International R&E Transit
Defining the Problem, Causes and Potential Solutions

Chris Robb
Indiana University Global Research NOC
TransPAC2 Engineer
Preview

- Technical Background
- The Problem
- Analysis
- Examples
- Potential Solutions
- References
Technical Background
Defining the Problem
Defining the Problem

- For the purposes of this discussion, it will be useful to compare the Commodity internet with the R&E internet
Defining the Problem

- For the purposes of this discussion, it will be useful to compare the Commodity internet with the R&E internet

- Commercial networks have been in existence for years and the economy has been a force that has created more formality and disciplined infrastructure
BGP Review
BGP Review

• BGP = Border Gateway Protocol
BGP Review

- BGP = Border Gateway Protocol
- AS = Autonomous System
  - Abilene is AS 11537
  - Indiana University is AS 87
BGP Review

- BGP = Border Gateway Protocol
- AS = Autonomous System
  - Abilene is AS 11537
  - Indiana University is AS 87
- Each AS controls a certain set of IP addresses as expressed by a network announcement
  - Indiana University owns 129.79.0.0/16
  - Abilene owns 198.32.18.0/22
BGP Review

• BGP = Border Gateway Protocol

• AS = Autonomous System
  • Abilene is AS 11537
  • Indiana University is AS 87

• Each AS controls a certain set of IP addresses as expressed by a network announcement
  • Indiana University owns 129.79.0.0/16
  • Abilene owns 198.32.18.0/22

• When networks interconnect, they use BGP announcements to tell other each other what networks they control
  • Abilene tells Indiana University that it owns 198.32.18.0/22
  • Indiana University tells Abilene that it owns 129.79.0.0/16
Definitions
Definitions

• BGP Peering: the act of exchanging your customer routes with the customer routes of another network
  
  • Typically done between networks of equal size and traffic demands

  • Not a transitive relationship (you only send your customer routes)

  • Also called Settlement-Free Interconnect (SFI) because no money exchanges hands

  • BGP peers are typically multihomed at four or more locations
Definitions

• BGP Peering: the act of exchanging your customer routes with the customer routes of another network
  • Typically done between networks of equal size and traffic demands
  • Not a transitive relationship (you only send your customer routes)
  • Also called Settlement-Free Interconnect (SFI) because no money exchanges hands
  • BGP peers are typically multihomed at four or more locations

• BGP Transit: purchasing access to a network’s entire routing table
  • Typically purchased by a smaller network from a larger network
  • This is a transitive relationship: you receive the other network’s routes, as well as the routes of it’s peer networks’ customers. Conversely, the other network advertises your routes to it’s peers as part if it’s customer base
  • Typically a much more regional relationship with single-homed networks
Traditional Types of Commercial Carriers
Traditional Types of Commercial Carriers

• **Tier 1**
  • Global Connectivity reached without purchasing transit from any other carriers
  • Reach all Internet destinations through SFI peerings with other Tier 1s and by selling service to customers
  • Examples
    • AOL Transit Data Network (ATDN)
    • AT&T
    • Global Crossing
    • Level3
    • Verizon Business
    • NTT
    • Qwest
    • SAVVIS
    • Sprint
Traditional Types of Commercial Carriers

• Tier 1
  • Global Connectivity reached without purchasing transit from any other carriers
  • Reach all Internet destinations through SFI peerings with other Tier 1s and by selling service to customers
  • Examples
    • AOL Transit Data Network (ATDN)
    • AT&T
    • Global Crossing
    • Level3
    • Verizon Business
    • NTT
    • Qwest
    • SAVVIS
    • Sprint

• Tier 2
  • Smaller network providers that need to purchase some or all of their network capacity to get to the global Internet
  • Purchase transit from Tier 1s and resell services to other, smaller Tier 2s or Content Providers
  • Examples
    • Cable companies
    • Regional Bell Operating Companies (RBOCs)
Traditional Types of Commercial Carriers

• Tier 1
  • Global Connectivity reached without purchasing transit from any other carriers
  • Reach all Internet destinations through SFI peerings with other Tier 1s and by selling service to customers
  • Examples
    • AOL Transit Data Network (ATDN)
    • AT&T
    • Global Crossing
    • Level3
    • Verizon Business
    • NTT
    • Qwest
    • SAVVIS
    • Sprint

• Tier 2
  • Smaller network providers that need to purchase some or all of their network capacity to get to the global Internet
  • Purchase transit from Tier 1s and resell services to other, smaller Tier 2s or Content Providers
  • Examples
    • Cable companies
    • Regional Bell Operating Companies (RBOCs)

• Content Providers
  • Network providers that purchase all connectivity from Tier 2s and have no interest in re-selling their capacity
  • Examples
    • Google
    • Microsoft
    • Apple
Traditional Commercial Peering Illustrated
Traditional Commercial Peering Illustrated
Traditional Commercial Peering Illustrated
Traditional Commercial Peering Illustrated
Traditional Commercial Peering Illustrated
Traditional Commercial Peering Illustrated

Content Provider A

Content Provider B

Content Provider C

AS 1

AS 2

AS 3

AS 4

AS 10

AS 11

AS 12

AS 13

AS 100
Traditional Commercial Peering Illustrated

Content Provider A
Content Provider B
Content Provider C

AS 1
AS 2
AS 3
AS 4
AS 10
AS 11
AS 12
AS 13
AS 100
Post-1999 Types of Commercial Carriers
Post-1999 Types of Commercial Carriers

- After 1999, there was a move to a much more distributed model of interconnection
Post-1999 Types of Commercial Carriers

- After 1999, there was a move to a much more distributed model of interconnection.
- Tier2s began to peer directly with each other.
  - especially prevalent in the broadband delivery Tier2s to save on the cost of file sharing traffic.
  - in some cases, those peerings were as pervasive as Tier1 interconnections had traditionally been.
Post-1999 Types of Commercial Carriers

- After 1999, there was a move to a much more distributed model of interconnection
- Tier2s began to peer directly with each other
  - especially prevalent in the broadband delivery Tier2s to save on the cost of file sharing traffic
  - in some cases, those peerings were as pervasive as Tier1 interconnections had traditionally been
- Content Providers began to peer directly with other Content Providers
  - As services and traffic loads increased, Content Providers found it advantageous to directly exchange traffic with other Content Providers
  - Example: Google may want to peer with Microsoft to exchange high volume of e-mail sent between their Gmail and Hotmail users
Post-1999 Commercial Peering Illustrated
Post-1999 Commercial Peering Illustrated

AS 10
AS 11
Content Provider A
Content Provider B
AS 1
Content Provider C
AS 2
AS 3
AS 4
AS 12
AS 13
AS 100
Post-1999 Commercial Peering Illustrated
Post-1999 Commercial Peering Illustrated
Commercial Motives
Commercial Motives
Commercial Motives
Commercial Motives

• Why would a network want to SFI Peer with another network?
  
  • Lowering transit costs: why pay for something you can get for next to nothing?
  
  • Increased control over routes: direct peering always yields more ability to shift traffic around
  
  • Decreased latency for your customers: the shorter distances involved cause TCP backoff algorithms to allow more bandwidth through, hence more traffic from your customers, whom you charge
Commercial Motives

• Why would a network want to SFI Peer with another network?
  • Lowering transit costs: why pay for something you can get for next to nothing?
  • Increased control over routes: direct peering always yields more ability to shift traffic around
  • Decreased latency for your customers: the shorter distances involved cause TCP backoff algorithms to allow more bandwidth through, hence more traffic from your customers, whom you charge

• Why would a network not want to SFI peer with another network?
  • Loss of potential revenue: why give something away that you could charge for?
  • Traffic load assymetry: if the other network has a lot of Content Providers, the assymetric nature of the traffic volume and hot-potato routing methodology may mean that you would bear the greater burden of creating a bidirectional data exchange
  • Peering may strengthen a potential competitor: you may lose customers!
Commercial Peering
Best Practices
Commercial Peering
Best Practices

• Monitor links aggressively for traffic imbalances
Commercial Peering
Best Practices

• Monitor links aggressively for traffic imbalances
• Much investigation into traffic symmetry before and after peering is established
Commercial Peering
Best Practices

- Monitor links aggressively for traffic imbalances
- Much investigation into traffic symmetry before and after peering is established
- Make sure a lawyer is involved. Better yet, a team of lawyers!
Commercial Peering
Best Practices

• Monitor links aggressively for traffic imbalances

• Much investigation into traffic symmetry before and after peering is established

• Make sure a lawyer is involved. Better yet, a team of lawyers!

• Only send your customer routes and not the routes of another peer network
Commercial Peering Best Practices

- Monitor links aggressively for traffic imbalances
- Much investigation into traffic symmetry before and after peering is established
- Make sure a lawyer is involved. Better yet, a team of lawyers!
- Only send your customer routes and not the routes of another peer network
- Do not send subsets of your customer routes to your peers
Commercial Peering
Best Practices

- Monitor links aggressively for traffic imbalances
- Much investigation into traffic symmetry before and after peering is established
- Make sure a lawyer is involved. Better yet, a team of lawyers!
- Only send your customer routes and not the routes of another peer network
- Do not send subsets of your customer routes to your peers
- Advertise the same sets of prefixes to your peers at all peering locations
R&E Peering Best Practices
R&E Peering Best Practices

- Monitor links and interconnects, but not from a contractual violation perspective
R&E Peering Best Practices

- Monitor links and interconnects, but not from a contractual violation perspective
- Flexible “gaming” of the network in the middle of the night to fix a performance problem
R&E Peering Best Practices

- Monitor links and interconnects, but not from a contractual violation perspective
- Flexible “gaming” of the network in the middle of the night to fix a performance problem
- Prefer R&E over Commodity
  - Since almost all R&E networks have both a commodity and high-performance connection, they tend to localpref R&E routes higher
  - Use other methods of route influence (BGP MEDs, localpref from BGP communities, AS path prepending, etc.)
R&E Peering Best Practices

- Monitor links and interconnects, but not from a contractual violation perspective
- Flexible “gaming” of the network in the middle of the night to fix a performance problem
- Prefer R&E over Commodity
  - Since almost all R&E networks have both a commodity and high-performance connection, they tend to localpref R&E routes higher
  - Use other methods of route influence (BGP MEDs, localpref from BGP communities, AS path prepending, etc.)
- Much more informal agreements
R&E Peering Best Practices

- Monitor links and interconnects, but not from a contractual violation perspective
- Flexible “gaming” of the network in the middle of the night to fix a performance problem
- Prefer R&E over Commodity
  - Since almost all R&E networks have both a commodity and high-performance connection, they tend to localpref R&E routes higher
  - Use other methods of route influence (BGP MEDs, localpref from BGP communities, AS path prepending, etc.)
- Much more informal agreements
- More flexibility in advertising networks you don’t own
Commercial vs R&E: Individual Applications
Commercial vs R&E: Individual Applications

- Commercial: “It’s all about the larger network announcement!”
  - Aggregate as much as possible and don’t break up announcements to fix end user problems
  - Don’t really care about asymmetrical paths since the low-performance applications don’t care either
Commercial vs R&E: Individual Applications

- Commercial: “It’s all about the larger network announcement!”
  - Aggregate as much as possible and don’t break up announcements to fix end user problems
  - Don’t really care about asymmetrical paths since the low-performance applications don’t care either
- R&E: “It’s all about the end user experience!”
  - High performance applications are sometimes more susceptible to asymmetrical network paths
  - Routing holes poked in import and export policies to influence a researcher’s traffic to a particular path. Sometimes those holes aren’t as granular as they could be
  - Poor documentation on changes with fewer resources to develop tools
Commercial vs R&E: Infrastructure
Commercial vs R&E: Infrastructure

- Commercial: “Never provision more than your ROI prediction model”
  - No prefixes smaller than /24
  - Larger routing table requires stricter analysis of prefix base
  - Lower speed links from smaller networks
  - For Tier 1s, presence in most major markets
  - Larger set of professional support resources
Commercial vs R&E: Infrastructure

- Commercial: “Never provision more than your ROI prediction model”
  - No prefixes smaller than /24
  - Larger routing table requires stricter analysis of prefix base
  - Lower speed links from smaller networks
  - For Tier 1s, presence in most major markets
  - Larger set of professional support resources

- R&E: “Always overprovision for the snowstorm (AKA, Supercomputing)”
  - Smaller routing table (<10,000) means smaller prefixes are OK and equipment can be more lightweight (which may lead to more bugs!)
  - Large interconnections that are typically underutilized by a factor of 10
  - Limited funding for professional support services (smaller, more efficient requirement for management)
  - Smaller networks peer directly with other smaller networks as well as the larger ones
Pre-1999 International Connection Model
Pre-1999 International Connection Model

- R&E Networks have always been more generous with their transit policies
  - send all or some of non-customer R&E routes at no cost to peers
Pre-1999 International Connection Model

• R&E Networks have always been more generous with their transit policies
  • send all or some of non-customer R&E routes at no cost to peers

• Pre-1999 model was built around network aggregators in each region that would serve as a R&E equivalent to the Commercial Tier1 providers. These aggregators had a few major differences
  • They didn’t have as many connections for multihoming (International circuit were still expensive)
  • They didn’t interconnect in a full mesh (e.g. GEANT didn’t directly connect with APAN)
Pre-1999 International Connection Model

- APAN
- Abilene
- GEANT
- CLARA

- AsiaNet A
- AsiaNet B
- USNet A
- USNet B
- EuroNet A
- EuroNet A
- LatinNet A
- LatinNet A
- LatinNet B
Post-1999 International Connection Model
Post-1999 International Connection Model

Diagram showing connections between various networks such as Abilene, APAN, SINET, JGN2, KREONET2, GLORIAD, CSTNET, CERNET, APAN, TWAREN, GEANT, CLARA, RBNET, TWARAK, and others, with bandwidth labels such as 1G, 10G, 2.5G, 622M, 155M.
Commercial vs R&E: Politics
Commercial vs R&E: Politics

- Commercial: “Money = Politics = Money”
- occasional user community pressure to peer (e.g. Level3 vs. Cogent)
- for the most part laissez-faire
Commercial vs R&E: Politics

- Commercial: “Money = Politics = Money”
  - occasional user community pressure to peer (e.g. Level3 vs. Cogent)
  - for the most part *laissez-faire*
- R&E: “There’s no funding like your funding!”
  - As projects and countries allocate funds for technology upgrades, a network is often built whether it’s needed or not!
  - Carriers often make aggressive discounts for sub-optimal designs in the name of visibility and government cooperation
  - Networks are sometimes provisioned based on regional politics (national pride, geopolitical aversion/affiliation, etc.) and not technical reasons
  - Closely affiliated networks form backup agreements that need to be honored by their peers
Commercial vs R&E: Accountability
Commercial vs R&E: Accountability

• Commercial: “If you break it, you will know”
  • There are a lot of people watching the network in a very distributed fashion
  • Sub-optimal network configuration will have trickle down effects on customers and legally binding documents to back them up
  • Large base of networks participate in NANOG and other regional commercial mailing lists that can be used to publicly debate the actions of a network
Commercial vs R&E:
Accountability

• Commercial: “If you break it, you will know”
  • There are a lot of people watching the network in a very distributed fashion
  • Sub-optimal network configuration will have trickle down effects on customers and legally binding documents to back them up
  • Large base of networks participate in NANOG and other regional commercial mailing lists that can be used to publicly debate the actions of a network

• R&E: “If you break it, someone might notice it...in a few months”
  • There are many fewer engineers watching the network
  • Most networks rely on user-complaints to track down sub-optimal routing
  • Traffic levels are low enough that routing changes are sometimes imperceptible without looking at the routing table by hand
  • Advanced services have a smaller base of tools to monitor them (multicast, IPv6, MPLS)
  • Example: until Abilene purchased IPv6 transit from commercial carriers in the US, it was using an Asian Pacific network to reach US-based IPv6 sites
The Problem
Recent Observations
Recent Observations

• Over the past few years, the Global Research NOC and our peers have noticed an increase in “troublesome” network paths
Recent Observations

- Over the past few years, the Global Research NOC and our peers have noticed an increase in “troublesome” network paths
- Multicast RPF failures on the rise
Recent Observations

- Over the past few years, the Global Research NOC and our peers have noticed an increase in “troublesome” network paths
- Multicast RPF failures on the rise
- More and more exceptions being made in the middle of the night
Recent Observations

- Over the past few years, the Global Research NOC and our peers have noticed an increase in “troublesome” network paths
- Multicast RPF failures on the rise
- More and more exceptions being made in the middle of the night
Problem Definition
Problem Definition

- The best practices laid out earlier may or may not cause significant problems
Problem Definition

• The best practices laid out earlier may or may not cause significant problems

• How do you define what the problem is before it changes?
Problem Definition

• The best practices laid out earlier may or may not cause significant problems

• How do you define what the problem is before it changes?

• A network is a collection of many other networks. How granular do you look at the problem
  • By Continent?
  • By Country?
  • By Organization?
  • By Institution?
  • By Application?
  • By User?
Analysis
Passive Data Collection
Passive Data Collection

• Once you’ve defined the scope, how do you get data?
Passive Data Collection

- Once you’ve defined the scope, how do you get data?
- Best: you have complete access to every network element in the R&E internet space
Passive Data Collection

- Once you’ve defined the scope, how do you get data?
- Best: you have complete access to every network element in the R&E internet space
- Better: you have looking glass views into one or two routers on a network
Passive Data Collection

• Once you’ve defined the scope, how do you get data?

• Best: you have complete access to every network element in the R&E internet space

• Better: you have looking glass views into one or two routers on a network

• Good: you have pervasive data feeds from each network
Passive Data Collection

• Once you’ve defined the scope, how do you get data?

• Best: you have complete access to every network element in the R&E internet space

• Better: you have looking glass views into one or two routers on a network

• Good: you have pervasive data feeds from each network

• Realistic: you have some data feeds and some looking glass views
Complete Access
Complete Access

- Raise your hand if you want to give me complete access to every network device in your network
Complete Access

• Raise your hand if you want to give me complete access to every network device in your network

• (If your hand is actually up) Raise your hand if you want to give every other person in this room complete access to every network device in your network
Looking Glass
Looking Glass

- Restricted, read-only view of one or two network devices
Looking Glass

• Restricted, read-only view of one or two network devices

• Non-programmatic access to content means that troubleshooting is done by hand
Looking Glass

- Restricted, read-only view of one or two network devices
- Non-programmatic access to content means that troubleshooting is done by hand
- Non-standard interface or network device CLI means that an engineer needs to be familiar with many different vendors
Looking Glass

- Restricted, read-only view of one or two network devices

- Non-programmatic access to content means that troubleshooting is done by hand

- Non-standard interface or network device CLI means that an engineer needs to be familiar with many different vendors

- Typically very slow
Data Feeds
Data Feeds

• Typically a eBGP feed from a single router
Data Feeds

- Typically a eBGP feed from a single router
- Only gives access to a subset of data
  - External BGP only expresses the *best* route, not *all* routes
- IETF draft in the works to fix that
Data Feeds

• Typically a eBGP feed from a single router
• Only gives access to a subset of data
  • External BGP only expresses the best route, not all routes
• IETF draft in the works to fix that
• Uniform data that can be scripted against
Data Feeds

• Typically a eBGP feed from a single router
• Only gives access to a subset of data
  • External BGP only expresses the best route, not all routes
• IETF draft in the works to fix that
• Uniform data that can be scripted against
• Typically in a central location with one interface to learn
The Reality
The Reality

• None of this is in place today
The Reality

• None of this is in place today

• Mixture of each methodology means everything needs to be done by hand
The Reality

- None of this is in place today
- Mixture of each methodology means everything needs to be done by hand
- Some networks completely opaque and can’t be analyzed directly
Active Measurements
Active Measurements

• Probe the network with ICMP traceroutes
  • difficult when networks block ICMP
  • need a good set of endpoints that fully represent a network (and don’t move)
  • large volume of data
Active Measurements

• Probe the network with ICMP traceroutes
  • difficult when networks block ICMP
  • need a good set of endpoints that fully represent a network (and don’t move)
  • large volume of data

• Active BGP announcements
  • Craft BGP announcements in a way that network convergence yields information about “hidden” links
  • Controversial
Analyzing the Data
Analyzing the Data

• Need to visualize the data
Analyzing the Data

• Need to visualize the data
• Who watches the network?
Analyzing the Data

- Need to visualize the data
- Who watches the network?
- Who stores the data?
Analyzing the Data

- Need to visualize the data
- Who watches the network?
- Who stores the data?
- How do you account for large topology changes?
Example
Michigan is in Korea?
Michigan is in Korea?

- Recent example from Wednesday, April 19th
Michigan is in Korea?

- Recent example from Wednesday, April 19th
- I added a new prefix for a Michigan network to the Abilene filters
Michigan is in Korea?

- Recent example from Wednesday, April 19th
- I added a new prefix for a Michigan network to the Abilene filters
- When I looked at the route on Abilene, the route was preferred via TransPAC2 rather than via the direct US-peering
Michigan is in Korea?

- Recent example from Wednesday, April 19th
- I added a new prefix for a Michigan network to the Abilene filters
- When I looked at the route on Abilene, the route was preferred via TransPAC2 rather than via the direct US-peering
- Routing table during problem:

```plaintext
chrobb@IPLSng-re0# run show route 198.110.96.0
inet.0: 9808 destinations, 15350 routes (9775 active, 0 holddown, 35 hidden)
+ = Active Route, - = Last Active, * = Both
198.110.96.0/20    *[BGP/170] 00:45:42, MED 100, localpref 160, from 198.32.8.198
    AS path: 22388 7660 9270 9270 9270 17579 22335 237 I
    > to 198.32.8.81 via so-3/2/0.0
198.110.96.0/20    *[BGP/170] 00:29:46, MED 10, localpref 140
    AS path: 237 I
    > to 192.122.183.9 via so-2/1/2.512
```
Michigan is in Korea? (cont.)
Michigan is in Korea? (cont.)

- MREN, the network provider for the Michigan network, also has a peering with KREONET2 in Chicago. They advertise the route to them, as they should.
• MREN, the network provider for the Michigan network, also has a peering with KREONET2 in Chicago. They advertise the route to them, as they should.

• KREONET2 has a peering with the KOREN network in Korea. They have a transit agreement in place where KOREN uses KREONET2 to get to the US.
Michigan is in Korea? (cont.)

- MREN, the network provider for the Michigan network, also has a peering with KREONET2 in Chicago. They advertise the route to them, as they should.

- KREONET2 has a peering with the KOREN network in Korea. They have a transit agreement in place where KOREN uses KREONET2 to get to the US.

- KOREN sends the route to the APAN network, who is only expecting to get KOREN customer routes
Michigan is in Korea? (cont.)

- MREN, the network provider for the Michigan network, also has a peering with KREONET2 in Chicago. They advertise the route to them, as they should.

- KREONET2 has a peering with the KOREN network in Korea. They have a transit agreement in place where KOREN uses KREONET2 to get to the US.

- KOREN sends the route to the APAN network, who is only expecting to get KOREN customer routes.

- APAN sends the route to TransPAC2 in Los Angeles. The route is given a BGP community tag of 11537:160, which is how APAN ensures that TransPAC2 is the preferred path for APAN prefixes. (see below)
Michigan is in Korea? (cont.)

- MREN, the network provider for the Michigan network, also has a peering with KREONET2 in Chicago. They advertise the route to them, as they should.

- KREONET2 has a peering with the KOREN network in Korea. They have a transit agreement in place where KOREN uses KREONET2 to get to the US.

- KOREN sends the route to the APAN network, who is only expecting to get KOREN customer routes

- APAN sends the route to TransPAC2 in Los Angeles. The route is given a BGP community tag of 11537:160, which is how APAN ensures that TransPAC2 is the preferred path for APAN prefixes. (see below)

- TransPAC2 passes the route to Abilene
MREN, the network provider for the Michigan network, also has a peering with KREONET2 in Chicago. They advertise the route to them, as they should.

KREONET2 has a peering with the KOREN network in Korea. They have a transit agreement in place where KOREN uses KREONET2 to get to the US.

KOREN sends the route to the APAN network, who is only expecting to get KOREN customer routes.

APAN sends the route to TransPAC2 in Los Angeles. The route is given a BGP community tag of 11537:160, which is how APAN ensures that TransPAC2 is the preferred path for APAN prefixes. (see below)

TransPAC2 passes the route to Abilene.

Abilene, seeing the tag 11537:160, localpreferences the route at 160, which is higher than the standard International route localpref of 100.
• MREN, the network provider for the Michigan network, also has a peering with KREONET2 in Chicago. They advertise the route to them, as they should.

• KREONET2 has a peering with the KOREN network in Korea. They have a transit agreement in place where KOREN uses KREONET2 to get to the US.

• KOREN sends the route to the APAN network, who is only expecting to get KOREN customer routes.

• APAN sends the route to TransPAC2 in Los Angeles. The route is given a BGP community tag of 11537:160, which is how APAN ensures that TransPAC2 is the preferred path for APAN prefixes. (see below)

• TransPAC2 passes the route to Abilene.

• Abilene, seeing the tag 11537:160, localpreferences the route at 160, which is higher than the standard International route localpref of 100.

• MREN had tagged the route with 11537:140, which makes Abilene localpref the route lower than the standard US-network route localpref of 200. This was done because MREN has backup agreements in place with other US networks and want this route preferred lower than other paths.
Michigan is in Korea? (cont.)
Michigan is in Korea? (cont.)
Michigan is in Korea? (cont.)
Michigan is in Korea? (cont.)

MREN peers with KREONET2 at Starlight in Chicago
Michigan is in Korea? (cont.)

MREN peers with KREONET2 at Starlight in Chicago

KREONET2 has a transit agreement with KOREN

AS 17579 KREONET2

AS 237 Merit

AS 2335 MREN

AS 22388 TransPAC2

AS 22335

AS 7660 APAN

AS 11537 Abilene

AS 9270 Koren

AS 7660

AS 1537
Michigan is in Korea? (cont.)

KREONET2 has a transit agreement with KOREN

KOREN sends customer and peer routes to APAN

KREONET2 peers with KOREN at Starlight in Chicago

KREONET2 has a transit agreement with KOREN

KOREN sends customer and peer routes to APAN

AS 237 Merit

AS 17579 KREONET2

MREN peers with KREONET2 at Starlight in Chicago

AS 22335 MREN

AS 9270 Koren

AS 7660 APAN

AS 22388 TransPAC2

AS 11537 Abilene
Michigan is in Korea? (cont.)

KREONET2 has a transit agreement with KOREN

AS 9270 Koren

KOREN sends customer and peer routes to APAN

AS 7660 APAN

APAN sends its peer routes to TransPAC2 with a BGP community of 11537:160

AS 22388 TransPAC2

MREN peers with KREONET2 at Starlight in Chicago

AS 17579 KREONET2

AS 22335 MREN

AS 237 Merit

AS 11537 Abilene

KREONET2 has a transit agreement with KOREN

KOREN sends customer and peer routes to APAN

APAN sends its peer routes to TransPAC2 with a BGP community of 11537:160
Michigan is in Korea? (cont.)

- **KREONET2** has a transit agreement with **KOREN**
- **AS 9270** Koren
- **AS 17579** KREONET2
  - MREN peers with KREONET2 at Starlight in Chicago
- **AS 22335** MREN
- **AS 22388** TransPAC2
  - TransPAC2 passes all APAN-learned routes to Abilene in Los Angeles
- **AS 7660** APAN
  - **AS 22388** TransPAC2
  - APAN sends its peer routes to TransPAC2 with a BGP community of 11537:160
- **AS 11537** Abilene
- **AS 237** Merit
  - KOREN sends customer and peer routes to APAN
  - APAN sends its peer routes to TransPAC2 with a BGP community of 11537:160
Michigan is in Korea? (cont.)
Michigan is in Korea? (cont.)

- So, what knowledge and access did I need to figure this out?
  - Access to Abilene and TransPAC2 routers
  - Detailed knowledge of their BGP communities, export policies and local preference settings
  - Knowledge of KREONET2 and KOREN’s transit policy
  - Knowledge that APAN trusts KOREN to send their customer routers
Michigan is in Korea? (cont.)

So, what knowledge and access did I need to figure this out?

- Access to Abilene and TransPAC2 routers
- Detailed knowledge of their BGP communities, export policies and localpreference settings
- Knowledge of KREONET2 and KOREN’s transit policy
- Knowledge that APAN trusts KOREN to send their customer routers

That’s a tall order for someone who doesn’t deal with these networks day-in, day-out
Solutions
Solutions
Solutions

• More rational interconnection agreements amongst international peers
Solutions

• More rational interconnection agreements amongst international peers
• Regional aggregation and fuller mesh of "Tier I" R&E peerings
Solutions

- More rational interconnection agreements amongst international peers
- Regional aggregation and fuller mesh of "Tier 1" R&E peerings
- R&E routing mailing list (NANOG-like)
Solutions

- More rational interconnection agreements amongst international peers
- Regional aggregation and fuller mesh of "Tier 1" R&E peerings
- R&E routing mailing list (NANOG-like)
- Stricter filtering (based on route registries?)
Solutions

- More rational interconnection agreements amongst international peers
- Regional aggregation and fuller mesh of "Tier 1" R&E peerings
- R&E routing mailing list (NANOG-like)
- Stricter filtering (based on route registries?)
- MPLS for traffic exceptions
Solutions

• Formalized backup agreements
• Reduction of assymetric routing policies (e.g. Internet2 Fednet-ITN transit)
• Removal of localpref controls in favor of BGP meds and AS-path prepending
• Formal communication and separation of customer routes from transit routes
• Responsive NOC-NOC communication
• MORE DEBUG TOOLS NEEDED!!
References
References

- "AOL Transit Data Network Settlement-Free Interconnection Policy" http://www.atdn.net/settlement_free_int.shtml