# Networking: Routing (BGP)

Oct. 18th, 2021

Alice Gao, Meng-Li Shih

# Outline

- Recap. “Interdomain Internet Routing”
  - Discuss related security and stability issues of BGP.
- Facebook Outage

# Internet Service Providers (ISP)

- Provides services for accessing, using, or participating in the Internet
- Categorized by size
  - Tier-3 ISP (Small):
    - Own a small number of localized end-customers
  - Tier-2 ISP (Medium):
    - Regional scope (e.g. State-wide, Region-wide)
  - Tier-1 ISP (Really huge):
    - Routing tables actually have routes to all currently reachable Internet
    - (e.g. AT&T, T-Mobile, NTT … etc.)
    - Own multiple **Autonomous Systems.**

# Autonomous Systems (AS)

- A collection of connected Internet Protocol (IP) routing prefixes
  - Under the control of one or more network operators
  - On behalf of a single administrative entity
- Autonomous Systems Number
  - A 16/32-bit number
  - Identify a certain AS
- Communicate between each other through **Border Gateway Protocol (BGP)**
- Relationship with other ASes
  - **Peering**
  - **Transit**

# Relationships between ASes

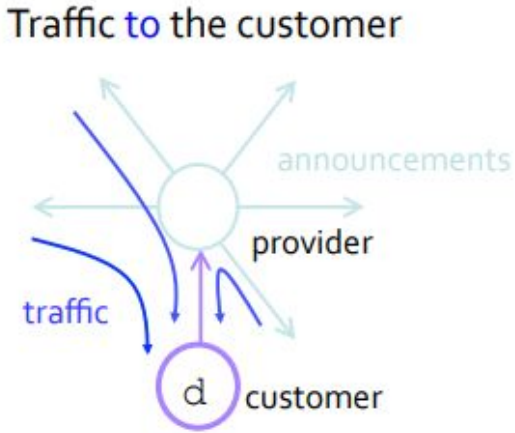
- Peering
  - An AS lets its peer reach (only) its customers
  - The relationship is settlement-free (i.e., no $$)
- Transit (Customer-Provider Relationship)
  - Customer needs to be reachable from and reach to everyone.
  - Provider → reachability → Customer
  - Provider ← $$ ← Customer

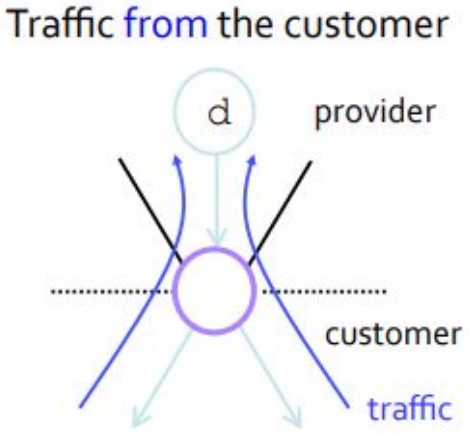
# Export Routes: Route Filtering

- Principle:
  - No ISP wants to act as transit for packets that it isn’t making money on.

# Export Routes: Route Filtering (Customer-Provider)

- Principle:
  - No ISP wants to act as transit for packets that it isn't somehow making money on.
- Customer needs to be reachable from everyone
  - Provider tells all its neighbors how to reach the customer
- Customer does not want to provide transit service
  - Customer does not let its providers route through it




18-PolicyRouting.pptx (umich.edu)

# Export Routes: Route Filtering (Peers)

- Peers exchange traffic between customers
  - AS exports only customer routes to a peer
  - AS exports a peer's routes only to its customers
  - Often the relationship is settlement-free

Discussion: https://tinyurl.com/cse550au21-bgp

# Importing Routes

- AS receive multiple routes, decide which route to install in forwarding table.
- One important factor (Preference):
  - Customer
    Ensure packets to the customer do not traverse additional ASes unnecessarily.
  - Peer
    Exchange reachability Information about mutual transit customers.
  - Provider
    No responsibility for a provider.

⇒ Customer > Peer > Provider (***Local Preference***)

How would you use import/export policies to influence routing?

Discussion: https://tinyurl.com/cse550au21-bgp

# Border Gateway Protocol

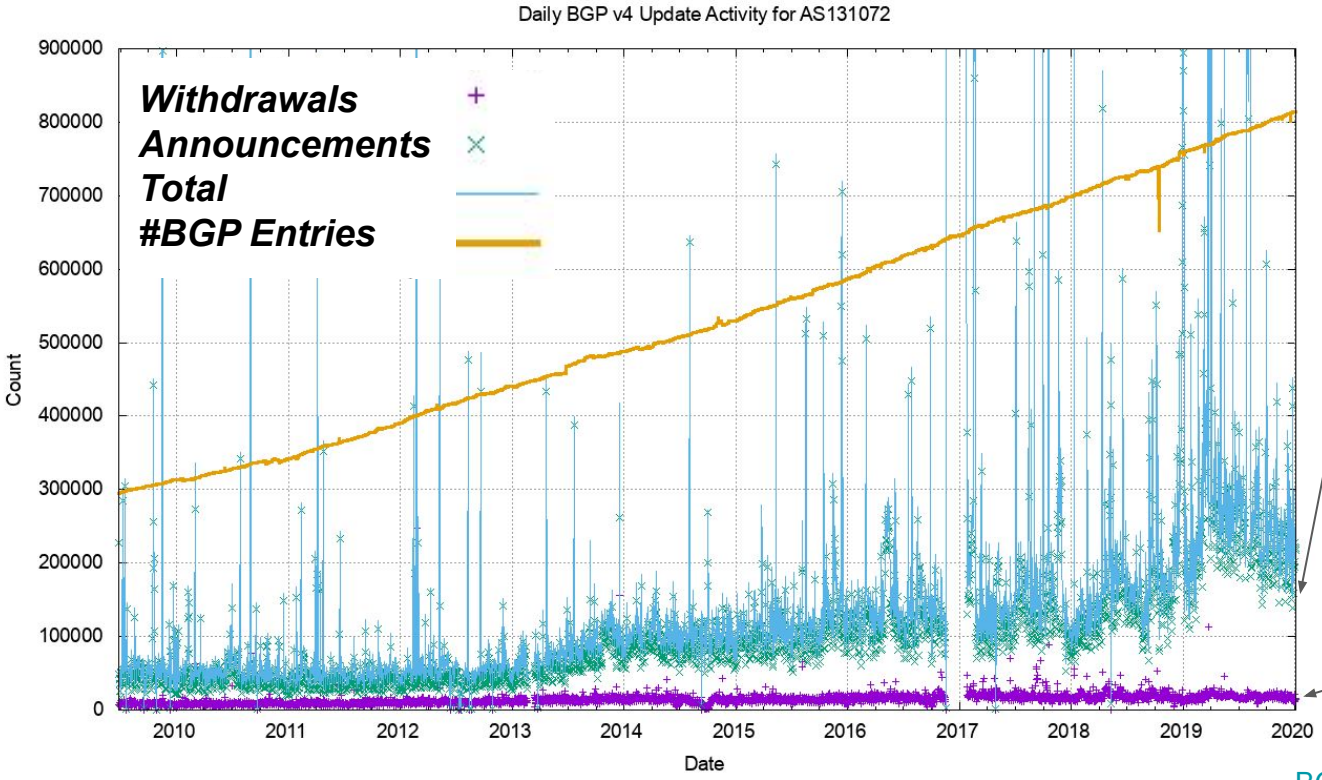
- How an AS communicates with another AS with respect to the relationship.
- Three important needs
  - **Scalability**
    - To ensure that the Internet routing infrastructure remained scalable as the number of connected networks increased
  - **Policy**
    - The ability for each AS to implement various forms of routing policy.
  - **Cooperation under competitive circumstances**
    - No single administrative entity
    - Should allow ASes to make purely local decisions on how to route packets, from among any set of choices

What are other important needs that are not mentioned ? Why ?

# Border Gateway Protocol (Protocol)

- Based on TCP
  - Port 179
- Initialization:
  - Send ***OPEN*** message to other routers
  - Exchange the tables of active routes
- Update:
  - Send “Update” message to other routers
    - ***Announcements***: Changes to existing routes / New routes
    - ***Withdrawals***: Named routes no longer exist
  - No need to be periodically announced
    - Instead, send ***KeepAlive*** message periodically to other routers.
- **Stability ?**

# Stability of BGP

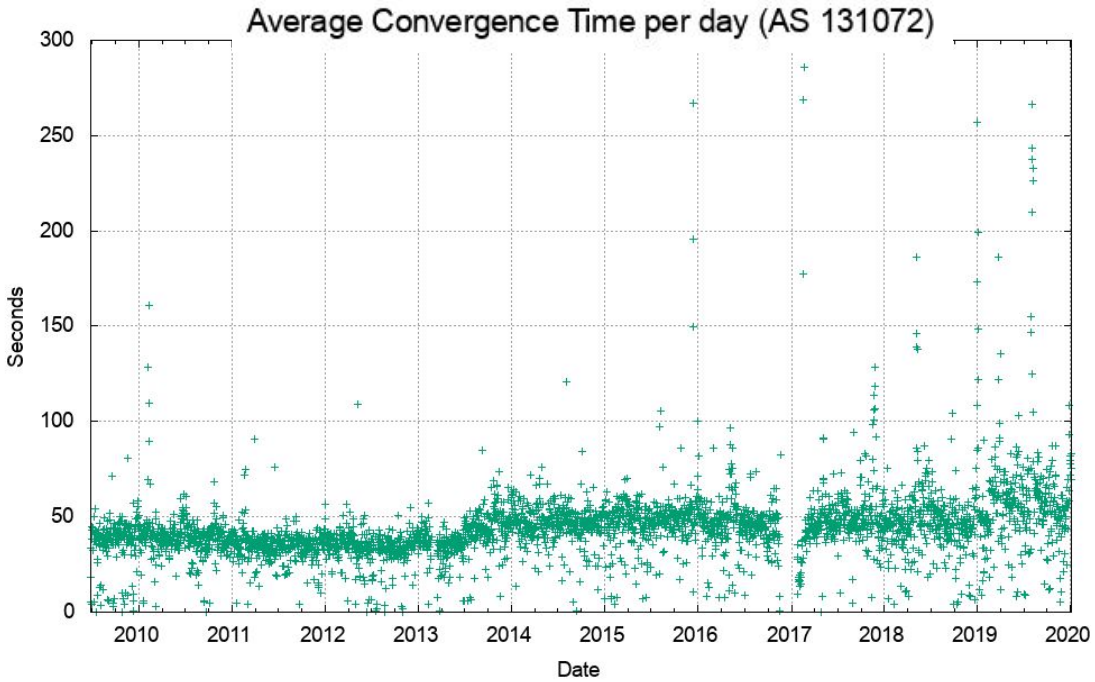



***#Announcements*** message is rising, but the increasing rate is lower than #BGP Entries.

***#Withdrawals*** message is very stable

[BGP in 2019 - BGP Churn | APNIC Blog]

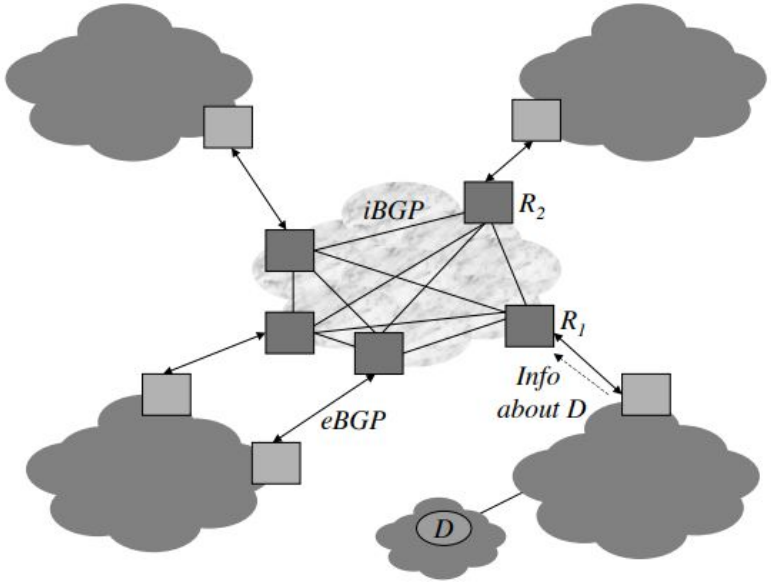
# Stability of BGP (Convergence Time)

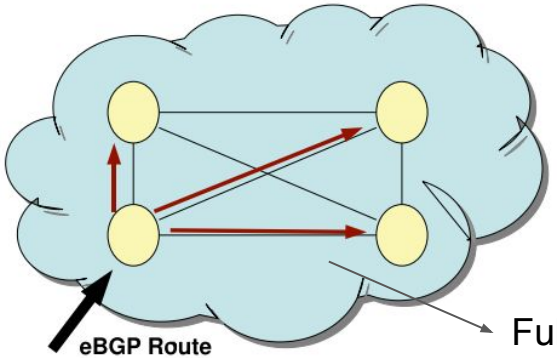



BGP in 2019 - BGP Churn | APNIC Blog

# eBGP and iBGP

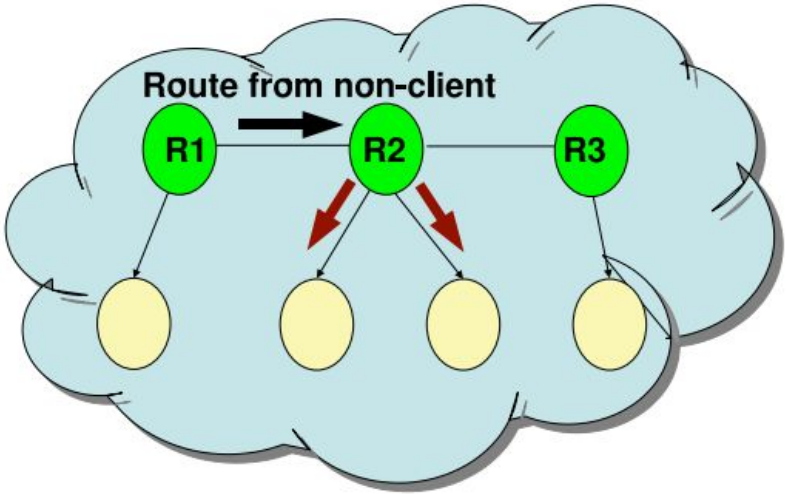
- eBGP:
  - BGP sessions between routers in different ASes
  - One-hop away in IP-level
- iBGP:
  - BGP sessions Between routers in the same AS
  - Loop-free forwarding
  - Complete visibility
  - Multiple hop in IP-level, require to use IGP



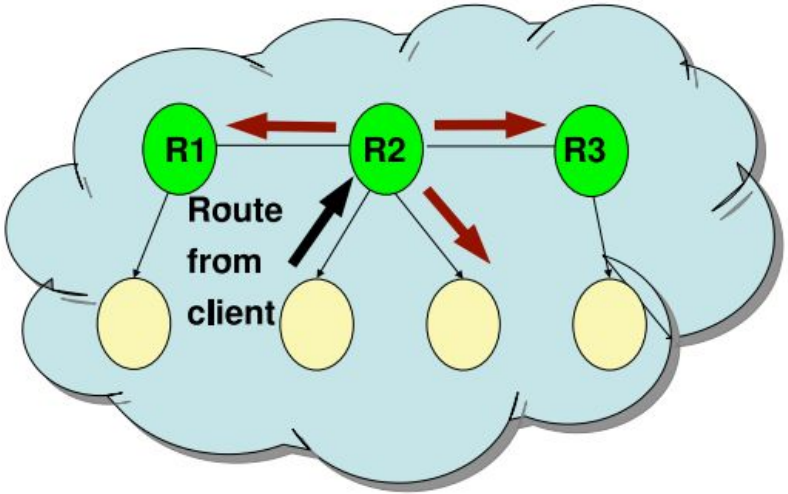




Full-Mesh → #Connections = e*(e-1)/2 → Route Reflector

# Router Reflector




(a) Routes learned from non-clients are re-advertised to clients only.

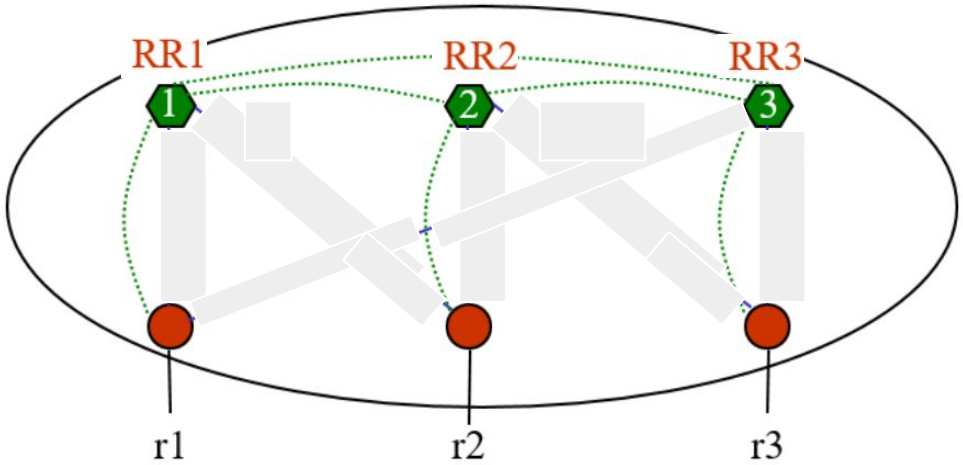



(b) Routes learned from clients are re-advertised over all iBGP sessions.

- Sub-optimal routing, Route oscillation, Increase of BGP convergence time, ...

# Problem of Router Reflector (RR)

- Protocol Oscillation
  - Inconsistency between the metric in IGP (distance) and the metric in BGP (***MED***).

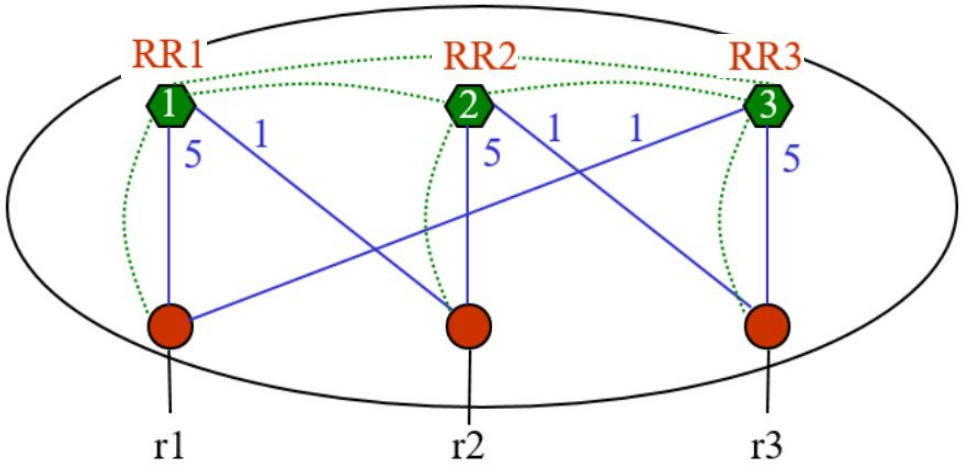



RR1 prefers r2 over r1
RR2 prefers r3 over r2
RR3 prefers r1 over r3

spring05-mar31.ppt (live.com)
BGP Persistent Oscillation – rule 11 reader

# Problems of Router Reflector (RR)

- Protocol Oscillation
  - Inconsistency between the metric in IGP (distance) and the metric in BGP (***MED***).




RR1 prefers r2 over r1
RR2 prefers r3 over r2
RR3 prefers r1 over r3

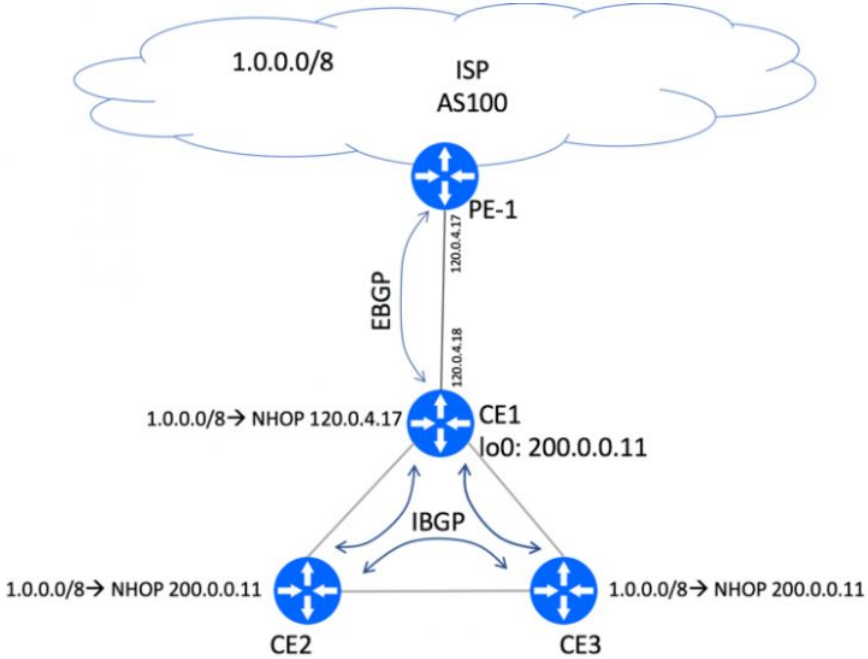
spring05-mar31.ppt (live.com)
BGP Persistent Oscillation – rule 11 reader

# BGP Attributes

- Network operators manipulate route **attributes** when disseminating routes
  - Control how a router ranks candidate routes and select paths to destinations
  - Control the “next hop” IP address for the advertised route to balance load.
- Attributes:
  - ***Next Hop***, ***ASPATH***, ***Local Pref***, ***Multiple-Exit Discriminator (MED)***, ...

# ***NEXT HOP*** Attribute

- IP address of the router to send the packet to




BGP Next-Hop Explained | BGP Help

# ***ASPATH*** Attribute

- A vector that lists all the ASes that this route announcement has been through.
- Loop avoidance:
  - Router checks if its own AS identifier is already in the ***ASPATH***.
    - If it is, discard this announcement
- Help pick a suitable path
  - No ***LOCAL_PREF*** is present → Shorter ***ASPATH*** lengths are preferred
- Security Issue

Discussion: https://tinyurl.com/cse550au21-bgp

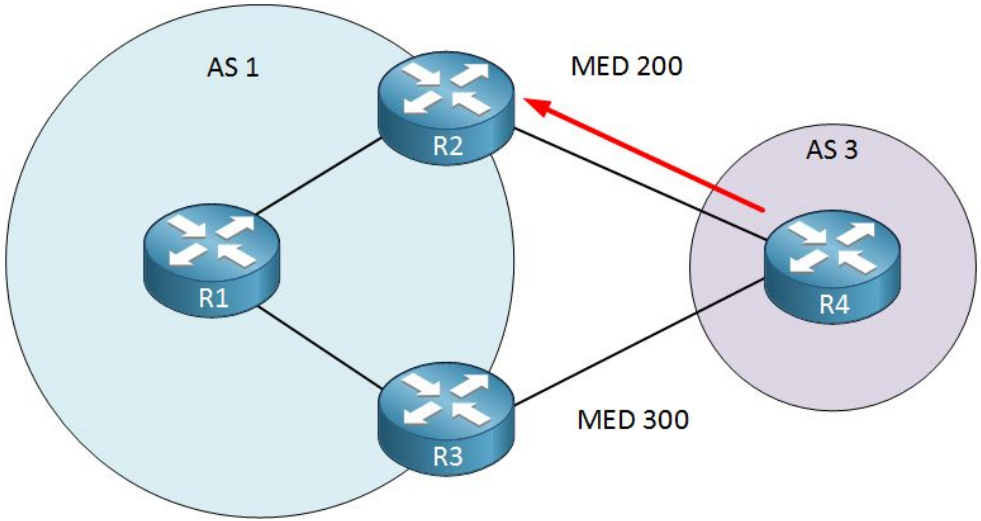
# ***ASPATH*** Attribute (Security Issue)

- Prefix Hijacking
  - An AS announces that it originates a prefix that it does not actually originate.
  - An AS announces a more specific prefix than what may be announced by the true originating AS.
  - An AS announces that it can route traffic to the hijacked AS through a shorter route than is already available, regardless of whether or not the route actually exists.

1. How to solve Prefix Hijacking ?
2. What are other security issues related to BGP ?

# ***MED*** Attributes

- Two ASes are linked at multiple locations
  - How to choose the transit point ?
    - (X) ***LOCAL PREF*** (Cannot distinguished)
    - (X) ***ASPATH*** (Length is equal)
    - (O) ***MED***




How to configure BGP MED Attribute (networklessons.com)

# ***MED*** Attributes

- Two ASes are linked at multiple locations
  - How to choose the transit point ?
    - (X) ***LOCAL PREF*** (Cannot distinguished)
    - (X) ***ASPATH*** (Length is equal)
    - (O) ***MED***
- Two ASes are in peer-peer relationship
  - Ignore ***MED***

    → Sometimes caused *hot-potato problem*

    → Provide incentive to tier-1 ISPs, ask them to carry cross-country packets

Discussion: https://tinyurl.com/cse550au21-bgp

# Put all attributes together

| Priority | Rule | Remarks |
|---|---|---|
| 1 | LOCAL PREF | Highest LOCAL PREF (§4.2.3).<br>*E.g.*, Prefer transit customer routes over peer and provider routes. |
| 2 | ASPATH | Shortest ASPATH length (§4.3.5)<br>*Not* shortest number of Internet hops or delay. |
| 3 | MED | Lowest MED preferred (§4.3.5).<br>May be ignored, esp. if no financial incentive involved. |
| 4 | eBGP > iBGP | Did AS learn route via eBGP (preferred) or iBGP? |
| 5 | IGP path | Lowest IGP path cost to next hop (egress router).<br>If all else equal so far, pick shortest internal path. |
| 6 | Router ID | Smallest router ID (IP address).<br>A random (but unchanging) choice; some implementations use a different tie-break such as the oldest route. |

If you can put one more attribute into BGP protocol,
- What kind of attribute are you going to put ? Why ?

# BGP Security Issues

A modern day horror story...

FACEBOOK

Facebook
@Facebook

We're aware that some people are having trouble accessing our apps and products. We're working to get things back to normal as quickly as possible, and we apologize for any inconvenience.

9:22 AM · Oct 4, 2021 · Twitter Web App

47.2K Retweets 23.4K Quote Tweets 174.9K Likes

Anil Kumar
@anilontwitter

Facebook Instagram, #WhatsApp down
Meanwhile Twitter:👇👇

9:07 AM · Oct 4, 2021 · Twitter Web App

1,184 Retweets 123 Quote Tweets 4,299 Likes


**Twitter** @Twitter

hello literally everyone

8:27 PM · Oct 4, 2021 · Sprinklr

**572.2K** Retweets **183.4K** Quote Tweets **3M** Likes

Tweet your reply

**Instagram** @instagram · 14h
Replying to @Twitter
Hi and happy Monday
9.9K 36.4K 268.1K

Show replies

**WhatsApp** @WhatsApp · 14h
Replying to @Twitter
hello!
30.1K 78.4K 319.5K

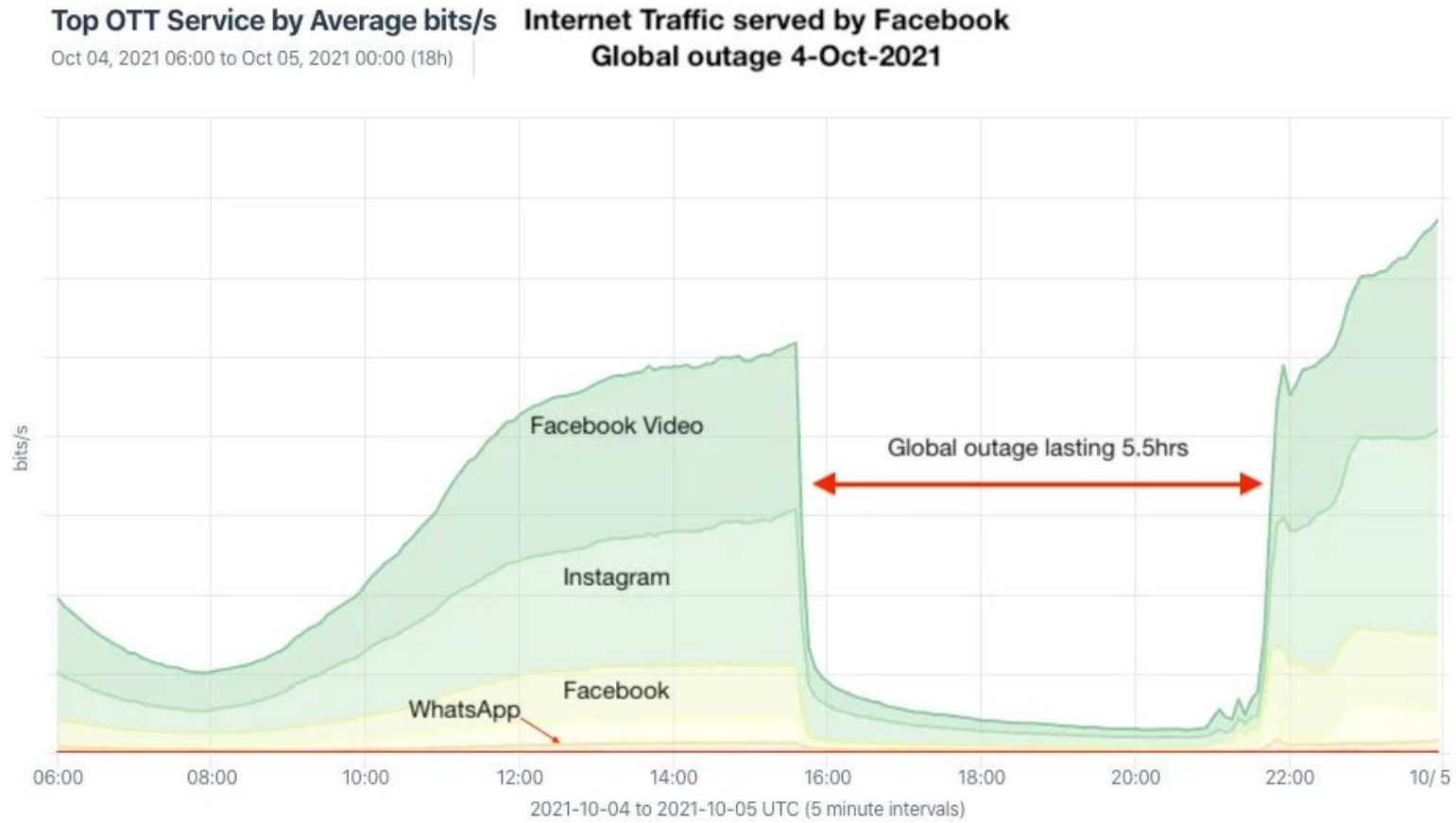
**Microsoft Teams** @MicrosoftTeams · 12h
Replying to @Twitter
*unmutes mic*

Hey everyone.
1K 8.6K 93.4K

**rain ; xiao zhan day !!** @wixearth · 7h
can u pls crash down i dont want to attend classes
15 241 4.3K

**McDonald's** @McDonalds · 14h
Replying to @Twitter
hi what can i get u
1.8K 14.3K 233.8K

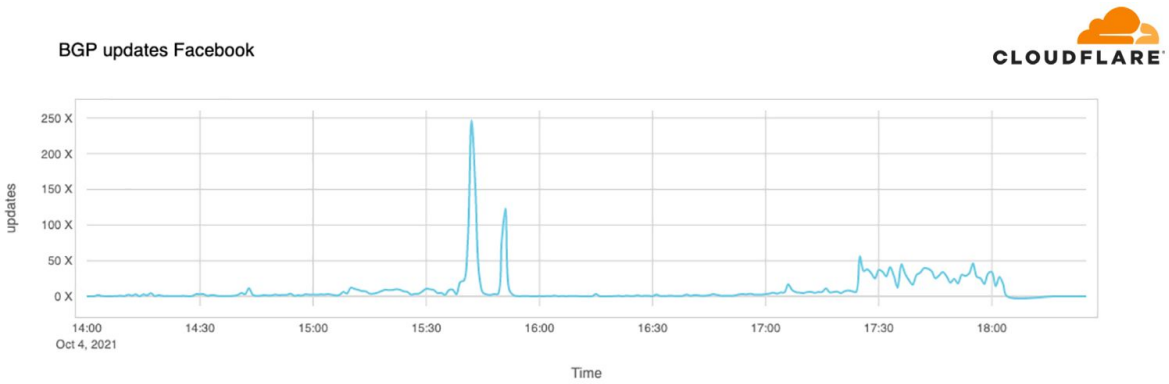
**Twitter** @Twitter · 13h
59.6 million nuggets for my friends
4.3K 33.6K 377.8K


Top OTT Service by Average bits/s
Oct 04, 2021 06:00 to Oct 05, 2021 00:00 (18h)
Internet Traffic served by Facebook
Global outage 4-Oct-2021
bits/s
Facebook Video
Global outage lasting 5.5hrs
Instagram
Facebook
WhatsApp
06:00
08:00
10:00
12:00
14:00
16:00
18:00
20:00
22:00
10/ 5
2021-10-04 to 2021-10-05 UTC (5 minute intervals)

So...what really happened?

# Causes?

- Caused by a loss of IP routes to Facebook DNS (Domain Name Systems)
  - Were all self-hosted at the time
- BGP routing wasn't restored until 21:50 UTC
- DNS services restored at 22:05 UTC
- Application-layer services gradually restored



POSTED ON OCTOBER 4, 2021 TO NETWORKING & TRAFFIC

# Update about the October 4th outage

https://engineering.fb.com/2021/10/04/networking-traffic/outage/

To all the people and businesses around the world who depend on us, we are sorry for the inconvenience caused by today's outage across our platforms. We've been working as hard as we can to restore access, and our systems are now back up and running. The underlying cause of this outage also impacted many of the internal tools and systems we use in our day-to-day operations, complicating our attempts to quickly diagnose and resolve the problem.

Our engineering teams have learned that configuration changes on the backbone routers that coordinate network traffic between our data centers caused issues that interrupted this communication. This disruption to network traffic had a cascading effect on the way our data centers communicate, bringing our services to a halt.

Our services are now back online and we're actively working to fully return them to regular operations. We want to make clear that there was no malicious activity behind this outage — its root cause was a faulty configuration change on our end. We also have no evidence that user data was compromised as a result of this downtime. *(Updated on Oct. 5, 2021, to reflect the latest information)*

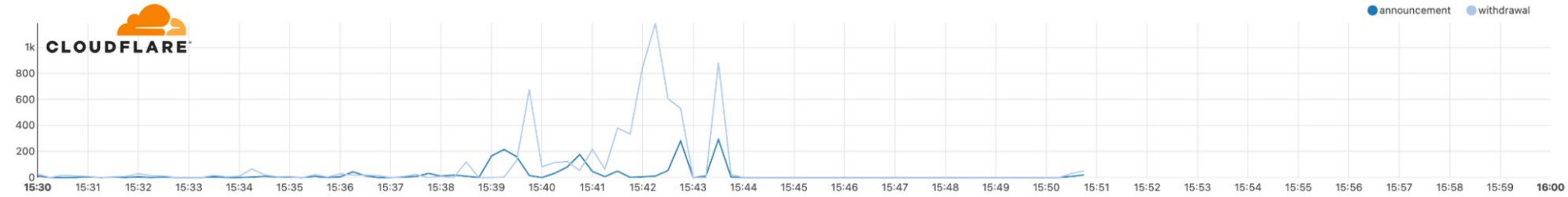
People and businesses around the world rely on us every day to stay connected. We understand the impact that outages like these have on people's lives, as well as our responsibility to keep people informed about disruptions to our services. We apologize to all those affected, and we're working to understand more about what happened today so we can continue to make our infrastructure more resilient.

# According to Facebook Engineering (cont)

The data traffic between all these computing facilities is managed by routers, which figure out where to send all the incoming and outgoing data. And in the extensive day-to-day work of maintaining this infrastructure, our engineers often need to take part of the backbone offline for maintenance — perhaps repairing a fiber line, adding more capacity, or updating the software on the router itself.

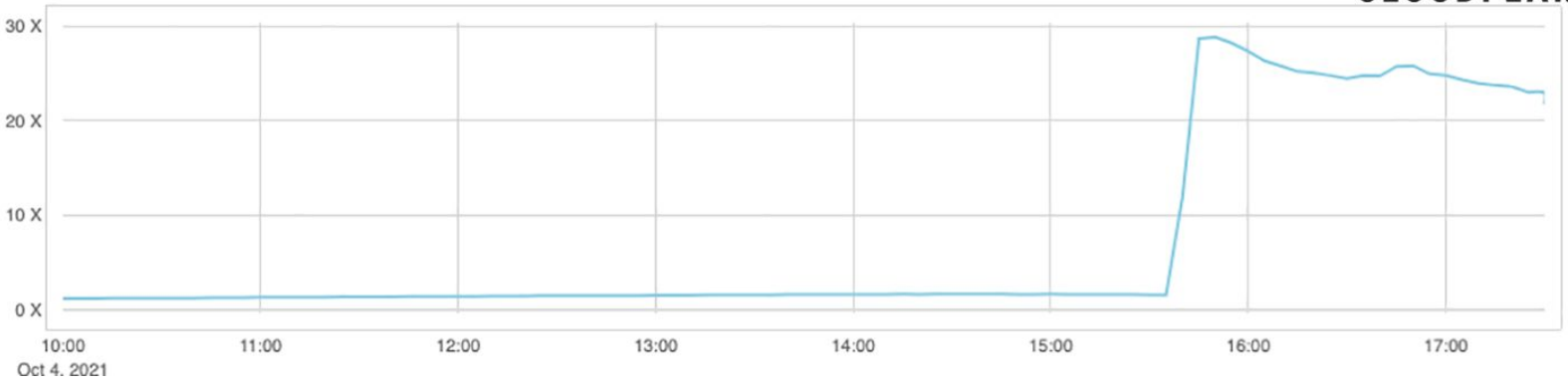
This was the source of yesterday's outage. During one of these routine maintenance jobs, a command was issued with the intention to assess the availability of global backbone capacity, which unintentionally took down all the connections in our backbone network, effectively disconnecting Facebook data centers globally. Our systems are designed to audit commands like these to prevent mistakes like this, but a bug in that audit tool prevented it from properly stopping the command.

https://engineering.fb.com/2021/10/05/networking-traffic/outage-details/

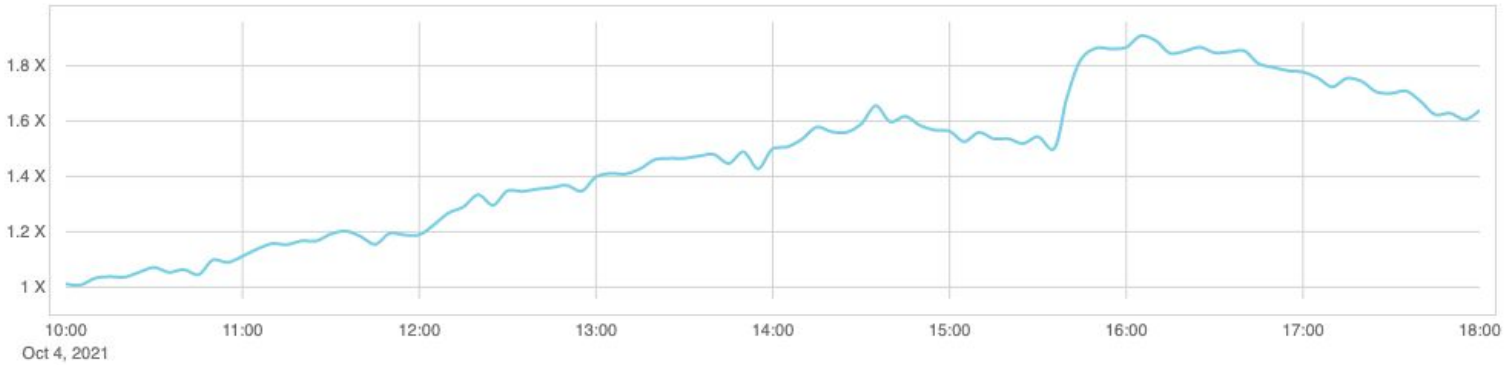



```
➜  ~ dig @1.1.1.1 facebook.com
;; ->>HEADER<<- opcode: QUERY, status: SERVFAIL, id: 31322
;facebook.com.                  IN  A
➜  ~ dig @1.1.1.1 whatsapp.com
;; ->>HEADER<<- opcode: QUERY, status: SERVFAIL, id: 31322
;whatsapp.com.                  IN  A
➜  ~ dig @8.8.8.8 facebook.com
;; ->>HEADER<<- opcode: QUERY, status: SERVFAIL, id: 31322
;facebook.com.                  IN  A
➜  ~ dig @8.8.8.8 whatsapp.com
;; ->>HEADER<<- opcode: QUERY, status: SERVFAIL, id: 31322
;whatsapp.com.                  IN  A
```

Queries for websites: facebook, whatsapp, messenger, instagram




Queries for websites: twitter, signal, telegram, tiktok



**Netflix**
@netflix

When Instagram & Facebook are down.


9:20 AM · Oct 4, 2021

683.5K 3.8K Copy link to Tweet

Tweet your reply

It's not DNS

There's no way it's DNS

It was DNS

# BGP Issues

- Instability - routing tables constantly adjusted to reflect actual changes in network structure
  - Route flapping
- Routing table growth
  - Routers can’t cope with resource requirements
- Load-balancing

# Solutions?

- Cryptographic techniques
  - Pairwise keying, message authentication codes, cryptographic hashes, etc
- Protecting connection between BGP routers
  - Need to protect TCP session
  - *Hop integrity*- peers can detect any modification
- S-BGP validates path attributes in updates

# ...Discuss…!!

- What kind of protocol or security measures (to BGP) could have prevented the Facebook outage of 2021?
  - What modifications to BGP can we have to prevent other catastrophes (Pakistan Youtube outage in 2008, Turkish ISP in 2004, etc)?
- What are some alternative ways to make BGP more secure?