BGP and Traffic Engineering with Akamai

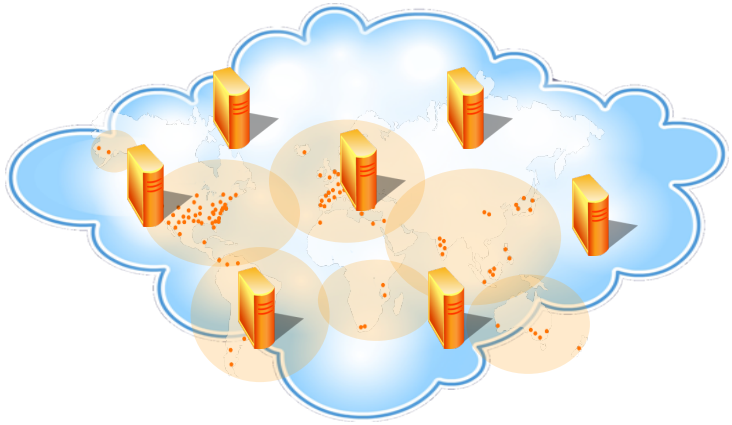
Christian Kaufmann
Akamai Technologies
MENOG 14

The Akamai Intelligent Platform

The world's largest on-demand, distributed computing platform delivers all forms of web content and applications

The Akamai Intelligent Platform:

147,000+ Servers	2,000+ Locations	1,200+ Networks	700+ Cities	92 Countries



Typical daily traffic:

- More than **2 trillion** requests served
- Delivering over **21 Terabits/second**
- **15-30%** of all daily web traffic

Basic Technology

Akamai mapping

How CDNs Work

When content is requested from CDNs, the user is directed to the optimal server

- This is usually done through the DNS, especially for non-network CDNs, e.g. Akamai
- It can be done through anycasting for network owned CDNs

Users who query DNS-based CDNs be returned different A (and AAAA) records for the same hostname

This is called “mapping”

The better the mapping, the better the CDN

How Akamai CDN Work

Example of Akamai mapping

- Notice the different A records for different locations:

```
[NYC]% host www.symantec.com
www.symantec.com    CNAME  e5211.b.akamaiedge.net.
e5211.b.akamaiedge.net.  A      207.40.194.46
e5211.b.akamaiedge.net.  A      207.40.194.49

[Boston]% host www.symantec.com
www.symantec.com    CNAME  e5211.b.akamaiedge.net.
e5211.b.akamaiedge.net.  A      81.23.243.152
e5211.b.akamaiedge.net.  A      81.23.243.145
```

Peering with Akamai

Why Akamai Peer with ISPs

Performance & Redundancy

- Removing intermediate AS hops seems to give higher peak traffic for same demand profile

Burstability

- During large events, having direct connectivity to multiple networks allows for higher burstability than a single connection to a transit provider

Peering reduces costs

Network Intelligence

Backup for on-net servers

- If there are servers on-net, the peering can act as a backup during downtime and overflow
- Allows serving different content types

Why ISPs peer with Akamai

Performance

- Akamai and ISPs are in the same business, just on different sides - we both want to serve end users as quickly and reliably as possible

Cost Reduction

- Transit savings
- Possible backbone savings

Marketing

- Claim performance benefits over competitors
- Keep customers from seeing “important” web sites through their second uplink

Because you are nice :-)

How Akamai use IXes

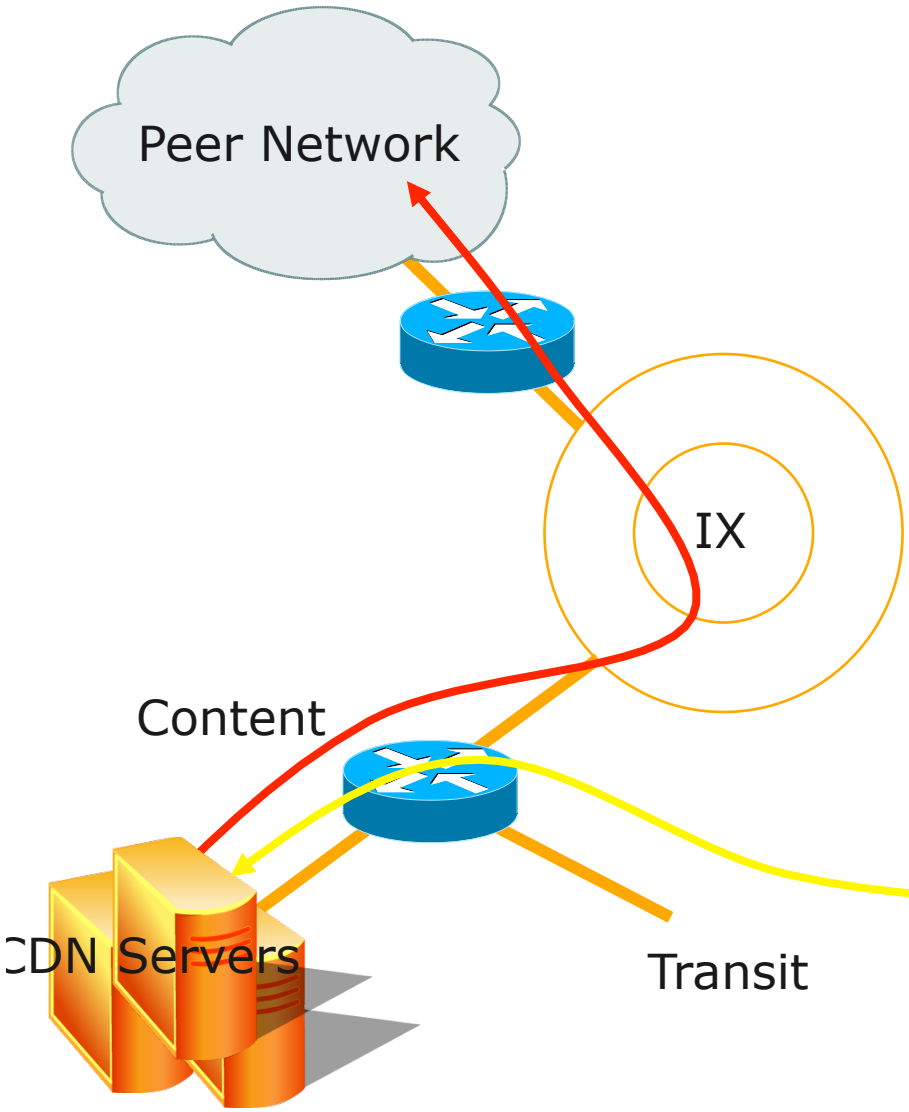
Akamai usually do not announce large blocks of address space because no one location has a large number of servers

- It is not uncommon to see a single /24 from Akamai at an IX

This does not mean you will not see a lot of traffic

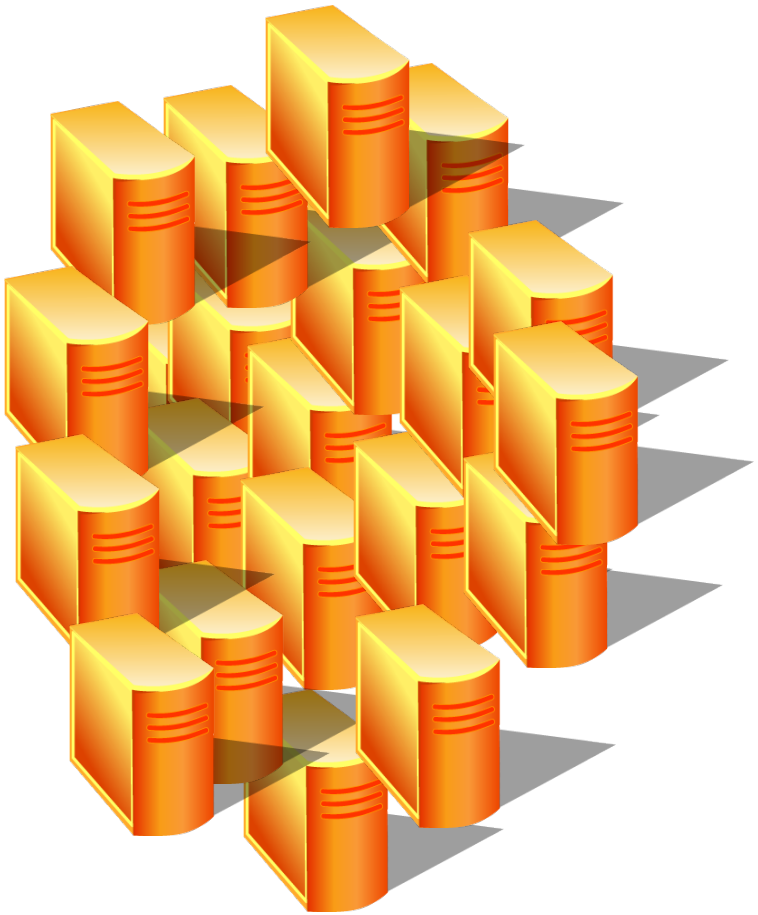
- How many web servers does it take to fill a gigabit these days?

How Akamai use IXes



- Akamai (Non-network CDNs) do not have a backbone, so each IX instance is independent
- Akamai uses transit to pull content into the servers
- Content is then served to peers over the IX
- After BGP is established, you might not see traffic until 24hrs
- Akamai Mapping System needs time to process new prefix

Why don't I get all the Akamai content via the Peering?



CDN Servers

- No single cluster can accommodate all Akamai content
- Peer with Akamai in different locations for accessing different Akamai Content
- ISP prefers Customer before Peers
- Akamai prefers on-net Cluster before Peer
- Do you want to host Akamai Cluster?

After Peering With Akamai….

DO and DON’T’s of Traffic Engineering

Akamai
FASTER FORWARD

The world uses…

AS Path Prepending

- **Before**

Akamai Router#sh ip b 100.100.100.100
BGP routing table entry for 100.100.100.0/20, version Paths: (1 available, best #1, table Default-IP-Routing-Table)
Multipath: eBGP
 Advertised to update-groups:
    2   7
 4635 **1001**
   202.40.161.1 from 202.40.161.1 (202.40.161.1)

- **After**

Akamai Router#sh ip b 100.100.100.100
BGP routing table entry for 100.100.100.0/20, version
Paths: (1 available, best #1, table Default-IP-Routing-Table)
Multipath: eBGP
 Advertised to update-groups:
    2   7
 4635 **1001 1001 1001 1001**
   202.40.161.1 from 202.40.161.1 (202.40.161.1)

Akamai
FASTER FORWARD

But it does not has the usual effect

Akamai
FASTER FORWARD

The world uses…

MED

- **Before**

Akamai Router#sh ip b 100.100.100.100
BGP routing table entry for 100.100.100.0/20, version Paths: (1 available, best #1, table Default-IP-Routing-Table)
Multipath: eBGP
 Advertised to update-groups:
    2   7
 4635 1001
   202.40.161.1 from 202.40.161.1 (202.40.161.1)
   Origin IGP, **metric 0**, localpref 100, valid, external, best

- **After**

Akamai Router#sh ip b 100.100.100.100
BGP routing table entry for 100.100.100.0/20, version Paths: (1 available, best #1, table Default-IP-Routing-Table)
Multipath: eBGP
 Advertised to update-groups:
    2   7
 4635 1001
   202.40.161.1 from 202.40.161.1 (202.40.161.1)
   Origin IGP, **metric 1000**, localpref 100, valid, external, best

Akamai
FASTER FORWARD

But it does not has the usual effect

Akamai
FASTER FORWARD

The world uses…

More Specific Route

- **Before**

Akamai Router#sh ip b 100.100.100.100
BGP routing table entry for 100.100.96.0**/20**, version
Paths: (1 available, best #1, table Default-IP-Routing-Table)
Multipath: eBGP
 Advertised to update-groups:
    2   7
 4635 1001
   202.40.161.1 from 202.40.161.1 (202.40.161.1)

- **After**

Akamai Router#sh ip b 100.100.100.100
BGP routing table entry for 100.100.100.0**/24**, version Paths: (1 available, best #1, table Default-IP-Routing-Table)
Multipath: eBGP
 Advertised to update-groups:
    2   7
 4635 1001
   202.40.161.1 from 202.40.161.1 (202.40.161.1)

Akamai
FASTER FORWARD

But it does not has the usual effect

Why doesn't it has the usual effect?

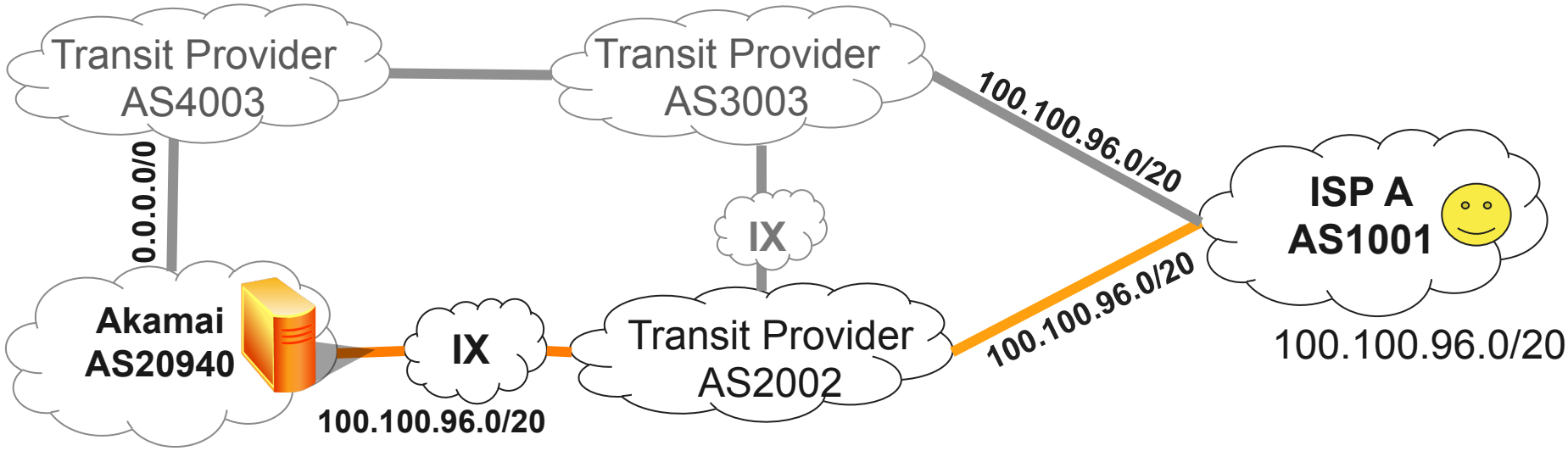
- Akamai uses Mapping, on top of the BGP routing
- Akamai Mapping is different from BGP routing
- Akamai uses multiple criteria to choose the optimal server
- These include standard network metrics:
  - Latency
  - Throughput
  - Packet loss

Typical Scenarios in Traffic Engineering

Scenario: In-consistent Route Announcement

Consistent Route Announcement of Multi-Home ISP A

- ISP A is multi-home to Transit Provider AS2002 and AS3003
- Transit Provider AS2002 peer with Akamai
- Transit Provider AS3003 do not peer with Akamai
- Akamai always sends traffic to ISP A via Transit Provider AS2002

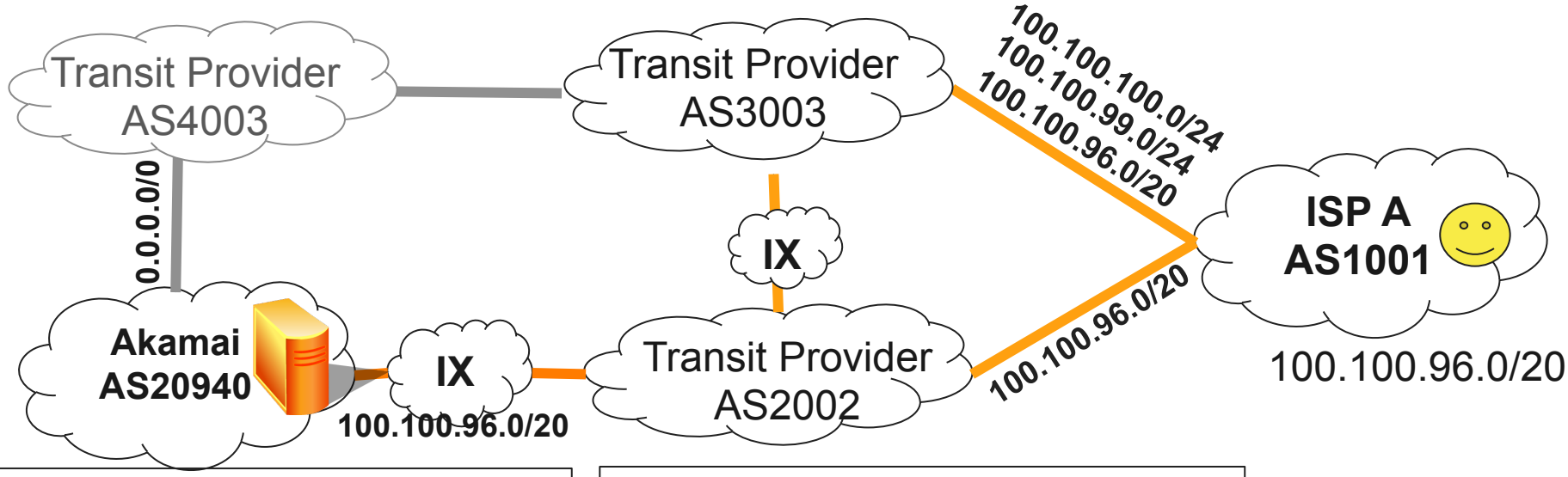


What will you do?

- ISP A would like to balance the traffic between two upstream providers
- ISP A prepend, MED to Transit Provider AS2002. Unfortunately, no effect on Akamai traffic…..
- Eventually, ISP A breaks the /20 and starts more specific & inconsistent route announcement
- What will happen?

ISP A Load Balance the Traffic Successfully

- ISP A announces more specific routes /24 to Transit Provider AS3003
- Transit Provider AS3003 announces new /24 to AS2002
- Akamai peer router do not have full routes like many other ISP, so traffic continue route to the superblock /20 of AS2002
- ISP A is happy with the balanced traffic on dual Transit Providers



Akamai AS20940 Routing Table

100.100.96.0/20	AS2002 AS1001
0.0.0.0/0	AS4003

AS2002 Routing Table

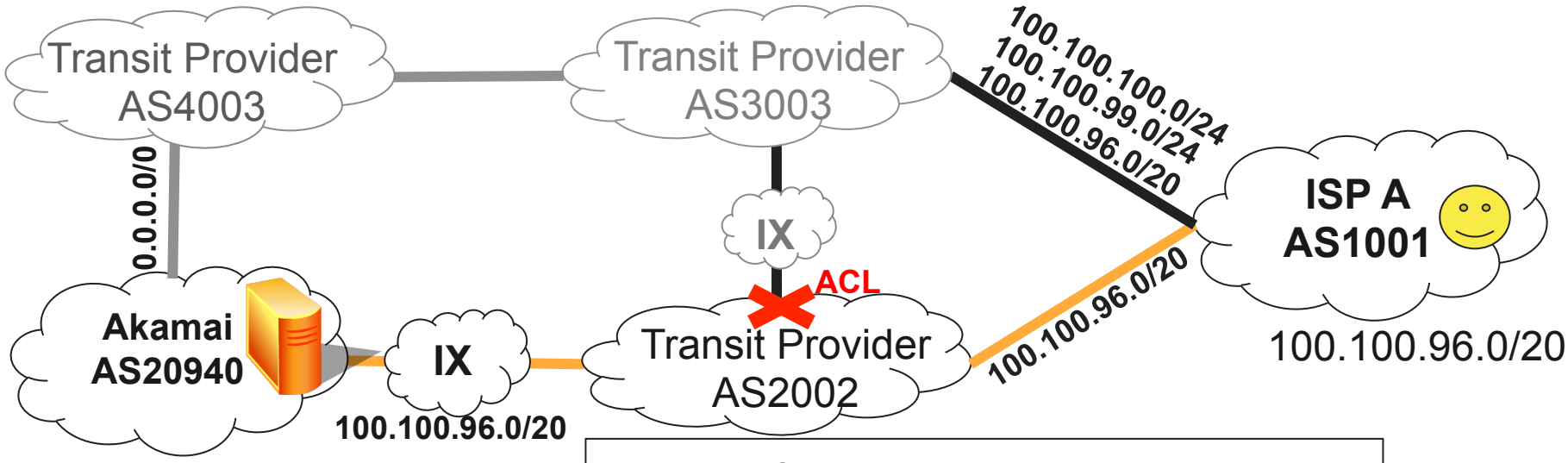
100.100.100.0/24	AS3003 AS1001
100.100.99.0/24	AS3003 AS1001
100.100.96.0/20	AS1001

What is the problem?

- Loss of revenue for Transit Provider AS2002 although their backbone is consumed
- What could happen if AS2002 does not like the peer-to-peer traffic?

AS2002 Filter Traffic on Peer Port

- In order to get rid of peer-to-peer traffic, Transit Provider AS2002 implement an ACL on IX port facing AS3003
- ISP A cannot access some websites due to traffic black hole




```
hostname AS2002-R1
!
interface TenGigabitEthernet1/1
ip access-group 101 out
!
access-list 101 deny ip any 100.100.100.0 0.0.0.255
access-list 101 deny ip any 100.100.99.0 0.0.0.255
access-list 101 permit ip any any
```

Is Traffic Filtering a good workaround?

- It is observed that some Transit Providers filter peer-to-peer traffic on IX port or Private Peer
- If you promised to carry the traffic of a block (eg./20), you should not have any holes (eg. /24) or drop any part of the traffic
- The end users connectivity will be impacted by your ACL!!!

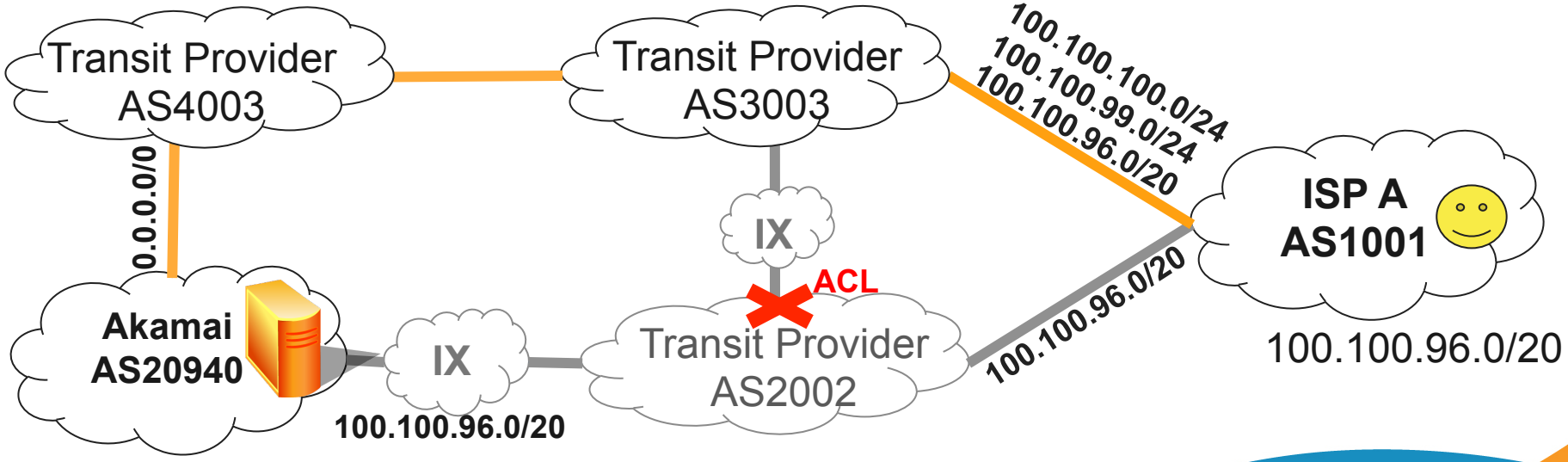
Your Promise

You break the promise!

Hong Kong

Akamai workaround on ISP Traffic Filtering

- Akamai observes ISP A user unable to access some websites
- Akamai blocks all prefix received from Transit Provider AS2002, so traffic shifts from IX to Transit AS4003
- ISP A can access all websites happily
- Transit Provider AS2002 observes traffic drop on IX

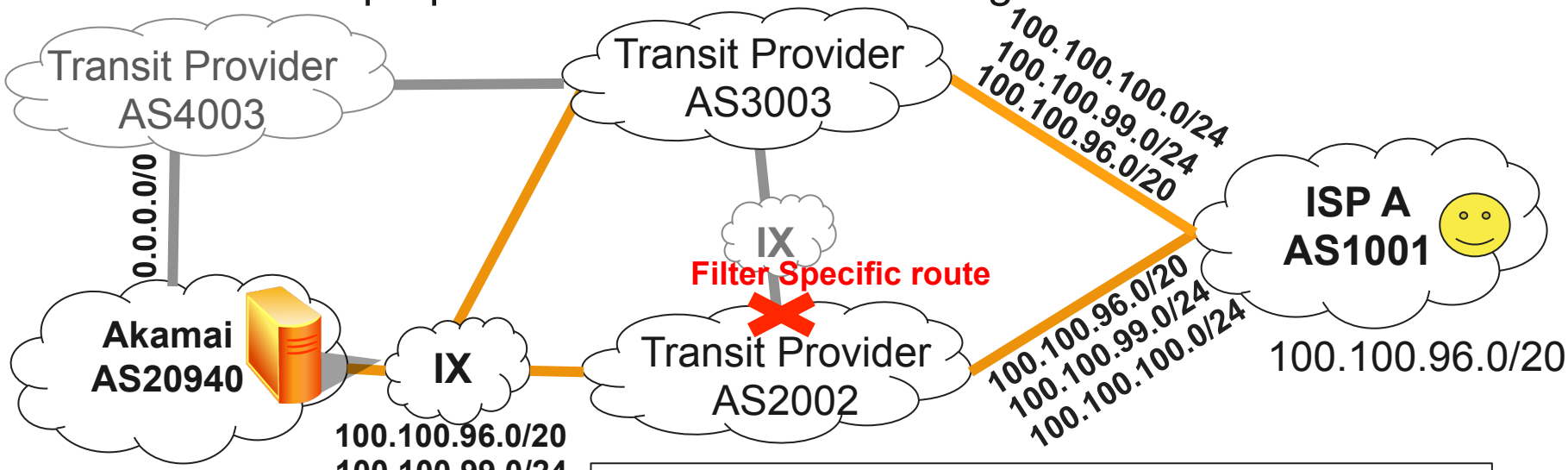


What is the result?

- ISP A results in imbalance traffic between two upstream providers
- We wish consistent route announcement
- Transit Provider AS2002 lost all Akamai traffic from peer because he breaks the promise of carrying the packet to destination
- Transit Provider AS2002 lost revenue due to the reducuction of traffic
- ISP should filter the specific routes rather than filter the traffic

Ideal solution

- Transit Provider AS2002 should filter the specific route rather than traffic
- ISP A can work with upstreams and Akamai together
- Transit Provider AS3003 can peer with Akamai
- ISP A can announces consistent /24 in both upstream
- ISP A can prepend the /24 for traffic tuning

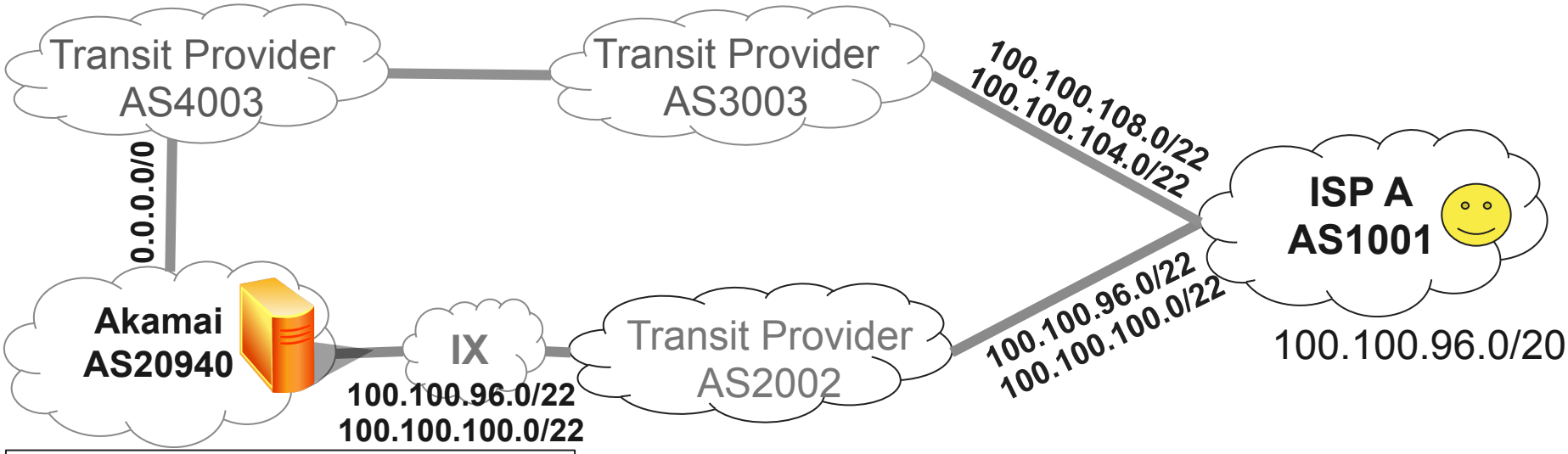



```
neighbor PEER-GROUP prefix-list DENY-SPECIFIC in
!
ip prefix-list DENY-SPECIFIC seq 5 deny 100.100.100.0/24
ip prefix-list DENY-SPECIFIC seq 10 deny 100.100.99.0/24
ip prefix-list DENY-SPECIFIC seq 100 permit 0.0.0.0/0 le 32
```

Scenario: In-complete Route Announcement

In-complete Route Announcement

- ISP A is multi-home to Transit Provider AS2002 and AS3003
- Transit Provider AS2002 peers with Akamai
- Transit Provider AS3003 does not peer with Akamai
- ISP A announces different prefix to different ISP
- ISP A can access full internet

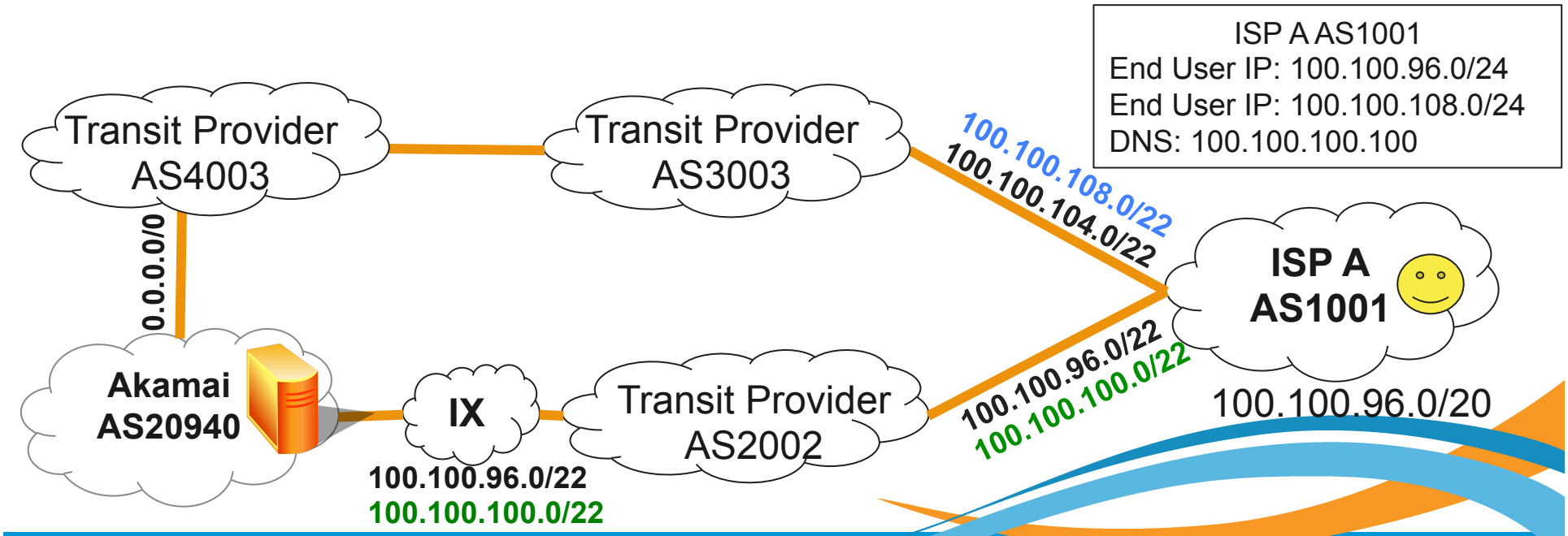


Akamai AS20940 Routing Table

100.100.96.0/22	AS2002 AS1001
100.100.100.0/22	AS2002 AS1001
0.0.0.0/0	AS4003

How will the traffic route to ISP A end users?

- End Users are using IP Address of 100.100.96.0/22, 100.100.100.0/22, 100.100.104.0/22, 100.100.108.0/22
- End Users are using ISP A DNS Server 100.100.100.100
- Akamai receives the DNS Prefix 100.100.100.0/22 from AS2002, so it maps the traffic of ISP A to this cluster
- 100.100.96.0/22 100.100.100.0/22 traffic is routed to AS2002 while 100.100.104.0/22 100.100.108.0/22 traffic is routed to AS3003 by default route

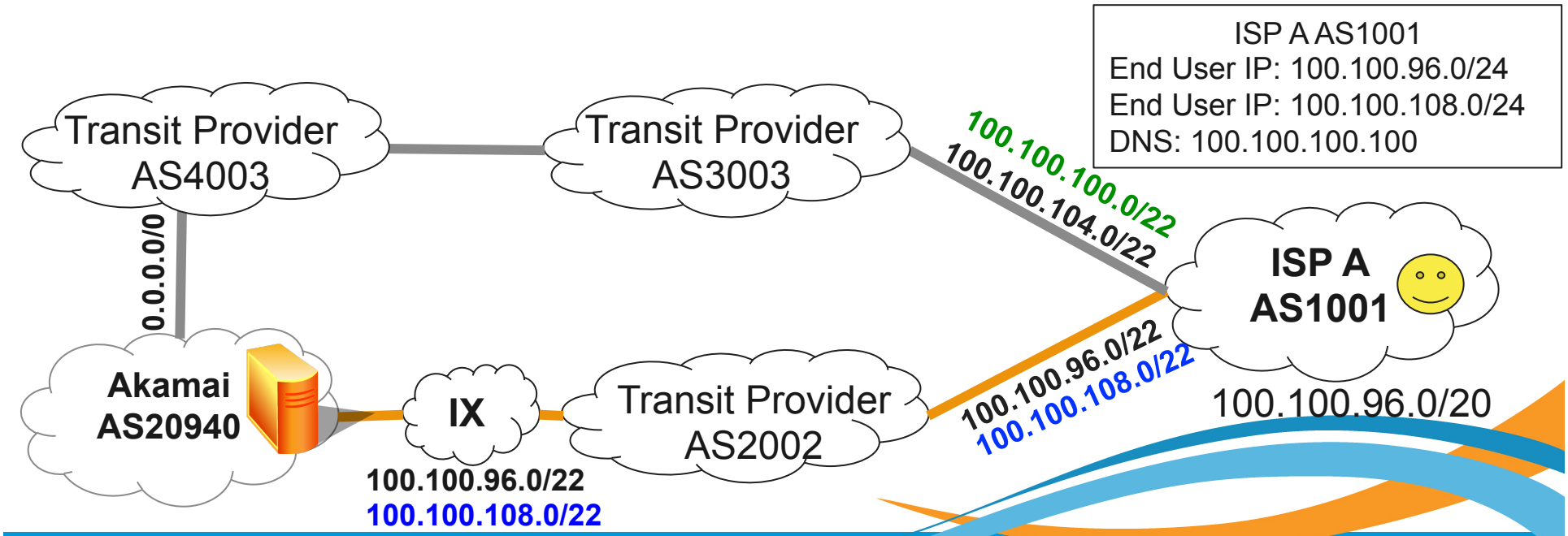


Does it cause problem?

- It is observed that some ISP performs in-complete route announcement (Eg. Announce different sub-set of prefix to different upstream)
- Some 100.100.100.108.0/22 end users have different performance than the others
- What will ISP A do if the user complains?

ISP A change the prefix announcement

- ISP A perceives AS3003 performance is lower than AS2002
- ISP A adjust the route announcement
- Both 100.100.96.0/22 and 100.100.108.0/22 are routed by AS2002 and end users have the same download speed
- ISP A end users are happy to close the complaint ticket



After 24hours

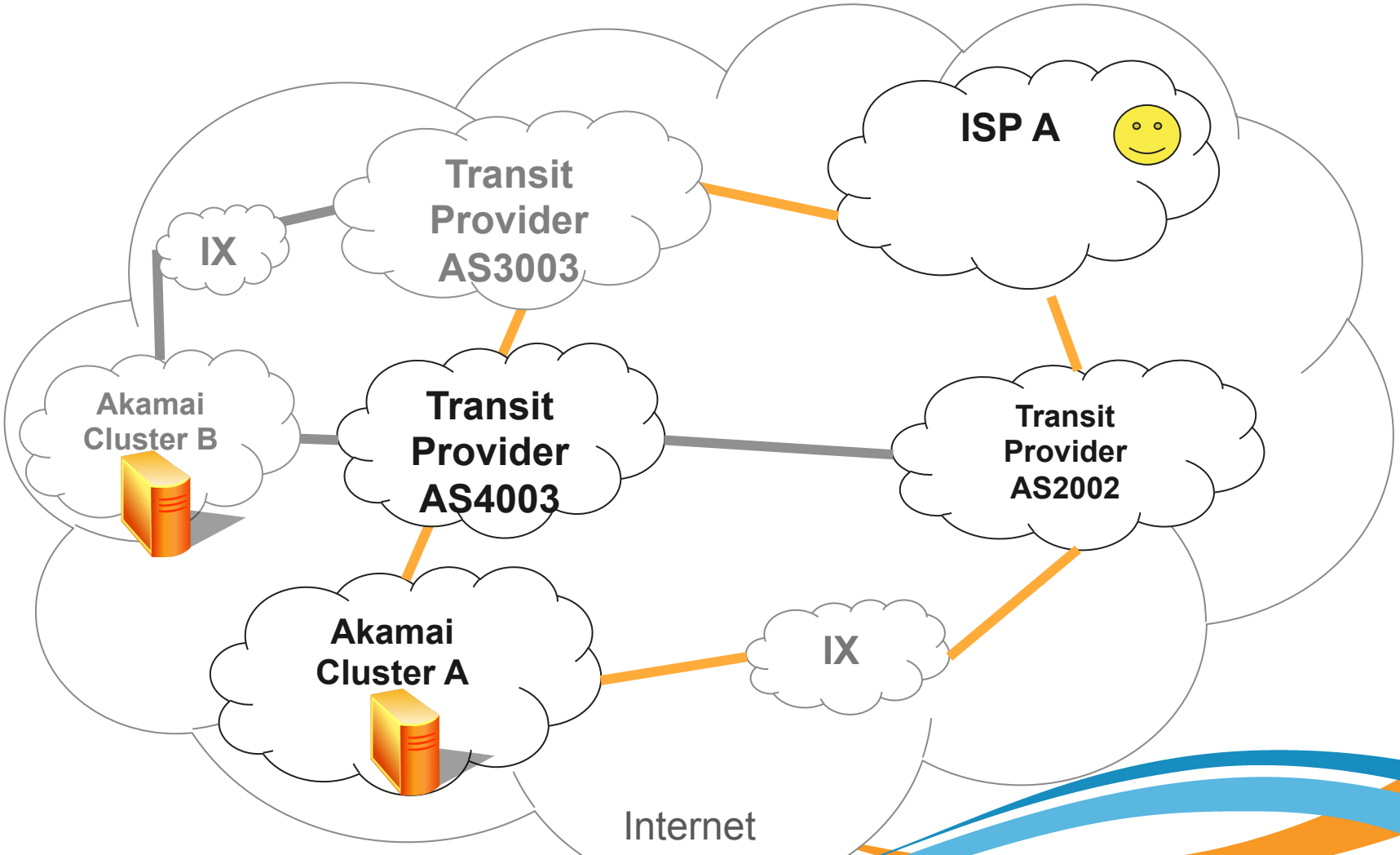
Akamai's Mapping System processes the change of prefix……

ISP A End Users complaints again

- Akamai no longer receive DNS prefix 100.100.100.0/22 from AS2002
- Akamai maps the traffic of ISP A to Cluster B instead of Cluster A
- ISP A still receives the traffic from both upstream
- ISP A End Users complain again ☹

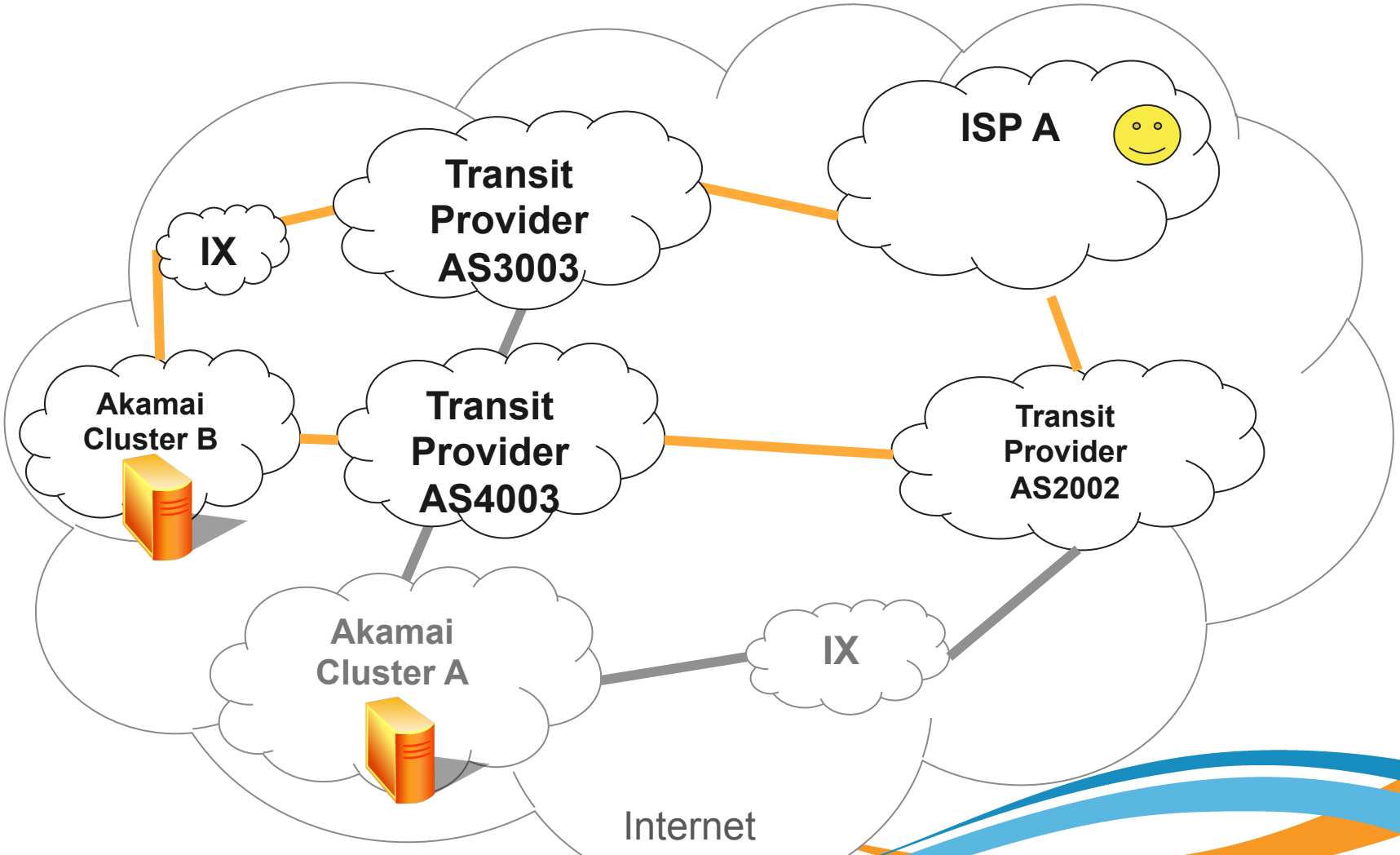
Before Akamai Mapping System refresh

- Akamai maps the traffic to Cluster A



After Akamai Mapping System refresh

- Akamai maps the traffic to Cluster B

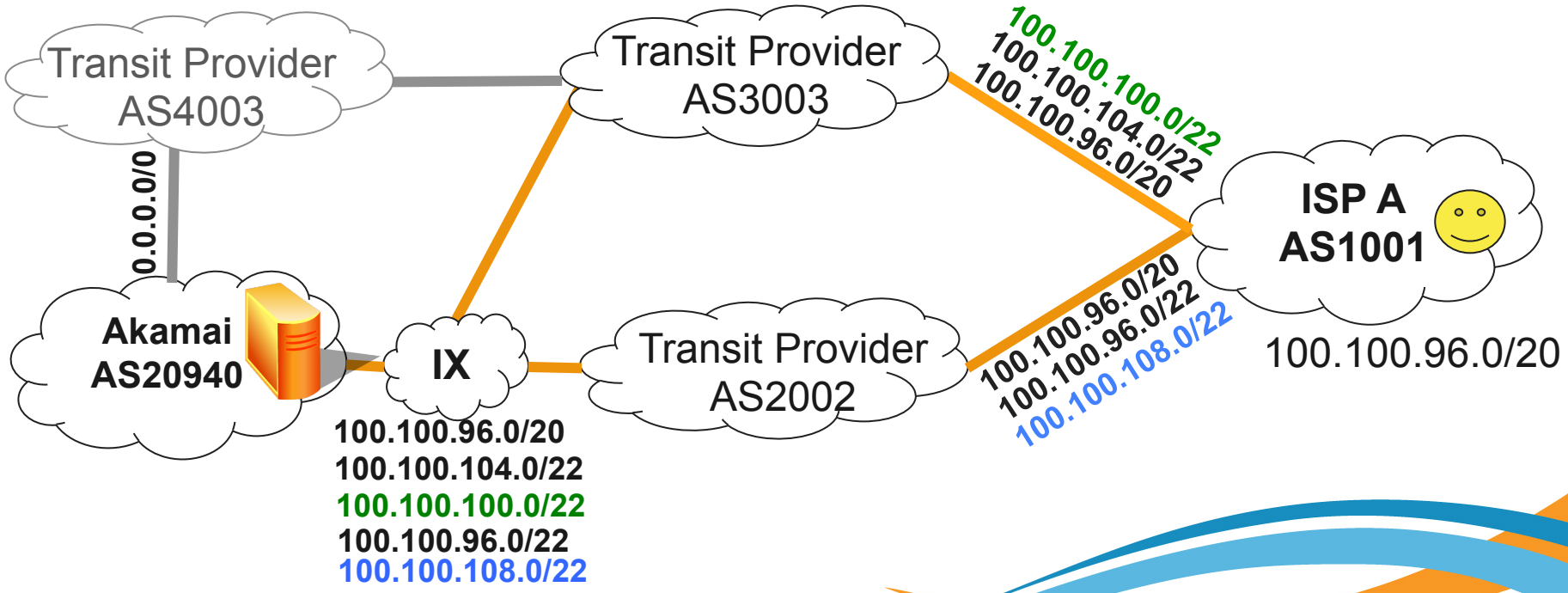


Our Recommendation

- Please maintain complete route announcement
- Talk to us if there are any traffic or performance issues
- We can work together for traffic engineering

Ideal solution

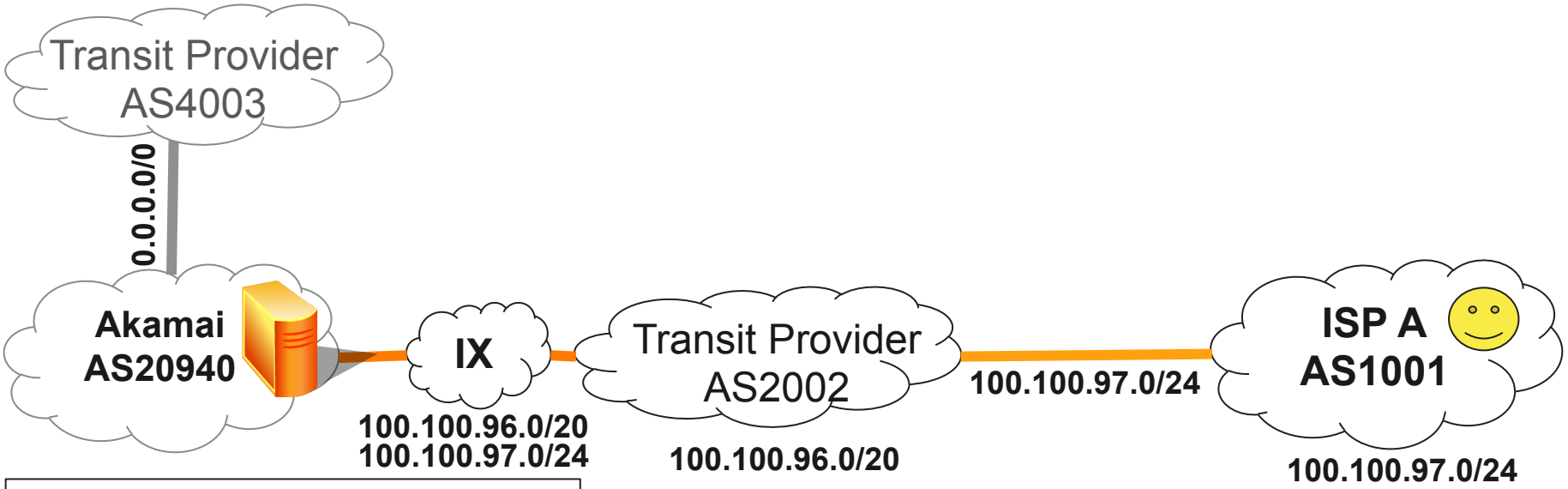
- ISP A should announces complete prefix in both upstream
- ISP A can work with upstream and Akamai together
- Transit Provider AS3003 can peer with Akamai



Scenario: Improper Prefix Announcement after customer left

Single Home ISP A

- ISP A is single home to Transit Provider AS2002
- ISP A obtains a /24 from Transit Provider AS2002
- Akamai always sends traffic to ISP A via Transit Provider AS2002

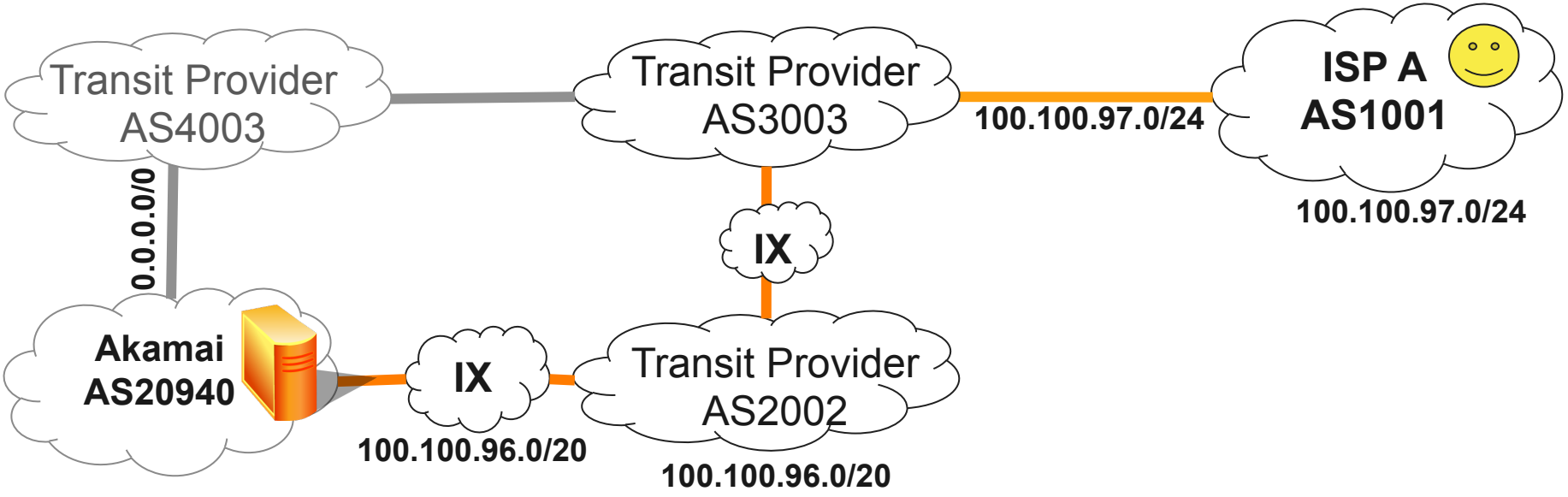


Akamai AS20940 Routing Table

100.100.96.0/20	AS2002
100.100.97.0/24	AS2002 AS1001
0.0.0.0/0	AS4003

Single Home ISP A changed upstream provider

- ISP A keeps using 100.100.97.0/24 from Transit Provider AS2002
- ISP A is changed upstream from AS2002 to AS3003
- Akamai always sends traffic to ISP A via Transit Provider AS2002 because the superblock /20 is received



Akamai AS20940 Routing Table	
100.100.96.0/20	AS2002
0.0.0.0/0	AS4003

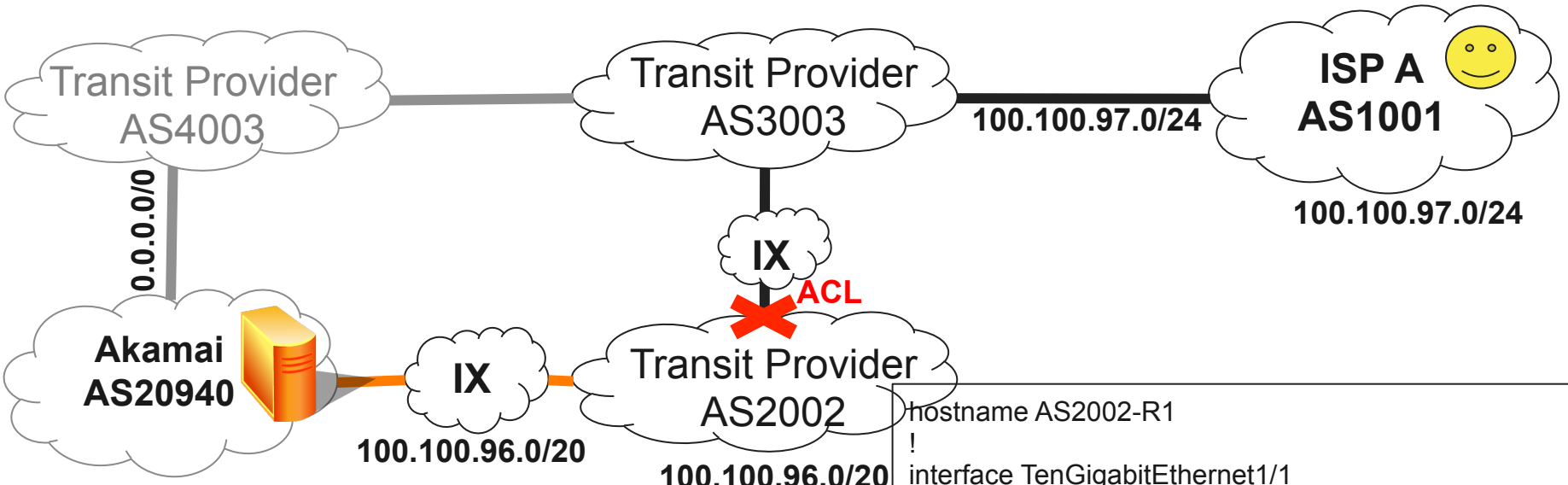
What is the problem?

- Lost of revenue for Transit Provider AS2002 although their backbone is consumed and customer left

- What could happen if AS2002 does not like the peer-to-peer traffic?

Transit Provider AS2002 Filter Traffic on Peer Link

- In order to get rid of peer-to-peer traffic, Transit Provider AS2002 implement an ACL on IX port facing AS3003
- ISP A cannot access some websites due to traffic black hole



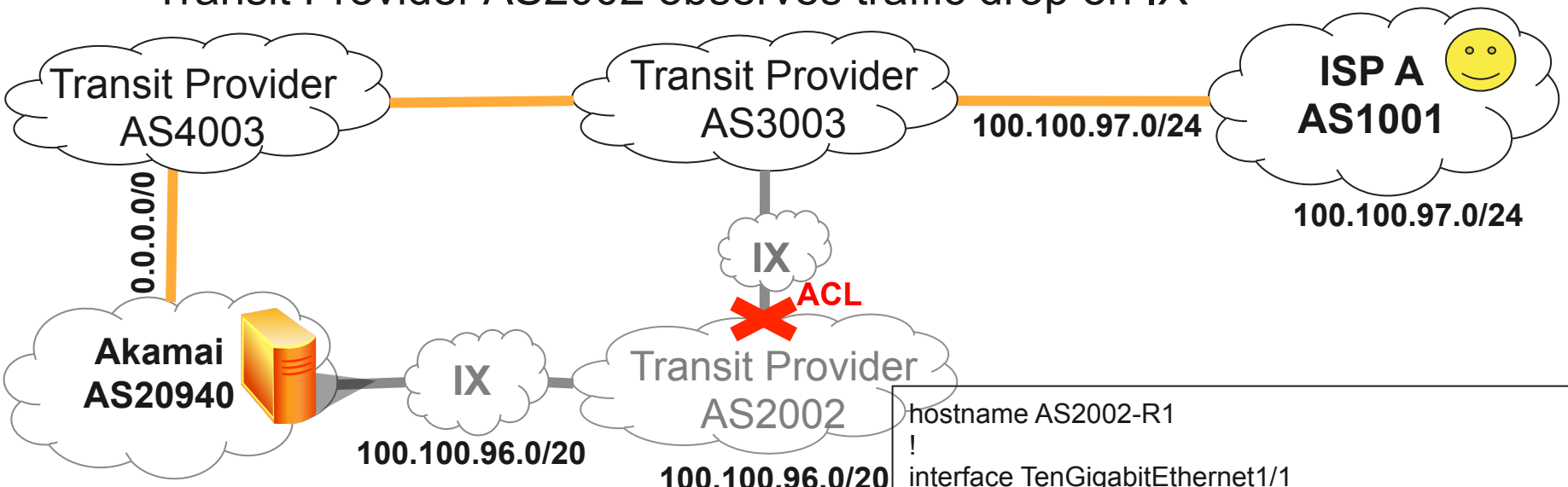
Akamai AS20940 Routing Table

100.100.96.0/20	AS2002
0.0.0.0/0	AS4003

```
hostname AS2002-R1
!
interface TenGigabitEthernet1/1
ip access-group 101 out
!
access-list 101 deny ip any 100.100.97.0 0.0.0.255
access-list 101 permit ip any any
```

Akamai workaround on ISP Traffic Filtering

- Akamai observes ISP A users unable to access some websites
- Akamai blocks all prefixes received from Transit Provider AS2002, so traffic shift from IX to Transit AS4003
- ISP A can access all websites happily
- Transit Provider AS2002 observes traffic drop on IX



Akamai AS20940 Routing Table

~~100.100.96.0/20~~	~~AS2002~~
0.0.0.0/0	AS4003

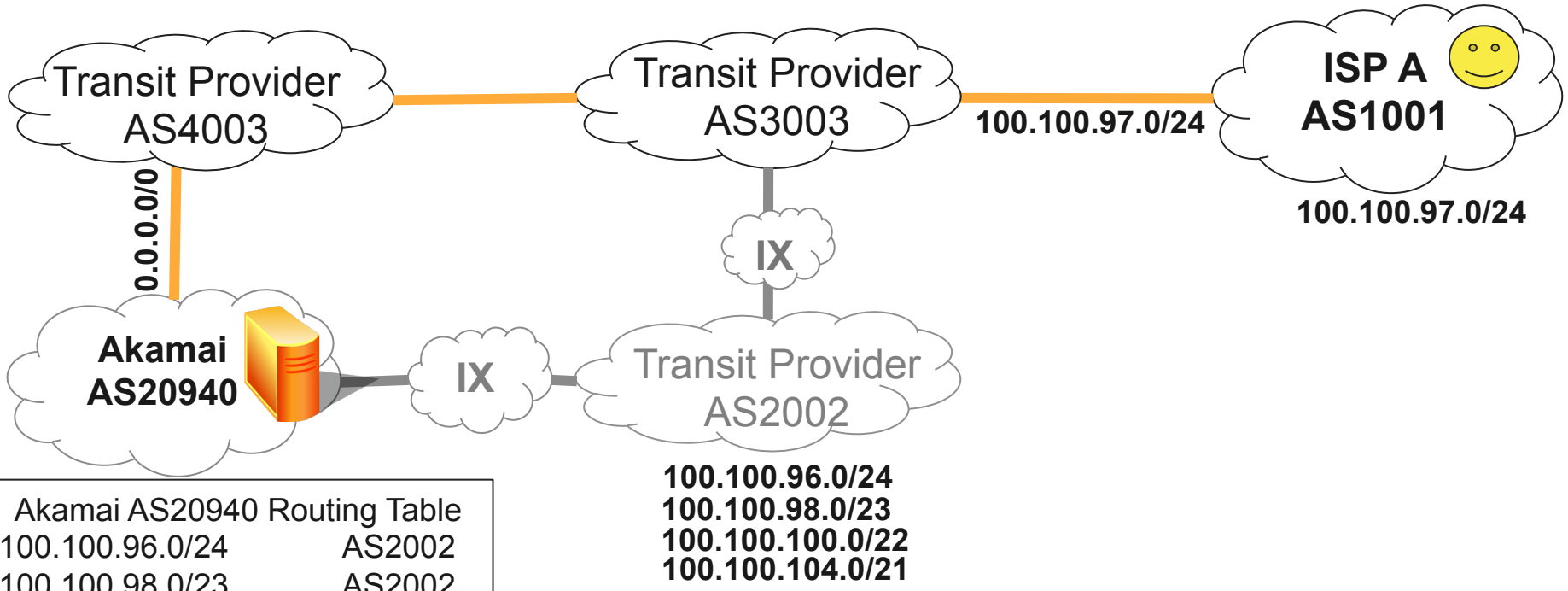
```
hostname AS2002-R1
!
interface TenGigabitEthernet1/1
ip access-group 101 out
!
access-list 101 deny ip any 100.100.97.0 0.0.0.255
access-list 101 permit ip any any
```

Is Traffic Filtering a good workaround?

- It is observed that some Transit Providers filter peer-to-peer traffic on IX port or Private Peer
- If you promised to carry the traffic of a block (eg./20), you should not have any holes (eg. /24) or drop any part of the traffic
- If you assign an IP block (eg. /24) to a customer permanently (eg. Assign Portable), you should not announce the superblock (eg. /20) after customer left
- The end users connectivity will be impacted by your ACL!!!

Ideal Solution

- AS2002 can break the superblock (/20) into sub-blocks
- AS2002 should not announce ISP A prefix



Akamai AS20940 Routing Table	
100.100.96.0/24	AS2002
100.100.98.0/23	AS2002
100.100.100.0/22	AS2002
100.100.104.0/21	AS2002
0.0.0.0/0	AS4003

Conclusions

Summary

Summary

- **Akamai Intelligent Platform**
  - Highly distributed edge servers
  - Akamai mapping is different from BGP routing
- **Peering with Akamai**
  - Improve user experience
  - Reduce transit/peering cost
- **DO and DONTS of Traffic Engineering**
  - Typical Traffic Optimization Techniques has no usual effect
  - Maintain consistent route announcement if possible
  - Maintain complete route announcement is a must
  - Do not filter traffic by ACL

Questions?

Christian Kaufmann <ck@akamai.com>

More information:

Peering: http://as20940.peeringdb.com

Akamai 60sec: http://www.akamai.com/60seconds