The Scourge of Excessive AS-Sets

Doug Madory (Kentik)
Job Snijders (*Looking for work!*)

kentik

🚨 Disclaimer: AS-SETs vs AS_SETs 🚨

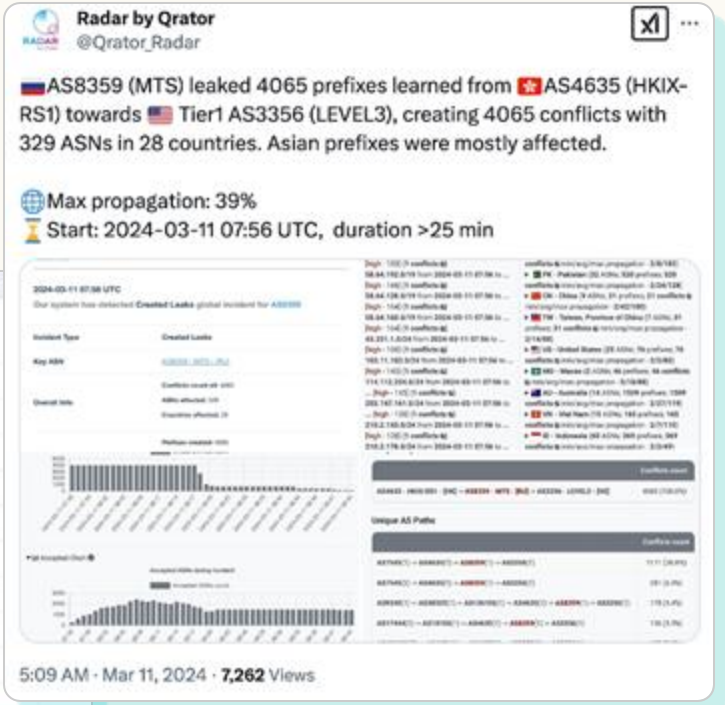
- NOTE: This talk discusses the **IRR AS-SET** object type!
  - A record in the IRR database that defines a group of ASNs used to simplify the management of routing policies by grouping multiple ASNs together.

- Not **BGP AS_SET** construct, which has been deprecated.
  - See *Deprecation of AS_SET and AS_CONFED_SET in BGP* (BCP 172, RFC 9774)
  - Aggregate AS_SETs appear in the AS_PATH of a BGP announcement as one or more ASNs surrounded by curly brackets.
  - Ex: 300 {200,100}. This set indicates that the aggregate summarizes routes that have passed through AS200 and AS100.

Let’s begin with a BGP leak

AS-SETs and BGP Leaks

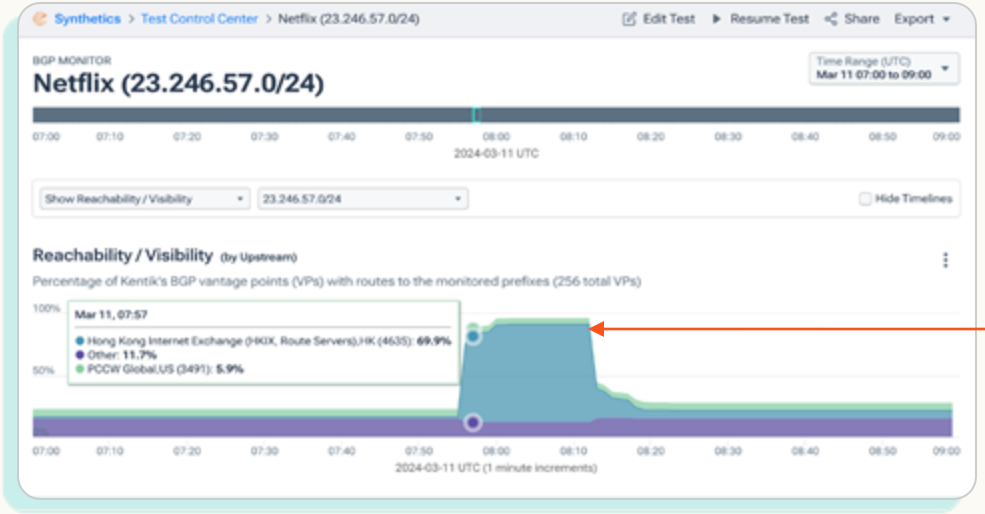
On 07:56 UTC on March 11, 2024, Russian mobile operator MTS (AS8359) mistakenly propagated over 30,000 routes from the Hong Kong Internet Exchange (HKIX, AS4635) to its transit providers Lumen (AS3356) and Arelion (AS1299).



AS-SETs and BGP Leaks

Propagation of one Netflix's BGP routes announced at HKIX.

- Normally circulated only regionally.
- During the leak, the leaked version via AS8359 propagated globally.



The bulge in the graphic represents the dramatic increase in the number of our BGP sources who had the leaked route in their table (with HKIX as the upstream).

What are AS-SETs?

- An AS-SET is a special database object type that represents a group of ASNs and other AS-SETs. It's primarily used for route filtering and policy control by ISPs and network operators.
- To build a prefix allowlist from an AS-SET, each member is recursively evaluated.
  - If the member is an ASN, the IRR is searched for route objects which contain that ASN in the origin field.
    - `route:` for IPv4, `route6:` for IPv6
  - Member AS-SETs are similarly recursively expanded into member ASNs, which are also expanded into their prefixes.
- The resulting prefix list can be loaded into the router's running configuration to be applied on the BGP session with the neighbor in question.

Note: IRR AS-SETs can be explored on the command line with Job's irrtree utility.

Big challenges for AS-SETs

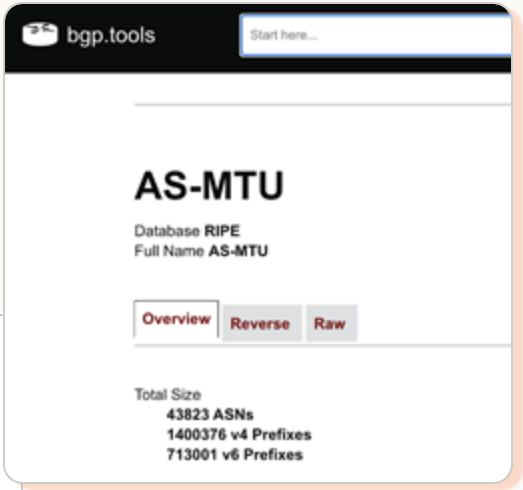
1. No inherent quality, integrity, and authenticity controls over content.
2. No limits to the number of AS members or AS-SETs that can be included.
3. No limits on the depth of the resulting recursion, which allows unlimited, unchecked growth.
4. No agreed upon understanding of different use cases of AS-SETs.
   Do they define lists of customers, peers, or IXP members?

If an AS defines an AS-SET which includes the AS-SETs of some customers who contain AS-SETs, the original AS can lose any awareness of what is contained in the resulting prefix list.

Excessively Large AS-SETs

- The leaker in the March 11 route leak, uses an AS-SET called **AS-MTU**.
- Web utility Bgp.tools lists the contents of AS-SETs.
  - Expands **AS-MTU** to 43,823 ASNs!
  - There are 83,617 ASNs in the global routing table.
  - Any network applying **AS-MTU** as a filter for an interface with AS8359 is creating an allowlist containing these.
- Some examples of prefixes contained in **AS-MTU**:

6.2.0.0/17	US Department of Defense
8.36.240.0/20	Rural Telephone Service Company, Lenora, Kansas
12.10.219.0/24	American Express, Phoenix, Arizona
23.20.0.0/14	AWS EC2 for us-east-1
41.76.175.0/24	National Government of Kenya



Excessively Large AS-SETs

- A popular tool for building BGP filter lists based on IRR data is bgpq4.
  - https://github.com/bgp/bgpq4

- For **AS-MTU**, bgpq4 "-J" returns a Junos router configuration that is almost 1.3 million lines long!

```
$ bgpq4 -Jl eltel AS-MTU | wc -l
1294200
```

- We can use the -A option to aggregate routes, reducing the lines of configuration to only a third of a million, but it is still a lot.

```
$ bgpq4 -Al eltel AS-MTU | wc -l
271171
```

- The routes contained in this AS-SET represent 1.8 billion unique IPv4 addresses out of a total possible 3 billion addresses currently in the IPv4 routing table.

Caution: AS-SETs vary by source IRR

- Note the difference in output length for the three variations of the command above when the source is set to APNIC, RIPE, and RADB.

```
$ bgpq4 -S APNIC -Al eltel AS-VOXILITY-SET | wc -l
4773
$ bgpq4 -S RIPE -Al eltel AS-VOXILITY-SET | wc -l
50961
$ bgpq4 -S RADB -Al eltel AS-VOXILITY-SET | wc -l
86630
```

- (Note: Default source for bgpq4 is NTT's IRR mirror service.)

This is why it is important to indicate the authoritative source of the AS-SET, in PeeringDB for example.

Excessively Large AS-SETs

***AS-MTU** is not alone, nor anywhere near the worst!*

- So, what are the internet's largest (and most absurd) AS-SETs?
- Ben Cartwright-Cox, creator of Bgp.tools, ran the numbers.
- The biggest AS-SETs contain more ASNs than are in the global routing table (~83k).

2,192 AS-SETs expand to over 1,000 ASNs!

RIR	AS-SET	ASNes
RIPE	AS39533:AS-PEERS	102479
RIPE	AS-CLARANETDE-PEERINGS	102335
RADB	AS-ST1-IXPS	102332
RIPE	AS-MERKEL-PEERS	102313
RIPE	as-cloud-ix-pro	102305
RIPE	AS3326:AS-PEERS-DEE	102301
RIPE	AS-DECIX-V6	102300
RIPE	AS12732:AS-UPSTREAMS	102299
RIPE	AS-NFON-DECIX-PEERS-v4	102298
RIPE	AS-NFON-DECIX-PEERS-v6	102298

Unraveling Nested AS-SETs

- The observant reader might notice that these ASN counts exceed the number of ASNs in the global routing table (~83k) and might ask where the unrouted ASNs are coming from?
- Well, they come from the myriad of downstream nested AS-SETs.
- To investigate this phenomenon, Ben Cartwright- Cox wrote a script to traverse the nesting from an excessive AS-SET to one of its component unrouted ASNs, and the journey is wild.

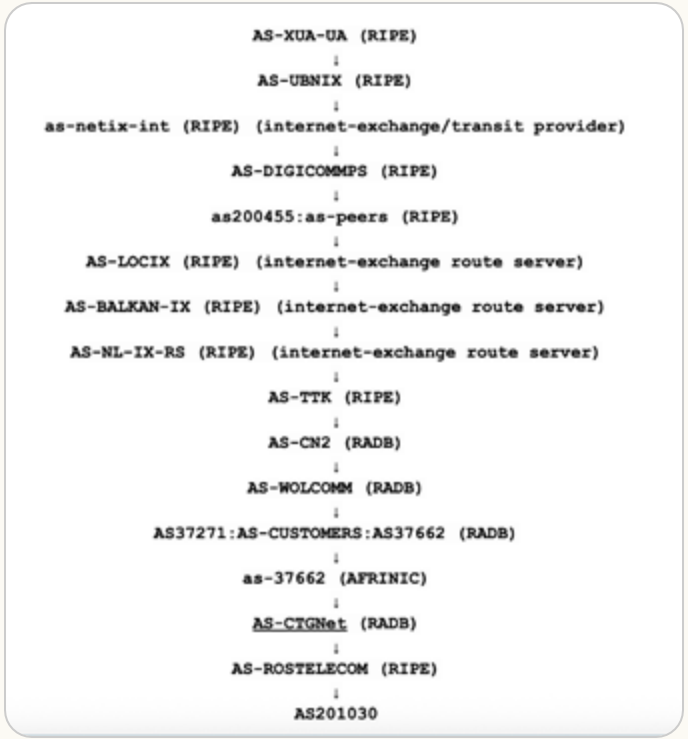
RIR	AS-SET	ASNes
RIPE	AS39533:AS-PEERS	102479
RIPE	AS-CLARANETDE-PEERINGS	102335
RADB	AS-ST1-IXPS	102332
RIPE	AS-MERKEL-PEERS	102313
RIPE	as-cloud-ix-pro	102305
RIPE	AS3326:AS-PEERS-DEE	102301
RIPE	AS-DECIX-V6	102300
RIPE	AS12732:AS-UPSTREAMS	102299
RIPE	AS-NFON-DECIX-PEERS-v4	102298
RIPE	AS-NFON-DECIX-PEERS-v6	102298

Unraveling Nested AS-SETs

- For example, AS-SET AS-XUA-UA from RIPE expands to almost 90,000 ASNs, including many unrouted ones.

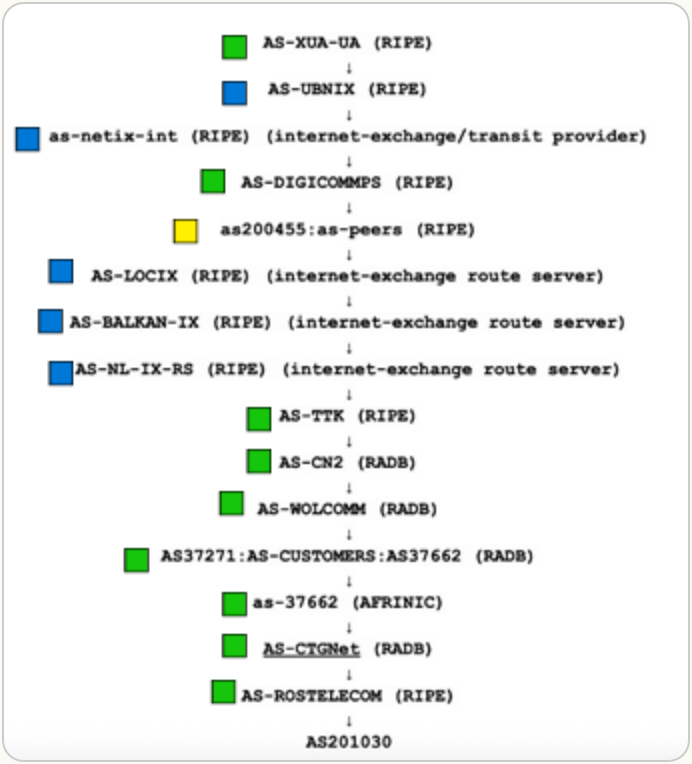
  How did an unrouted ASN wind up here?

- This circuitous sequence of AS-SETs begins in Ukraine and includes, among other countries, Bulgaria (Balkan-IX), the Netherlands (NL-IX), Russia (TTK), China (CN2), South Africa (WOLCOMM), China again (CTGNet), and Russia again (Rostelecom), before landing on AS201030, the unrouted AS of "Public corporation for organisation of air traffic in the Russian Federation." Phew!



Unraveling Nested AS-SETs

- How did this list come to be?
- There are (unofficially) different types of AS-SETs:
  - Customers
  - Peers
  - IXP members
- The semantics of these are not really defined.
- The sequence on the right contains all three types.
- This mixing of types explodes the size of an AS-SET.



Why is this a problem?

- The March 2024 leak by AS8359 could have been exacerbated by its excessive AS-SET. It certainly didn't help.
- Our only hope to reduce harm from BGP mishaps is automation.
  - IRR data enables automated generation of allowlists (BGP session filters).
- Excessively large AS-SETs defeats the purpose of an allowlist.
- Excessively large AS-SETs also breaks automation!
  - Requires large amount of data to be repeatedly transferred and stored.
  - Generates extremely large (and unusable) router configurations.
- Providers have had to create workarounds to deal with this IRR pollution.

What's the solution?

1. Owners of AS-SET objects need to review any AS-SETs they define and ensure that they contain the minimum amount of ASes and AS-SETs to facilitate the creation of effective allowlists.
2. Ideally, AS-SET recursion is avoided where possible or at least kept to a minimum.

   The downside of this "solution" is that it requires full and cognizant cooperation of all AS-SET holders, which is unrealistic in the global internet routing system.

What's the long(er) term solution?

- Ultimately, the issue of BGP route leaks needs to be addressed through something better than unwieldy self-asserted allowlists.
- Instead, the industry should use a combination of:
  a. In-band BGP signaling, such as described in RFC 9234,
  b. RPKI-based signaling using ASPA verification (work-in-progress),
  c. Perhaps future RPKI extensions such as Signed Prefix Lists (work-in-progress).
- IRR-based AS-SETs simply lack a degree of precision and contextual awareness to mitigate route leaks at scale.
- Better to use a combination of in-band and out-of-band signals to ascertain whether a given BGP route announcement is a leak or not.

Thanks!

Special thanks to **Ben Cartwright-Cox** of bgp.tools for his AS-SET analysis.

Also, a big thanks to **Tony Tauber** (Comcast), **Aaron Weintraub** (Cogent), and **Anees Shaikh** (Google), **Lasse Jarskov** (Telia Company) for their feedback and suggestions.

kentik

Thank you!

Doug Madory dmadory@kentik.com
Job Snijders job@sobornost.net