
# Best Practices in IPv4 Anycast Routing

SANOG17

Colombo, Sri Lanka

Jonny Martin

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.
- Mentioned, although not described in detail, in numerous RFCs since time immemorial.
- It's been the basis for large-scale content-distribution networks since at least 1995.
- It's gradually taking over the core of the DNS infrastructure, as well as much of the periphery of the world wide web.

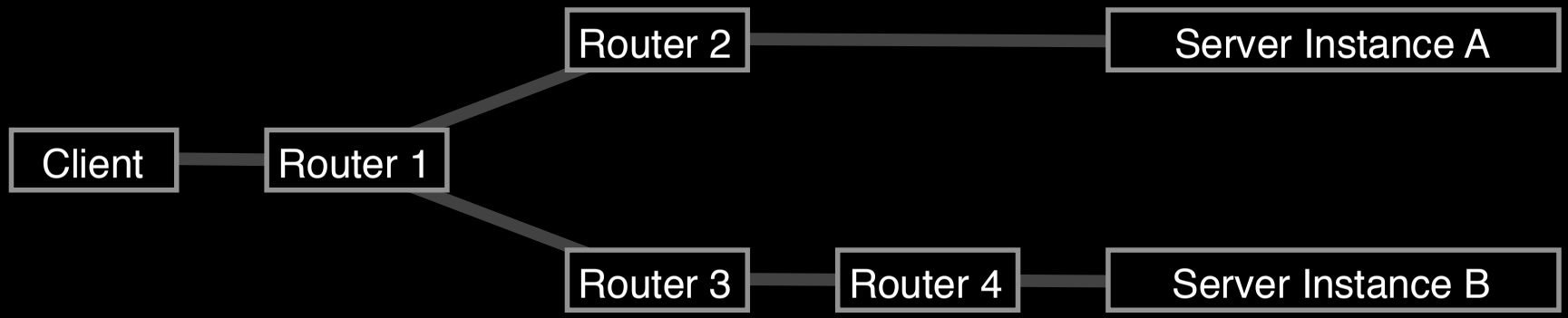
PCH
Packet Clearing House

# 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.

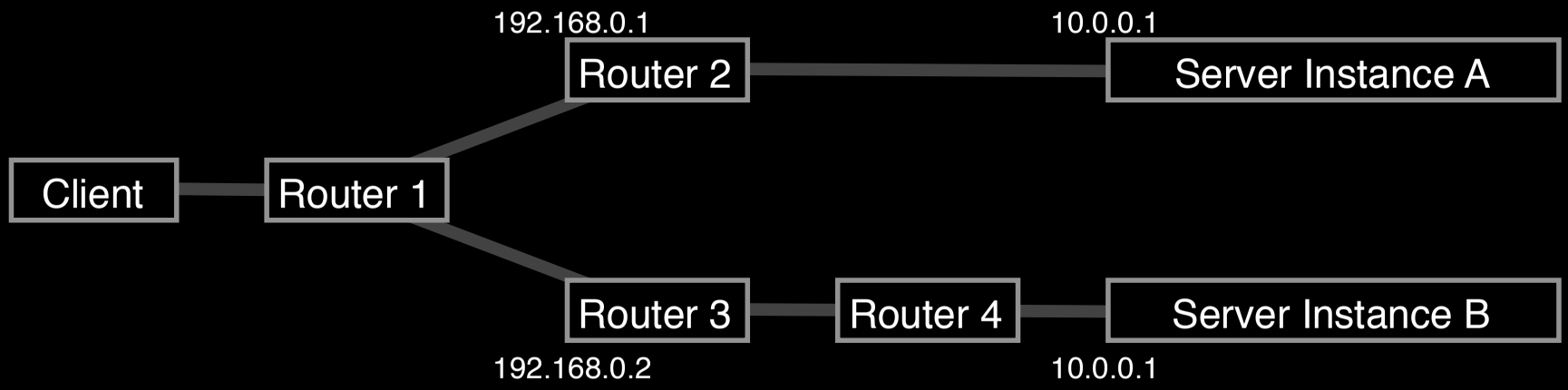
PCH
Packet Clearing House

# Example



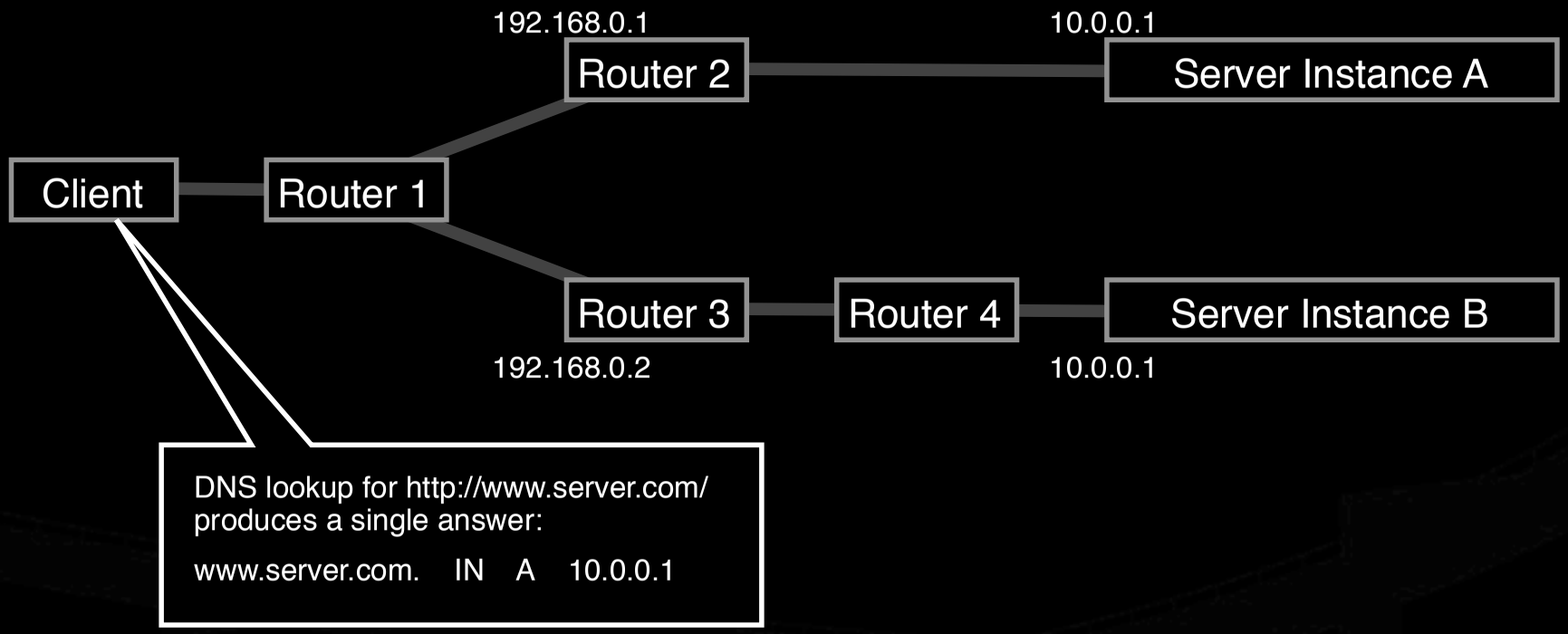
Packet Clearing House

# Example



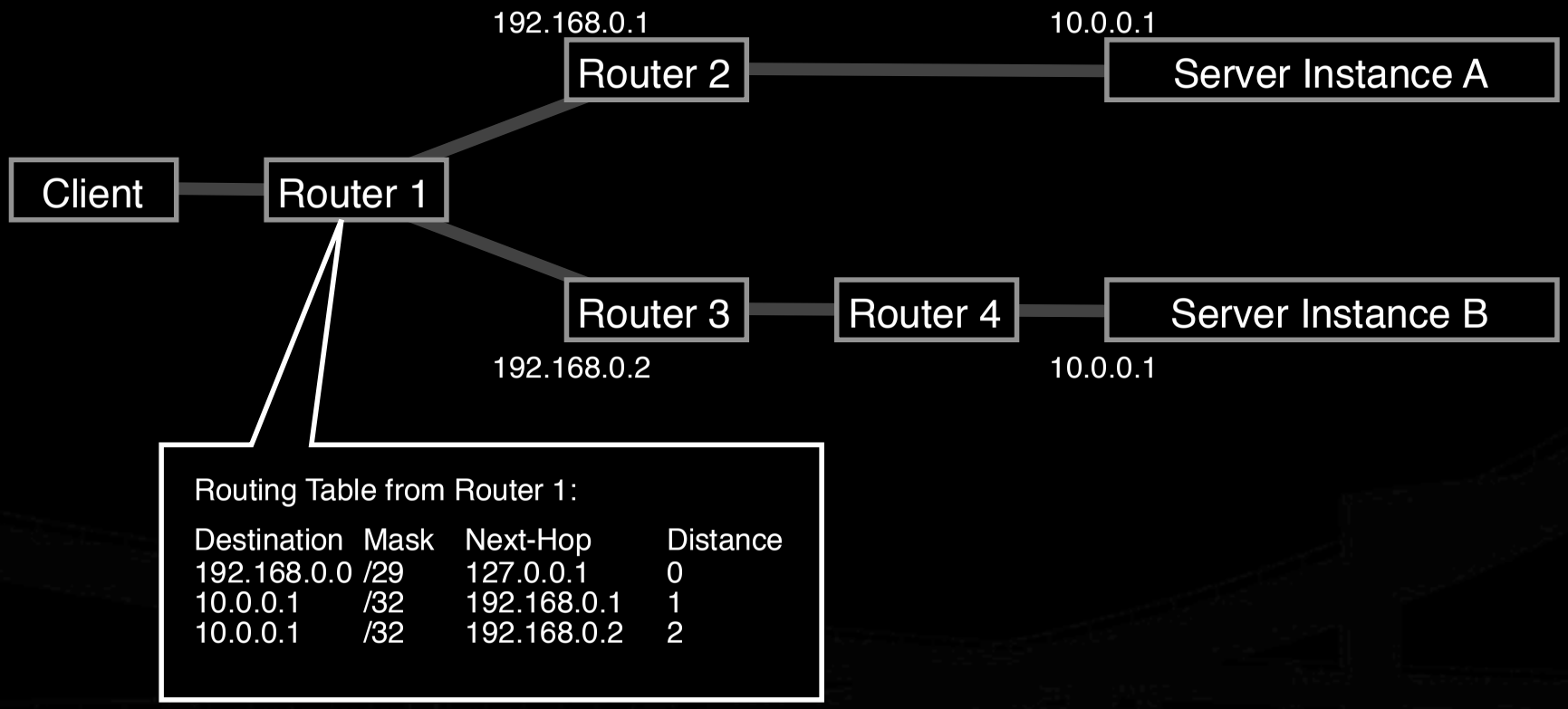
Packet Clearing House

# Example



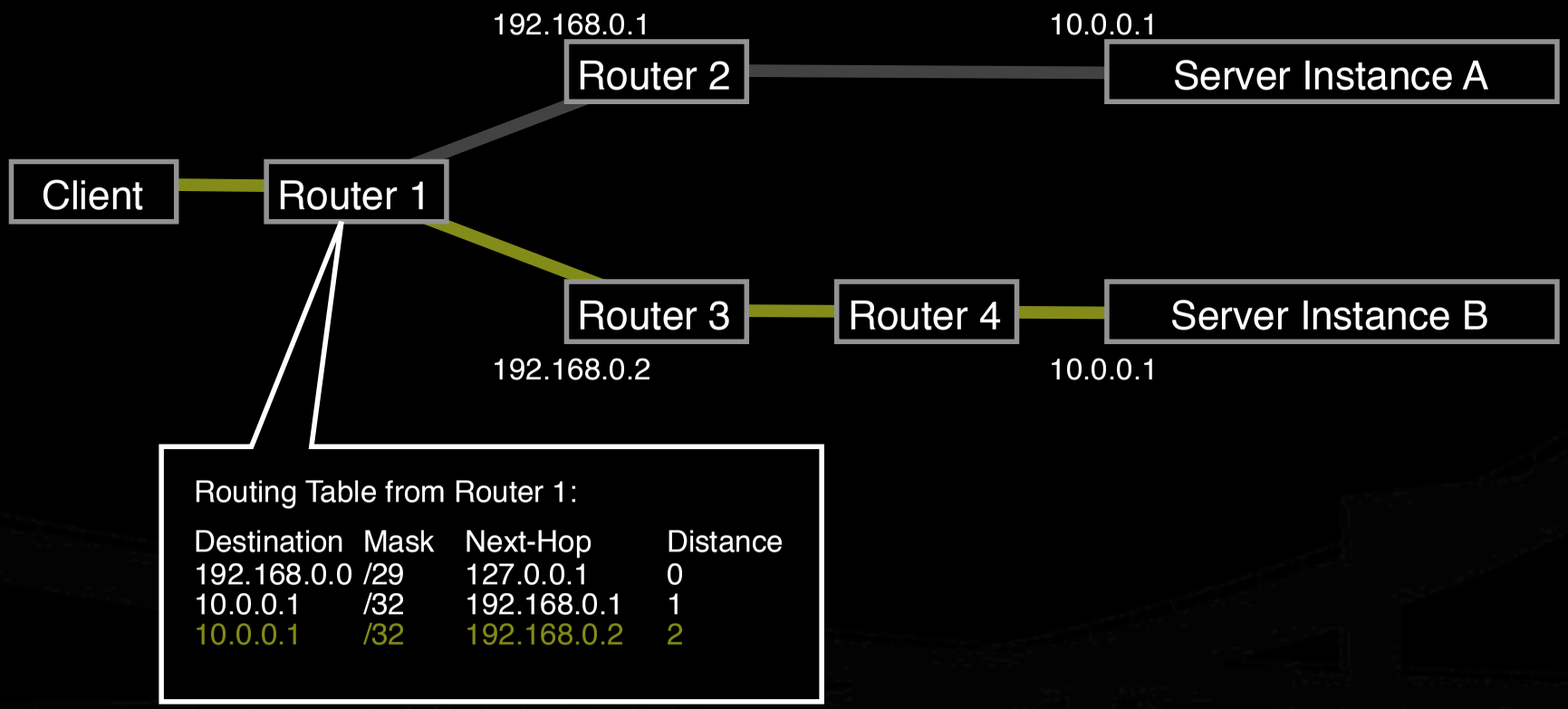
PCH
Packet Clearing House

# Example



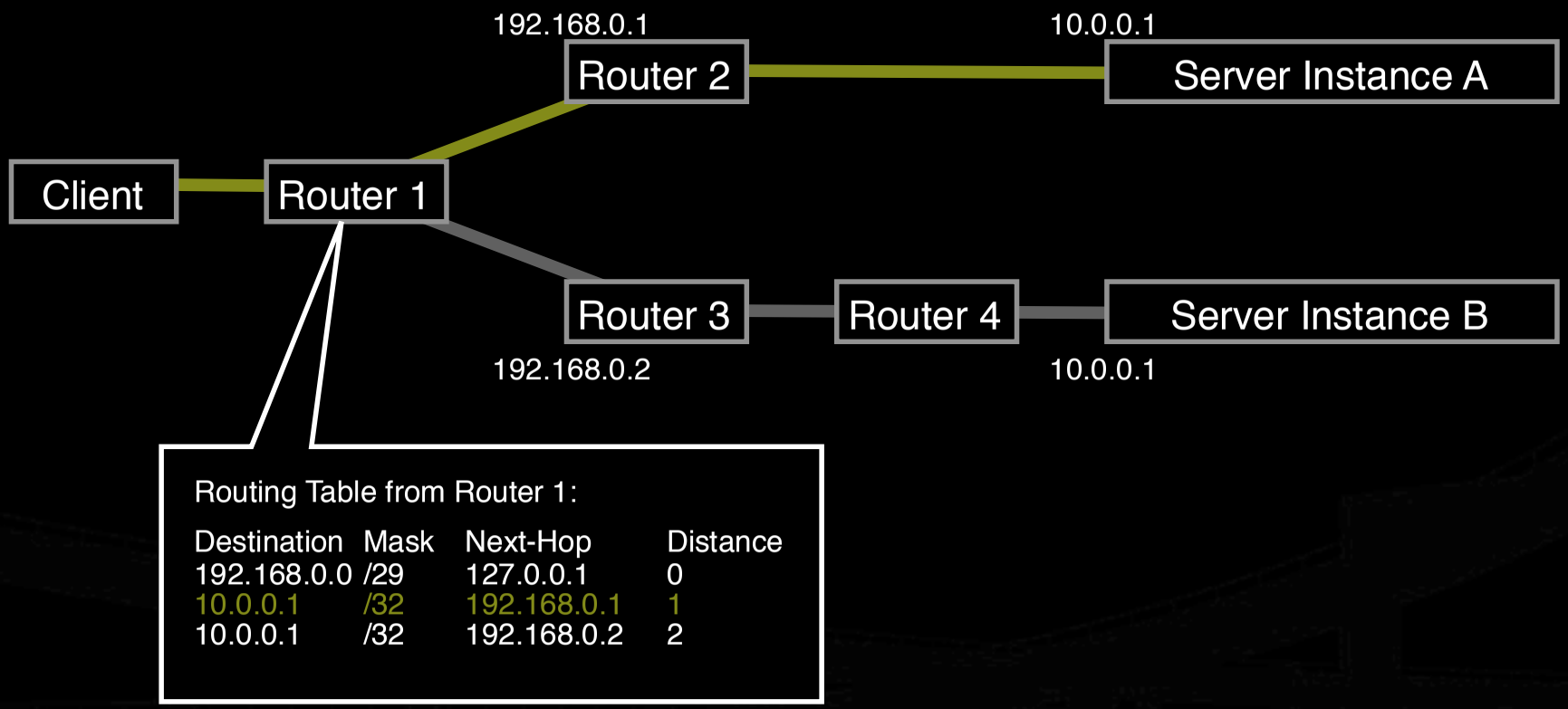
PCH
Packet Clearing House

# Example



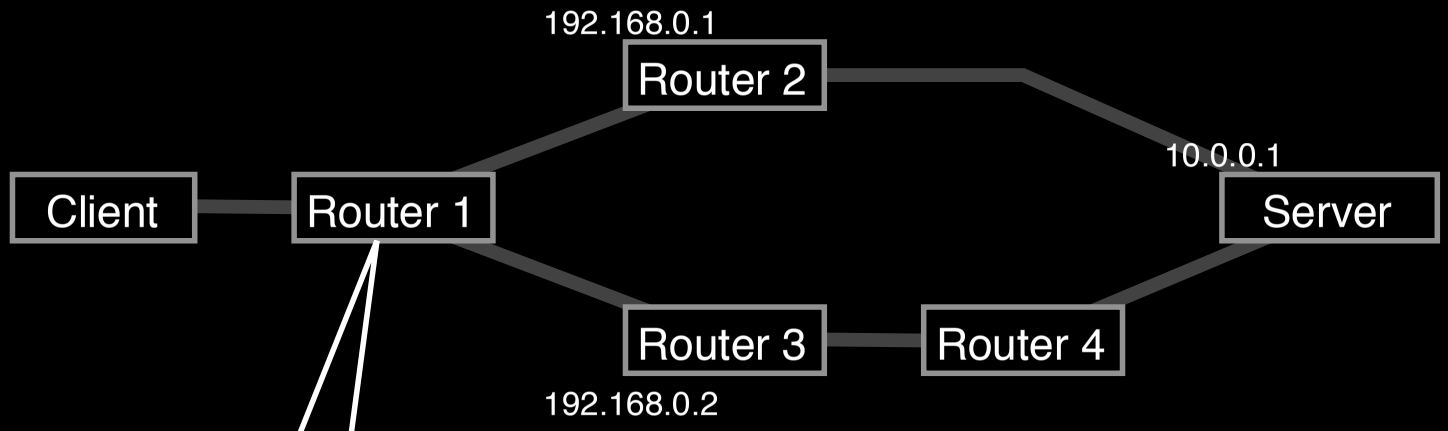
PCH
Packet Clearing House

# Example



# 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 |

PCH
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.
- F-root is a local anycast server cluster, for instance.

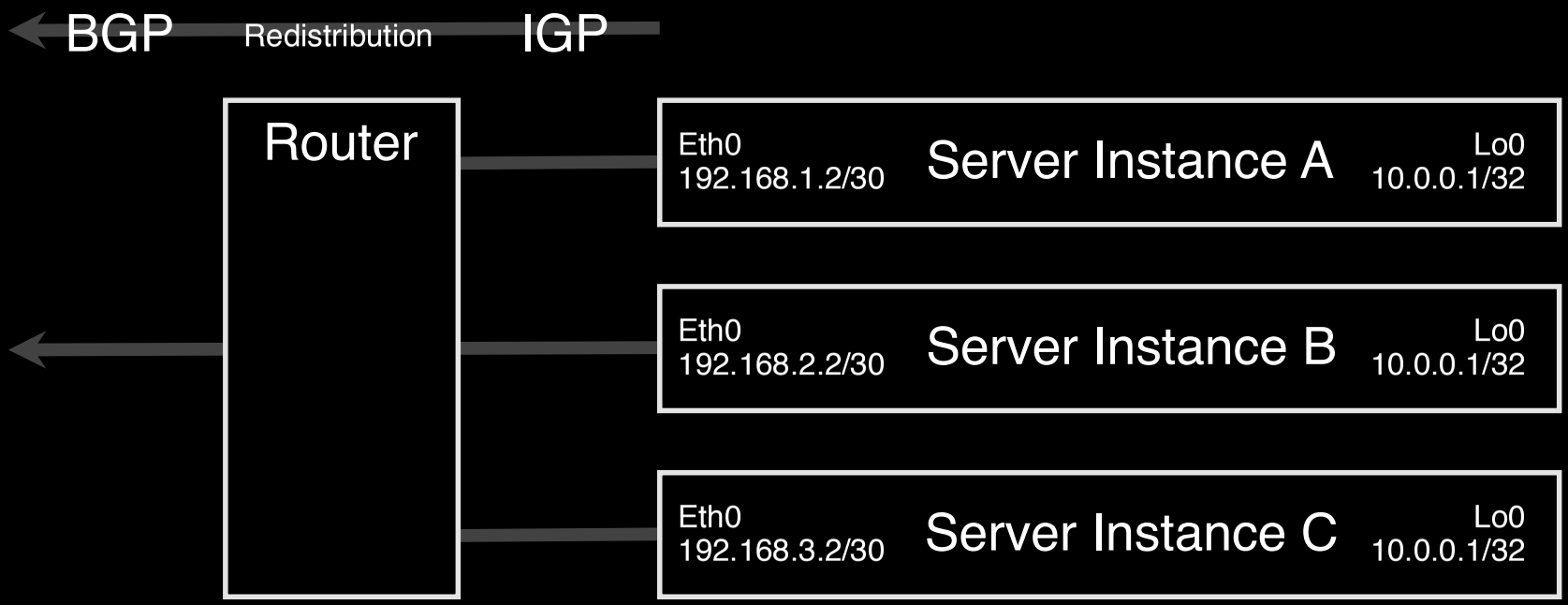
PCH
Packet Clearing House

# 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.

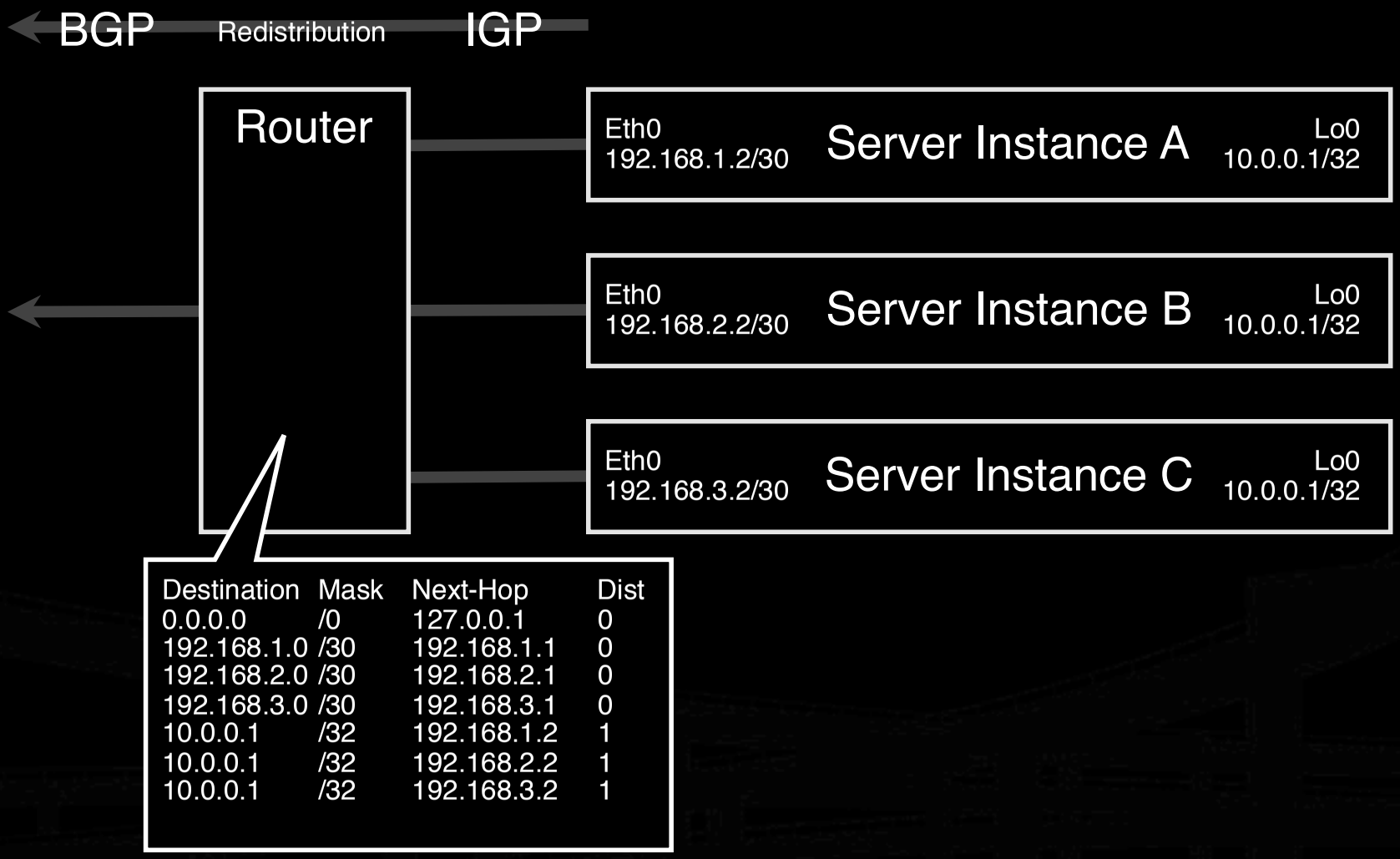
Packet Clearing House

# Example



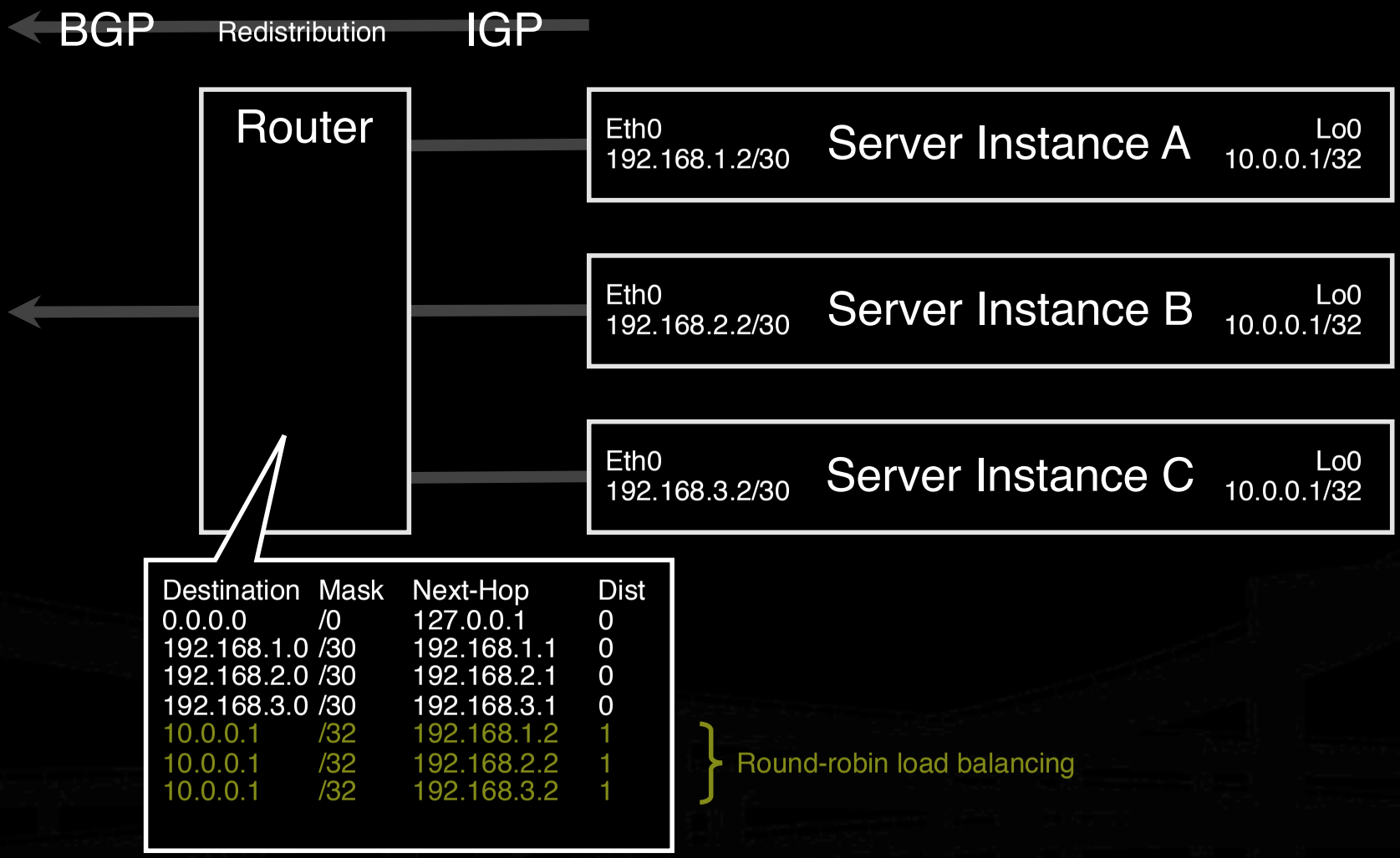
PCH
Packet Clearing House

# Example



PCH
Packet Clearing House

# Example



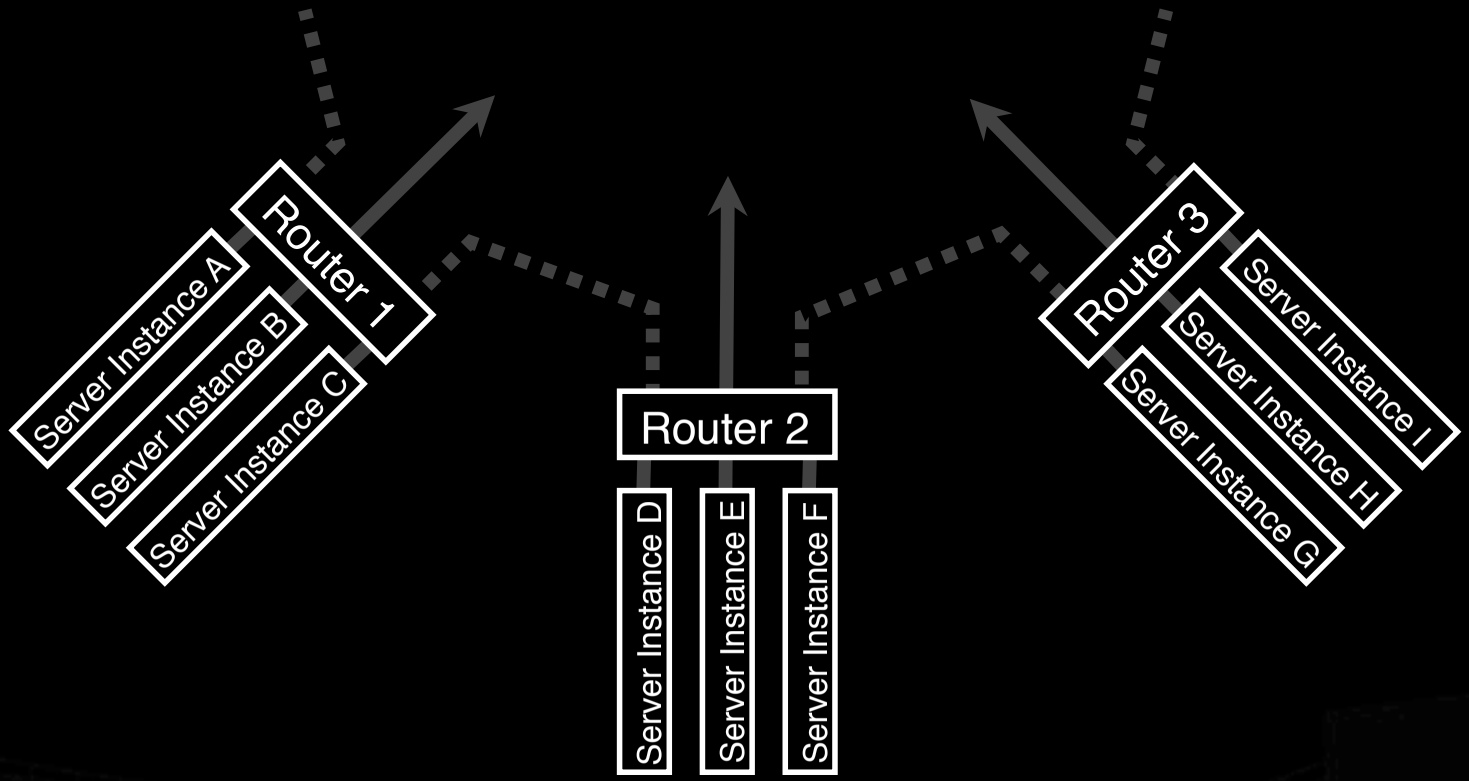
PCH
Packet Clearing House

# 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.

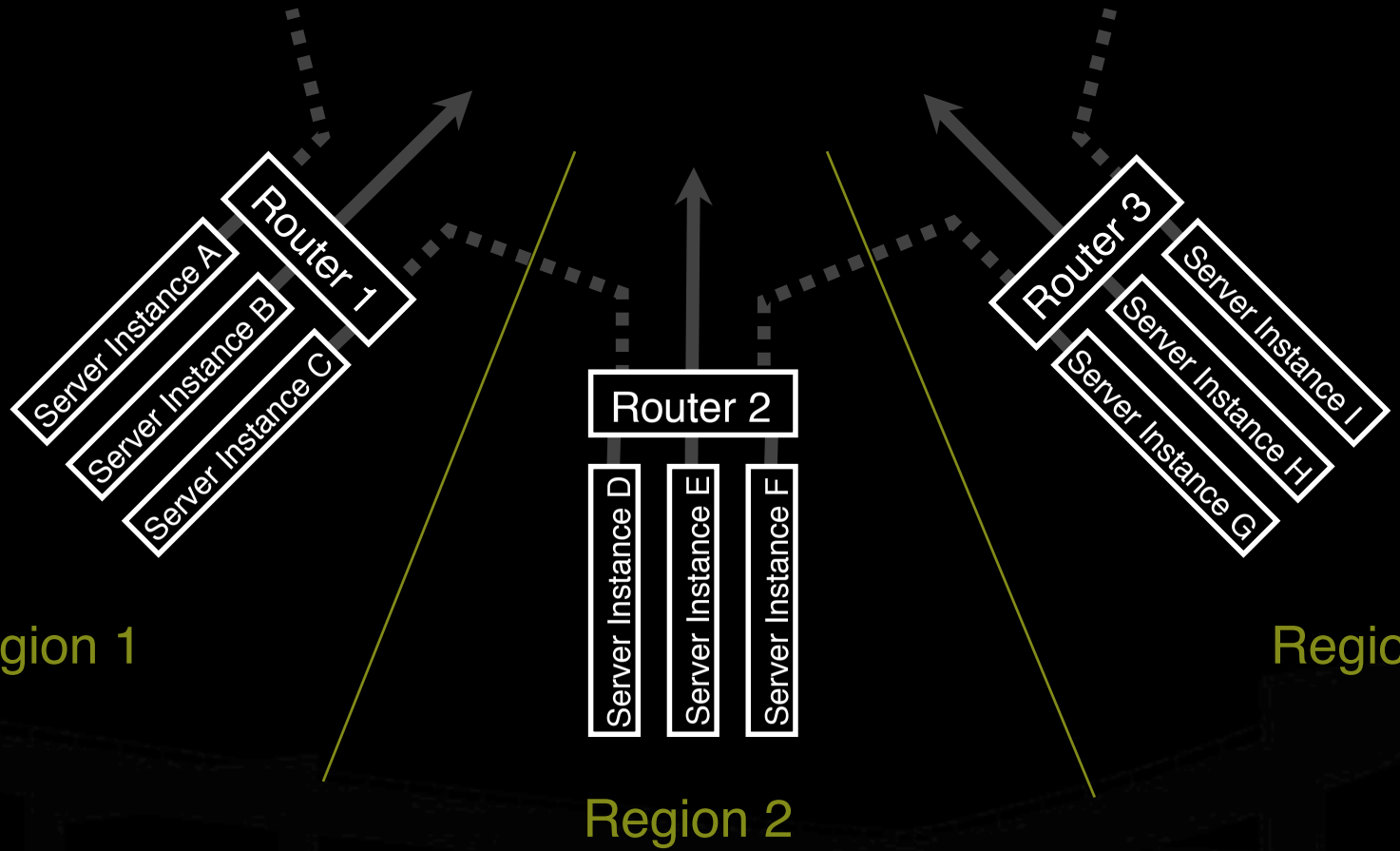
PCH
Packet Clearing House

# Example



PCH
Packet Clearing House

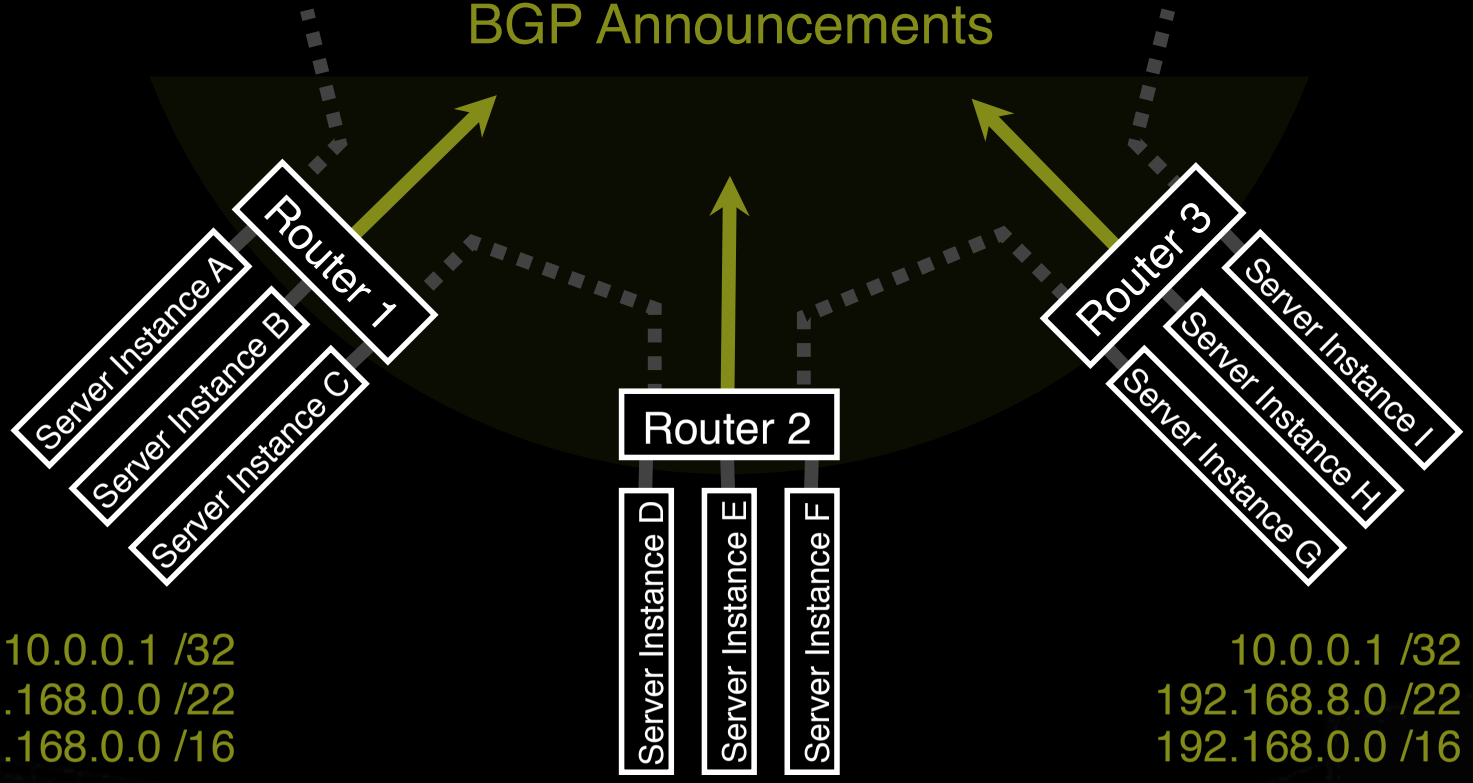
# Example



PCH
Packet Clearing House

# Example

## BGP Announcements




10.0.0.1 /32
192.168.0.0 /22
192.168.0.0 /16

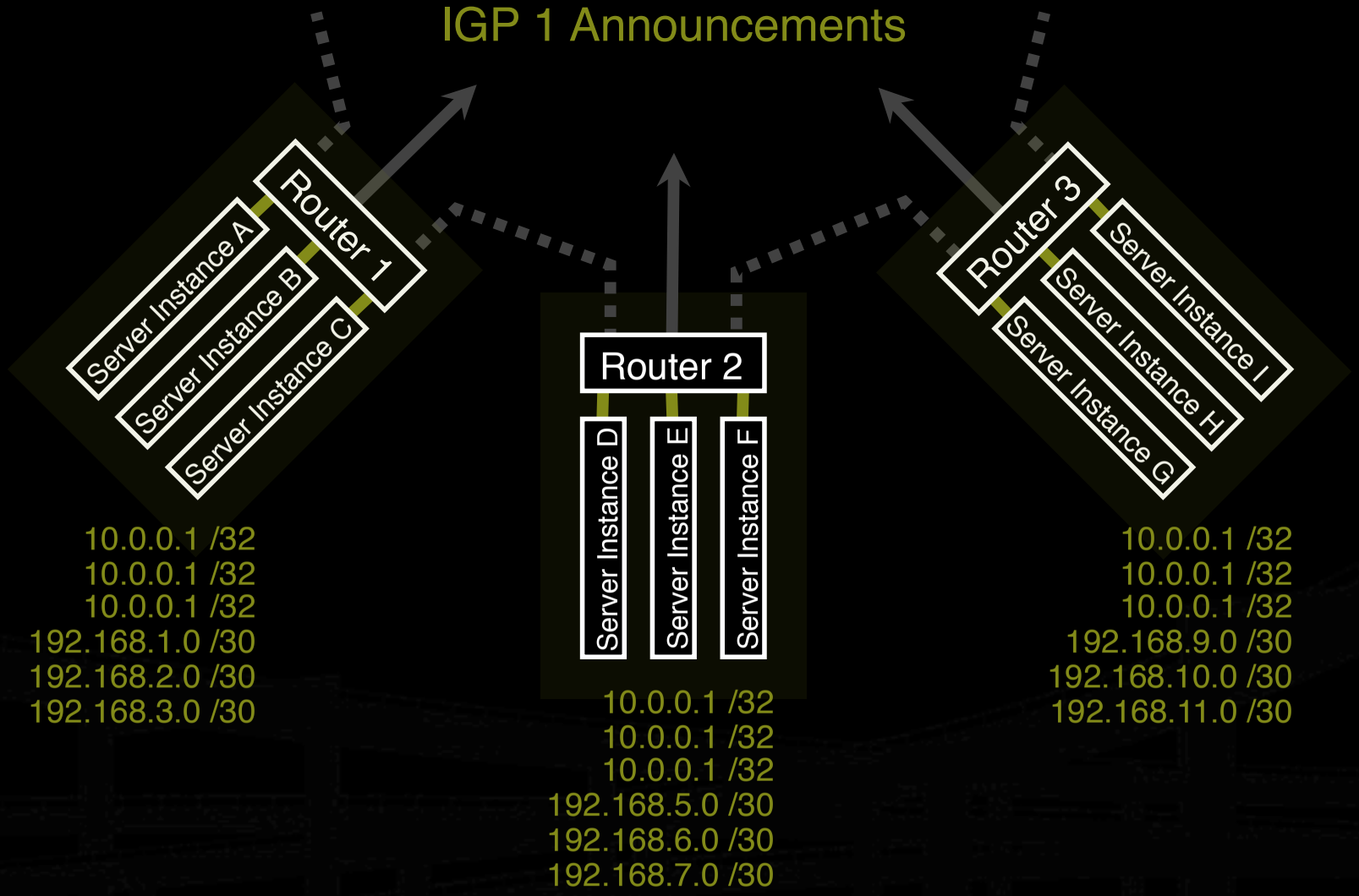
10.0.0.1 /32
192.168.4.0 /22
192.168.0.0 /16

10.0.0.1 /32
192.168.8.0 /22
192.168.0.0 /16

Packet Clearing House

# Example

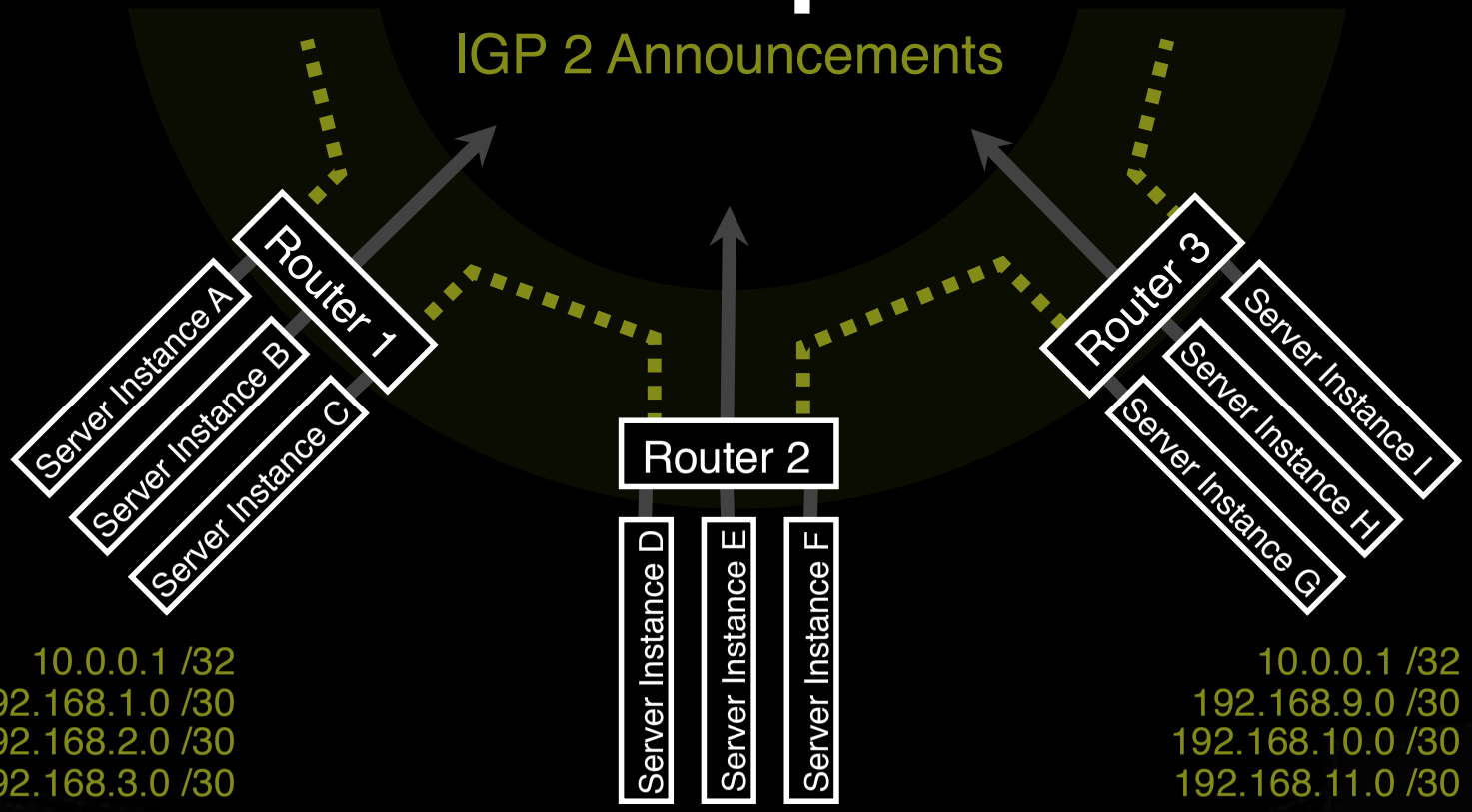
## IGP 1 Announcements



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

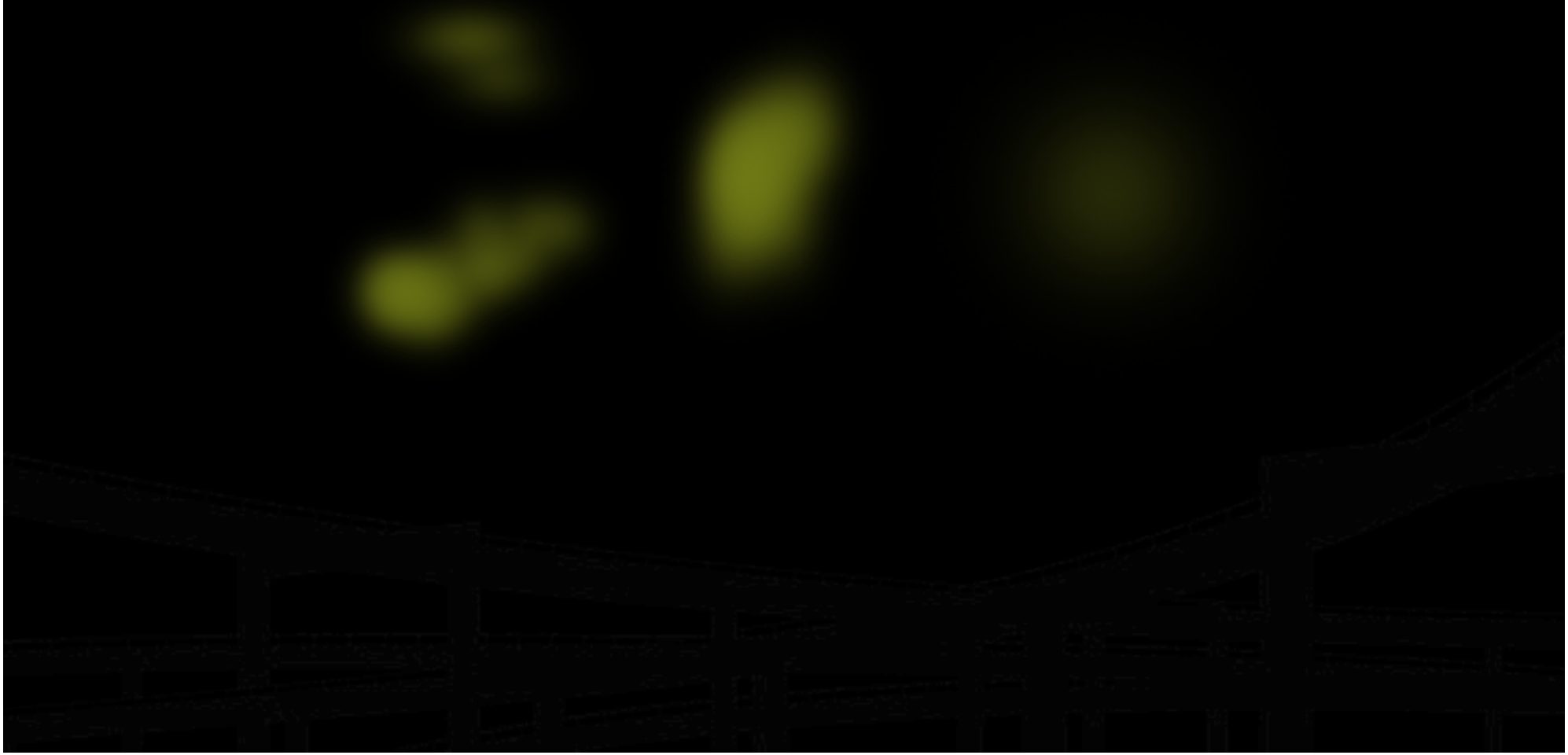
# Performance-Tuning Anycast Networks

- 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.

PCH
Packet Clearing House

# Example

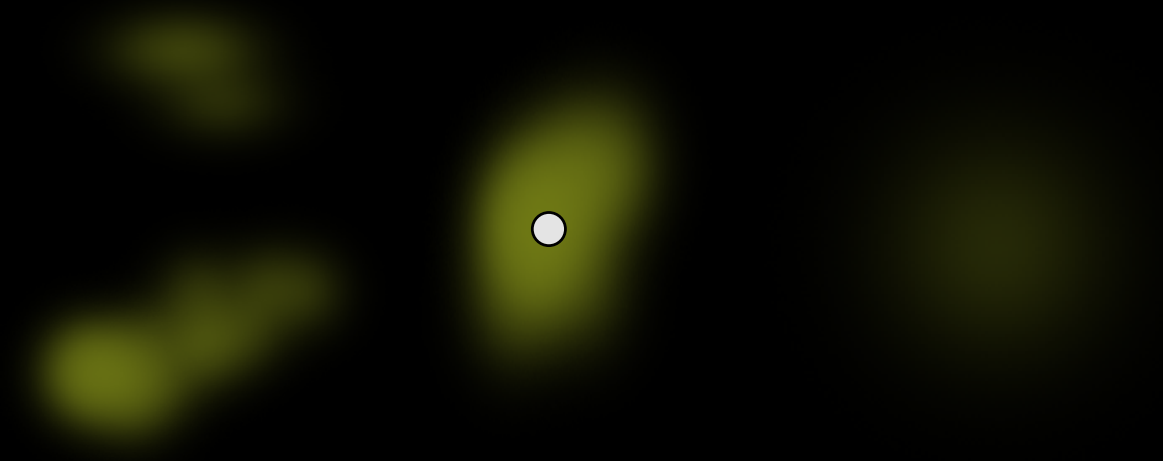
Geographic plot of user population density



PCH
Packet Clearing House

# Example

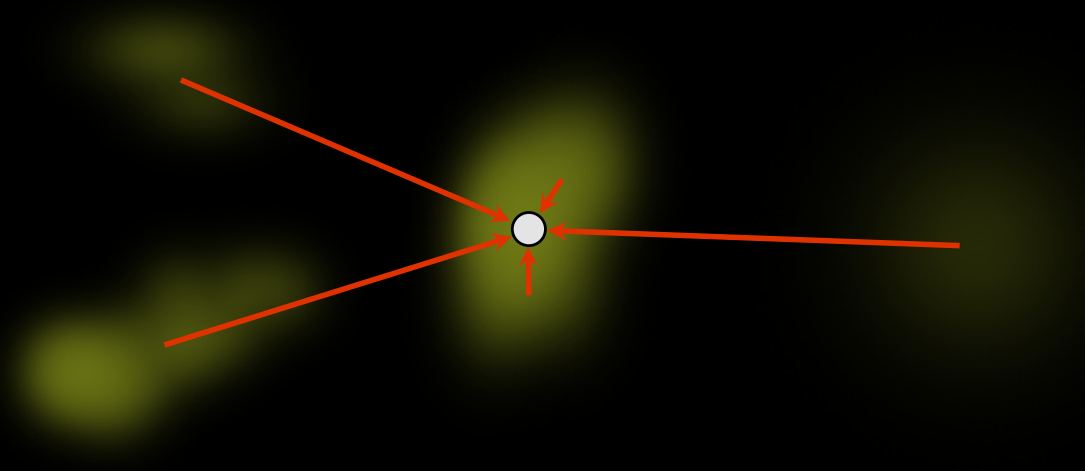
Geographic plot of user population density



Server deployment

# Example

Geographic plot of user population density



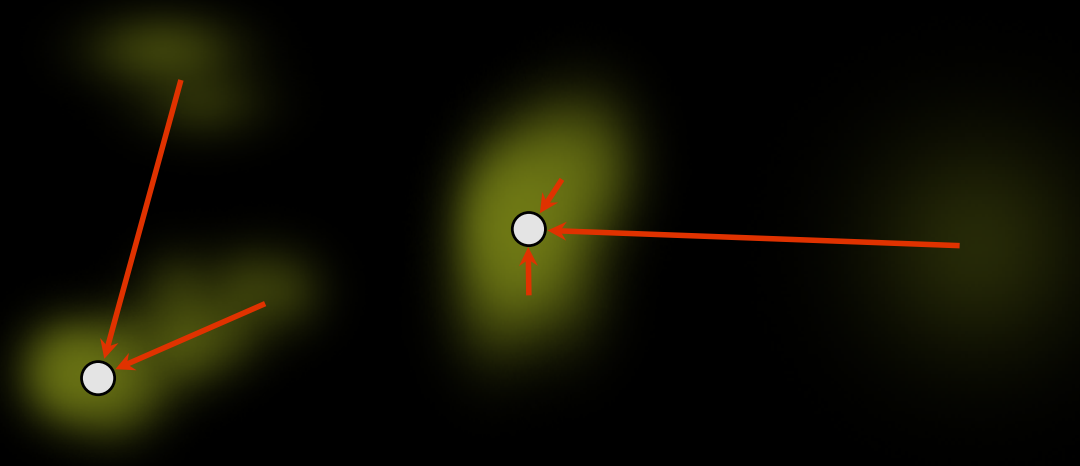
Server deployment

Traffic Flow

Packet Clearing House

# Example

Geographic plot of user population density



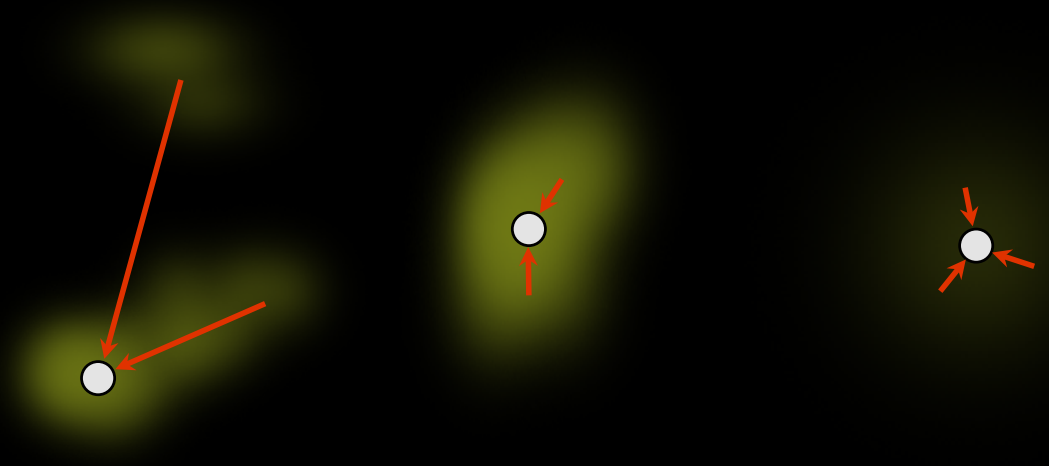
Server deployment

Traffic Flow

Packet Clearing House

# Example

Geographic plot of user population density



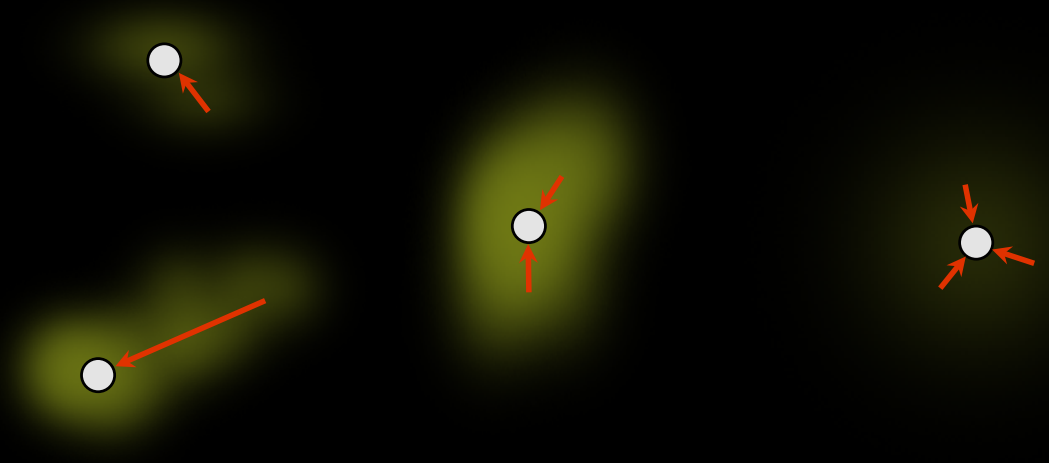
Server deployment

Traffic Flow

Packet Clearing House

# Example

Geographic plot of user population density



Server deployment

Traffic Flow

PCH
Packet Clearing House

# Example

Drawing traffic growth away from a hot-spot

Packet Clearing House

# Example

Drawing traffic growth away from a hot-spot

PCH
Packet Clearing House

# Example

Drawing traffic growth away from a hot-spot

Packet Clearing House

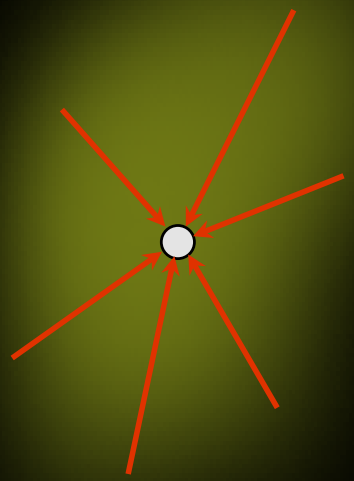
# Example

Drawing traffic growth away from a hot-spot

Packet Clearing House

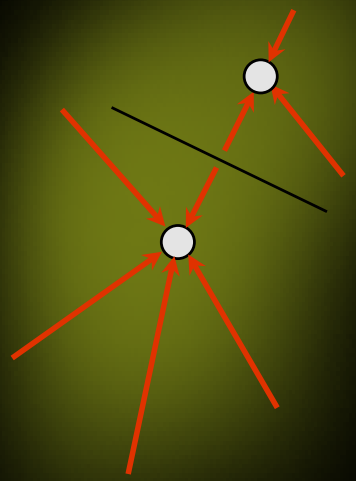
# Example

Drawing traffic growth away from a hot-spot



# Example

Drawing traffic growth away from a hot-spot

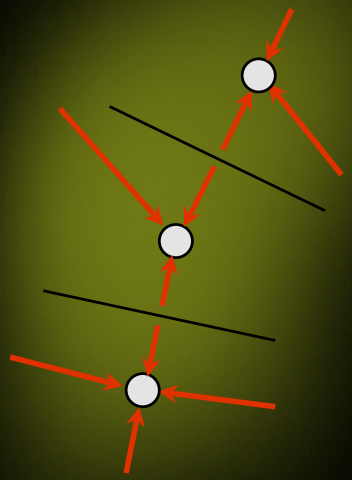


Topological watershed

PCH
Packet Clearing House

# Example

Drawing traffic growth away from a hot-spot



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.

PCH
Packet Clearing House

# A Security Ramification

- Anycast server clouds have the useful property of sinking DOS attacks at the instance nearest to the source of the attack, leaving all other instances unaffected.
- This is still of some utility even when DOS sources are widely distributed.


# Thanks, and Questions?

Copies of this presentation can be found
in Keynote, PDF, QuickTime and PowerPoint formats at:

**http:// www.pch.net / resources / tutorials / anycast**

Jonny Martin
Internet Analyst
Packet Clearing House
**jonny@pch.net**

PCH
Packet Clearing House

# Best Practices in DNS Service-Provision Architecture

SANOG17

Colombo, Sri Lanka

Jonny Martin

Packet Clearing House

PCH
Packet Clearing House

# It's all Anycast

Large ISPs have been running production anycast DNS for more than a decade.

Which is a very long time, in Internet years.

95% of the root nameservers are anycast.

The large gTLDs are anycast.

PCH
Packet Clearing House

# Reasons for Anycast

Transparent fail-over redundancy

Latency reduction

Load balancing

Attack mitigation

Configuration simplicity (for end users)
or lack of IP addresses (for the root)

PCH
Packet Clearing House

# No Free Lunch

The two largest benefits, fail-over redundancy and latency reduction, both require a bit of work to operate as you'd wish.

PCH
Packet Clearing House

# Fail-Over Redundancy

DNS resolvers have their own fail-over mechanism, which works... um... okay.

Anycast is a very large hammer.

Good deployments allow these two mechanisms to reinforce each other, rather than allowing anycast to foil the resolvers' fail-over mechanism.

PCH
Packet Clearing House

# Resolvers' Fail-Over Mechanism

DNS resolvers like those in your computers, and in referring authoritative servers, can and often do maintain a ***list*** of nameservers to which they'll send queries.

Resolver implementations differ in how they use that list, but basically, when a server doesn't reply in a timely fashion, resolvers will try another server from the list.

PCH
Packet Clearing House

# Anycast Fail-Over Mechanism

Anycast is simply layer-3 routing.

A resolver's query will be routed to the topologically nearest instance of the anycast server visible in the routing table.

Anycast servers govern their own visibility.

Latency depends upon the delays imposed by that topologically short path.

PCH
Packet Clearing House

# Conflict Between These Mechanisms

Resolvers measure by latency.

Anycast measures by hop-count.

They don’t necessarily yield the same answer.

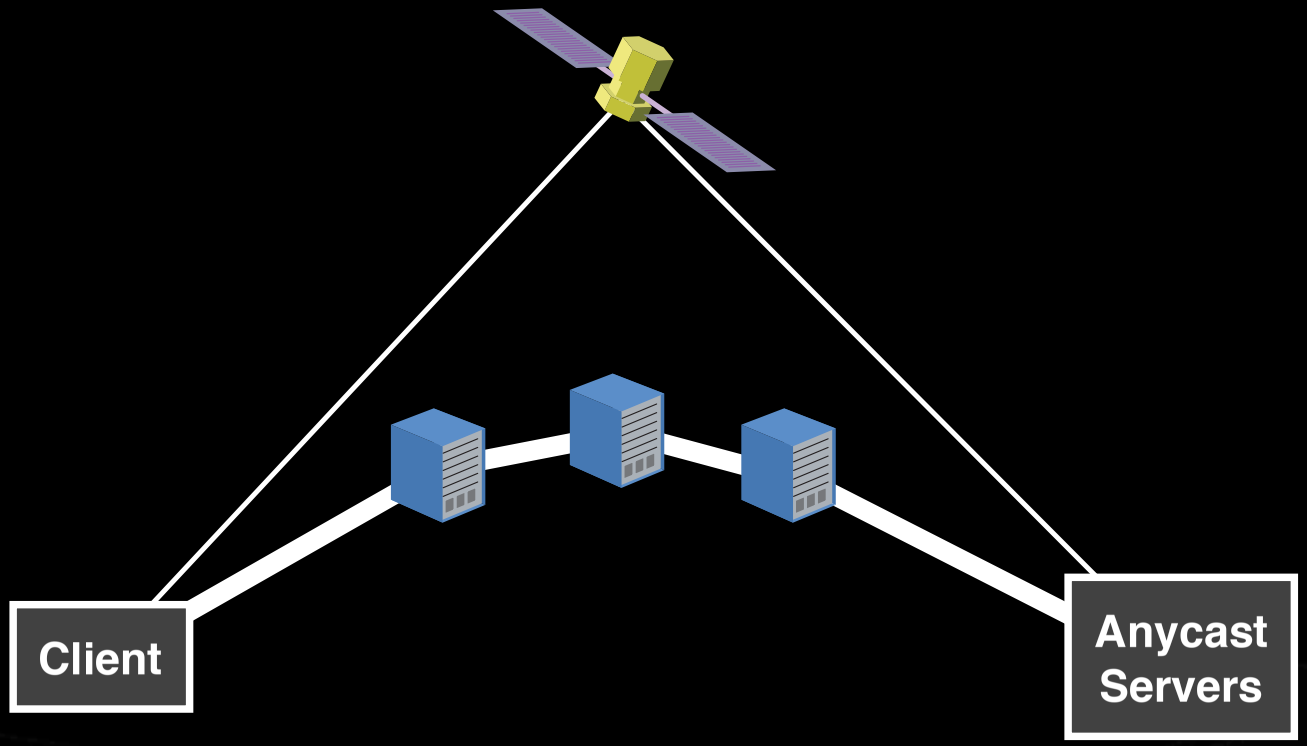
Anycast always trumps resolvers, if it’s allowed to.

Neither the DNS service provider nor the user are likely to care about hop-count.

Both care a great deal about latency.

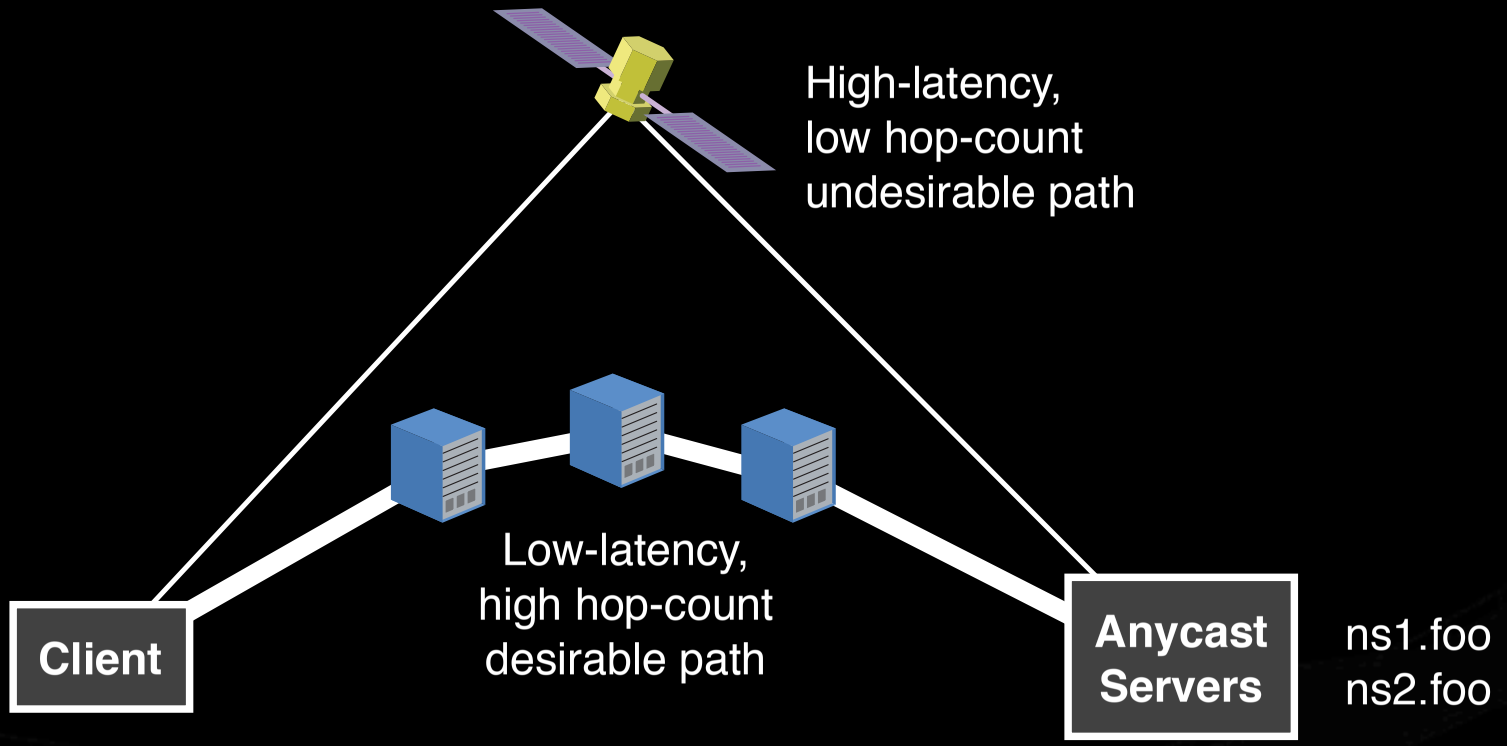
PCH
Packet Clearing House

# How The Conflict Plays Out



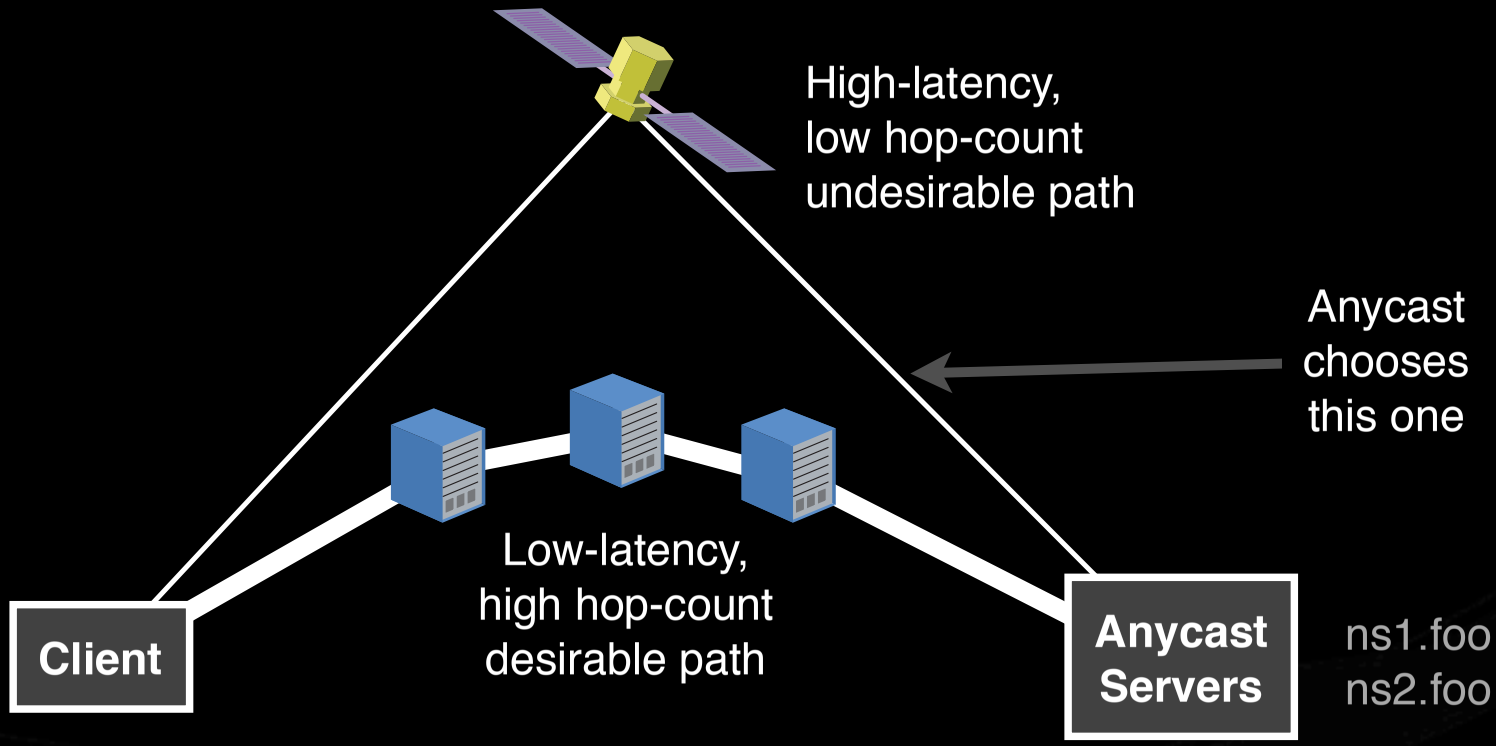
PCH
Packet Clearing House

# How The Conflict Plays Out



PCH
Packet Clearing House

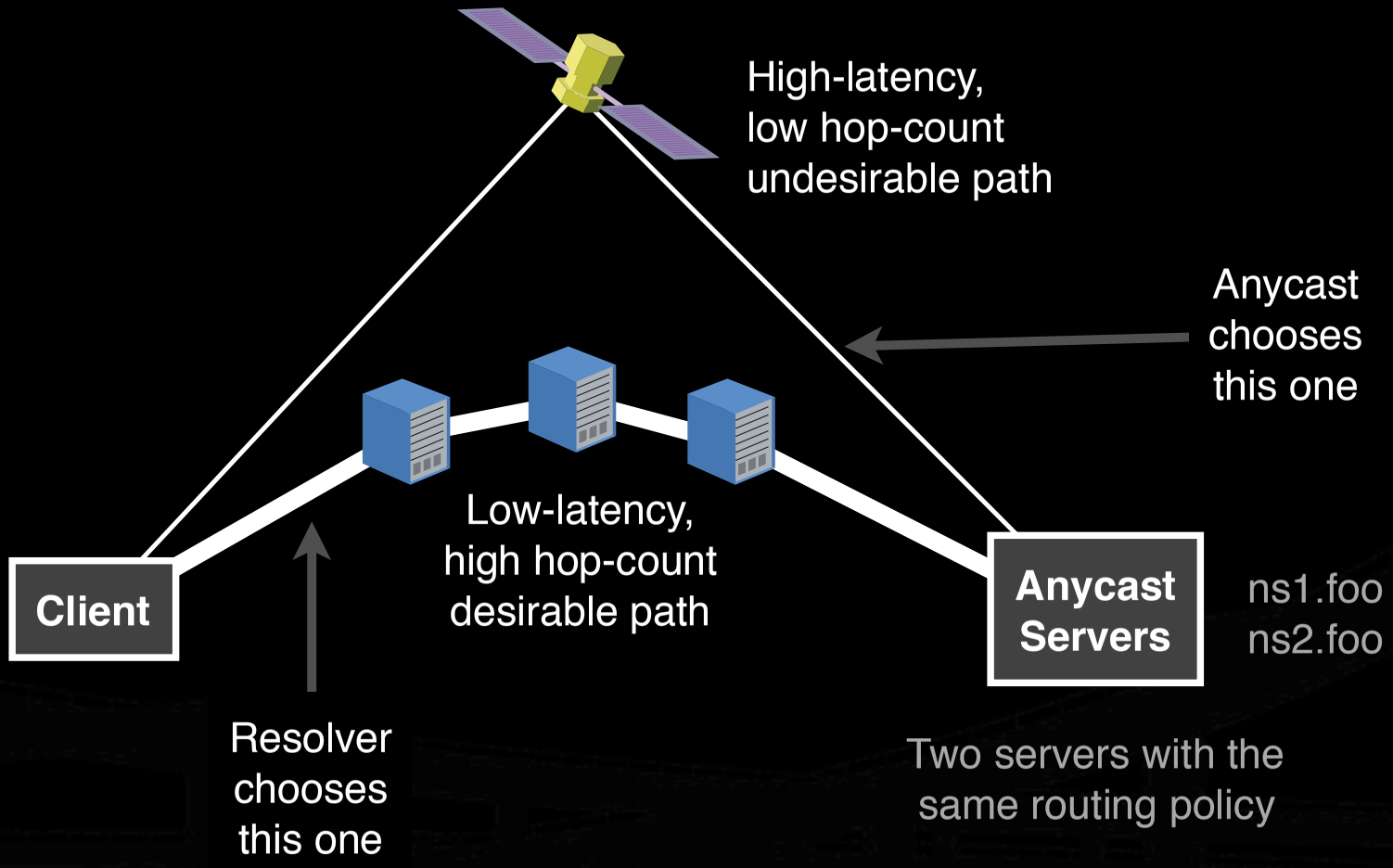
# How The Conflict Plays Out




Two servers with the
same routing policy

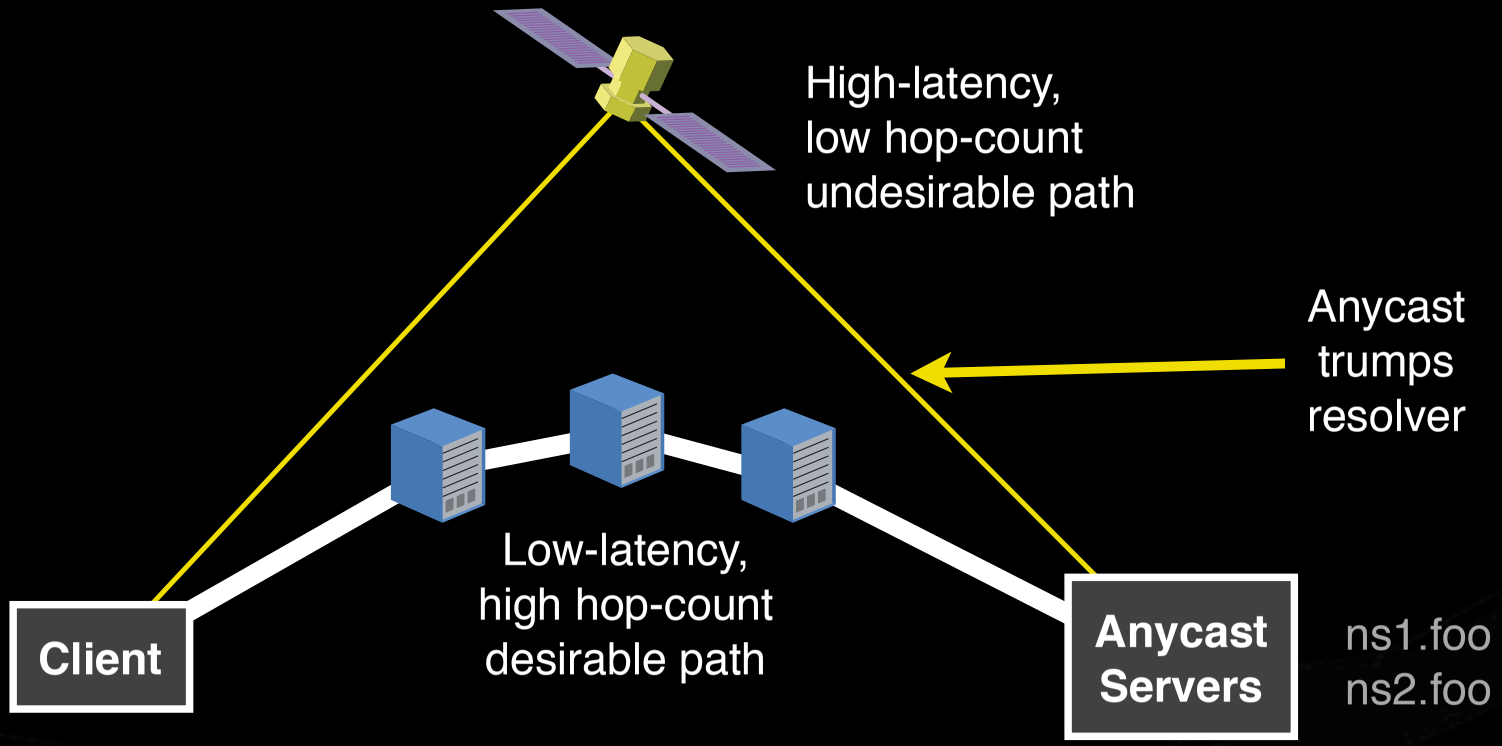
PCH
Packet Clearing House

# How The Conflict Plays Out



PCH
Packet Clearing House

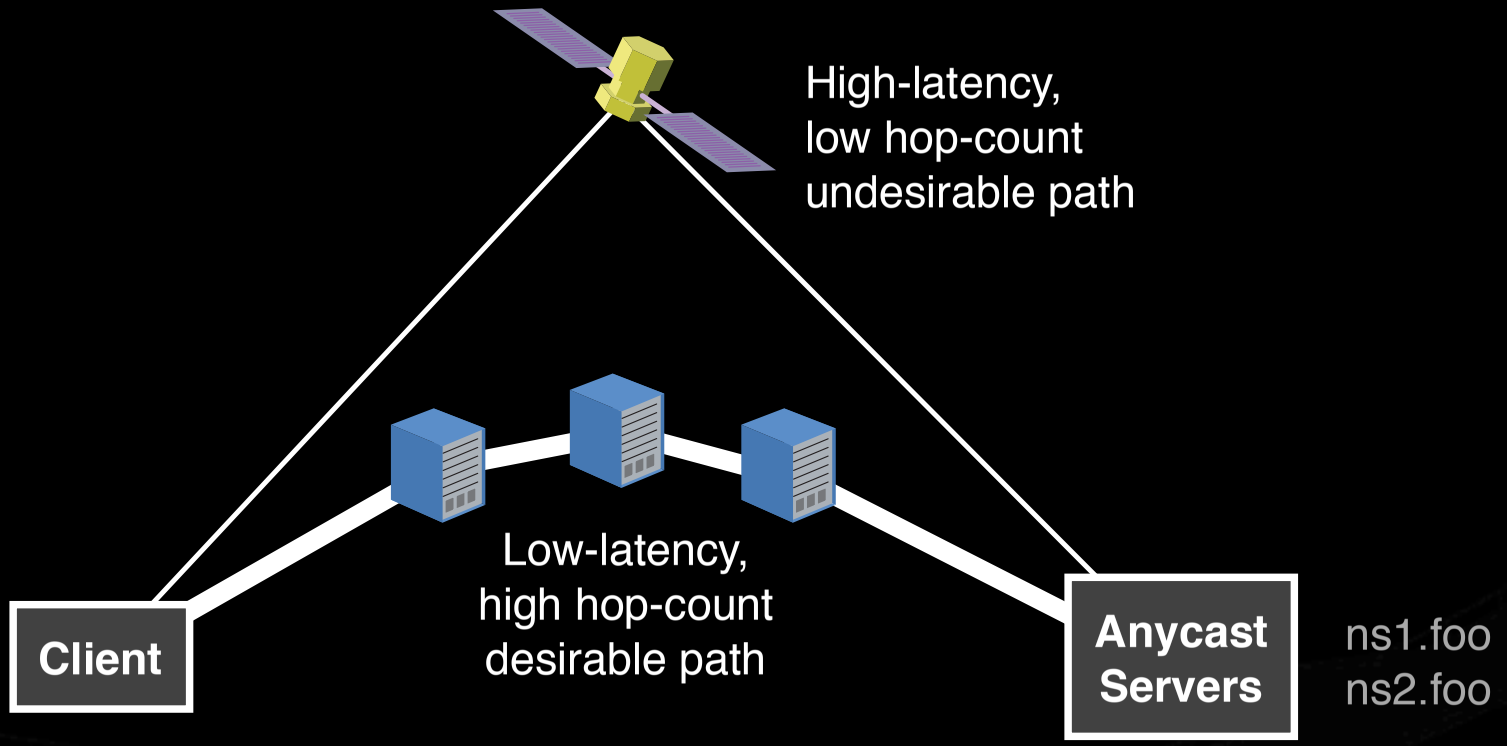
# How The Conflict Plays Out




Two servers with the
same routing policy

PCH
Packet Clearing House

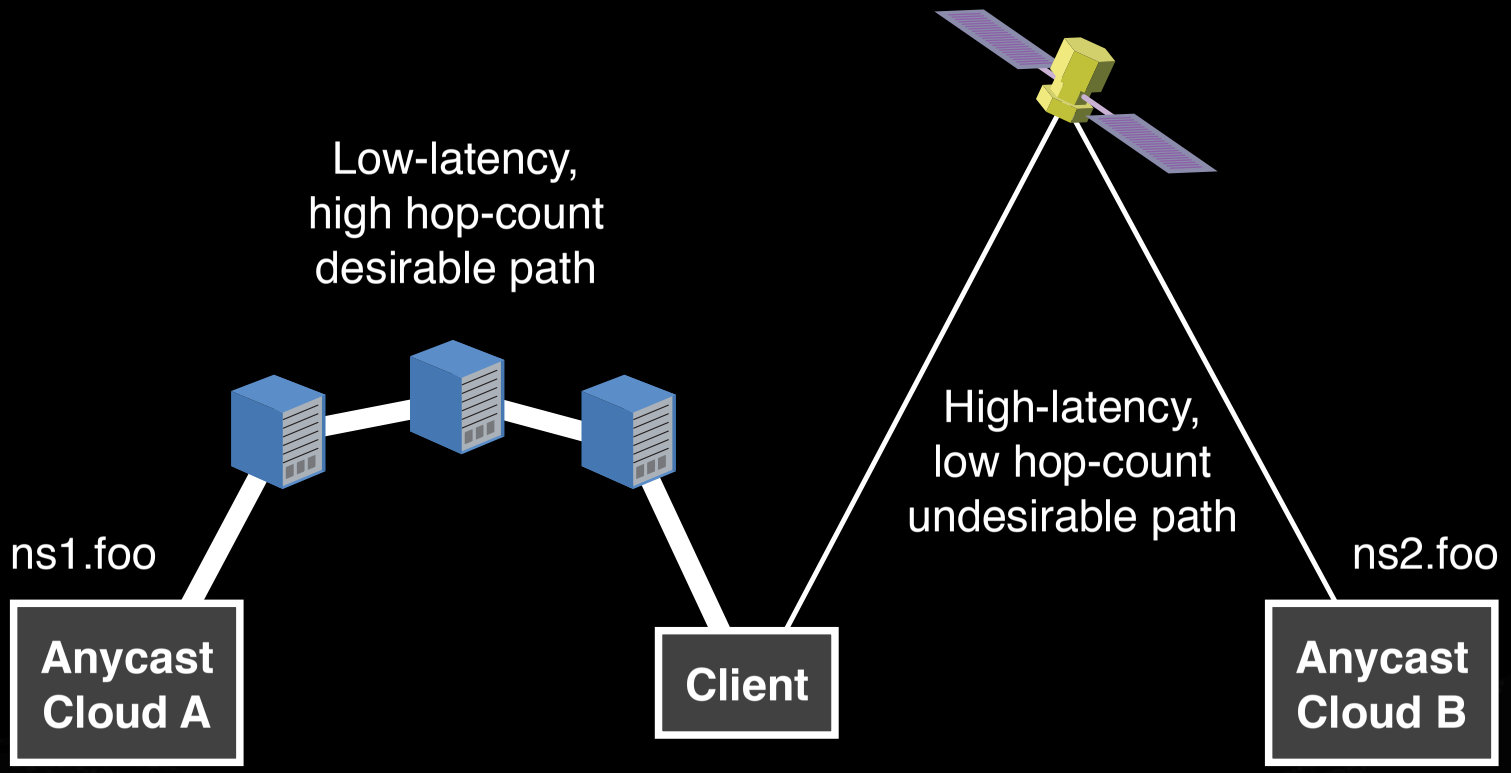
# Resolve the Conflict




The resolver uses different IP addresses for its fail-over mechanism, while anycast uses the same IP addresses.

PCH
Packet Clearing House

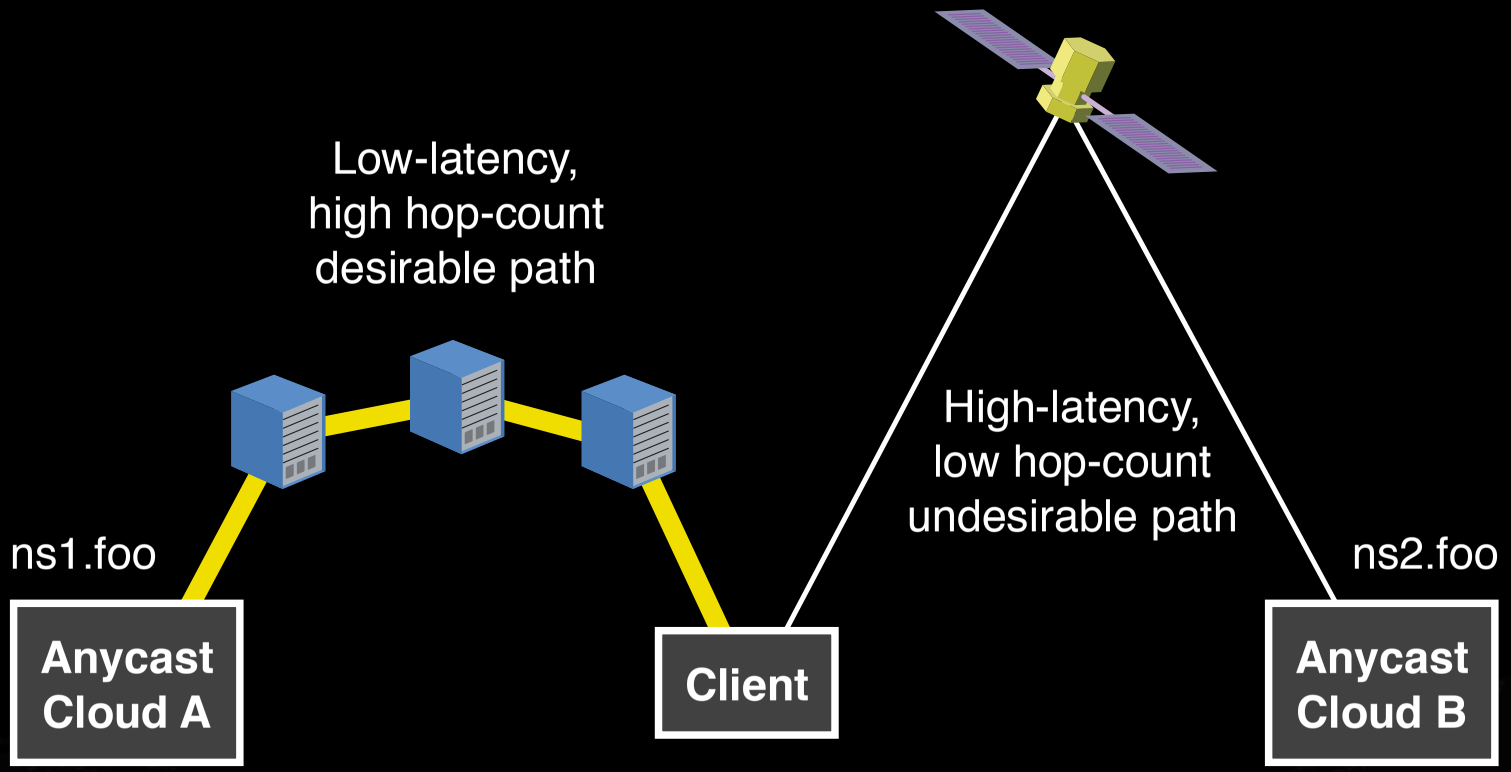
# Resolve the Conflict




Split the anycast deployment into “clouds” of locations, each cloud using a different IP address and different routing policies.

PCH
Packet Clearing House

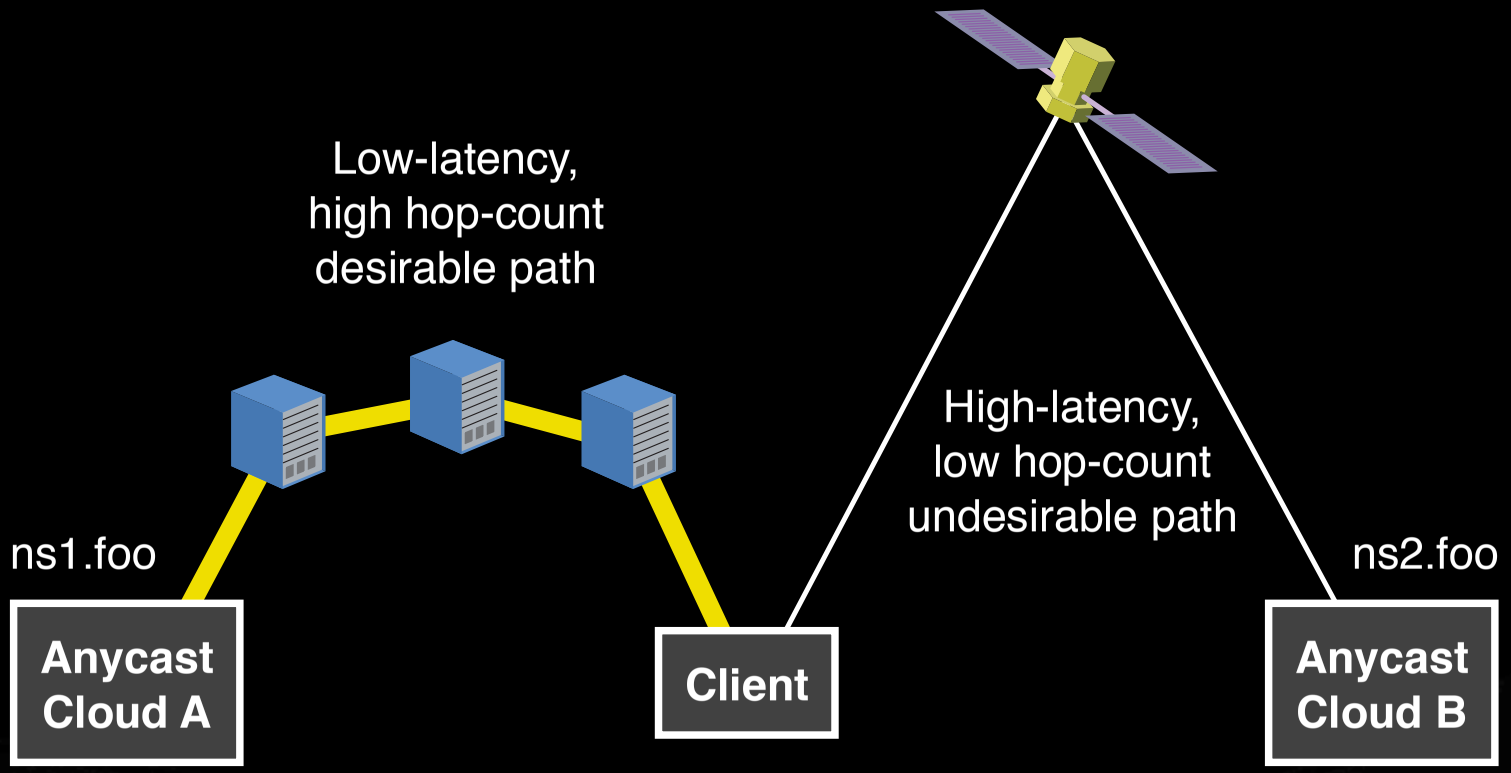
# Resolve the Conflict




This allows anycast to present the nearest servers,
and allows the resolver to choose the one which performs best.

PCH
Packet Clearing House

# Resolve the Conflict




These clouds are usually referred to as “A Cloud” and “B Cloud.”
The number of clouds depends on stability and scale trade-offs.

PCH
Packet Clearing House

# Latency Reduction

Latency reduction depends upon the native layer-3 routing of the Internet.

The theory is that the Internet will deliver packets using the shortest path.

The reality is that the Internet will deliver packets according to ISPs' policies.

PCH
Packet Clearing House

# Latency Reduction

ISPs' routing policies differ from shortest-path where there's an economic incentive to deliver by a longer path.

PCH
Packet Clearing House

# ISPs' Economic Incentives (Grossly Simplified)

ISPs have  high cost to deliver traffic through transit.

ISPs have a low cost to deliver traffic through their peering.

ISPs receive money when they deliver traffic to their customers.

PCH
Packet Clearing House

# ISPs' Economic Incentives (Grossly Simplified)

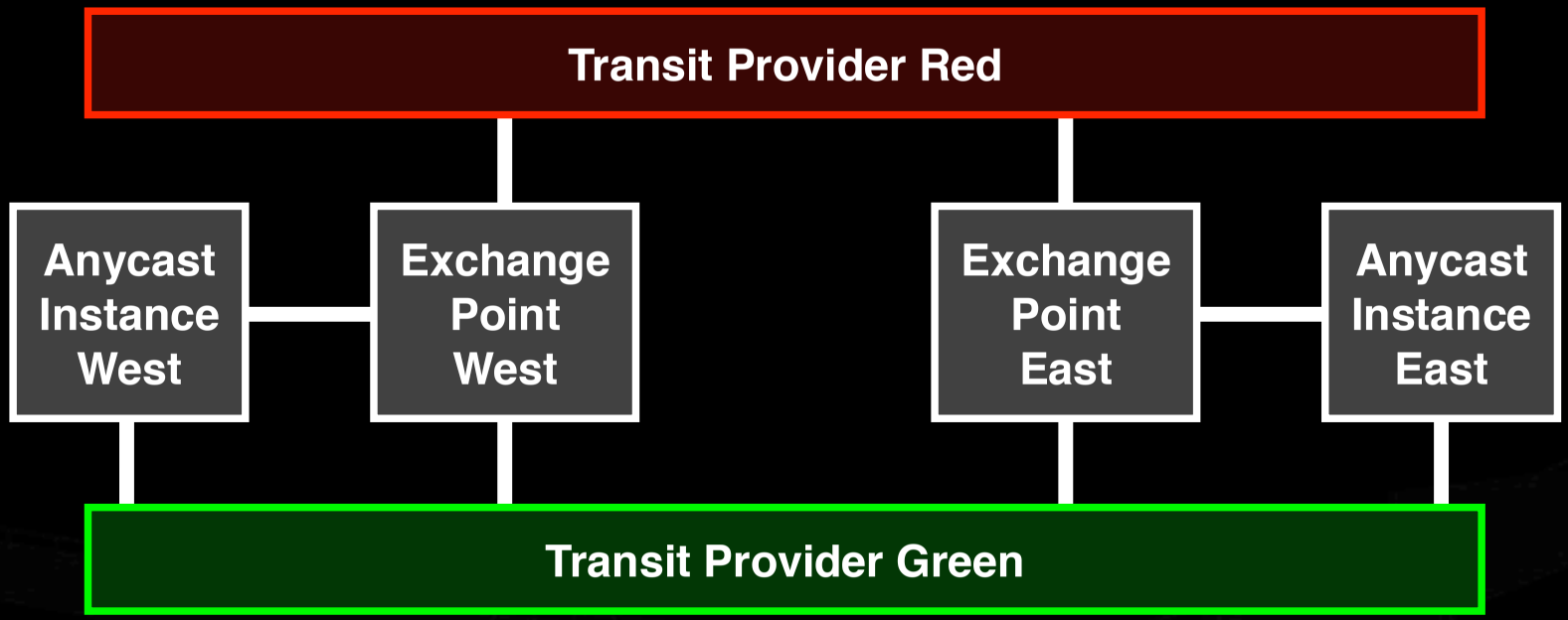
Therefore, ISPs will deliver traffic to a customer across a longer path, before by peering or transit across a shorter path.

If you are both a customer, and a customer of a peer or transit provider, this has important implications.

PCH
Packet Clearing House

# Normal Hot-Potato Routing

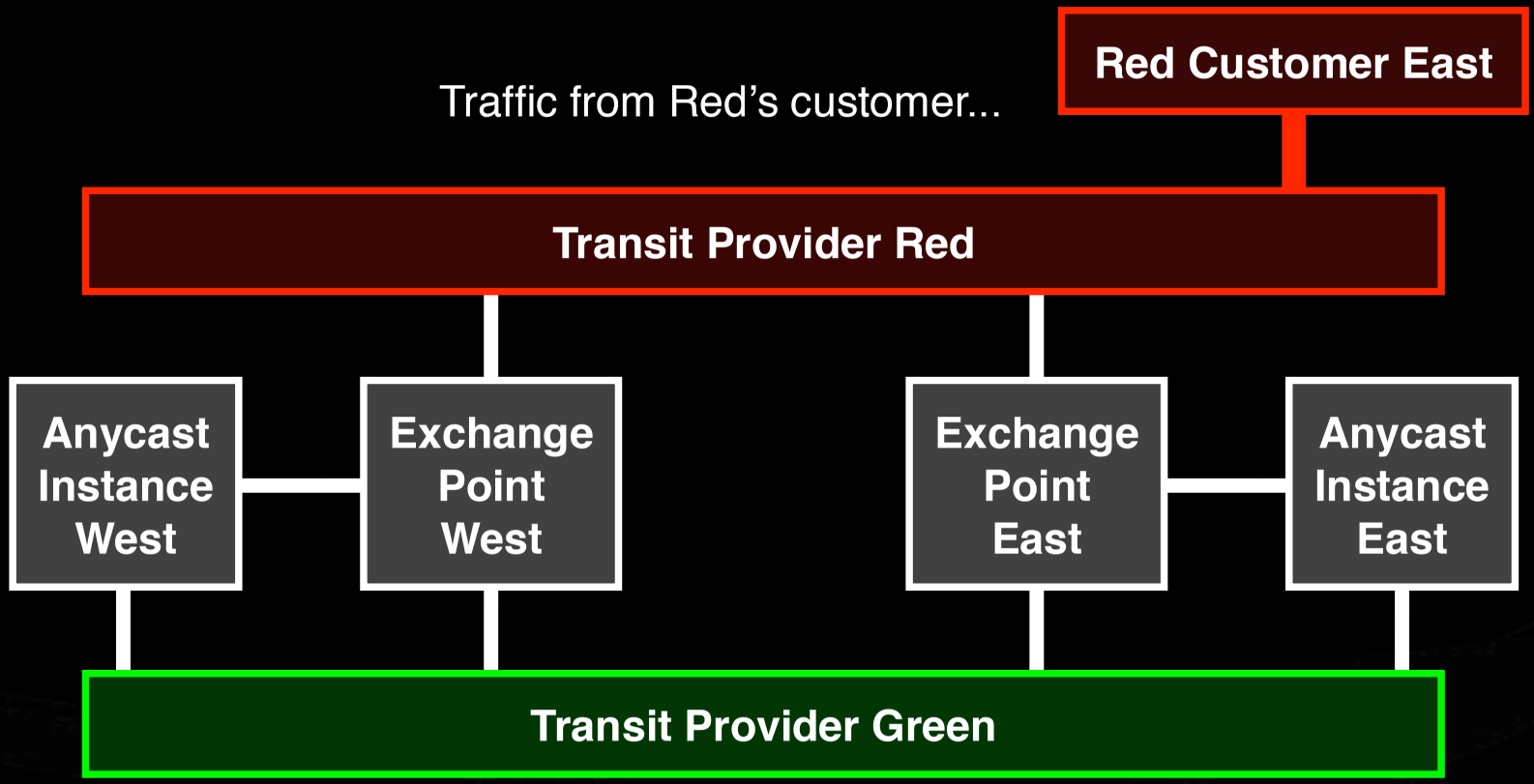
If the anycast network is not a customer of large Transit Provider Red...




...but is a customer of large Transit Provider Green...

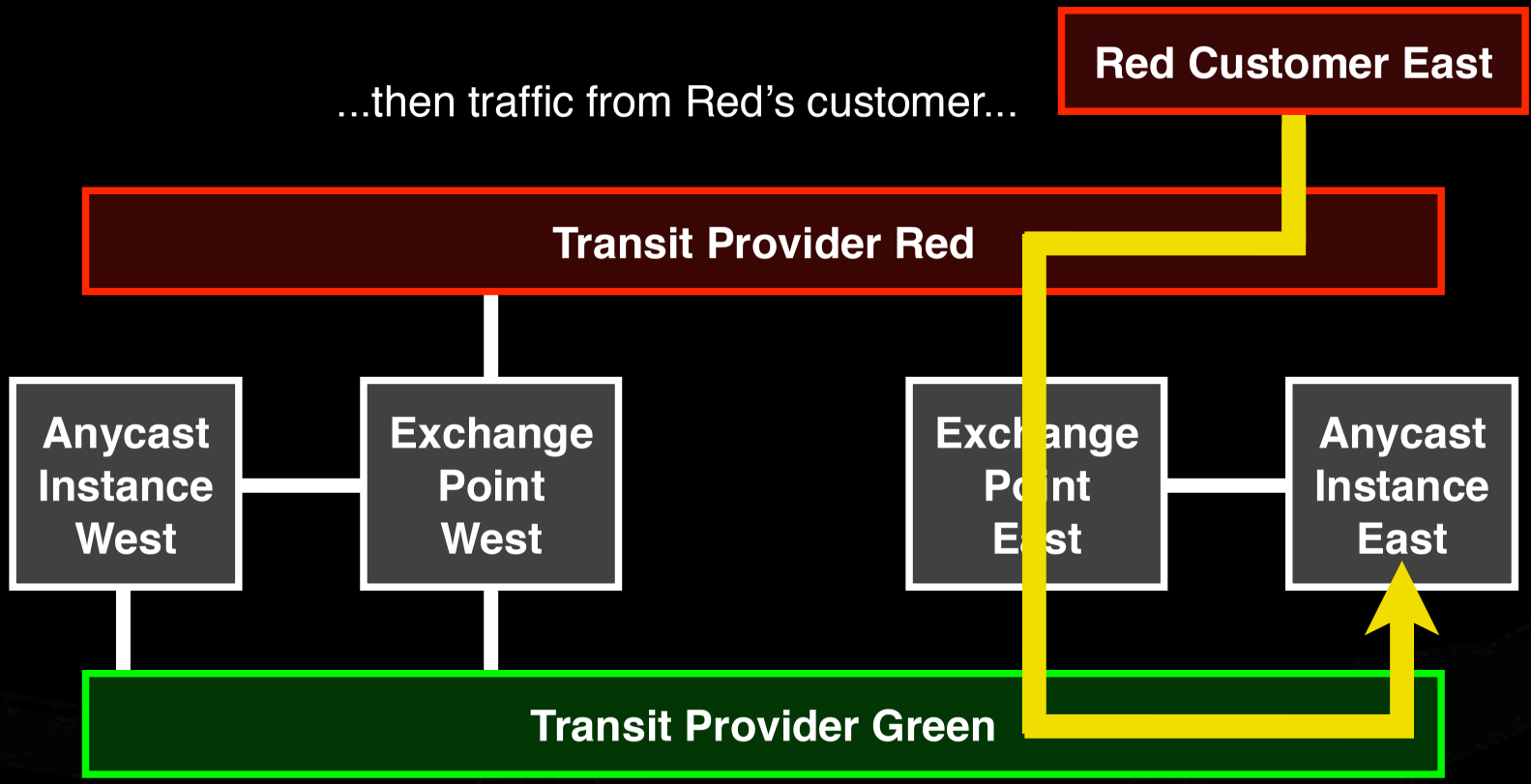
PCH
Packet Clearing House

# Normal Hot-Potato Routing



PCH
Packet Clearing House

# Normal Hot-Potato Routing

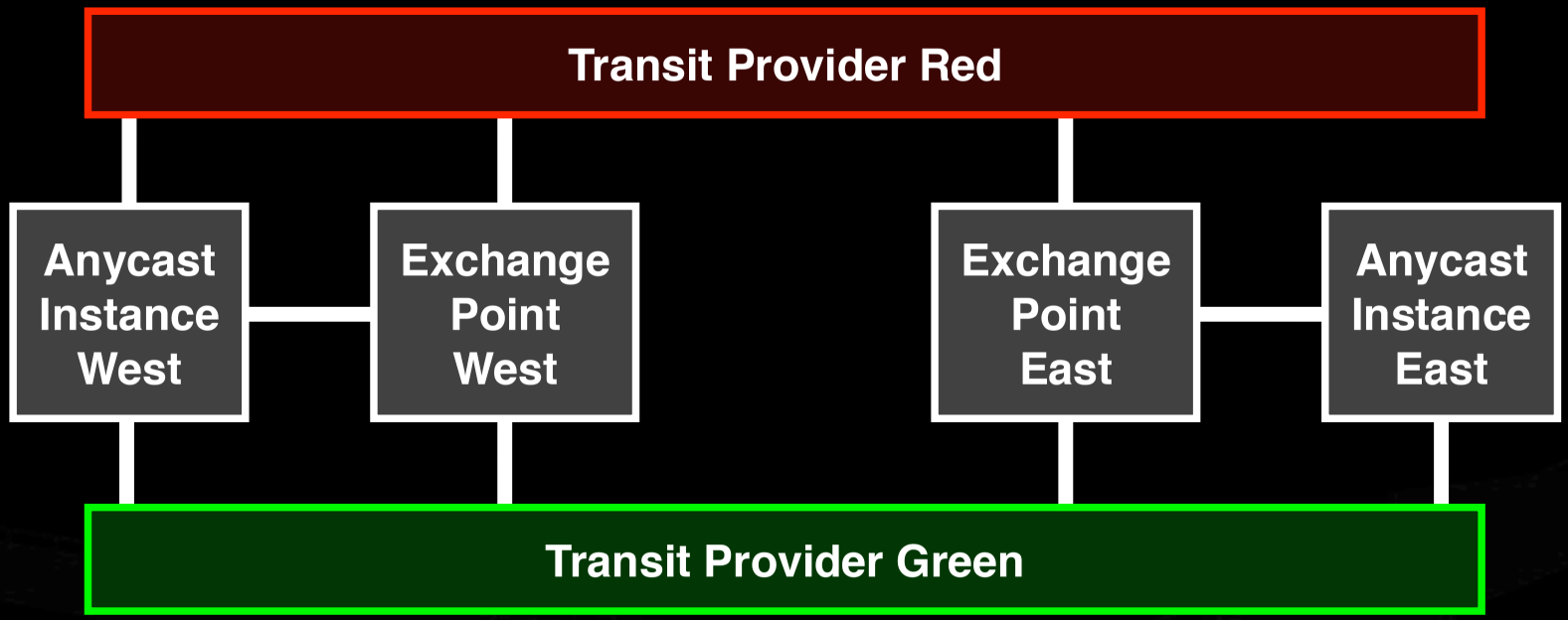



...is delivered from Red to Green via local peering, and reaches the local anycast instance.

PCH
Packet Clearing House

# How the Conflict Plays Out

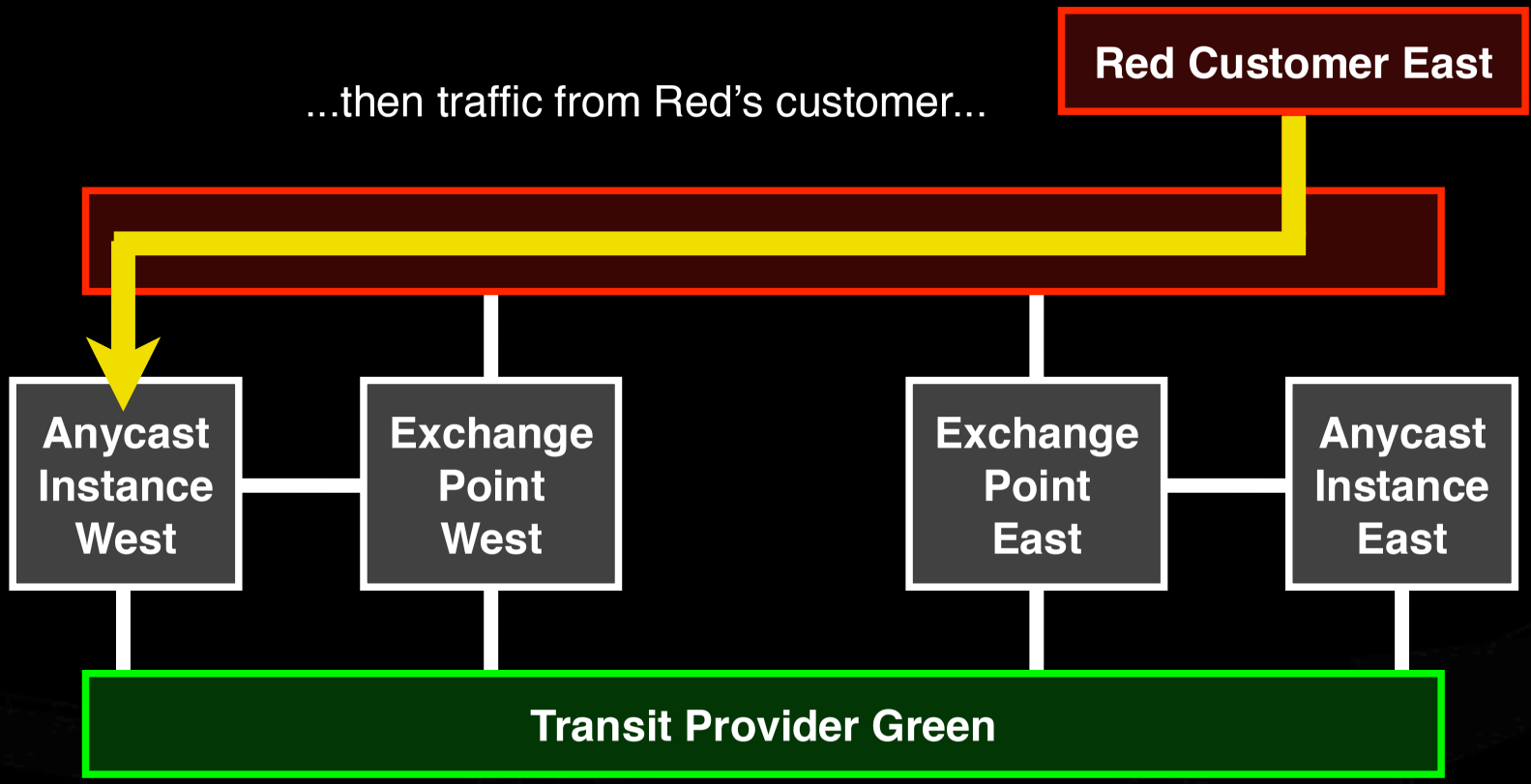
But if the anycast network is a customer of both large Transit Provider Red...




...and of large Transit Provider Green, but not at all locations...

PCH
Packet Clearing House

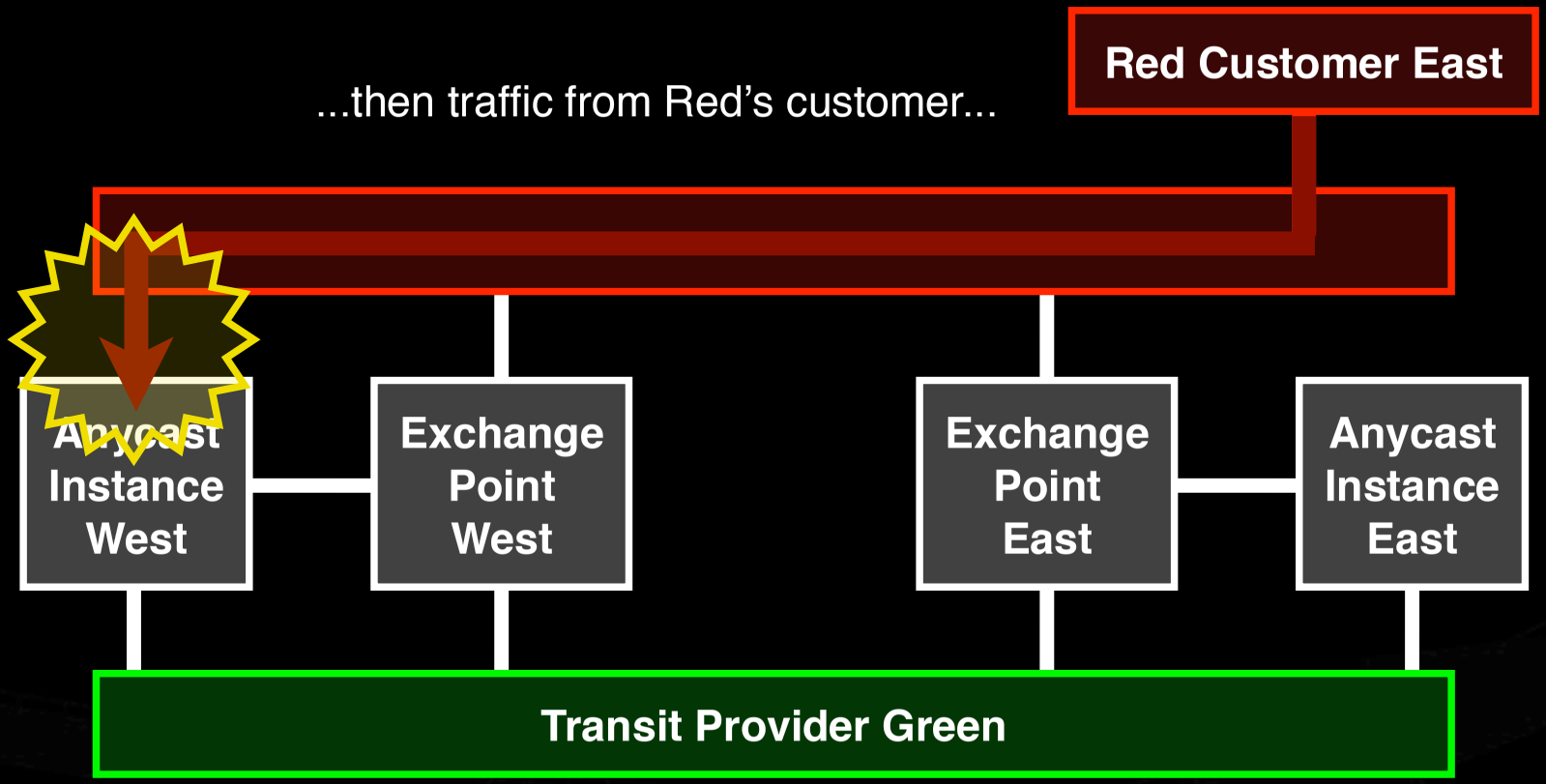
# How the Conflict Plays Out




...will be misdelivered to the remote anycast instance...

PCH
Packet Clearing House

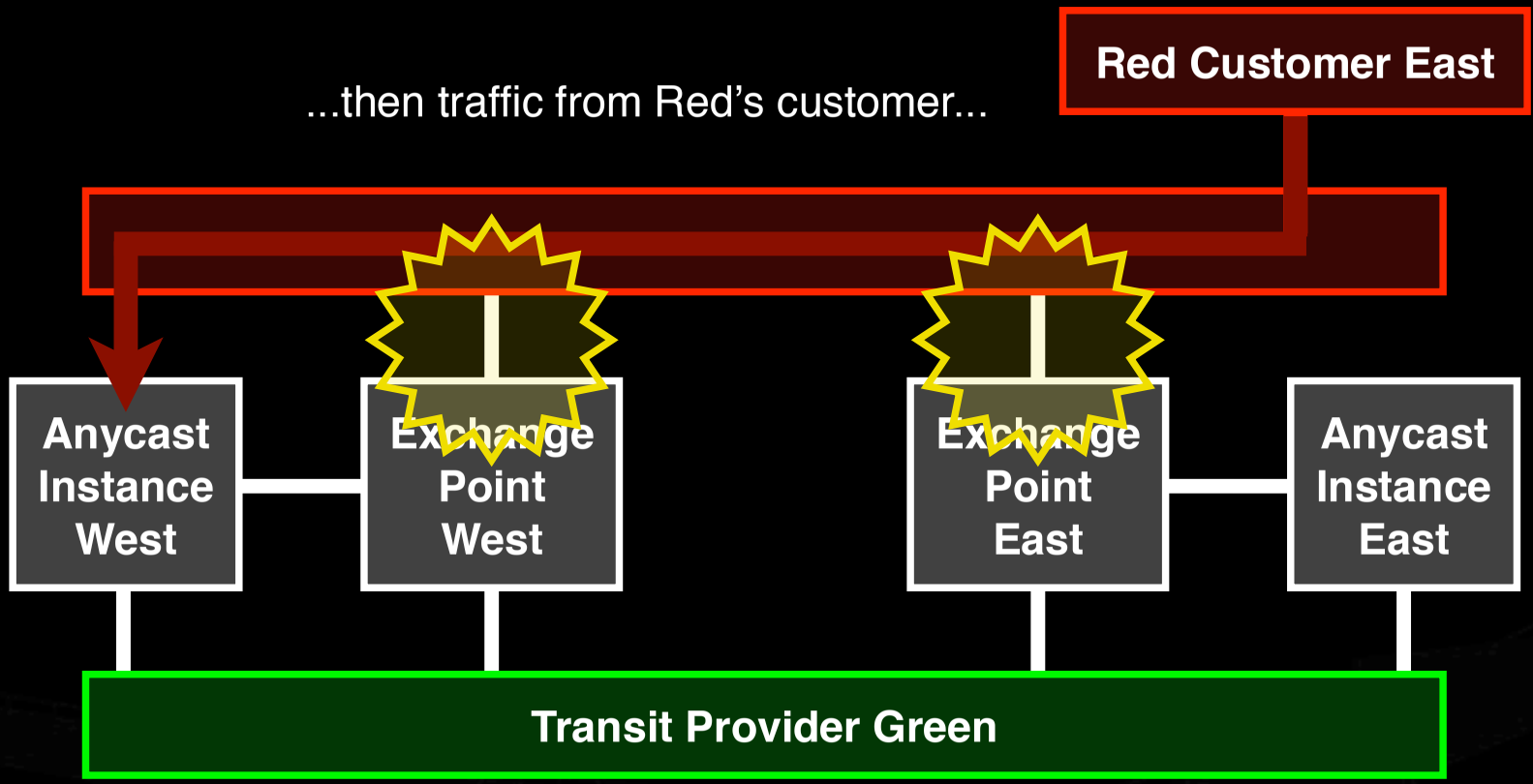
# How the Conflict Plays Out




...will be misdelivered to the remote anycast instance, because a customer connection...

PCH
Packet Clearing House

# How the Conflict Plays Out

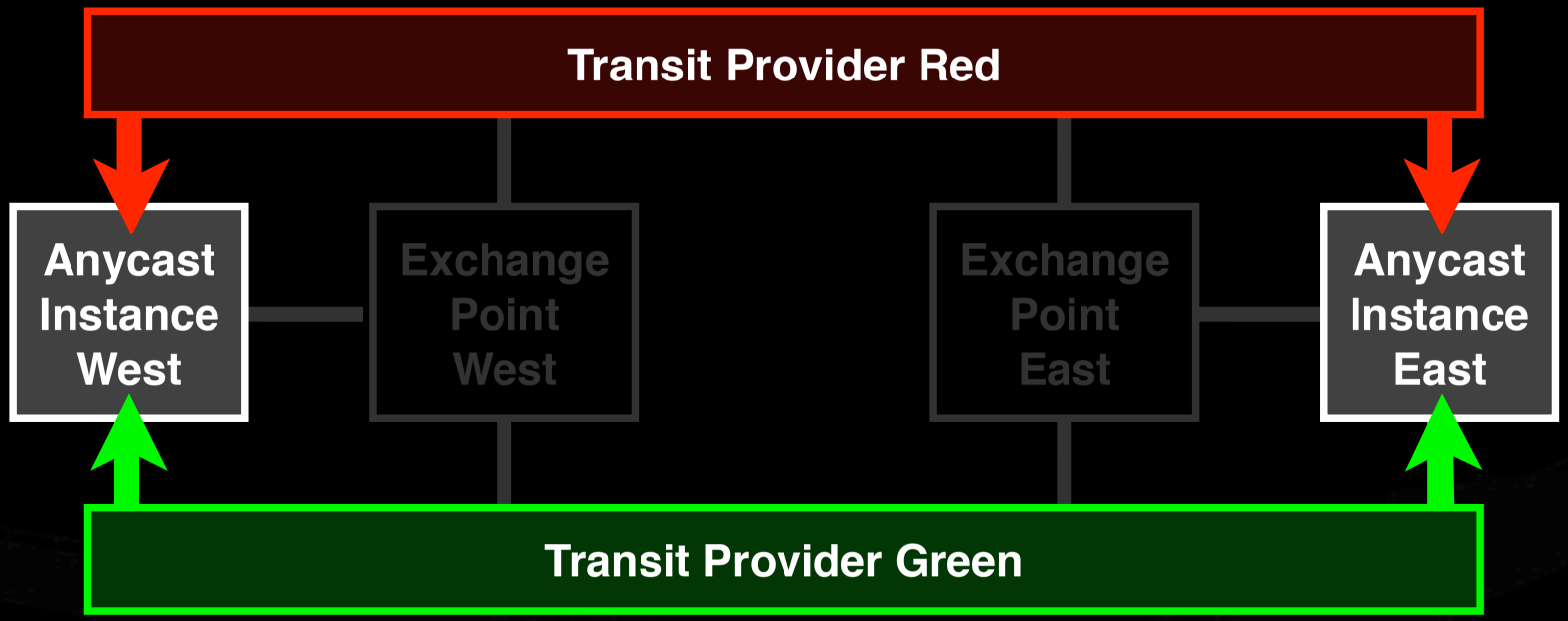



...will be misdelivered to the remote anycast instance, because a customer connection is preferred for economic reasons over a peering connection.

PCH
Packet Clearing House

# Resolve the Conflict

Any two instances of an anycast service IP address must have the same set of large transit providers at all locations.



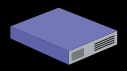

This caution is not necessary with small transit providers who don't have the capability of backhauling traffic to the wrong region on the basis of policy.

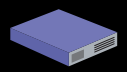
PCH
Packet Clearing House

# Putting the Pieces Together

- We need an A Cloud and a B Cloud.
- We need a redundant pair of the same transit providers at most or all instances of each cloud.
- We need a redundant pair of hidden masters for the DNS servers.
- We need a network topology to carry control and synchronization traffic between the nodes.

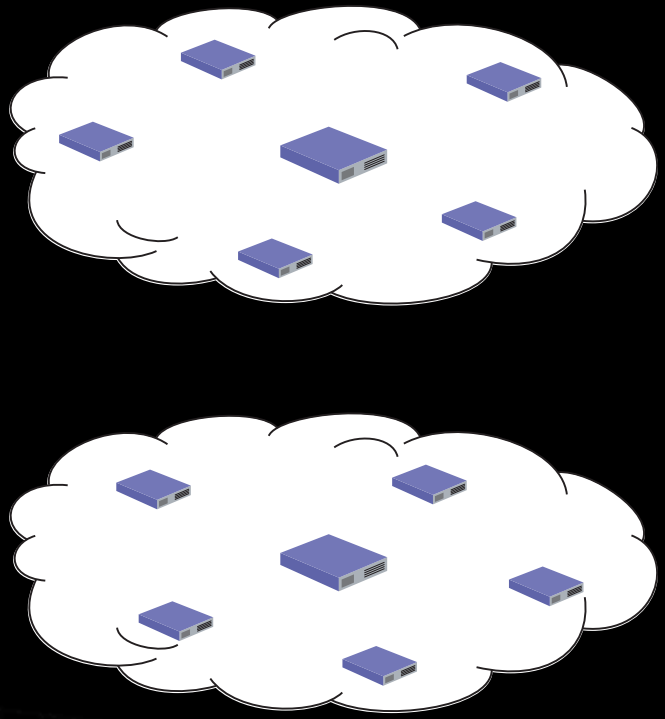
# Redundant Hidden Masters





Packet Clearing House

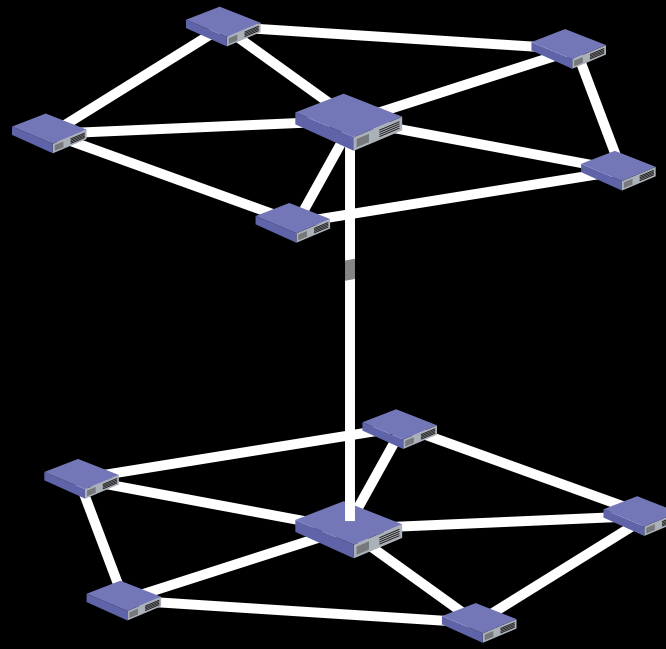
# An A Cloud and a B Cloud



PCH
Packet Clearing House

# A Network Topology

## “Dual Wagon-Wheel”



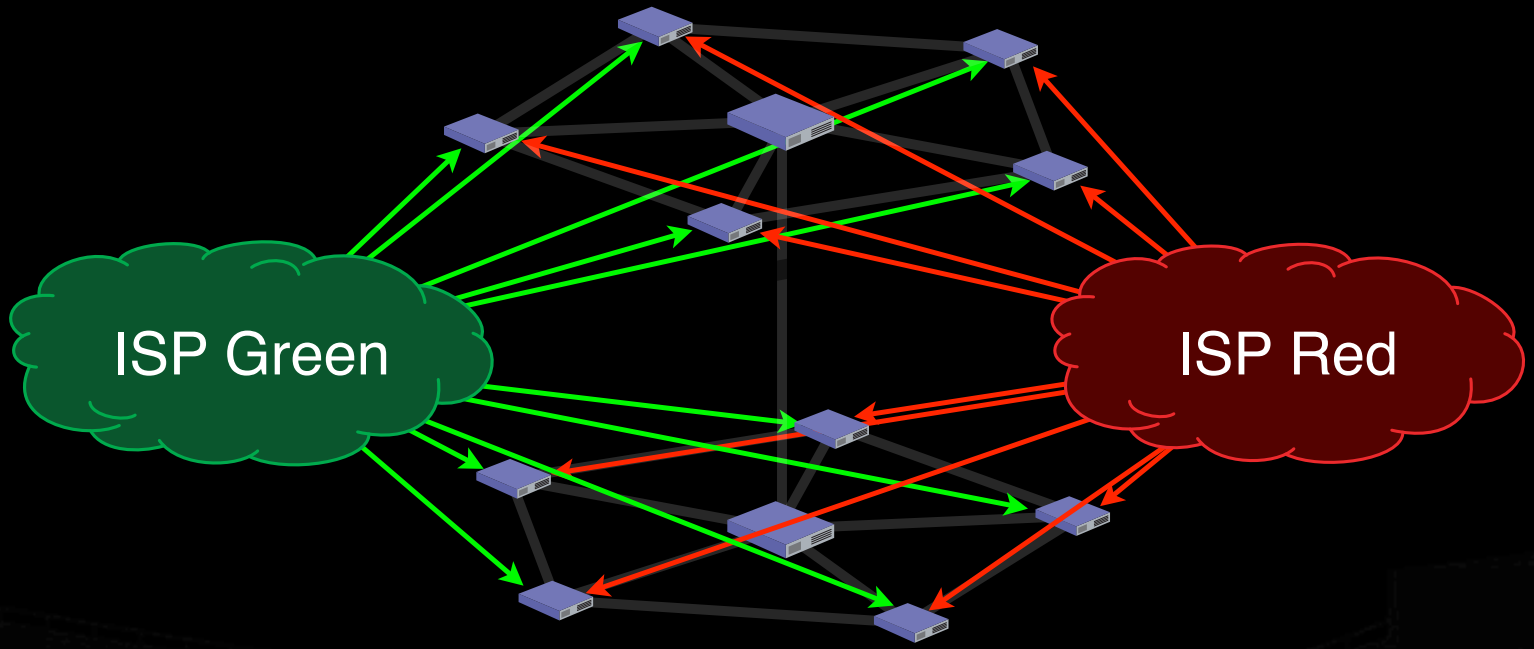
A Ring

B Ring

Packet Clearing House

# Redundant Transit

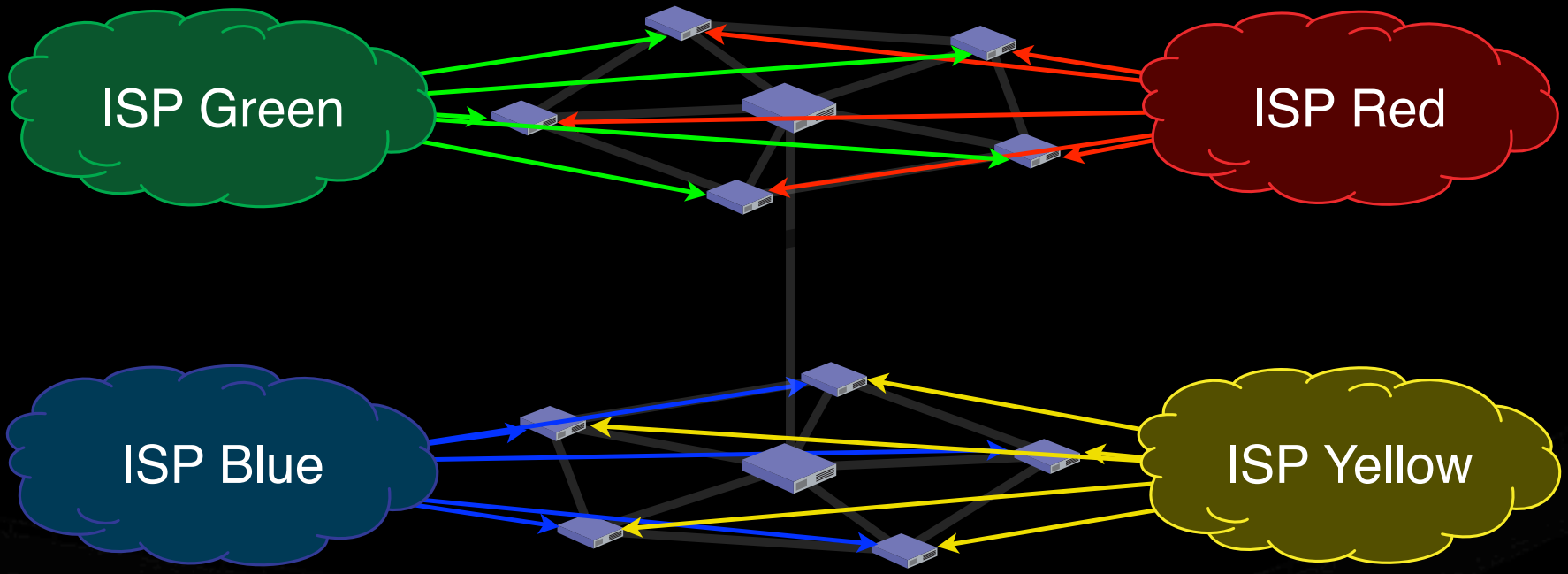
## Two ISPs



PCH
Packet Clearing House

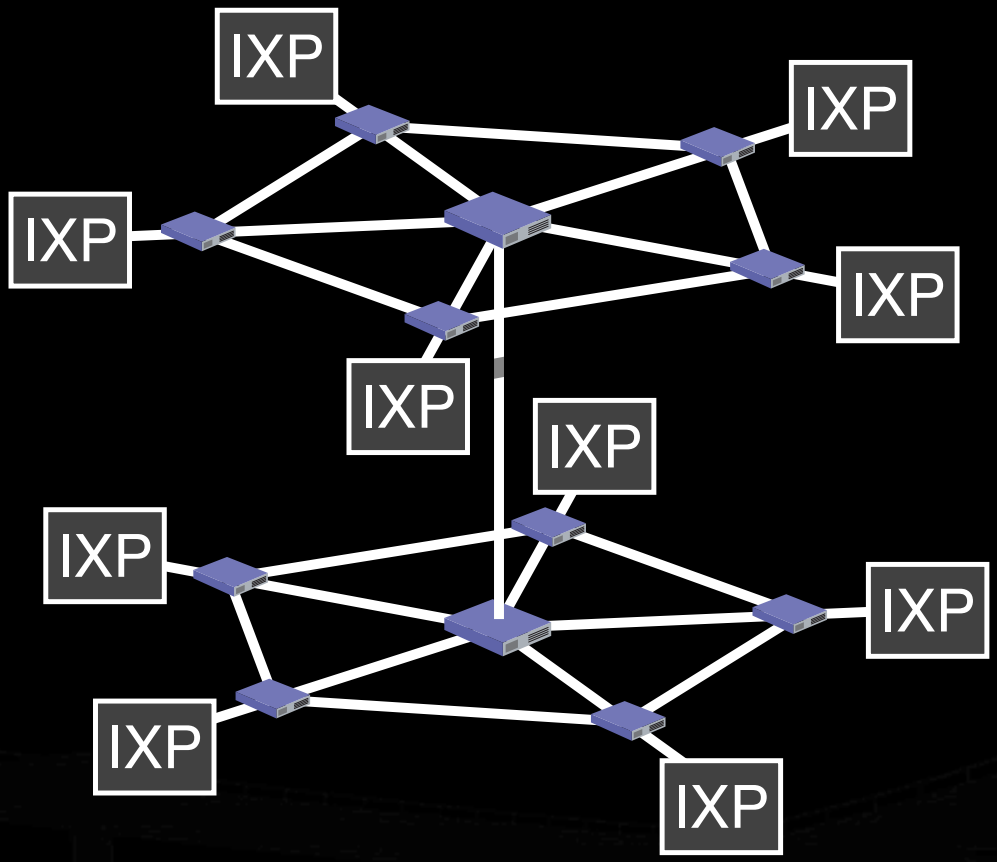
# Redundant Transit

## Or four ISPs



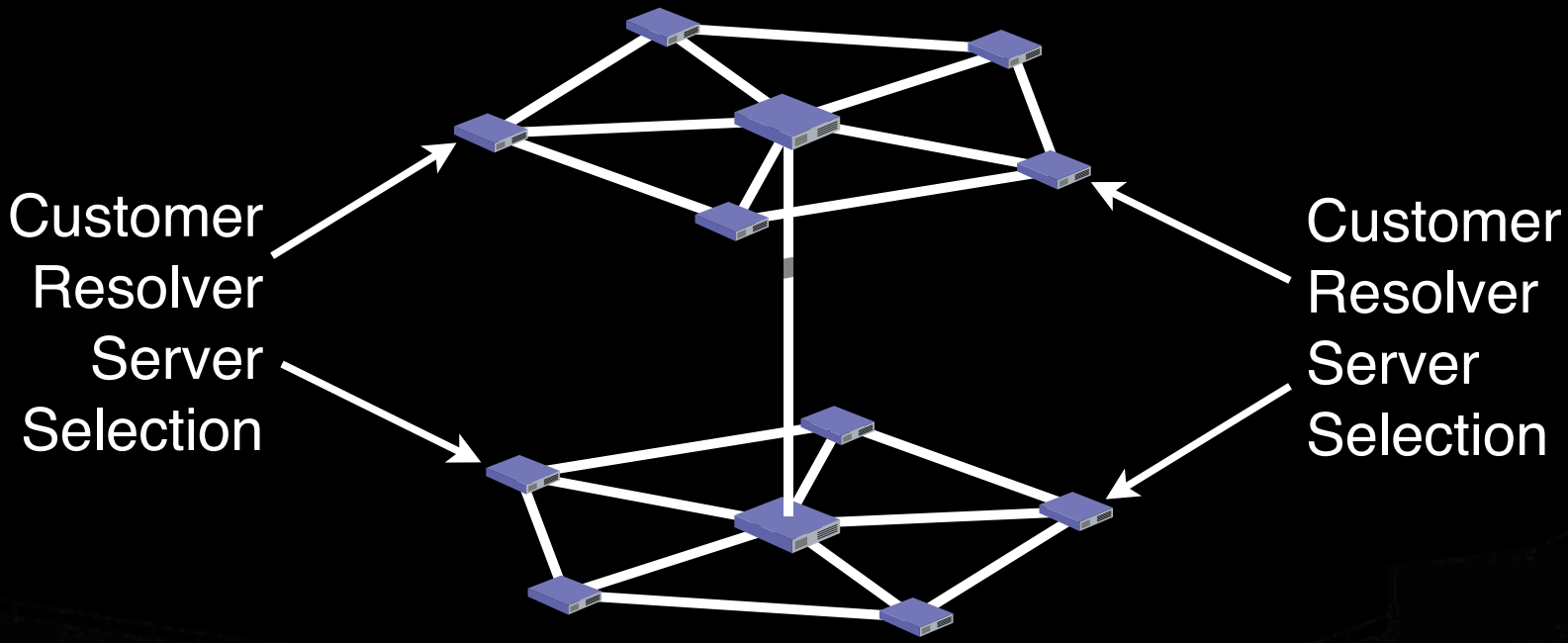
PCH
Packet Clearing House

# Local Peering



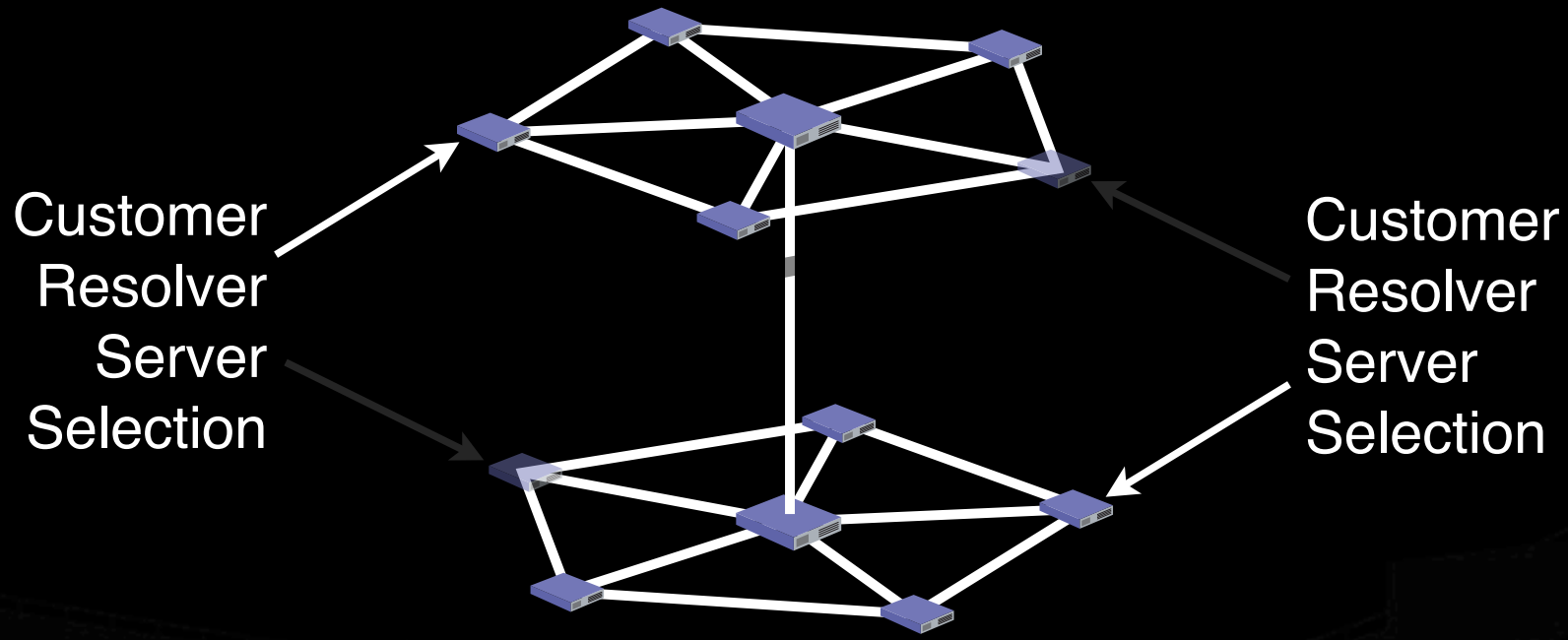
PCH
Packet Clearing House

# Resolver-Based Fail-Over



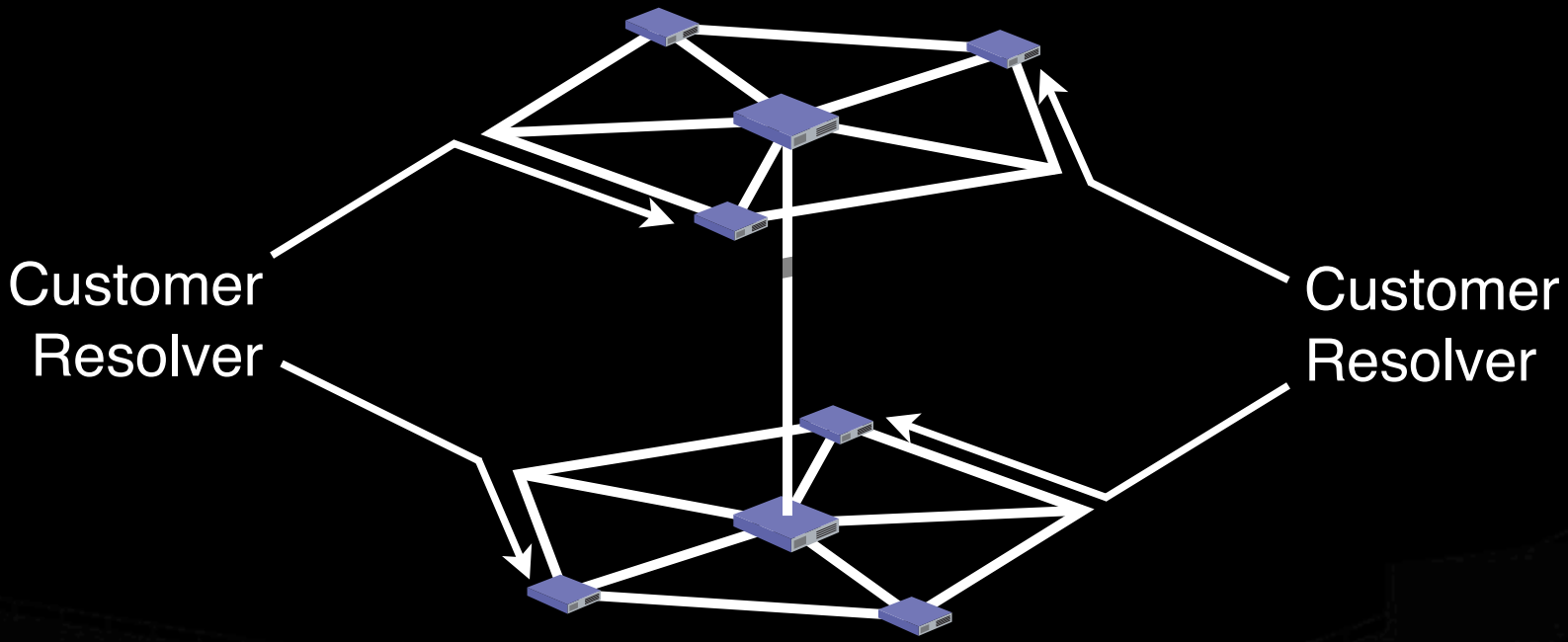
PCH
Packet Clearing House

# Resolver-Based Fail-Over



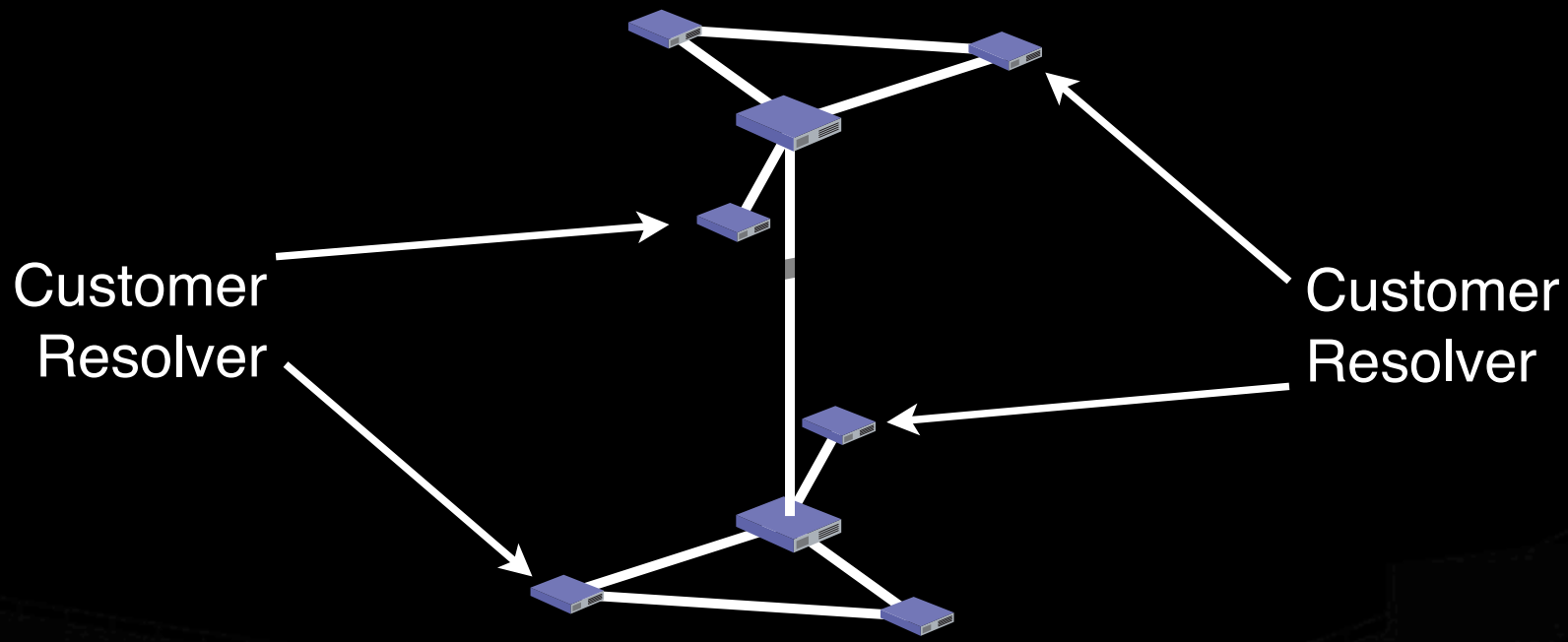
PCH
Packet Clearing House

# Internal Anycast Fail-Over



Packet Clearing House

# Global Anycast Fail-Over




# Thanks, and Questions?

Copies of this presentation can be found
in Keynote, PDF, and QuickTime formats at:

**http:// www.pch.net / resources / papers / dns-service-architecture**

Jonny Martin
Internet Analyst
Packet Clearing House
**jonny@pch.net**