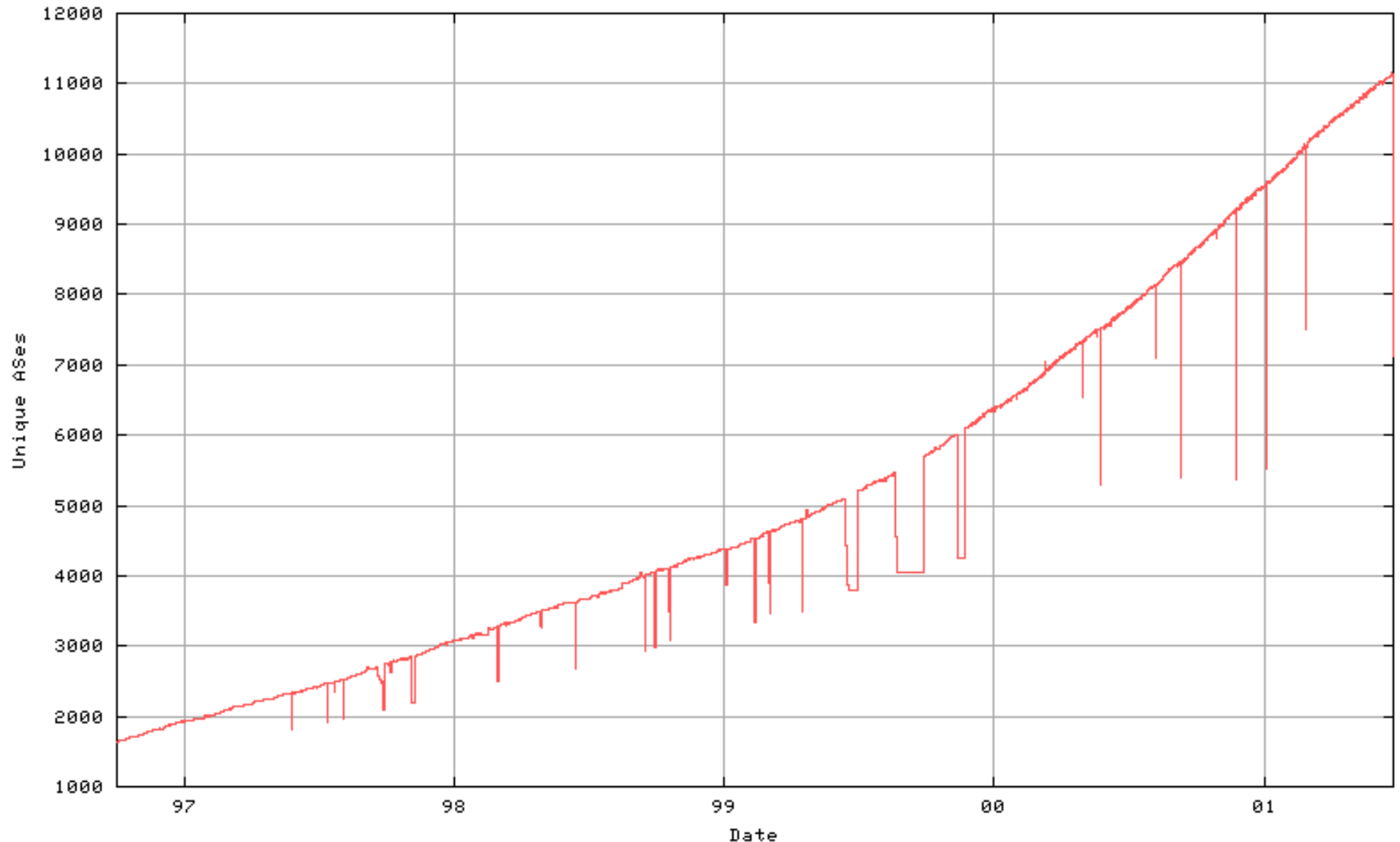
ASNs are 16 bit values.
64512 through 65535 are “private”

Currently over 12,000 in use.

- Yale: 29
- MIT: 3
- Harvard: 11
- Genuity: 1
- AT&T: 7018, 6341, 5074, ...
- UUNET: 701, 702, 284, 12199, ...
- Sprint: 1239, 1240, 6211, 6242, ...
- ...

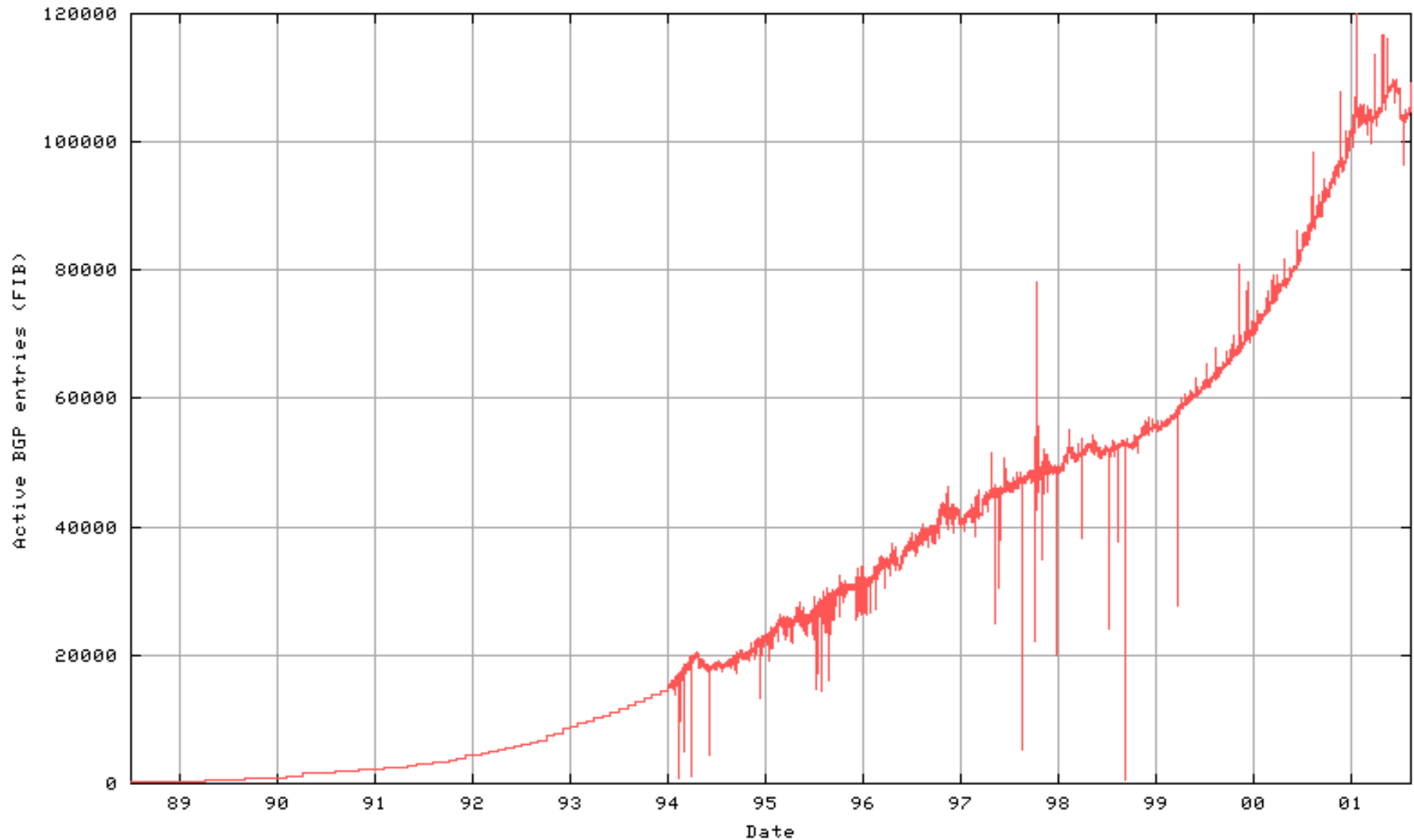
ASNs represent units of routing policy

How Many ASNs are there?



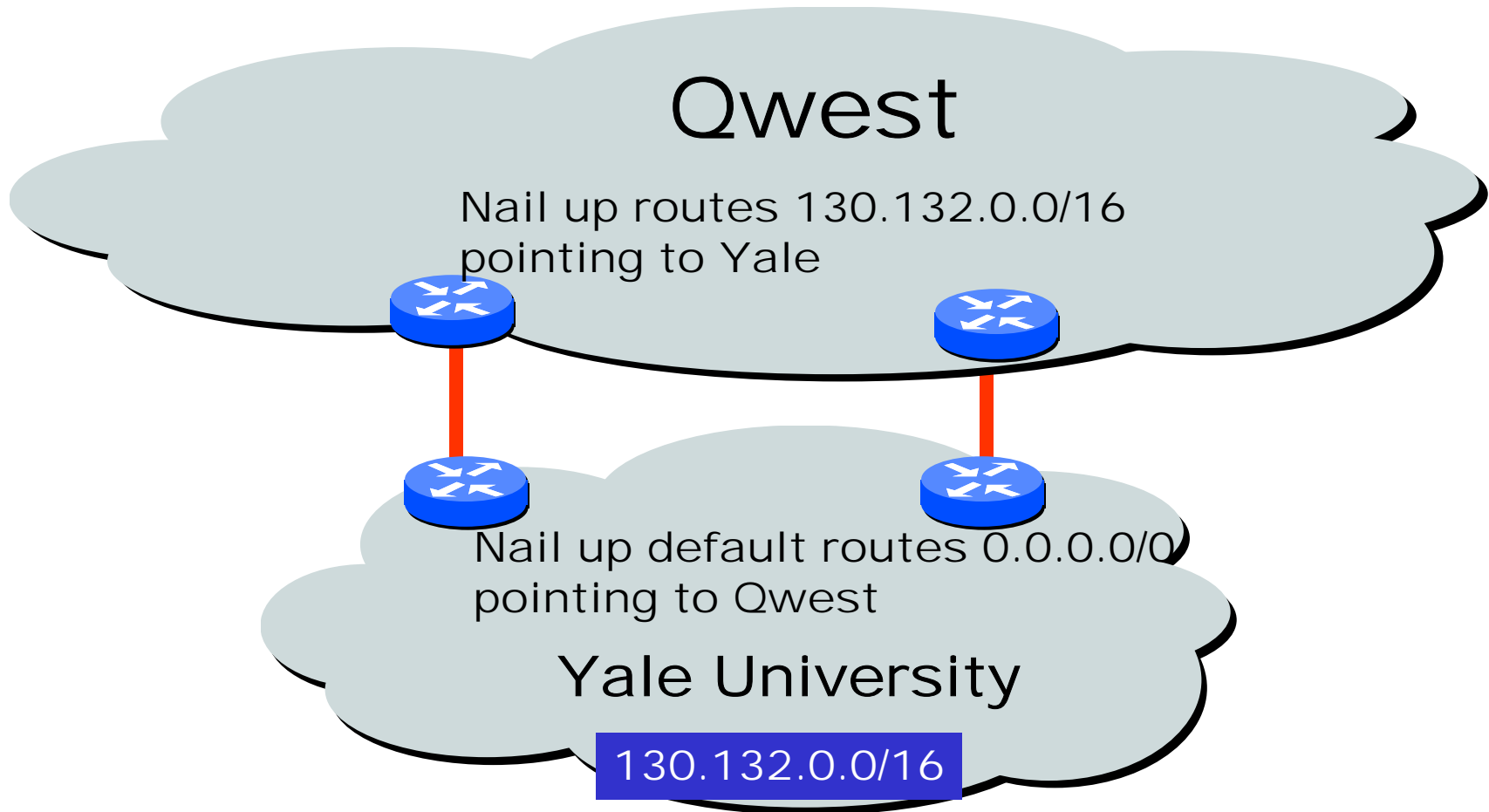
Thanks to Geoff Huston. <http://www.telstra.net/ops> on June 23, 2001

BGP Table Growth



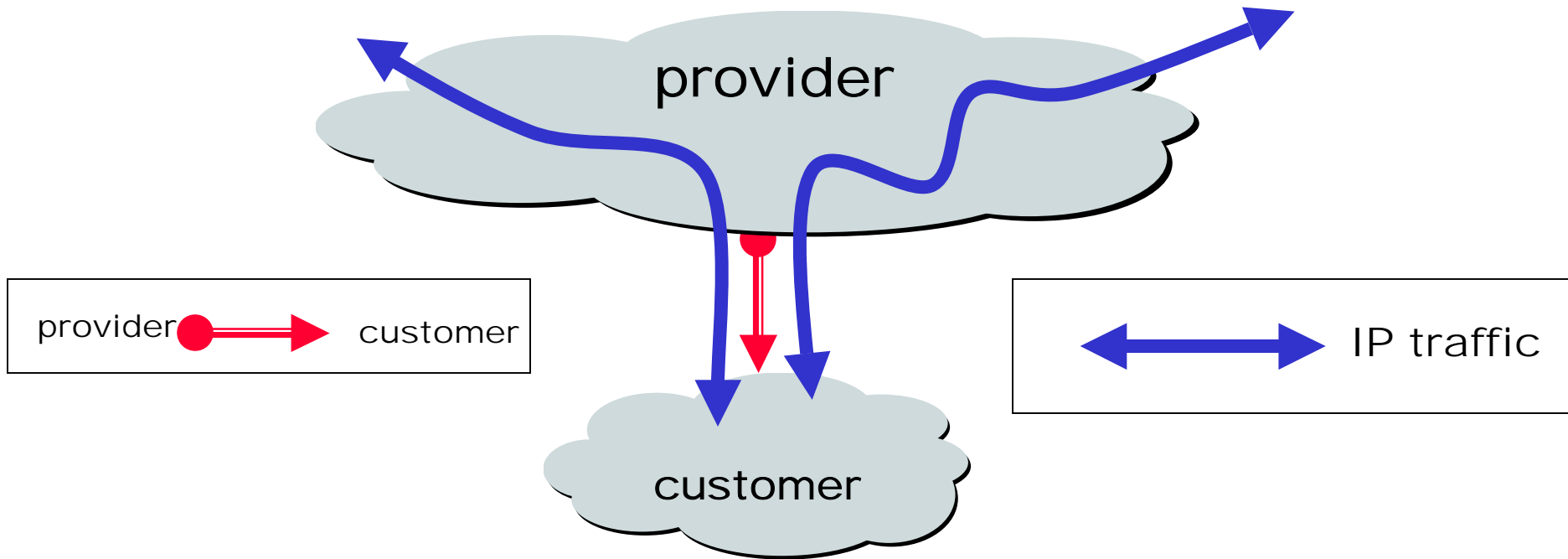
Thanks to Geoff Huston. <http://www.telstra.net/ops/bgptable.html> on August 8, 2001

Autonomous Routing Domains Don't Always Need BGP or an ASN



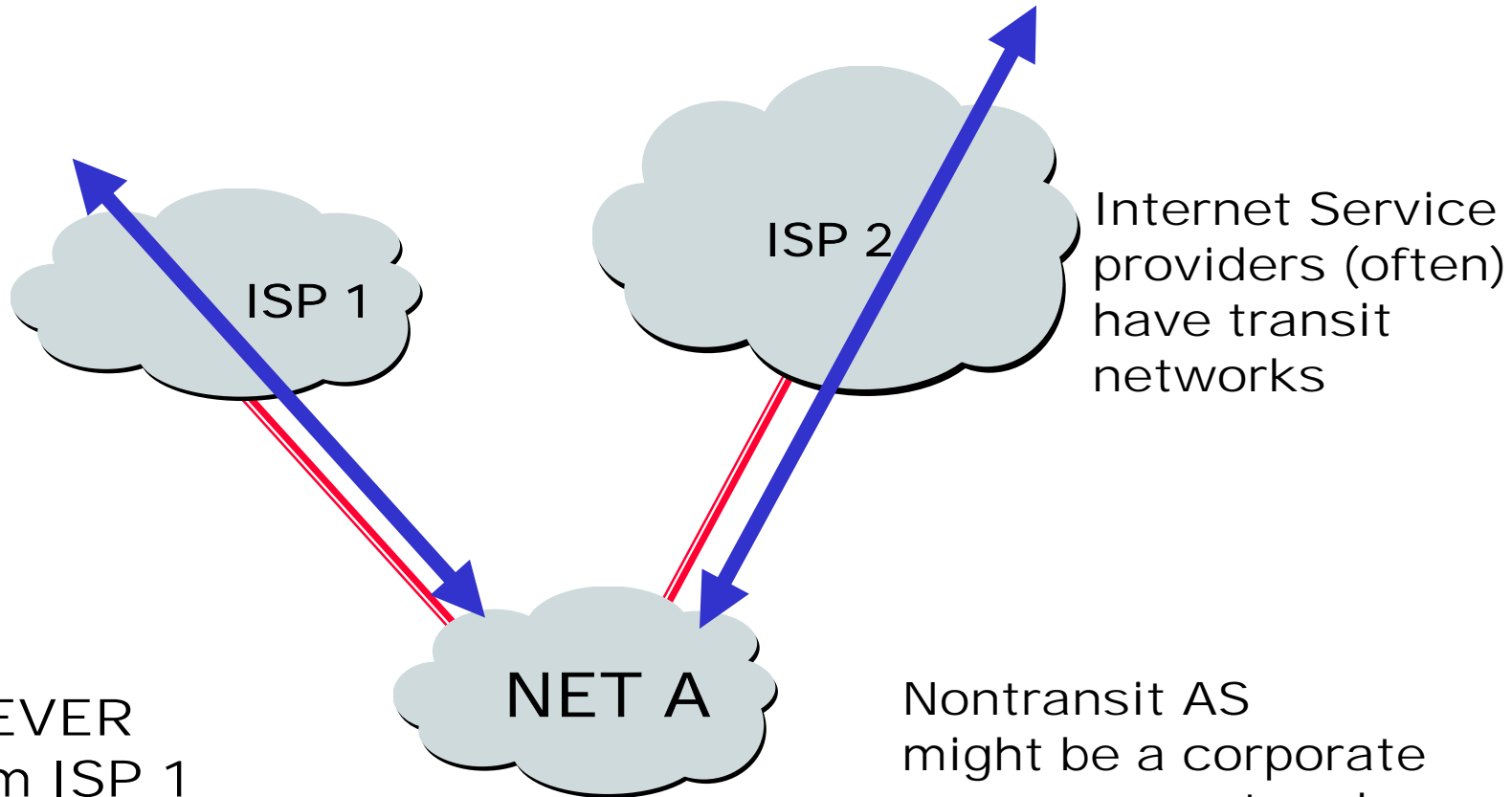
Static routing is the most common way of connecting an autonomous routing domain to the Internet. This helps explain why BGP is a mystery to many ...

Customers and Providers

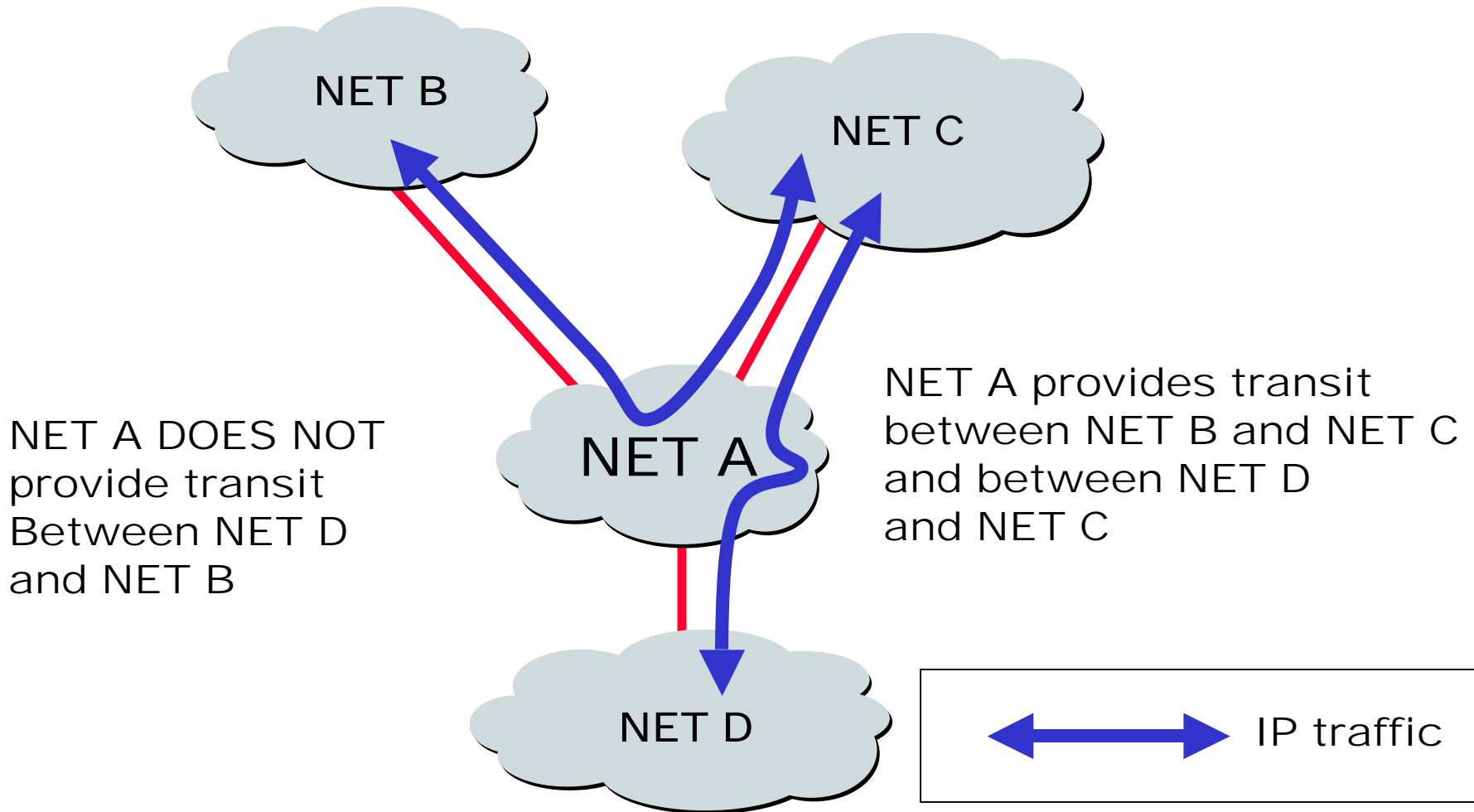


Customer pays provider for access to the Internet

Nontransit vs. Transit ASes

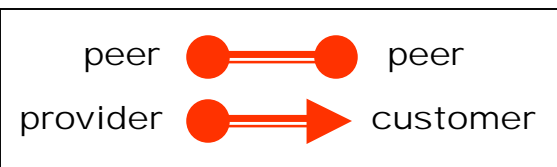
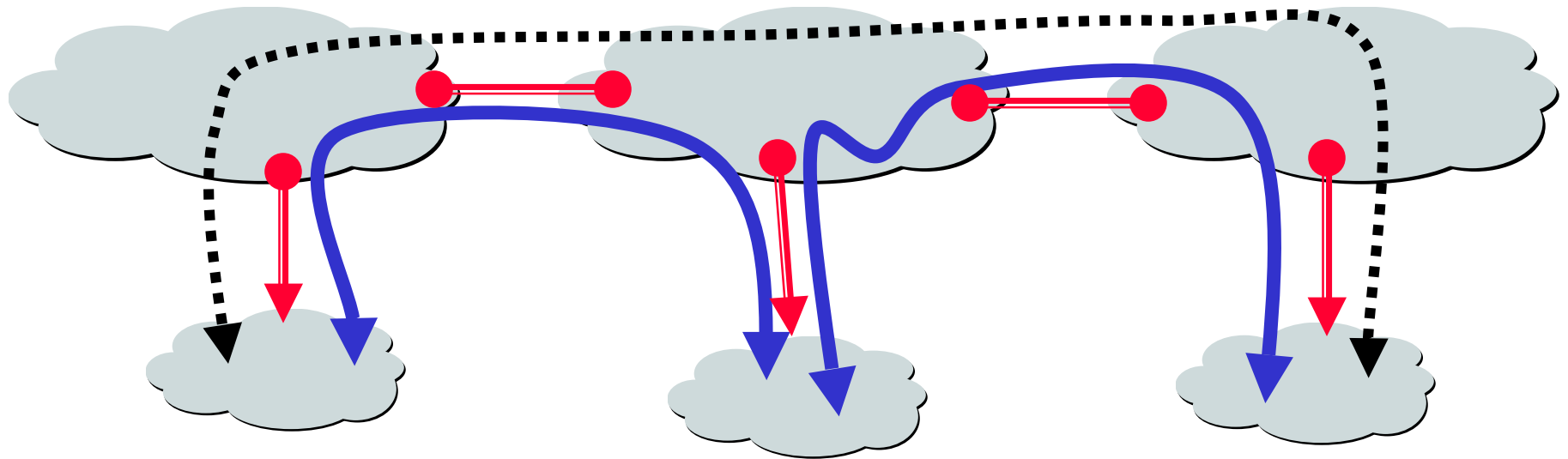


Selective Transit



Most transit networks transit in a selective manner...

The Peering Relationship



traffic
allowed



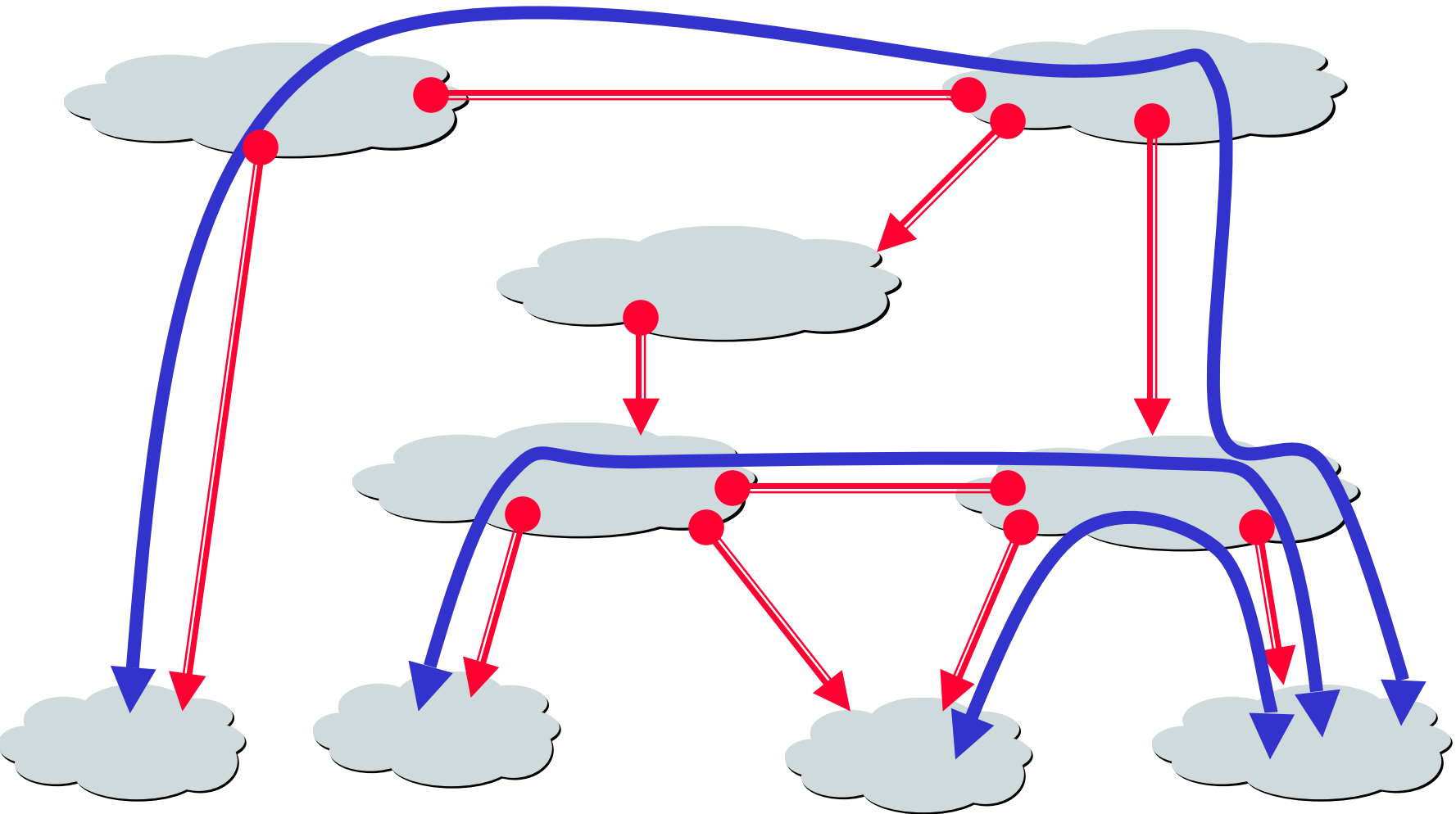
traffic NOT
allowed

Peers provide transit between their respective customers

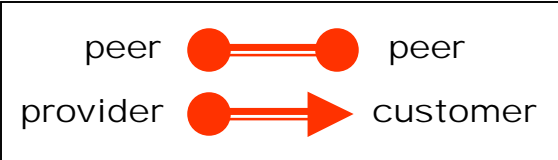
Peers do not provide transit between peers

Peers (often) do not exchange \$\$\$

Peering Provides Shortcuts



Peering also allows connectivity between the customers of "Tier 1" providers.



Technology of Distributed Routing

Link State

- Topology information is flooded within the routing domain
- Best end-to-end paths are computed locally at each router.
- **Best end-to-end paths determine next-hops.**
- Based on minimizing some notion of distance
- Works only if policy is shared and uniform
- Examples: OSPF, IS-IS

Vectoring

- Each router knows little about network topology
- Only best next-hops are chosen by each router for each destination network.
- **Best end-to-end paths result from composition of all next-hop choices**
- Does not require any notion of distance
- Does not require uniform policies at all routers
- Examples: RIP, BGP

The Gang of Four

Link State

Vectoring

IGP

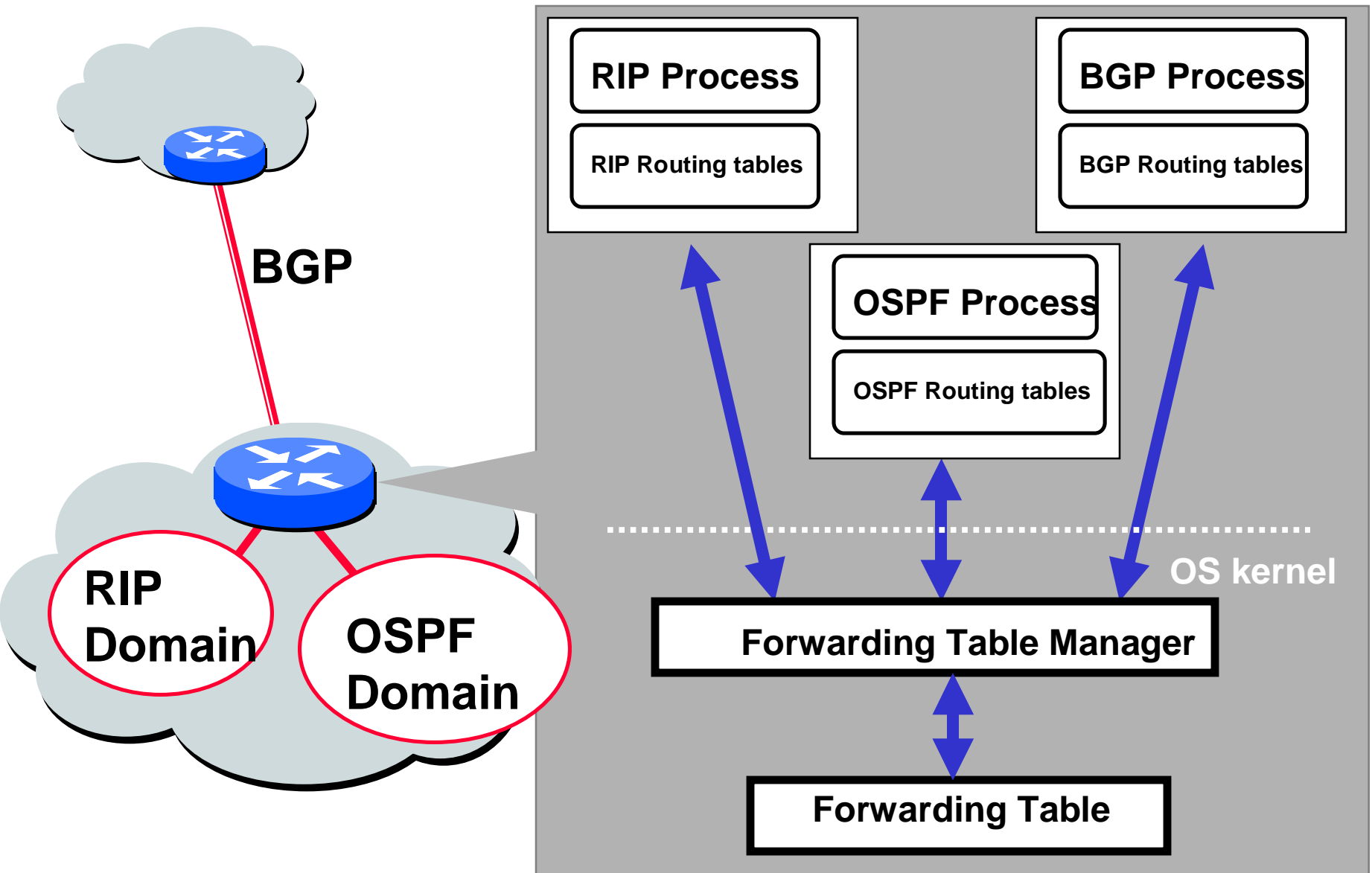
OSPF
IS-IS

RIP

EGP

BGP

Many Routing Processes Can Run on a Single Router



BGP-Based Computational Model (1)

- Follow abstract BGP model of Griffin and Wilfong:
Network is a graph with nodes corresponding to ASes and bidirectional links; intradomain-routing issues are ignored.
- Each AS has a routing table with LCPs to all other nodes:

Dest.	LCP				LCP cost
AS1	AS3	AS5	AS1		3
AS2	AS7	AS2			2

Entire paths are stored, not just next hop.

BGP-Based Computational Model (2)

- An AS “advertises” its routes to its neighbors in the AS graph, whenever its routing table changes.
- The computation of a single node is an infinite sequence of stages:



- Complexity measures:
 - Number of stages required for convergence
 - Total communication

BGP Attributes

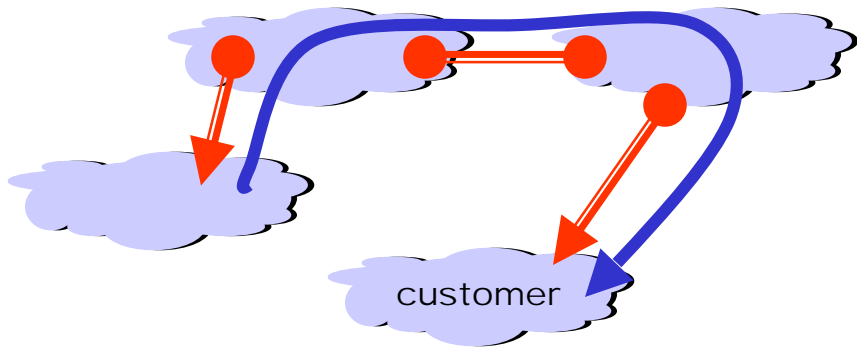
Value	Code	Reference
1	ORIGIN	[RFC1771]
2	AS_PATH	[RFC1771]
3	NEXT_HOP	[RFC1771]
4	MULTI_EXIT_DISC	[RFC1771]
5	LOCAL_PREF	[RFC1771]
6	ATOMIC_AGGREGATE	[RFC1771]
7	AGGREGATOR	[RFC1771]
8	COMMUNITY	[RFC1997]
9	ORIGINATOR_ID	[RFC2796]
10	CLUSTER_LIST	[RFC2796]
11	DPA	[Chen]
12	ADVERTISER	[RFC1863]
13	RCID_PATH / CLUSTER_ID	[RFC1863]
14	MP_REACH_NLRI	[RFC2283]
15	MP_UNREACH_NLRI	[RFC2283]
16	EXTENDED_COMMUNITIES	[Rosen]
...		
255	reserved for development	

Most
important
attributes

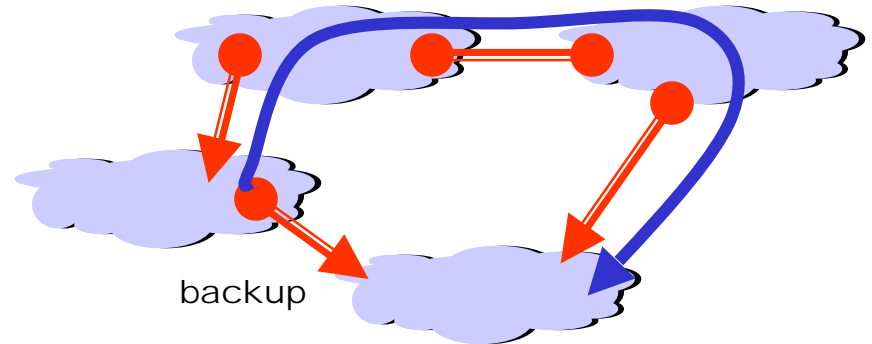
From IANA: <http://www.iana.org/assignments/bgp-parameters>

Not all attributes
need to be present in
every announcement

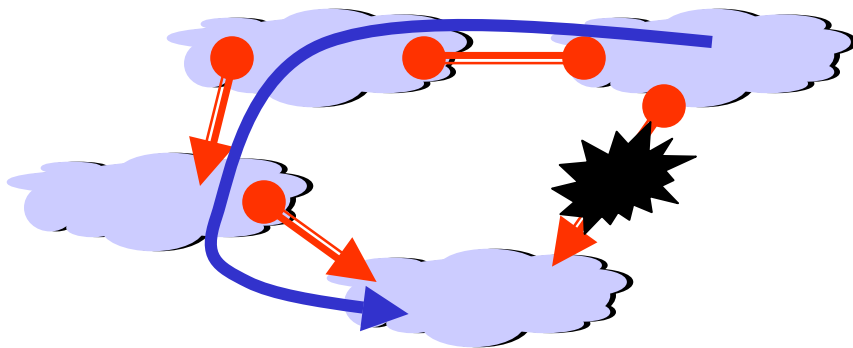
Policies Can Interact Strangely ("Route Pinning" Example)



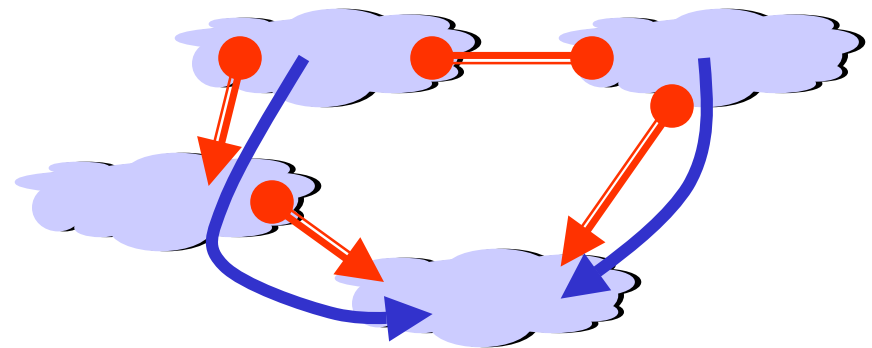
1



2 Install backup link using community



3 Disaster strikes primary link
and the backup takes over



4 Primary link is restored but some
traffic remains *pinned* to backup

News at 11:00h

- **BGP is not guaranteed to converge on a stable routing. Policy interactions could lead to “livelock” protocol oscillations.**
[See “Persistent Route Oscillations in Inter-domain Routing” by K. Varadhan, R. Govindan, and D. Estrin. ISI report, 1996](#)
- **Corollary: BGP is not guaranteed to recover from network failures.**

Can we model BGP?

Underlying problem

Shortest Paths

X?

**Distributed means of
computing a solution.**

RIP, OSPF, IS-IS

BGP

What Problem is BGP solving?

X could :

- aid in the design of policy analysis algorithms and heuristics
- aid in the analysis and design of BGP and extensions
- help explain some BGP routing anomalies
- provide a fun way of thinking about the protocol

Separate dynamic and static semantics

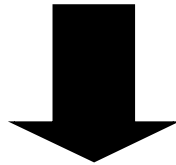
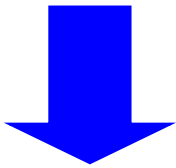
**static
semantics**

**dynamic
semantics**

BGP Policies

BGP

**Booo Hooo,
Many, many
complications...**



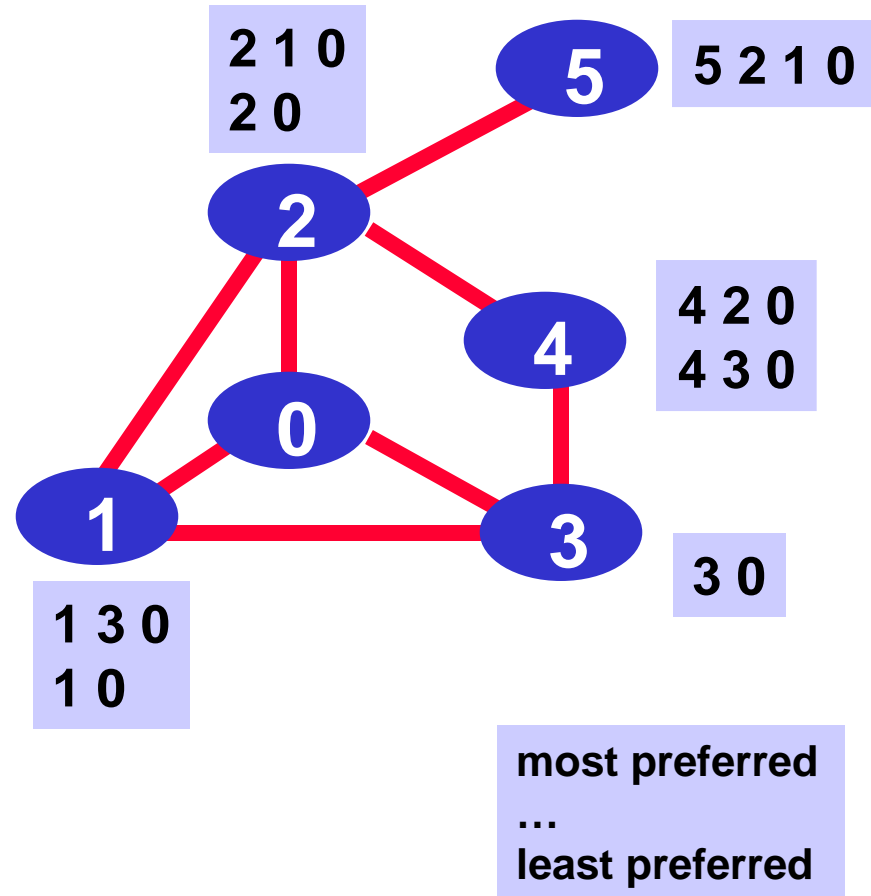
**Stable Paths
Problem (SPP)**

SPVP

**SPVP = Simple Path
Vector Protocol = a
distributed
algorithm for
solving SPP**

An instance of the *Stable Paths Problem* (SPP)

- A graph of nodes and edges,
- Node 0, called *the origin*,
- For each non-zero node, a set or permitted paths to the origin. This set always contains the “null path”.
- A ranking of permitted paths at each node. Null path is always least preferred. (Not shown in diagram)



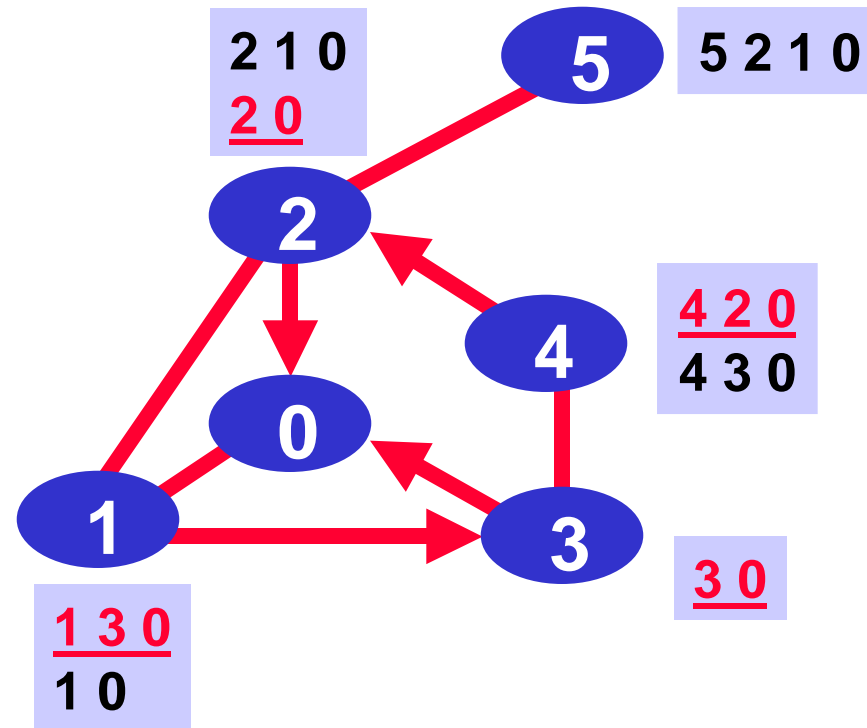
When modeling BGP : nodes represent BGP speaking routers, and 0 represents a node originating some address block

Yes, the translation gets messy!

A Solution to a Stable Paths Problem

A solution is an assignment of permitted paths to each node such that

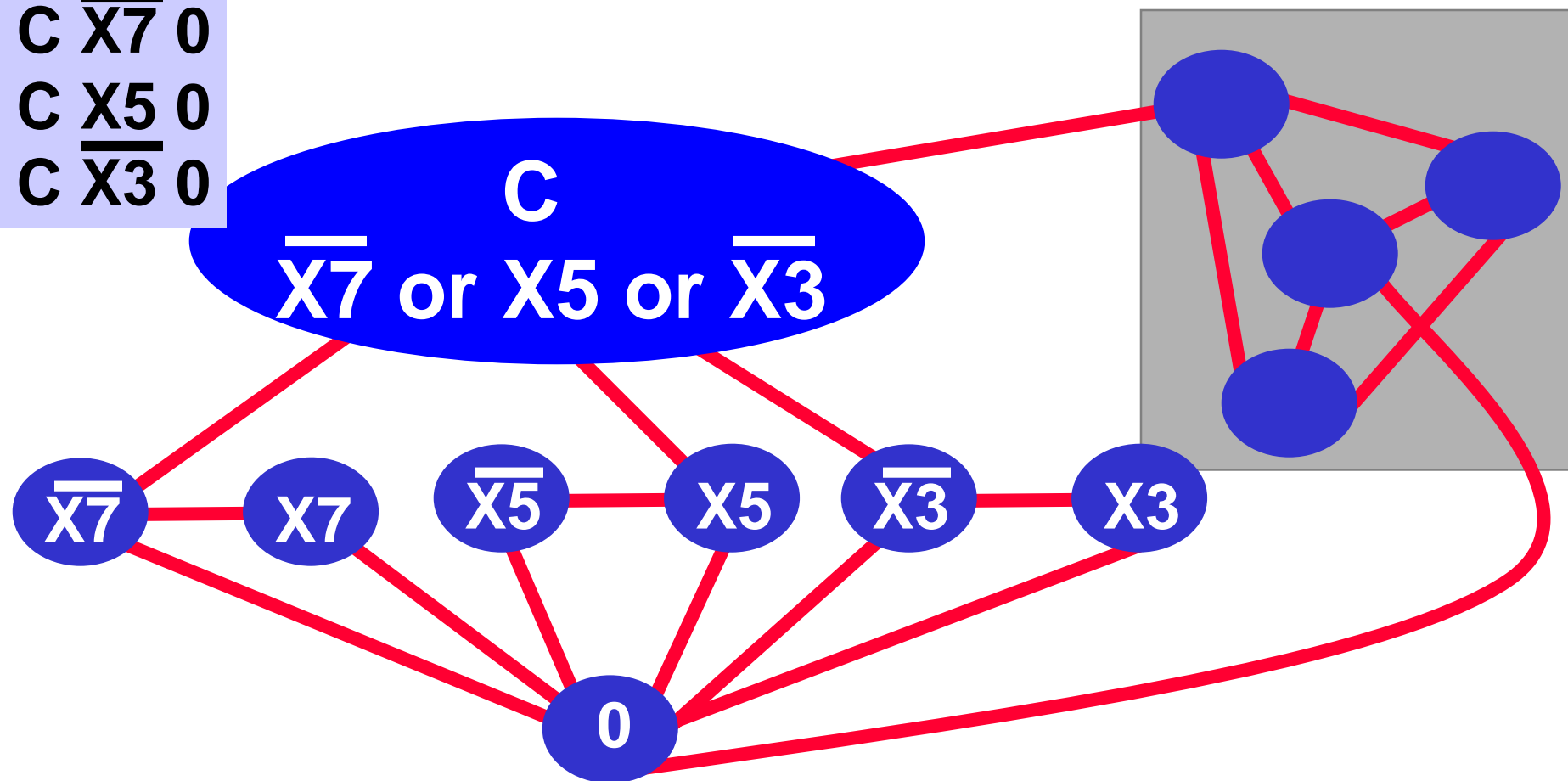
- node u 's assigned path is either the null path or is a path uwP , where wP is assigned to node w and $\{u,w\}$ is an edge in the graph,
- each node is assigned the highest ranked path among those consistent with the paths assigned to its neighbors.



A Solution need not represent a shortest path tree, or a spanning tree.

SPP solvability is NP-complete

C $\overline{X7}$ 0
C X5 0
C $\overline{X3}$ 0



SPVP protocol

Pick the best path available at any given time...

```
process spvp[u]
{
  receive P from w →
  { rib-in(u←w) := u P
    if rib(u) != best(u) {
      rib(u) := best(u)
      foreach v in peers(u) {
        send rib(u) to v
      }
    }
  }
}
```

A Sufficient Condition for Sanity

If an instance of SPP has an **acyclic dispute digraph**, then

Static (SPP)

solvable

unique solution

**all sub-problems
uniquely solvable**

Dynamic (SPVP)

safe (can't diverge)

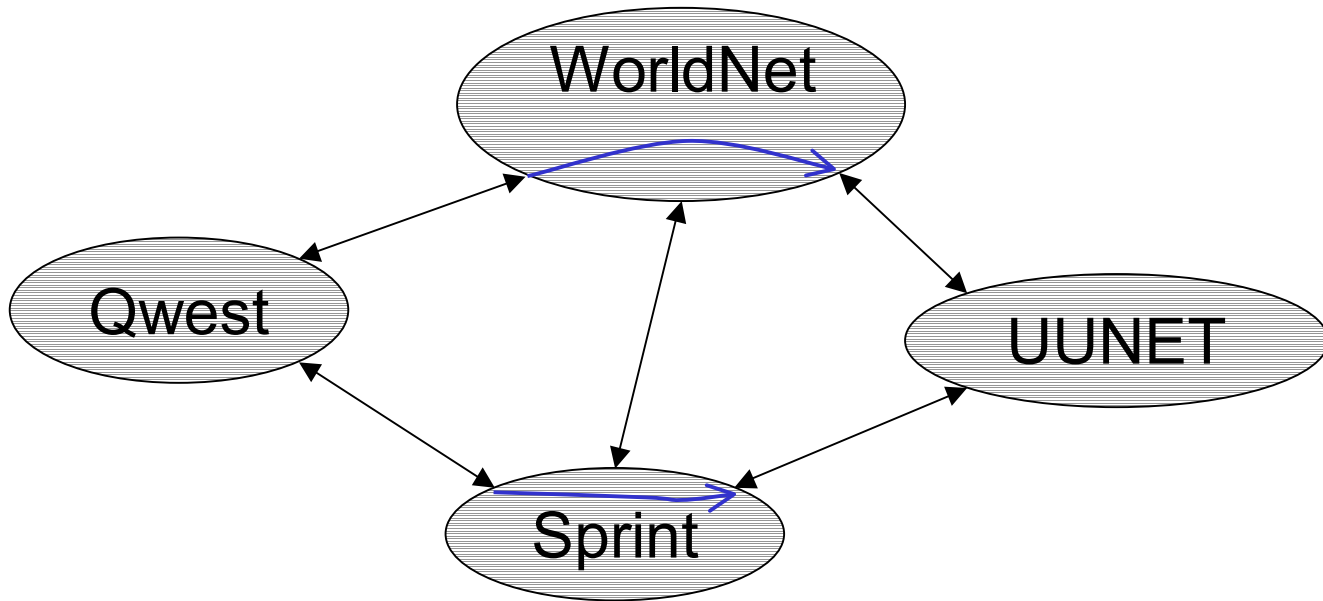
predictable restoration

**robust with respect to
link/node failures**

Path Vector Systems

- The **path vector system** is another model of the interdomain routing problem.
 - Preferences are specified at each node by functions on adjacent edges, instead of rankings on complete paths to the origin.
 - Any path vector system can be converted to an instance of SPP, and vice versa.
 - **Local condition “equivalent” to an acyclic dispute digraph:** preference functions must be increasing.
- Can we design a local language to describe routing policies that enforces these “nice” **local conditions**?

Lowest-Cost (LCP) Routing Mechanism-Design Problem



Agents: Transit ASs

Inputs: Transit costs

Outputs: Routes, Payments

Some Results

- If the network is biconnected, there is a unique strategyproof mechanism computing the LCP that does not pay nodes not on the LCP.
- There exists a BGP-based (distributed) algorithm to compute the above prices:
 - using routing tables that are only a constant factor larger than those required for BGP; and
 - converging as fast as BGP, with an additive penalty, in the worst case, equal to the maximum length of a lowest-cost path when one of the nodes on the original LCP is removed.

Ongoing Work on Incentive-Compatible Routing

- **Reconciling Computational and Strategic Models**
- **Handling Adversarial *and* Strategic Agents**
- **Improved Cost Models**
- **Mechanism-Design Goals other than Lowest Cost**
 - **Minimize delay?**
 - **General route preferences**

Security and BGP

- No scalable means to verify the authenticity and legitimacy of BGP control traffic.
- BGP is vulnerable to attack.
 - [S-BGP IETF draft]: **An attack is anything that causes abnormal operation**, e.g., as a result of messages or malfunctioning BGP speakers.
 - Update messages could be constructed and sent maliciously (*not* by a functioning BGP speaker), e.g., operational routes could be withdrawn
 - Messages could be processed by unintended recipients

S-BGP

- **S-BGP** attempts to deploy countermeasures against malicious adversaries.
 - Establish a Public Key Infrastructure (PKI)
 - Use IPSec
 - Add new attributes to BGP messages
 - Digital signatures protect BGP attributes
- **It does not protect** BGP speakers from behavior inconsistent with intended routing policies.

S-BGP Has Not Been Deployed

- **Are all of those digital signatures too time-consuming?**
- **Is S-BGP solving “the wrong problem?”**
- **Griffin: “What’s needed is an interdomain trust model.”**

“Foundations” of Interdomain Routing

- **Griffin: “Beyond BGP: When will it break, and what will replace it?”**
- **Potential SPYCE activity**
 - **What *is* interdomain routing?
(we need a better answer than “what BGP does”)**
 - **Inherent tension between “autonomy” (the A in AS) and the ability to make global guarantees (connectivity, stability, robustness, trustworthiness)?**
 - **Find the right tradeoffs and develop incentive-compatible protocols and enforcement mechanisms.**