Fighting Route Leaks at Cloudflare

Bryton Herdes
Principal Network Engineer

Mingwei Zhang
Senior Systems Engineer

Agenda

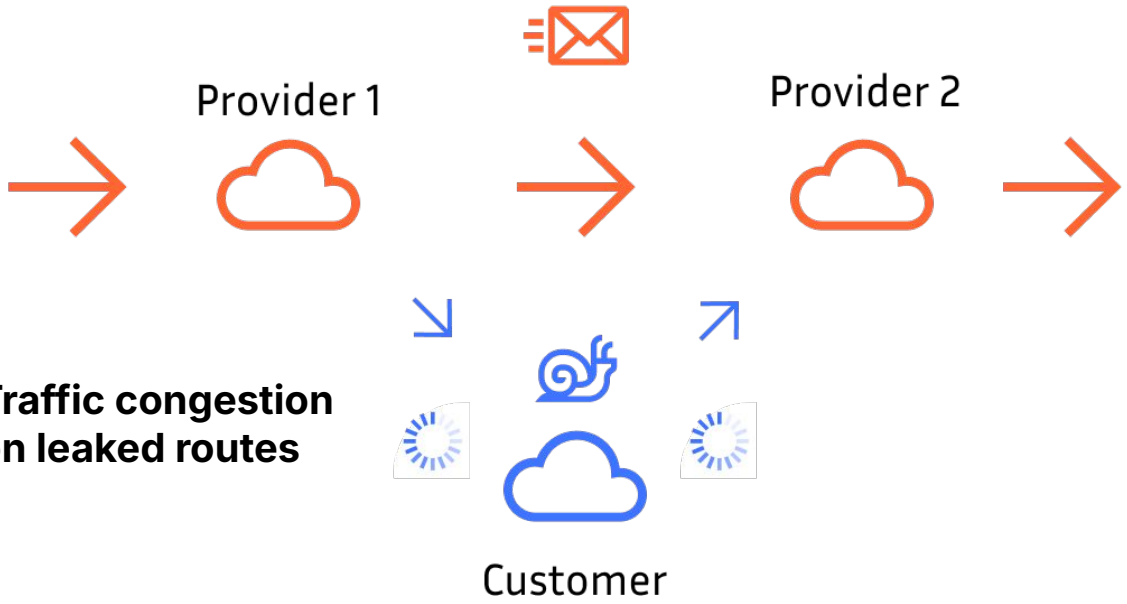
1. What’s a route leak?
2. Complex peering relationships
3. The leak detection pipeline
4. Future impact prevention measures
5. Q&A

What’s a route leak?

Route leaks

- RFC7908
- “A route leak is the propagation of routing announcement(s) beyond their **intended scope**”

Traffic impact



Traffic congestion on leaked routes

Cloudflare's perspective of the October 30 (How [illegible] Optimal [illegible] 1.1.1.1 incident on [illegible] 2024-10-30

[illegible] a BGP [illegible] Large Parts [illegible] Offline Today

Cloudflare 1.1.1.1 June 27, 2024

2024-07-04

Route leak incident on October 2, 2014

2014-10-02

Complex peering relationships

335 cities

Added 19 new cities since Jan 2024. Have 713 data centers, in 128 countries/regions, and AI inference enabled in 197 cities

13,000 networks

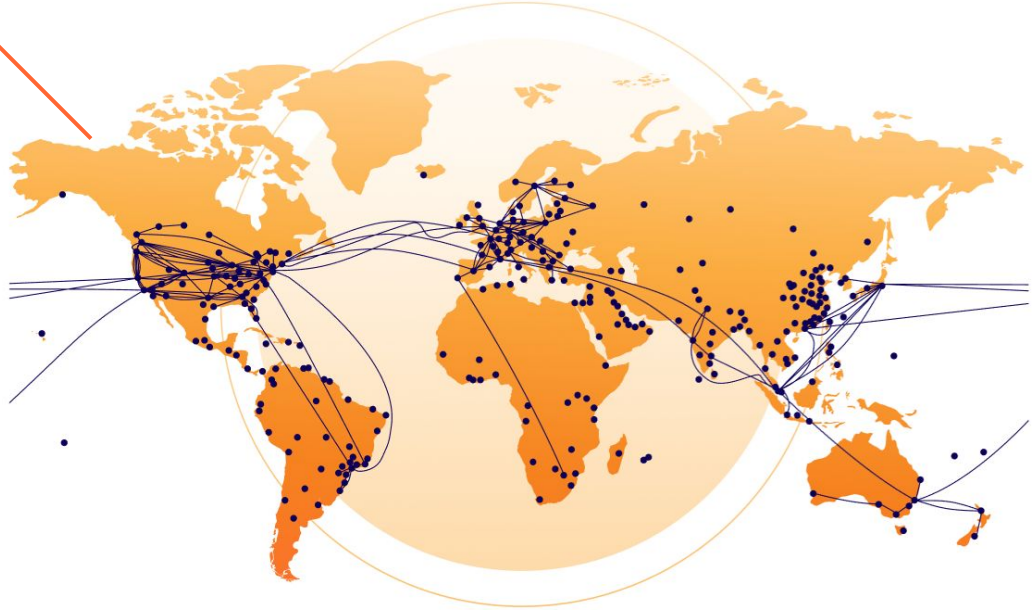
directly connect to Cloudflare, including most major ISPs, cloud providers, and enterprises

348 Tbps

global network edge capacity, consisting of transit connections, peering, and private network interconnects; added 30% capacity in 2024

~50 ms

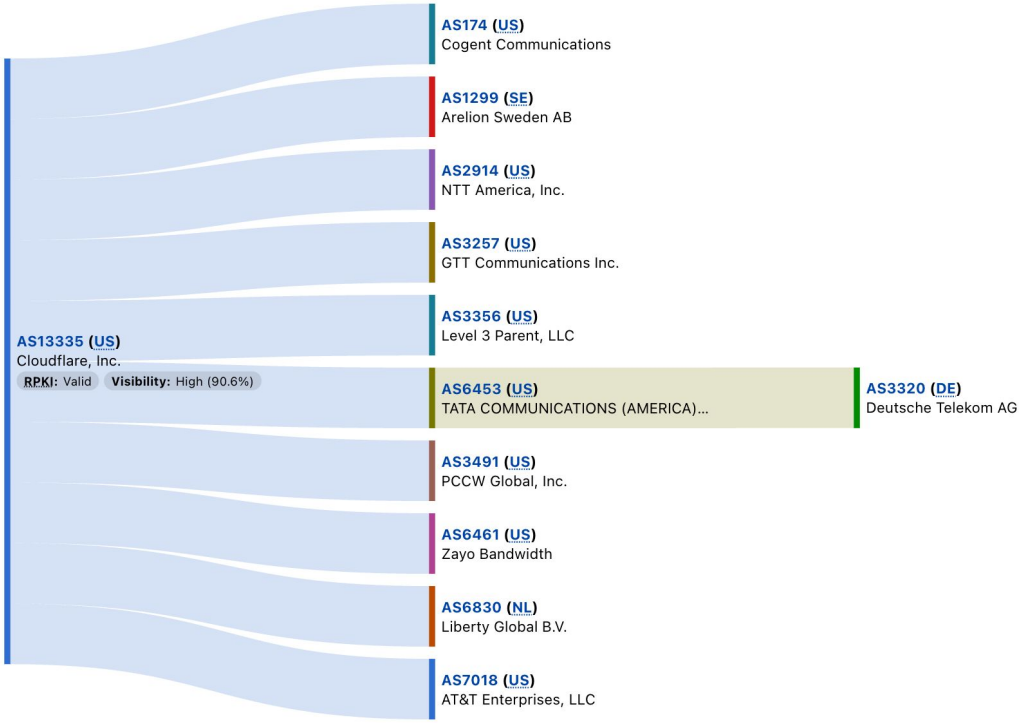
from ~95% of the world's Internet connected population



Cloudflare city (as of Q1 2025)

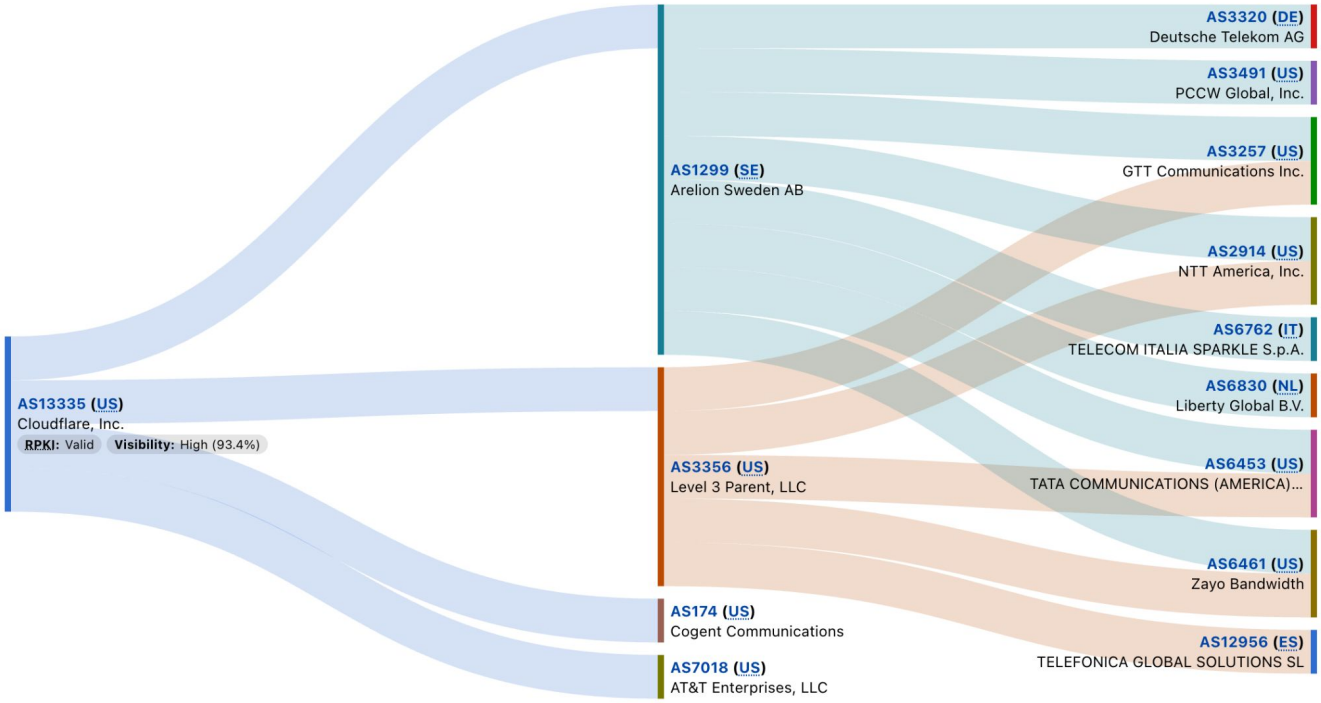
Cloudflare backbone (as of Q1 2025)

Anycast



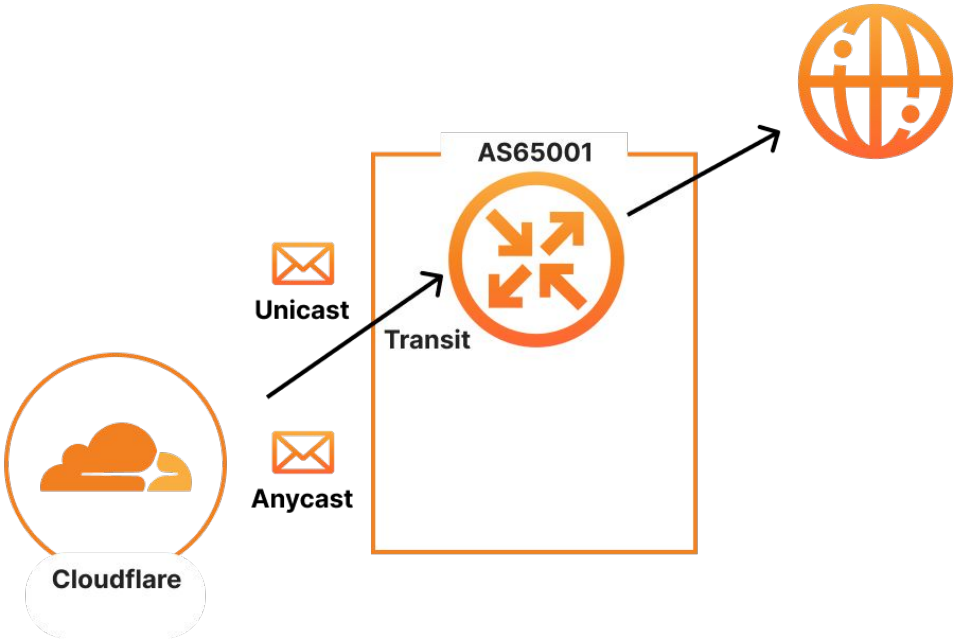
- **Advertised everywhere**
- **Routed to nearest data center**
- **Directly shared with almost every tier-1**

Unicast



- **Originated from single location**
- **Routed to single data center and server**

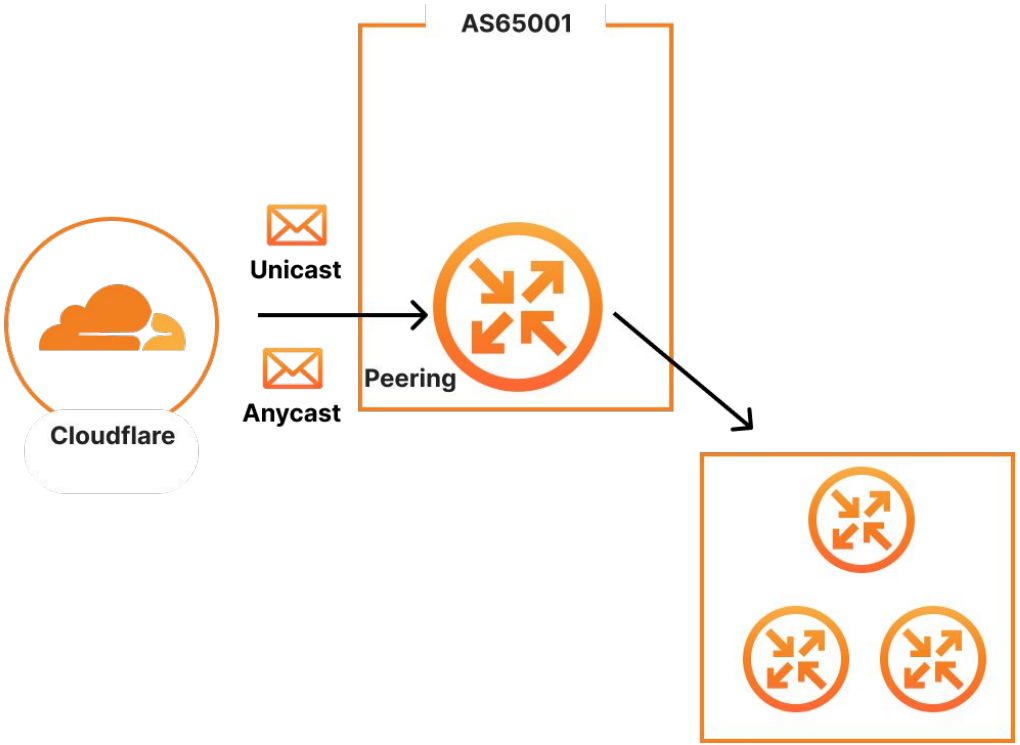
Transit



- Typical provider relationship per data center
- AS65001 advertises our prefixes anywhere and everywhere*

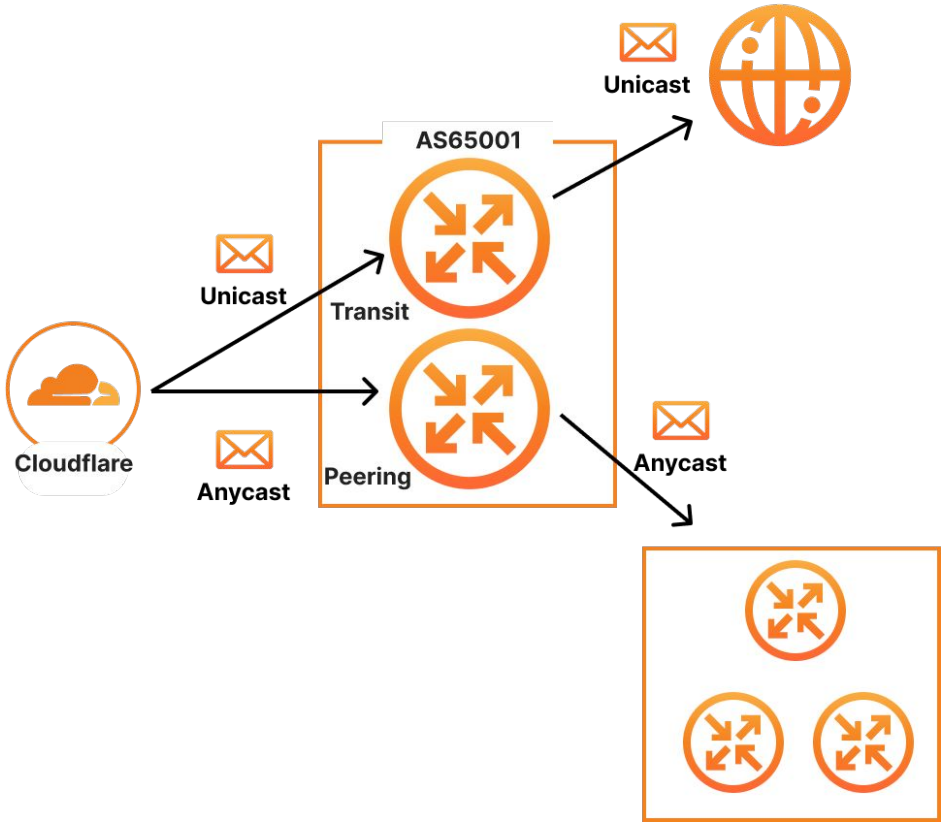
  * - kind of

Peering



- Typical peering relationship
- Advertise our routes only to AS65001 customers
- Peer→Provider propagation is a leak

Mixed transit and peering



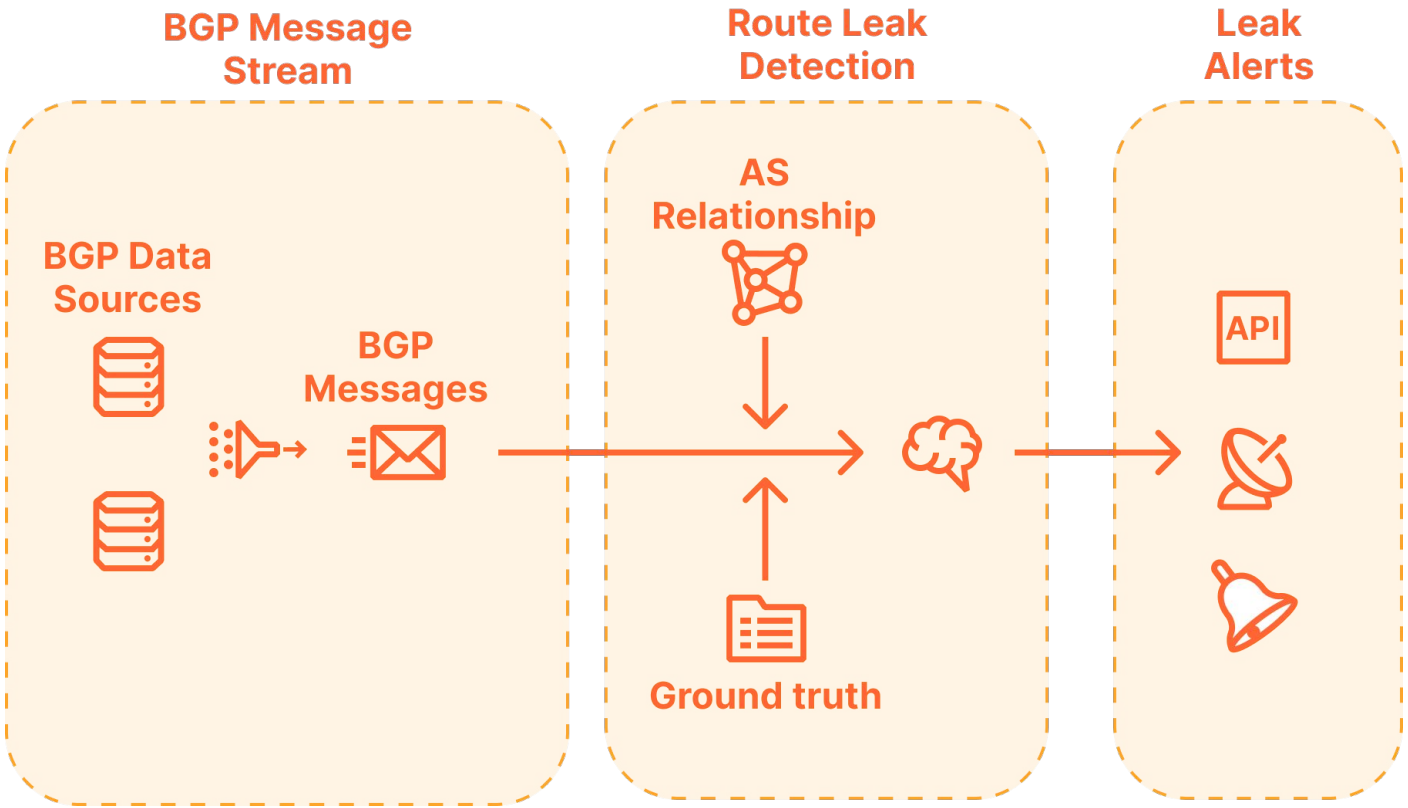
- Send *local* unicast prefixes upstream
- Share anycast prefixes with customers
- Anycast peer→provider propagation is a leak
- Common for embedded cache

Variables to account for

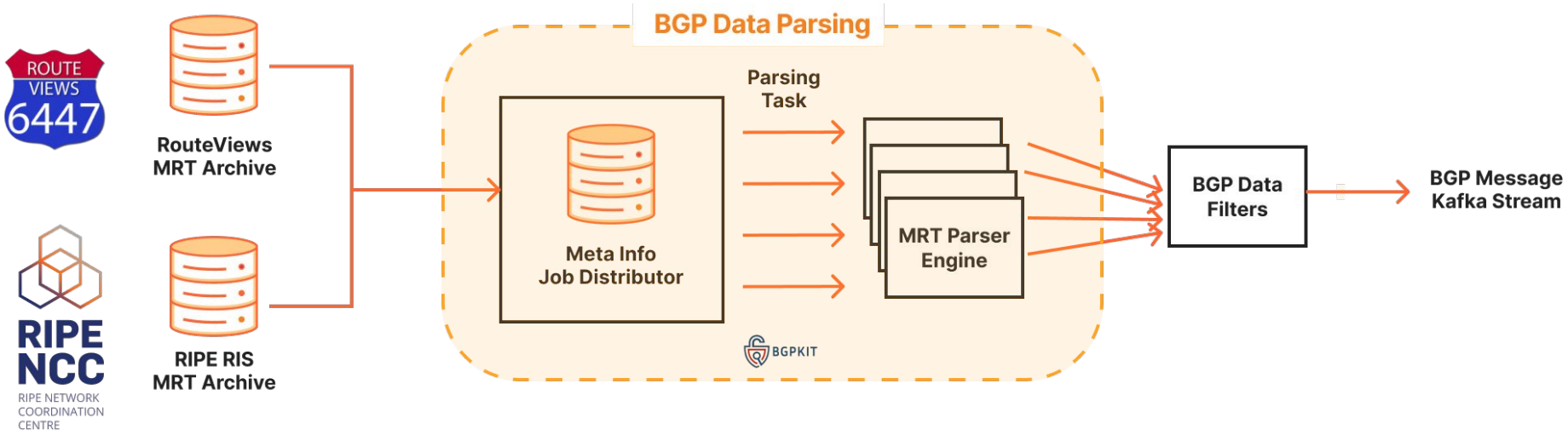
- Leak detection relies on accurate AS-level relationship inference
- AS relationship varies **per prefix**
  - anycast vs. unicast
- AS relationship varies **per location**
  - A transit somewhere may be a peer elsewhere

Detection Pipeline

Pipeline overview

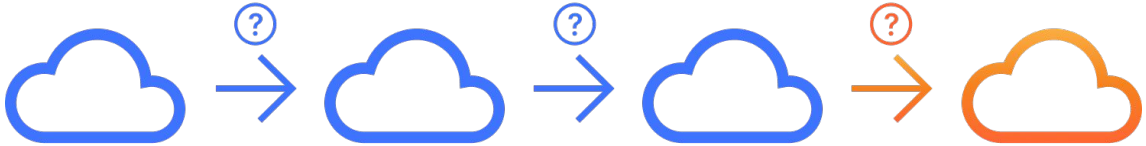


BGP message stream



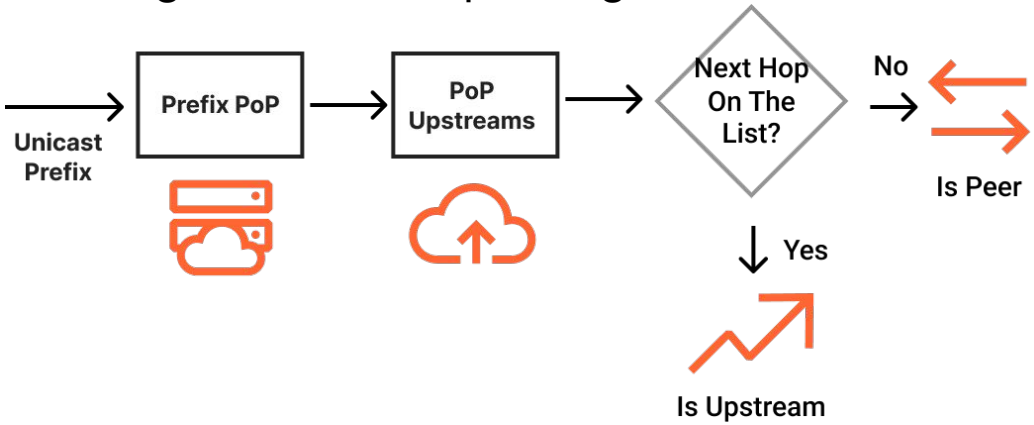
AS relationship inference

- Peer-peer or upstream-downstream?
- Combination of data sources
  - CAIDA/UCSD's AS relationship data
  - BGPKIT AS relationship data
  - Internal inference results
- Inference can be unreliable, especially with complex relationships



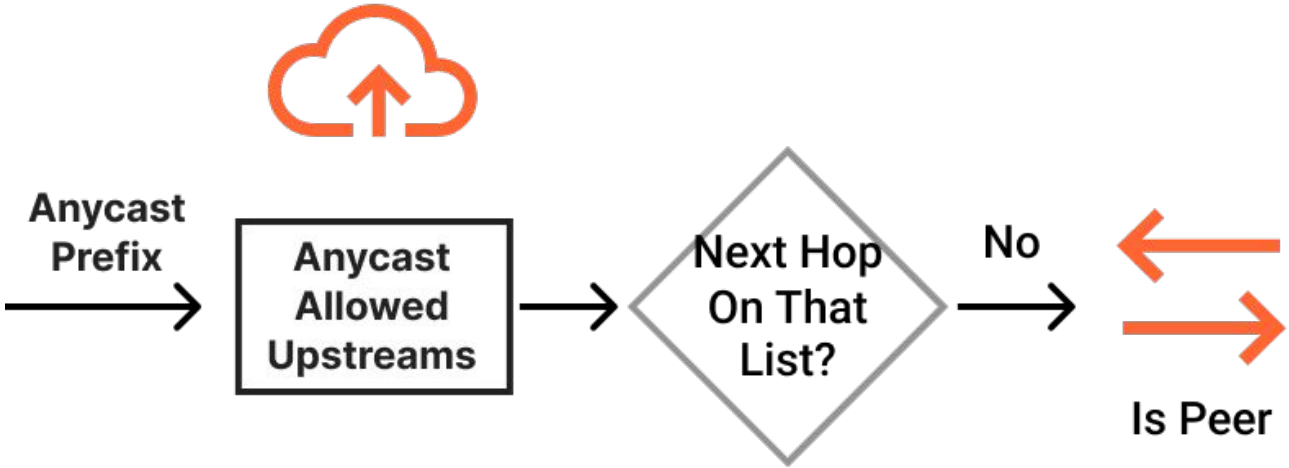
Prefix-level Ground-truth: Unicast Prefix

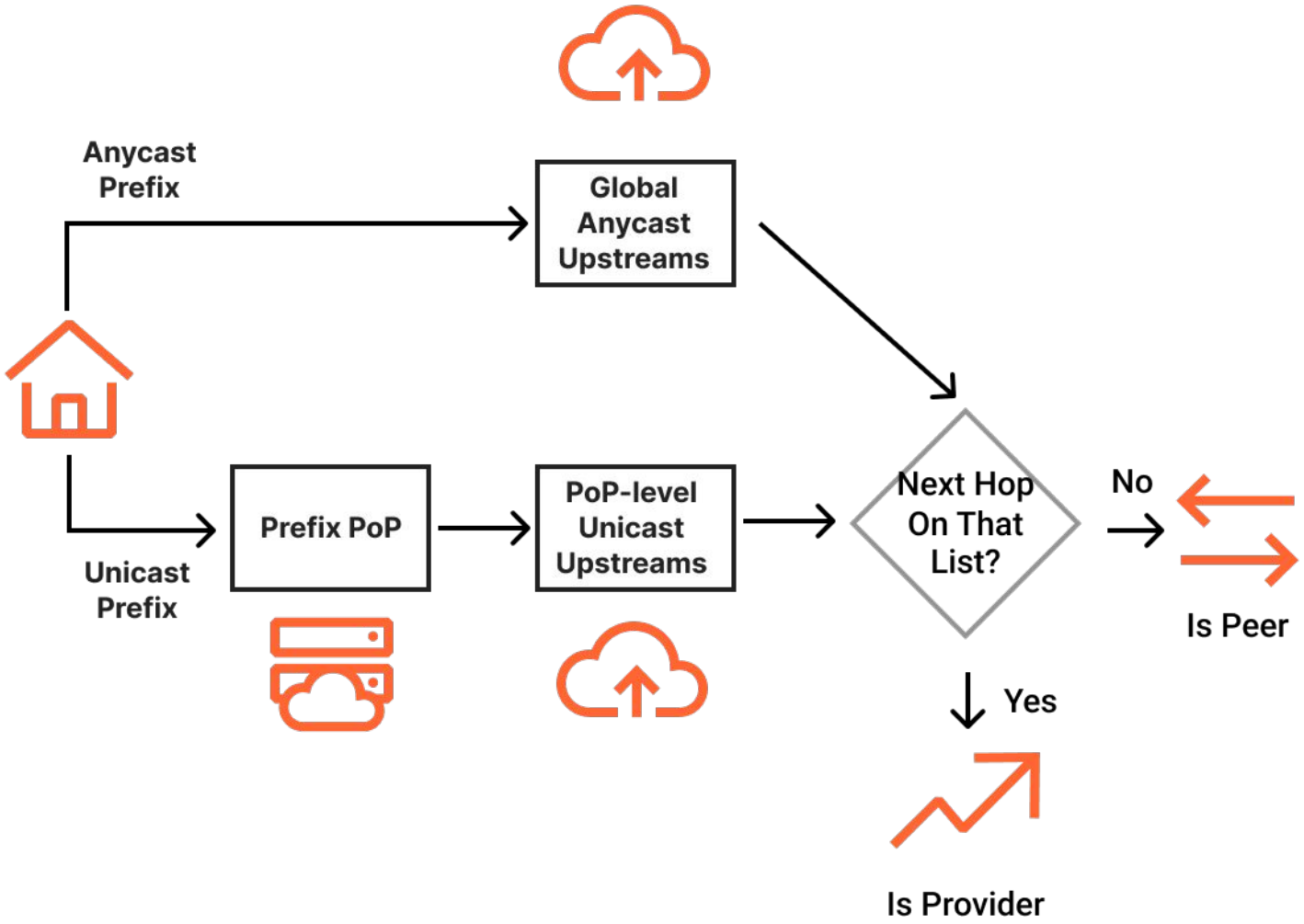
- Each unicast prefix should only be announced via one PoP
- Each PoP have a number upstreams
- Next hop on the upstream list?
  - Yes: treating AS-rel to be upstream
  - No: treating AS-rel to be peering

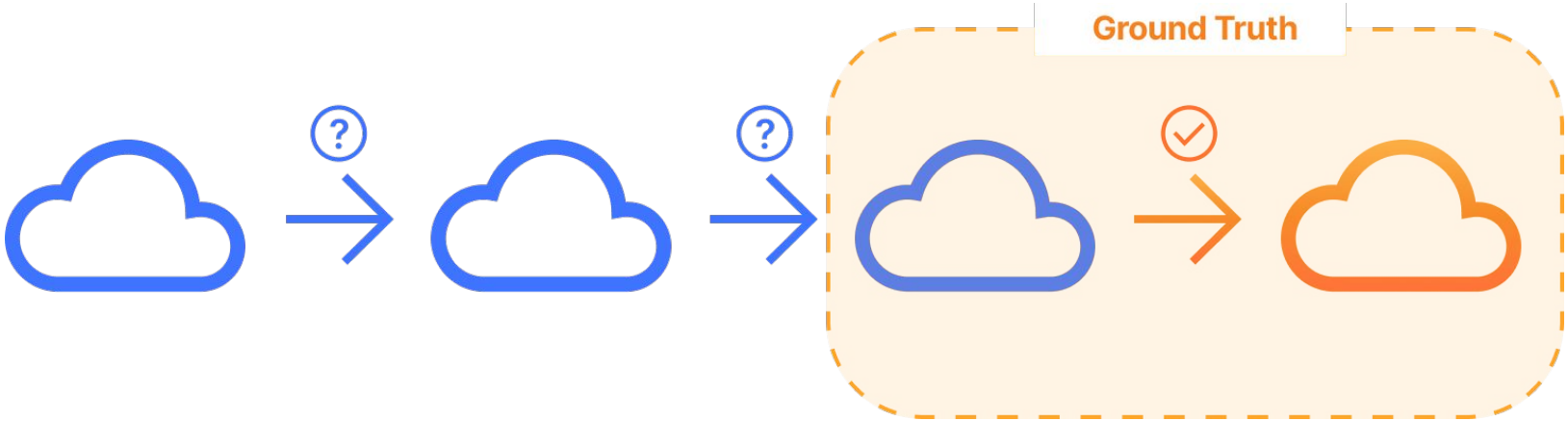


Prefix-level Ground-truth: Anycast Prefix

- Only a handful of ASNs should be allow to provide transit for anycast prefixes
- If next-hop is not one of them, we force treating it as peering relationship




Anycast Prefix
Global Anycast Upstreams
Unicast Prefix
Prefix PoP
PoP-level Unicast Upstreams
Next Hop On That List?
No
Is Peer
Yes
Is Provider


Ground Truth

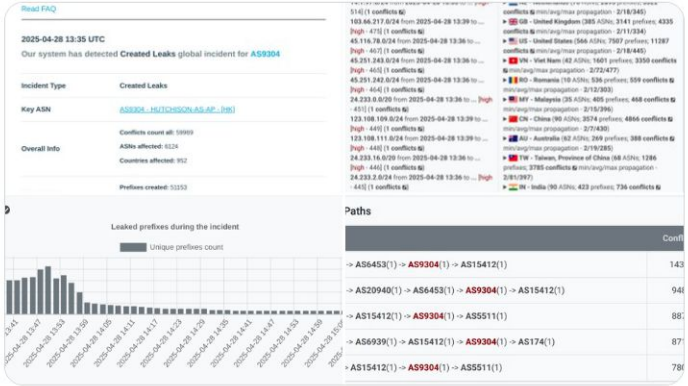
Example internal alerts

Radar by Qrator
@Qrator_Radar

🚨 Leak at 2025-04-28 13:36 UTC

🇭🇰AS9304 (HUTCHISON) leaked 51,153 prefixes from 🇬🇧AS15412 (FLAG), 🇺🇸AS6453, 🇭🇰AS4635 (HKIX-RS1) towards 🇫🇷AS5511 (ORANGE), 🇬🇧AS15412 (FLAG), 🇭🇰AS4635 (HKIX-RS1) creating conflicts with 3510 ASNs in 131 countries

🌐Prop: 100%
⏱Duration: Ongoing



9:45 AM · Apr 28, 2025 · **130** Views

Detected route leak event: https://.
AS rel 0: https
AS rel 1: https:.
Event type: **t4 Peer-Cust-Prov**
Detected time: **2025-04-28T13:46:43 UTC**
Leak ASN: 🇭🇰 **9304 HGC Global Communications Limited; Hong Kong**
Leak segment: 🇺🇸 **174** 🇭🇰 **9304** 🇺🇸 **13335**
Origins Count: **1**
Peer Count: **9**
Prefix Count: **4**

Event 307217

T4: Peer-Cust-Prov

Leak Path Segment	174 9304 13335
Leak ASN	9304 HGC Global Communications Limited
Earliest Leak Time	2025-04-28T13:45:19Z
Latest Leak Time	2025-04-28T14:16:56Z
`monocle` command	`monocle search --start-ts 2025-04-28T20:44:19Z --end-ts 2025-04-28T21:17:56Z --as-path "[ \d]*174[ \d]*9304[ \d]*13335[ \d$]*"`

Leak Msgs Count	Affected Prefixes Count	Affected Origin ASes Count
108	25	1

16[illegible] (lg) gone

colo: [illegible]

transits: [illegible]

transit mismatch!

tags: sitelocal, unicast

CfColoMismatch CfUnicastPrefix AsRelForcePeerPeer

39122 174 15412 9304 13335

Future impact prevention measures

BGP Autonomous System Provider Authorization (ASPA)

- draft-ietf-sidrops-aspa-verification
- Create signed ASPA objects on RPKI
- List of authorized transit upstream providers per ASN
- Validate paths, and invalidate route leaks
- Implementation status
  - OpenBGPD, BIRD, FreeRTR, BGP-SRx

Limitations of ASPA

- **No prefix level granularity**
- Not so great for current state of AS13335

Getting the most out of ASPA

- Express BGP intent at AS-level **if possible** for primary ASN (13335)
- Use of alternative origin ASN ??
- Bonus: clean up AS-SET memberships

RFC9234 Roles and Only To Customer Attribute

- BGP roles assigned to peering and communicated in OPEN
- OTC attribute (Only To Customer)
- Implementation status
  - OpenBGPD, BIRD, FRR, Mikrotik RouterOS (partial)

RFC9234 Roles and Only To Customer Attribute

6. Additional Considerations

Roles **MUST NOT** be configured on an eBGP session with a Complex peering relationship. If multiple eBGP sessions can segregate the Complex peering relationship into eBGP sessions with normal peering relationships, BGP Roles **SHOULD** be used on each of the resulting eBGP sessions.

An operator may want to achieve an equivalent outcome by configuring policies on a per-prefix basis to follow the definitions of peering relations as described in Section 3.1. However, in this case, there are no in-band measures to check the correctness of the per-prefix peering configuration.

CLOUDFLARE

Thank you

Questions?