PCH
Packet Clearing House

Best Practices in Anycast Routing

Version 0.6

June, 1996

Bill Woodcock
Packet Clearing House

PCH
Packet Clearing House

What *isn't* Anycast?

- Not a protocol, not a different version of IP, nobody's proprietary technology.
- Doesn't require any special capabilities in the servers, clients, or network.
- Doesn't break or confuse existing infrastructure.

PCH
Packet Clearing House

What *is* Anycast?

- Just a configuration methodology.
- Described in concept, if not in detail, in RFC 1546, “Host Anycasting Service”.
- Used for topological load-balancing of Internet-connected services since at least 1989.
- It’s the proposed basis for expanding the number of root nameservers beyond the current thirteen.

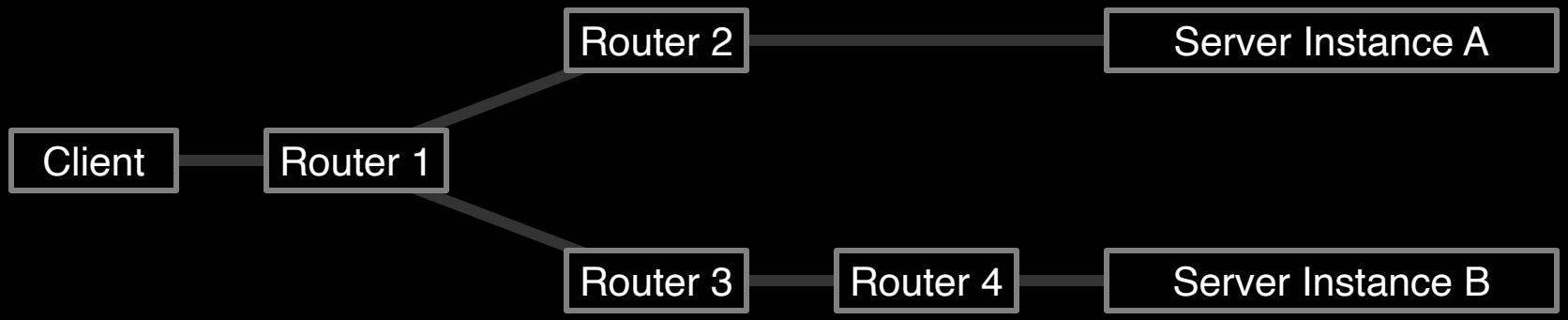
PCH
Packet Clearing House

How Does Anycast Work?

- The basic idea is extremely simple:
- Multiple instances of a service share the same IP address.
- The routing infrastructure directs any packet to the topologically nearest instance of the service.
- What little complexity exists is in the optional details.

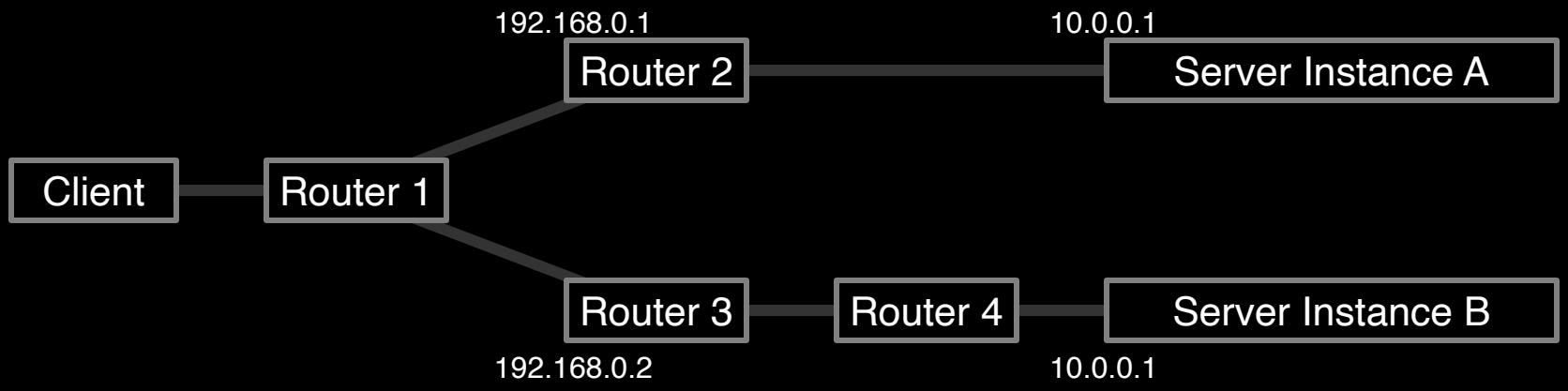
Packet Clearing House

Example



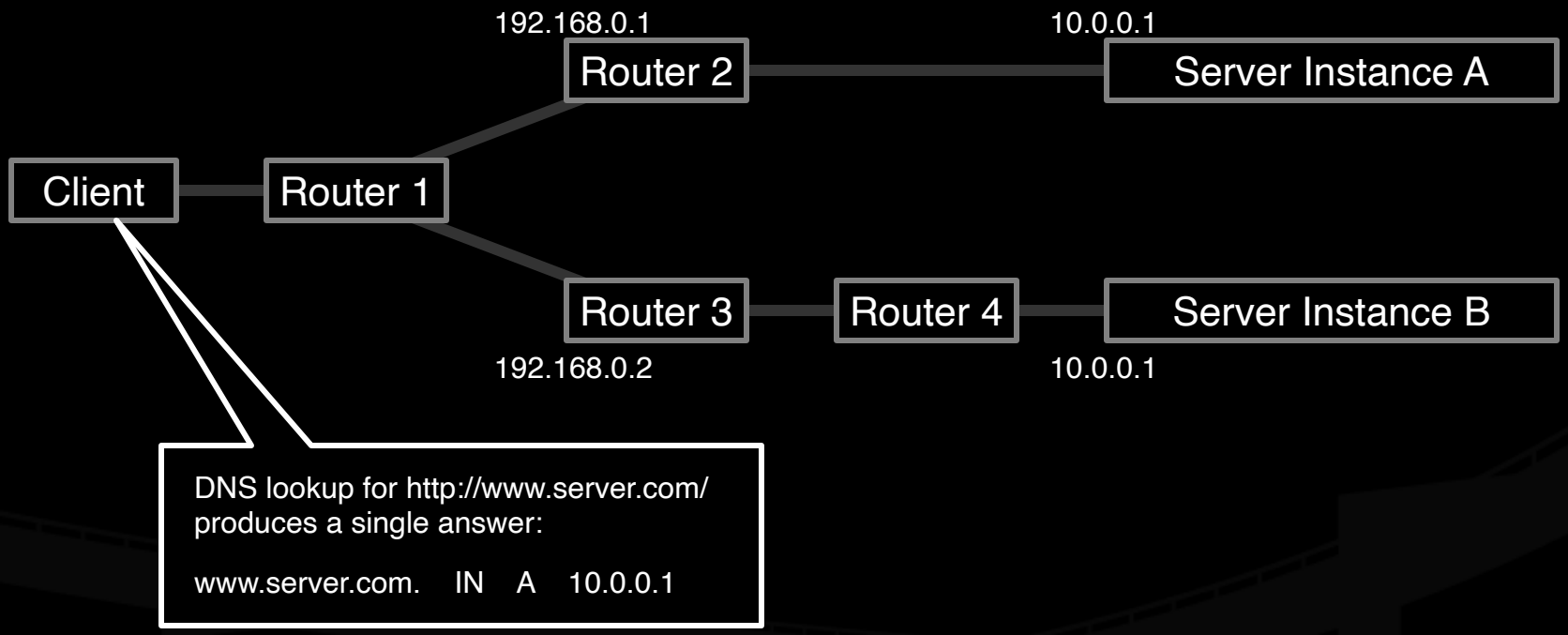
Packet Clearing House

Example

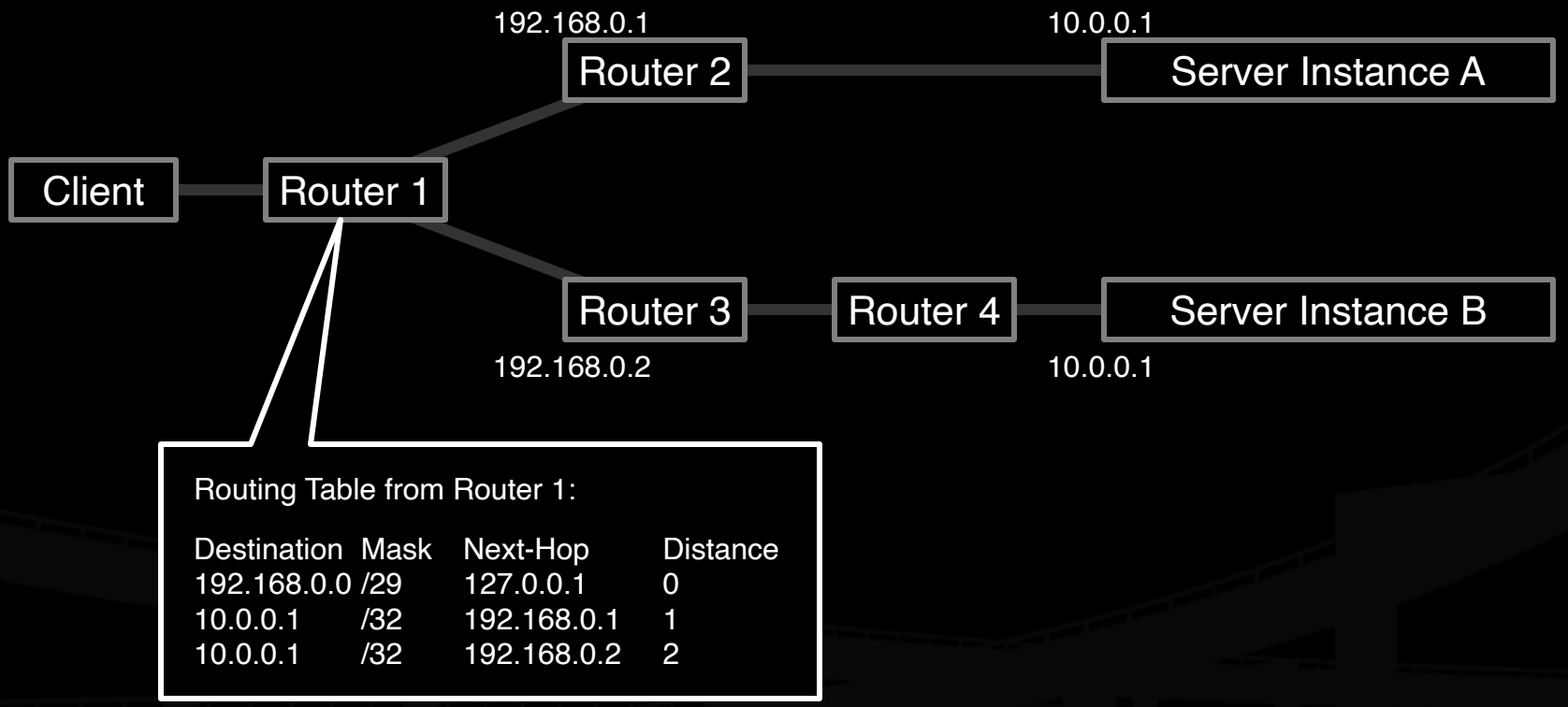


Packet Clearing House

Example

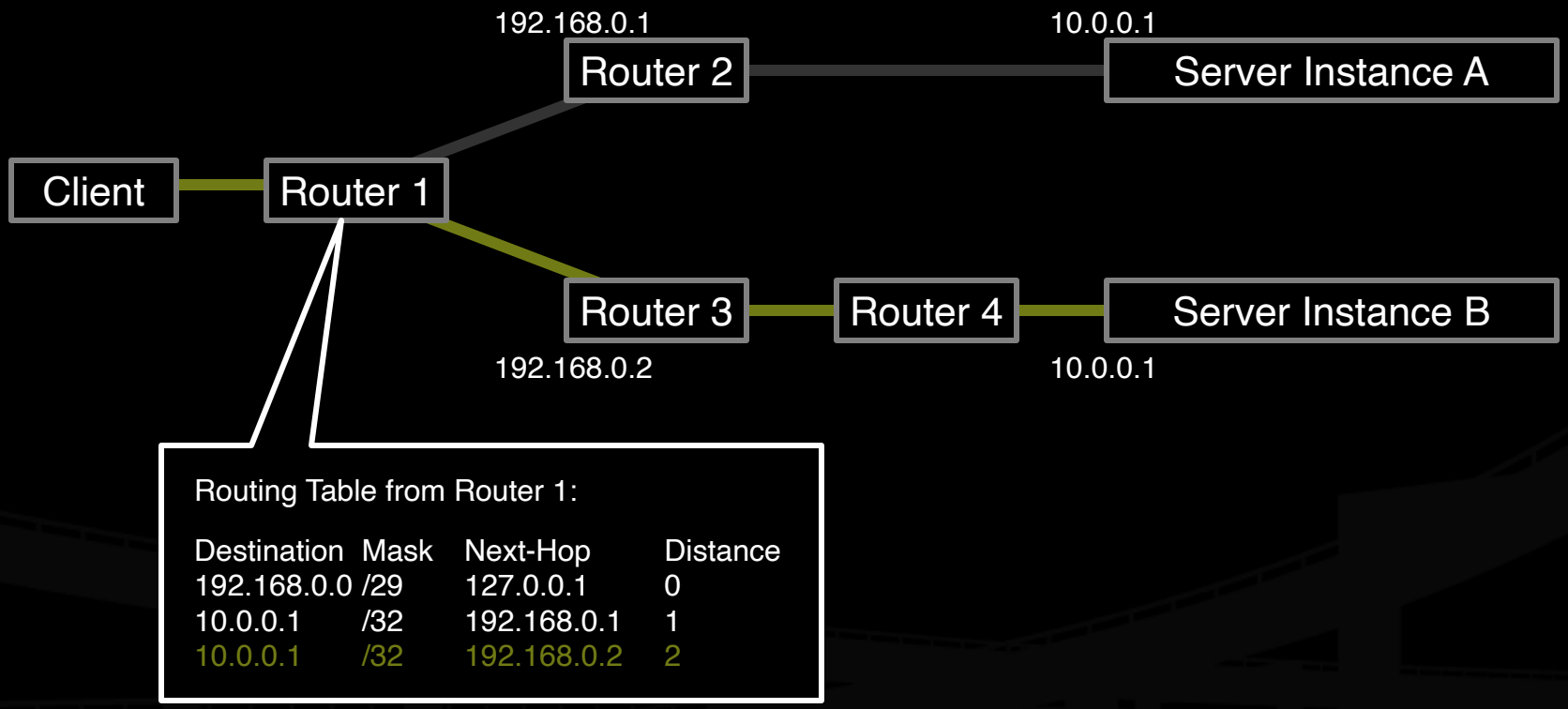


Example

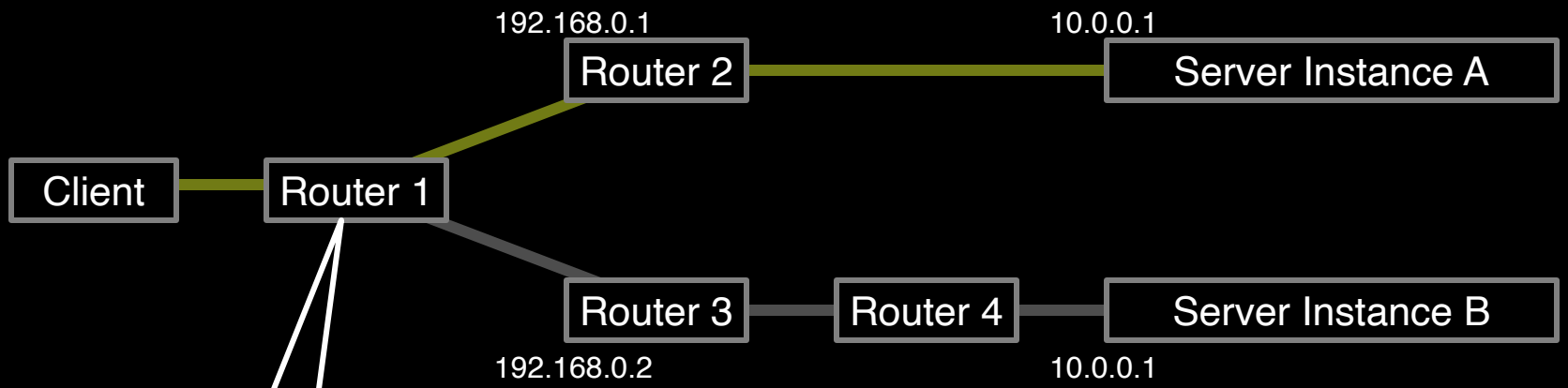

192.168.0.1
10.0.0.1
Router 2
Server Instance A
Client
Router 1
Router 3
Router 4
Server Instance B
192.168.0.2
10.0.0.1
Routing Table from Router 1:
Destination Mask Next-Hop Distance
192.168.0.0 /29 127.0.0.1 0
10.0.0.1 /32 192.168.0.1 1
10.0.0.1 /32 192.168.0.2 2

Packet Clearing House

Example



Example

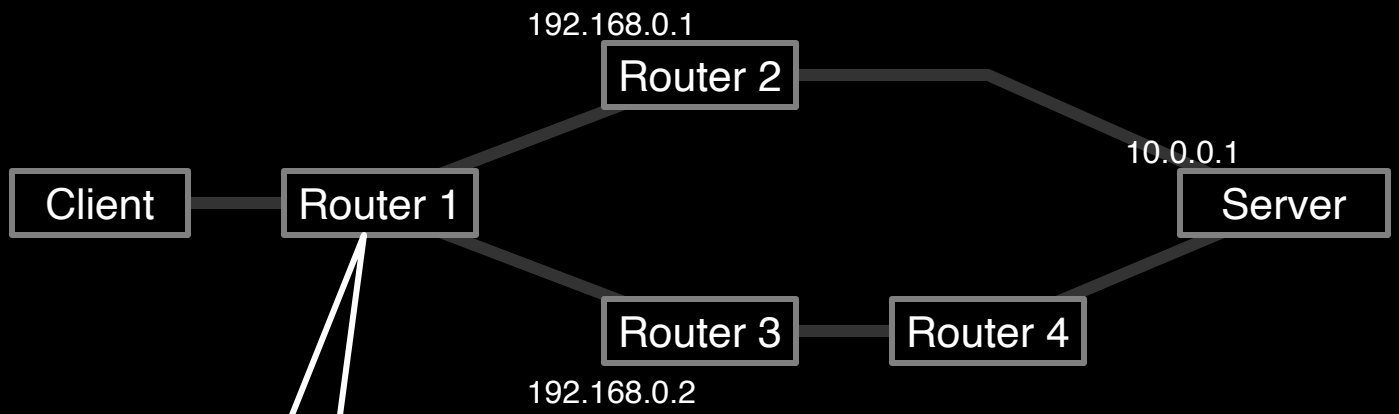


Routing Table from Router 1:

Destination	Mask	Next-Hop	Distance
192.168.0.0	/29	127.0.0.1	0
10.0.0.1	/32	192.168.0.1	1
10.0.0.1	/32	192.168.0.2	2

Example

What the routers think the topology looks like:



Routing Table from Router 1:

Destination	Mask	Next-Hop	Distance
192.168.0.0	/29	127.0.0.1	0
10.0.0.1	/32	192.168.0.1	1
10.0.0.1	/32	192.168.0.2	2

Packet Clearing House

Building an Anycast Server Cluster

- Anycast can be used in building either local server clusters, or global networks, or global networks of clusters, combining both scales.

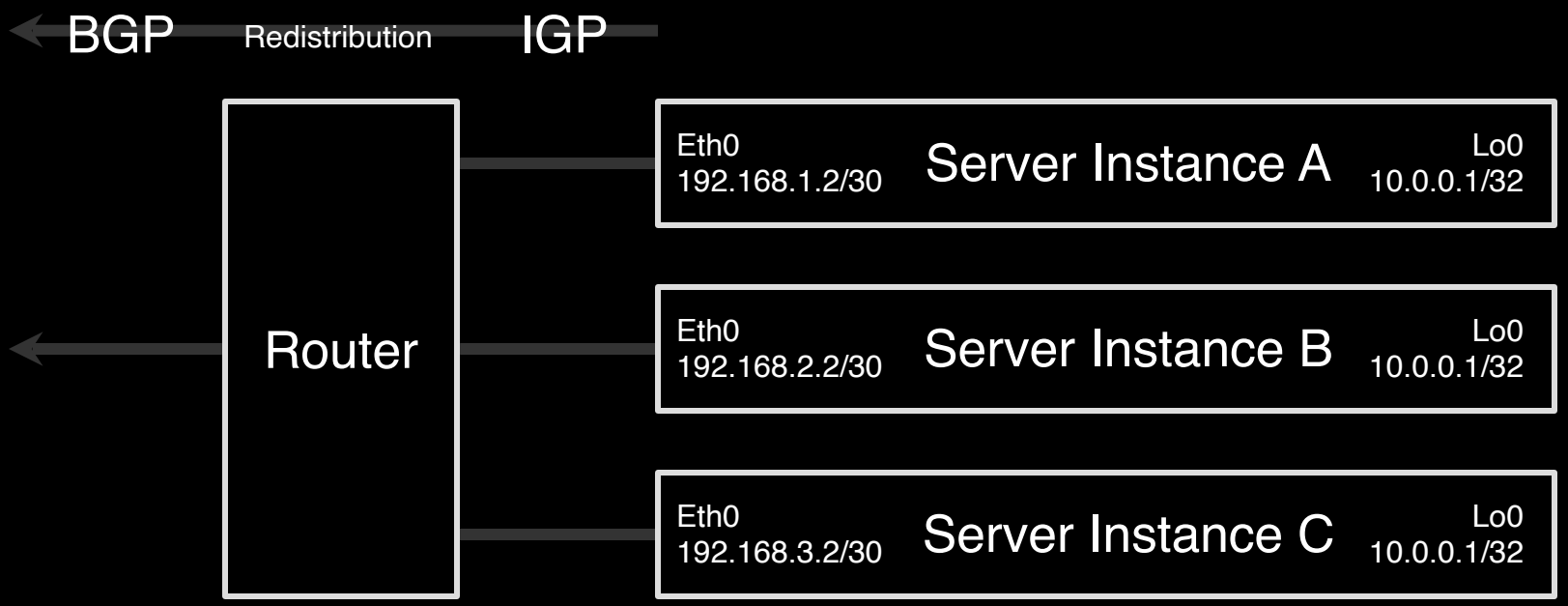
PCH
Packet Clearing House

Building an Anycast Server Cluster

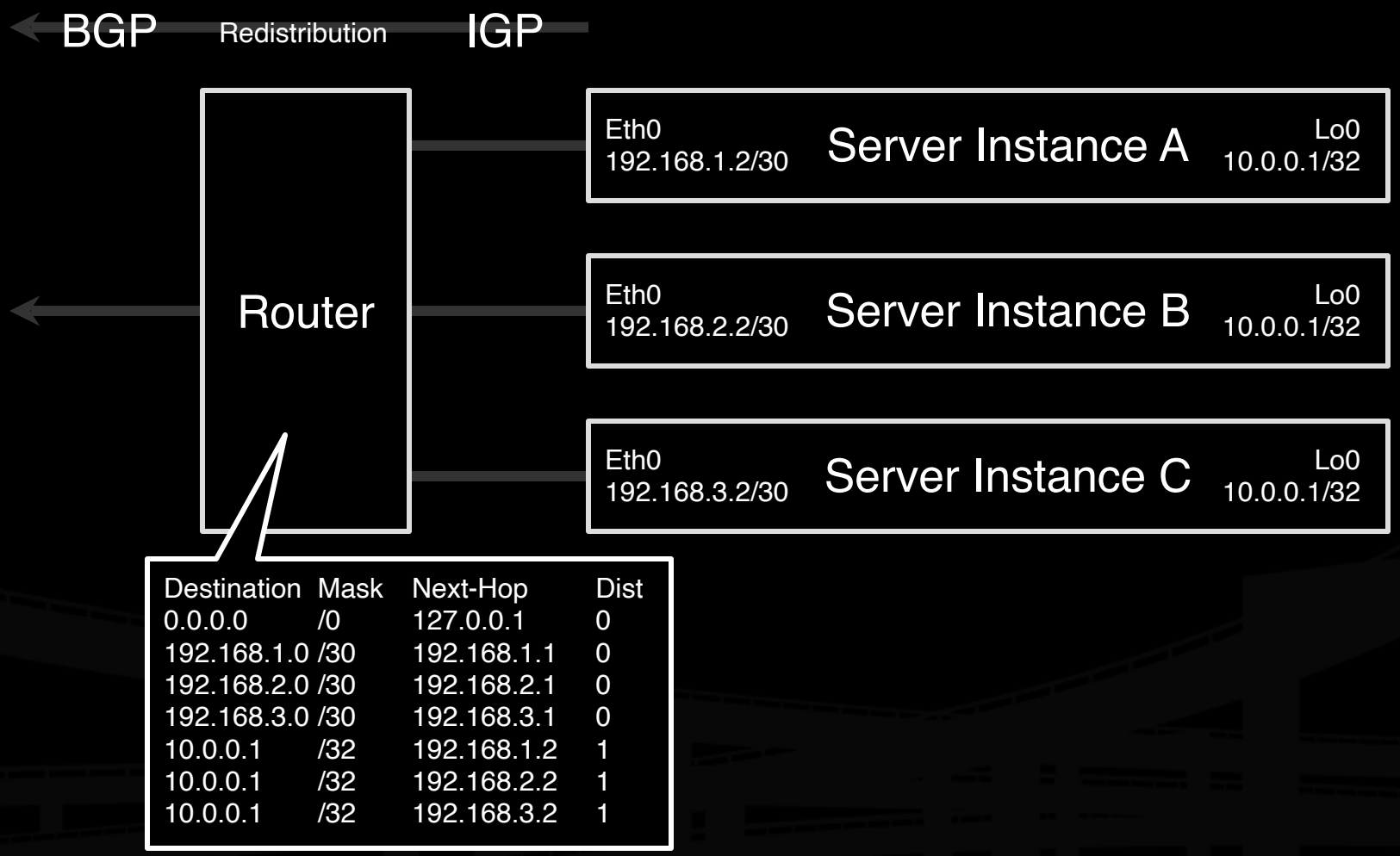
- Typically, a cluster of servers share a common virtual interface attached to their loopback devices, and speak an IGP routing protocol to an adjacent BGP-speaking border router.
- The servers may or may not share identical content.

Packet Clearing House

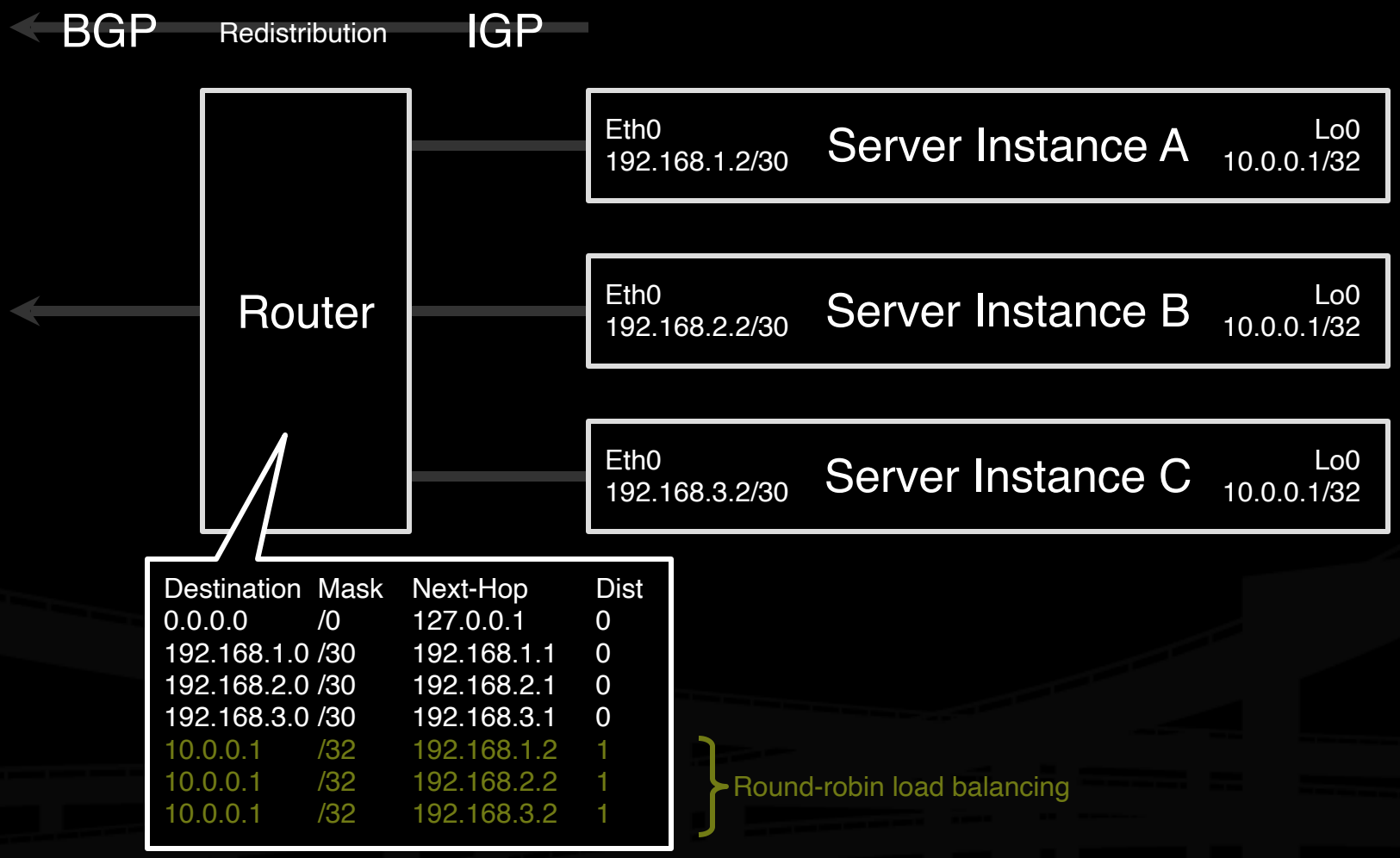
Example



Example


BGP
Redistribution
IGP
Router
Eth0
192.168.1.2/30
Server Instance A
Lo0
10.0.0.1/32
Eth0
192.168.2.2/30
Server Instance B
Lo0
10.0.0.1/32
Eth0
192.168.3.2/30
Server Instance C
Lo0
10.0.0.1/32
Destination Mask Next-Hop Dist
0.0.0.0 /0 127.0.0.1 0
192.168.1.0 /30 192.168.1.1 0
192.168.2.0 /30 192.168.2.1 0
192.168.3.0 /30 192.168.3.1 0
10.0.0.1 /32 192.168.1.2 1
10.0.0.1 /32 192.168.2.2 1
10.0.0.1 /32 192.168.3.2 1

Example


BGP
Redistribution
IGP
Router
Eth0
192.168.1.2/30
Server Instance A
Lo0
10.0.0.1/32
Eth0
192.168.2.2/30
Server Instance B
Lo0
10.0.0.1/32
Eth0
192.168.3.2/30
Server Instance C
Lo0
10.0.0.1/32
Destination Mask Next-Hop Dist
0.0.0.0 /0 127.0.0.1 0
192.168.1.0 /30 192.168.1.1 0
192.168.2.0 /30 192.168.2.1 0
192.168.3.0 /30 192.168.3.1 0
10.0.0.1 /32 192.168.1.2 1
10.0.0.1 /32 192.168.2.2 1
10.0.0.1 /32 192.168.3.2 1
Round-robin load balancing

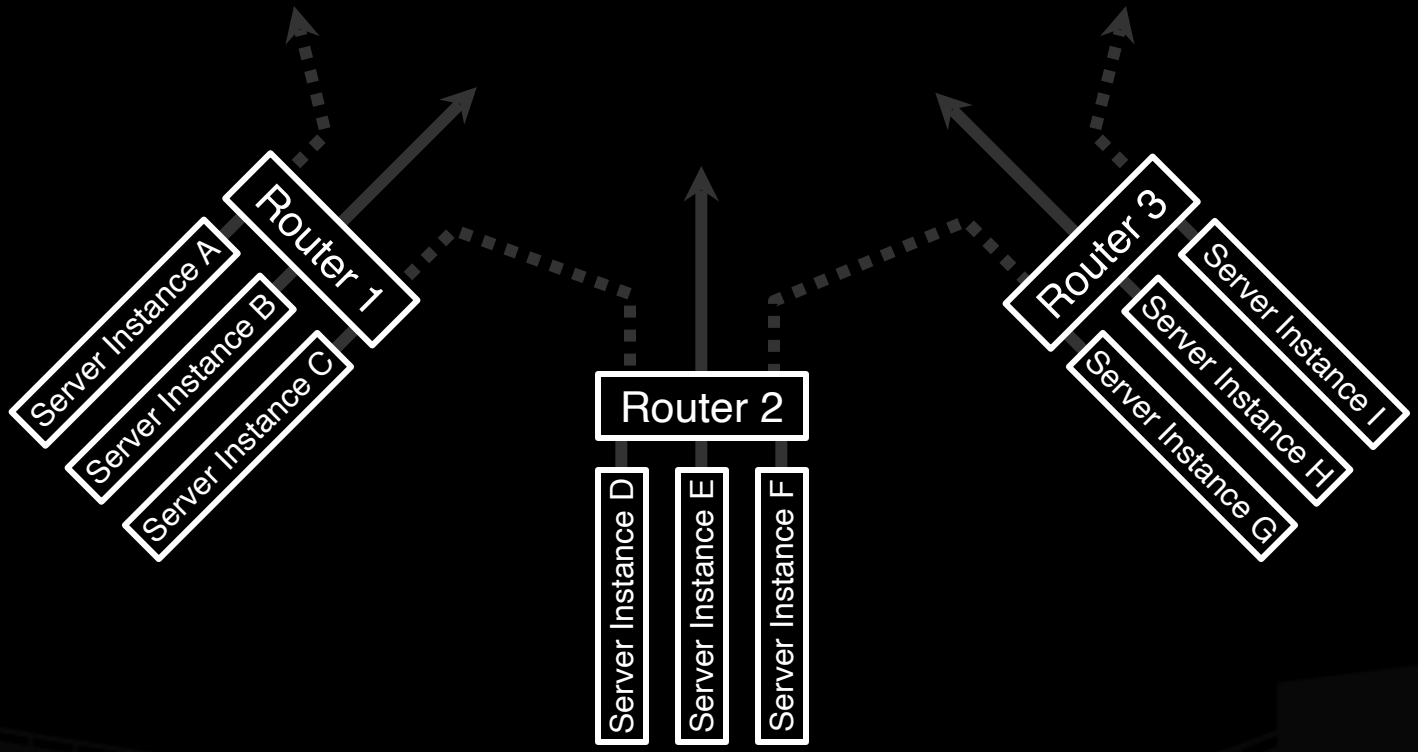
PCH
Packet Clearing House

Building a Global Network of Clusters

- Once a cluster architecture has been established, additional clusters can be added to gain performance.
- Load distribution, fail-over between clusters, and content synchronization become the principal engineering concerns.

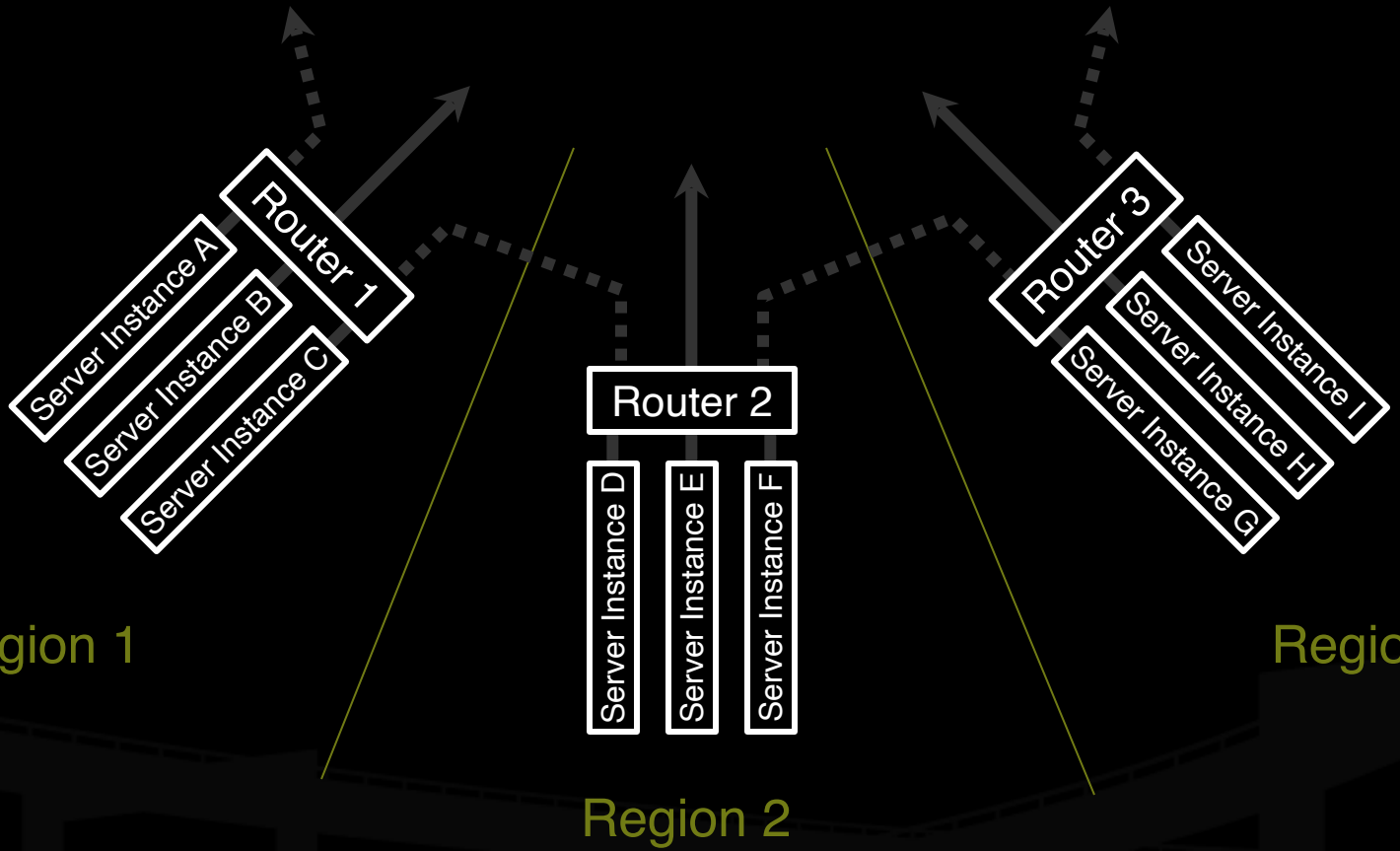
Packet Clearing House

Example



PCH
Packet Clearing House

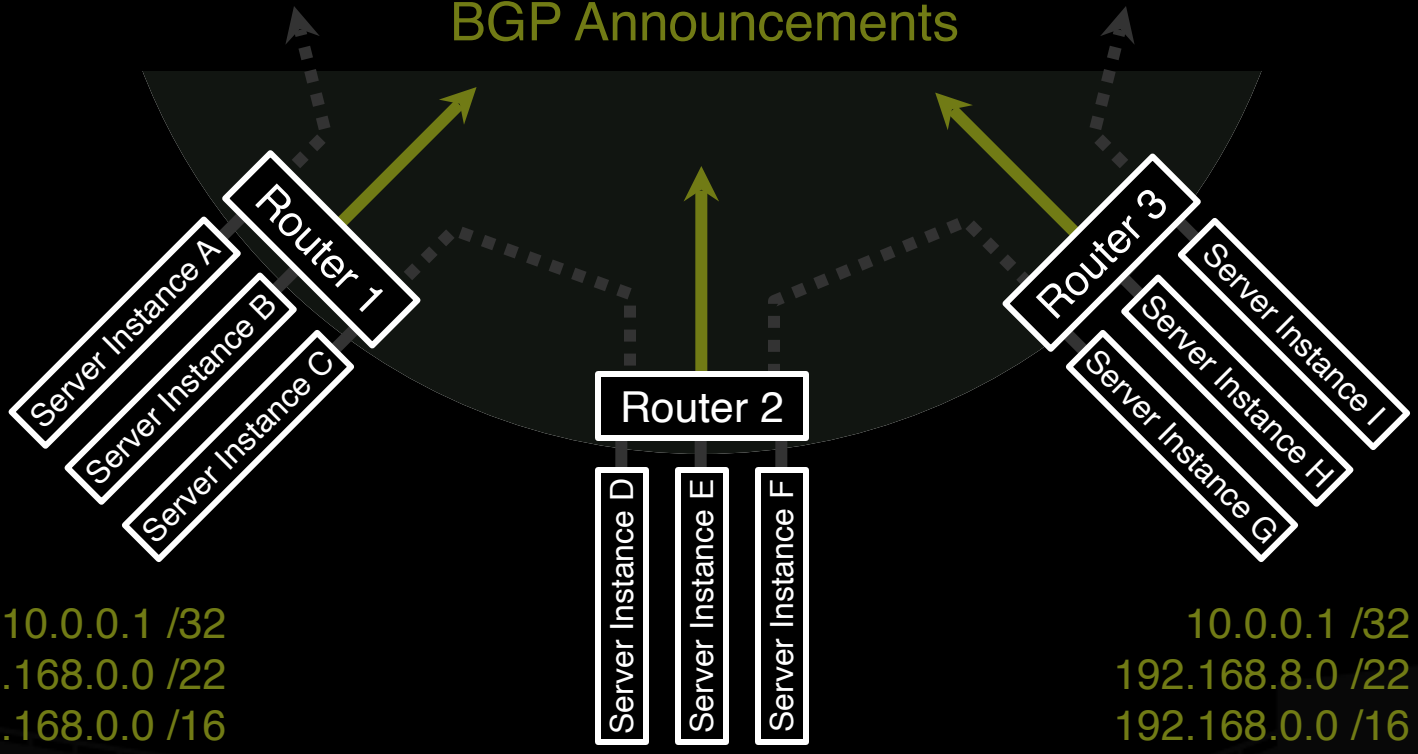
Example



Packet Clearing House

Example

BGP Announcements



10.0.0.1 /32
192.168.0.0 /22
192.168.0.0 /16

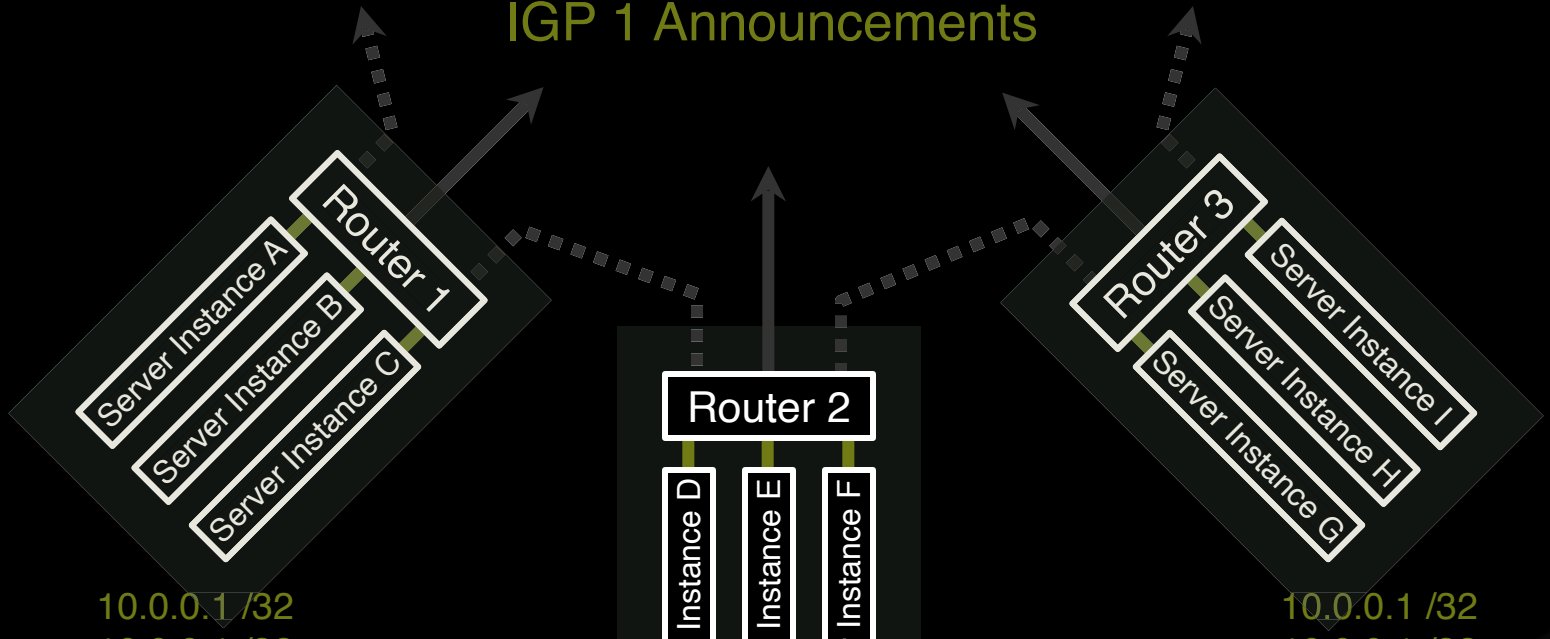
10.0.0.1 /32
192.168.4.0 /22
192.168.0.0 /16

10.0.0.1 /32
192.168.8.0 /22
192.168.0.0 /16

Packet Clearing House

Example

IGP 1 Announcements



10.0.0.1 /32
10.0.0.1 /32
10.0.0.1 /32
192.168.1.0 /30
192.168.2.0 /30
192.168.3.0 /30

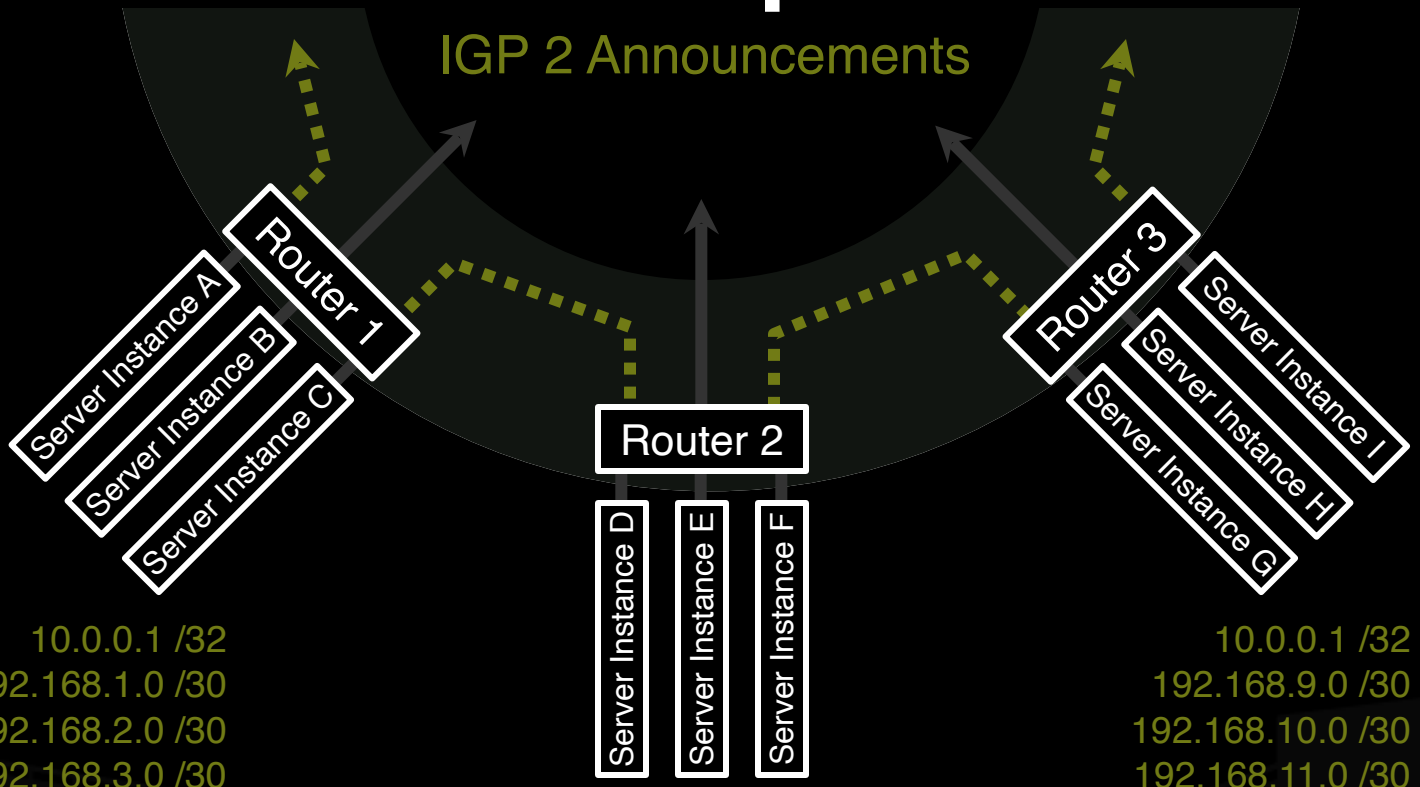
10.0.0.1 /32
10.0.0.1 /32
10.0.0.1 /32
192.168.5.0 /30
192.168.6.0 /30
192.168.7.0 /30

10.0.0.1 /32
10.0.0.1 /32
10.0.0.1 /32
192.168.9.0 /30
192.168.10.0 /30
192.168.11.0 /30

Packet Clearing House

Example

IGP 2 Announcements



10.0.0.1 /32
192.168.1.0 /30
192.168.2.0 /30
192.168.3.0 /30

10.0.0.1 /32
192.168.5.0 /30
192.168.6.0 /30
192.168.7.0 /30

10.0.0.1 /32
192.168.9.0 /30
192.168.10.0 /30
192.168.11.0 /30

PCH
Packet Clearing House

Performance-Tuning Anycast Networks

- Server deployment in anycast networks is always a tradeoff between absolute cost and efficiency.
- The network will perform best if servers are widely distributed, with higher density in and surrounding high demand areas.
- Lower initial cost sometimes leads implementers to compromise by deploying more servers in existing locations, which is less efficient.

Packet Clearing House

Caveats and Failure Modes

- DNS resolution fail-over
- Long-lived connection-oriented flows
- Identifying which server is giving an end-user trouble

PCH
Packet Clearing House

DNS Resolution Fail-Over

- In the event of poor performance from a server, DNS servers will fail over to the next server in a list.
- If both servers are in fact hosted in the same anycast cloud, the resolver will wind up talking to the same instance again.
- Best practices for anycast DNS server operations indicate a need for two separate overlapping clouds of anycast servers.

PCH
Packet Clearing House

Long-Lived Connection-Oriented Flows

- Long-lived flows, typically TCP file-transfers or interactive logins, may occasionally be more stable than the underlying Internet topology.
- If the underlying topology changes sufficiently during the life of an individual flow, packets could be redirected to a different server instance, which would not have proper TCP state, and would reset the connection.
- This is not a problem with web servers unless they're maintaining stateful per-session information about end-users, rather than embedding it in URLs or cookies.
- Web servers HTTP redirect to their unique address whenever they need to enter a stateful mode.
- Limited operational data shows underlying instability to be on the order of one flow per ten thousand per hour of duration.

PCH
Packet Clearing House

Identifying Problematic Server Instances

- Some protocols may not include an easy in-band method of identifying the server which persists beyond the duration of the connection.
- Traceroute always identifies the ***current*** server instance, but end-users may not even have traceroute.

Bill Woodcock
woody@pch.net

www.pch.net/documents/tutorials/anycast