Maximum Prefix Tripping: A potential workaround for leaking on the Internet

Tom Scholl, AT&T Labs

NANOG 38
October 9th, 2006

Understanding leaks

- BGP leaks will **always** continue to happen between transit providers, peers and customers

Yes, proper filtering could solve all these issues, the reality is not everyone will apply filtering at all the right locations. Nothing can be done about networks failure to apply filters on customers or peers, it may be best to get over it and come up with an alternative solution.

This presentation is not an excuse to not filter. You should always filter customers, peers and transit networks 

Understanding the impact of leaks

The impact of BGP leaks are:

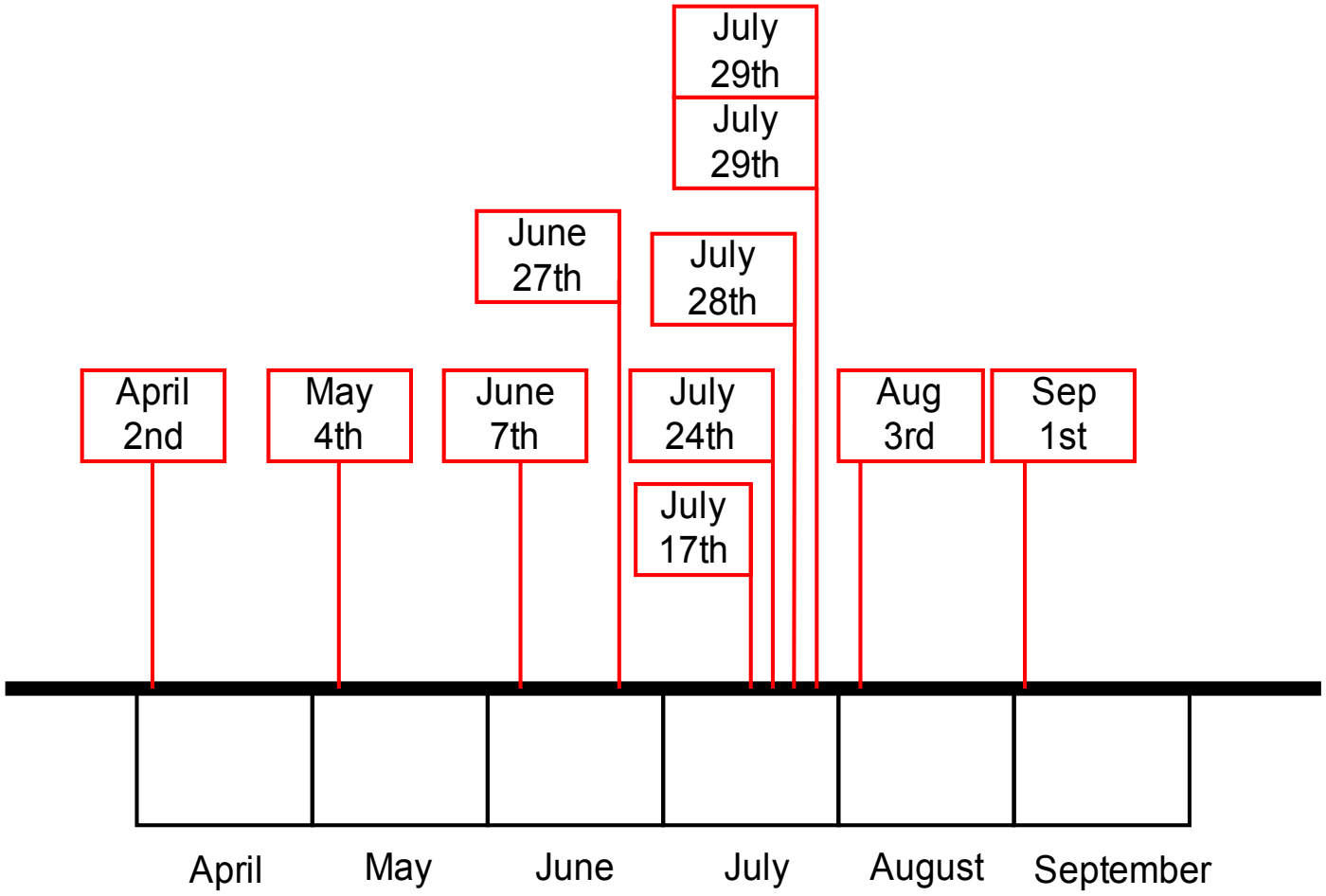
- Tripping another networks maximum-prefix safety mechanism

  Resulting in the BGP session to that adjacent network to be severed. Results in manual intervention that could take hours. Remember, not all NOC's are permitted to touch peering routers (must escalate).

- People routing traffic the “wrong way”

  Another network accepting the leaked prefixes and forwarding traffic towards you based upon it

A spring + summer of leaks



...and then there is maximum-prefix

- The feature is a good intention
- However, maximum-prefix is the equivalent of the US power grid:

  Overload resulting in shut down

- If you exceed a peers maximum-prefix limit and the session is closed, you will not know based upon logging information. Further attempts to establish the session return vague errors (malformed update/attribute)

Thoughts on making maximum-prefix better

- If maximum-prefix is exceeded, suppress any routes learned within the last X amount of minutes before the maximum-prefix was exceeded.
- Its sort of like dampening. Think BGP neighbor received-routes suppression.
  - Keeps the BGP session up, keeps existing prefixes installed, keeps everyone happy (hopefully).
- The assumption is that a leak is a sudden burst over a short period of time.
- When prefix received count decreases below the maximum-prefix value, toss out the suppressed routes and force the peer to refresh.

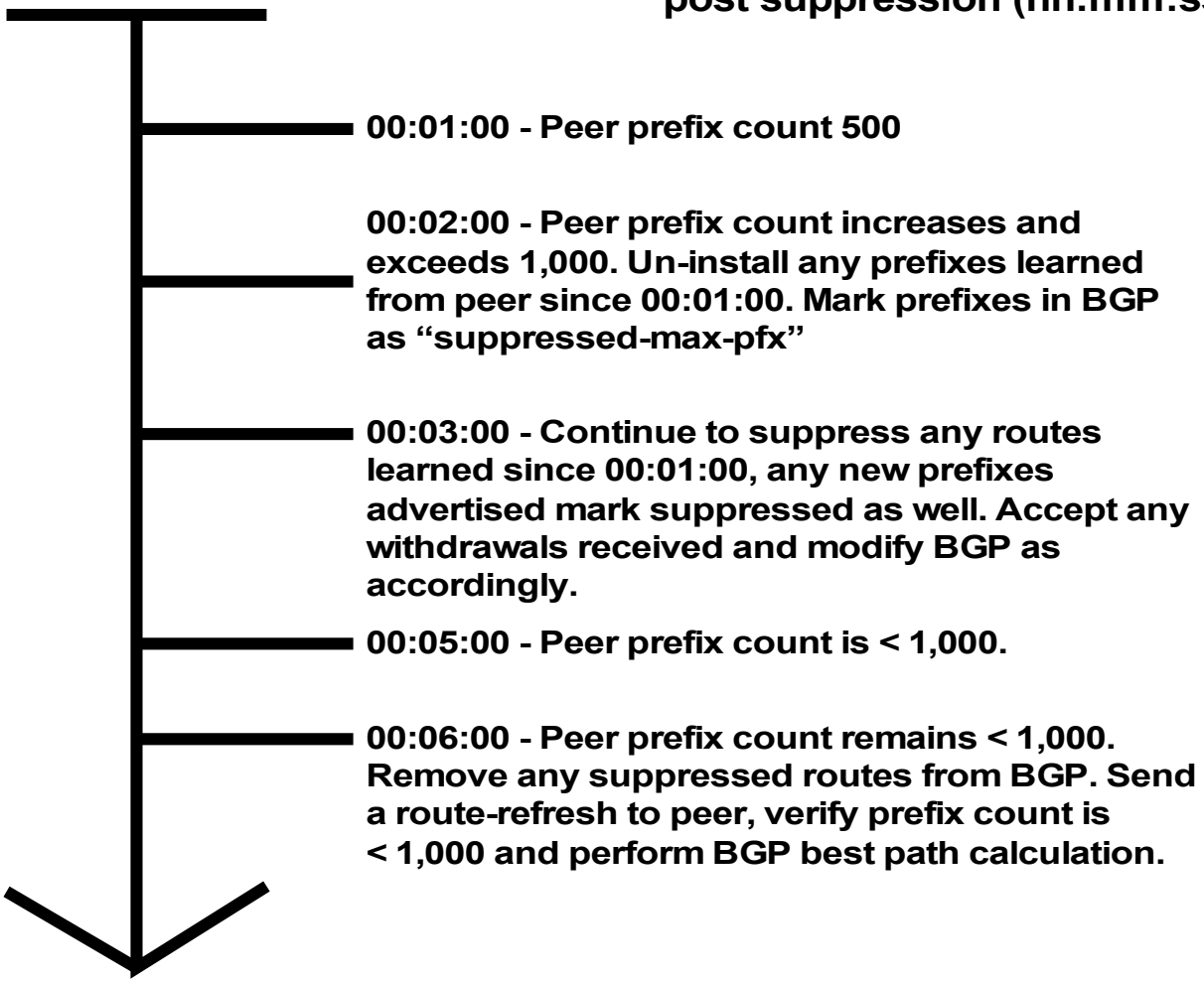
How would it work?

- Each peer/group has a defined maximum-prefix value based upon current advertisements, recommendations from peer, etc.
- A value is defined to determine the length of time before maximum-orefix exceeding action to suppress (ie. 1 minute).
- Upon any BGP peer exceeding maximum-prefix, the router marks any routes learned after that period as suppressed.
- Any additional advertisements received are marked as suppressed (continue to accept withdrawals).

How would it work? (cont'd)

- Once received routes drops below Maximum-Prefix, wait for a defined time period (ie. 1 minute?), then transmit a route refresh message to the peer to re-advertise prefixes.
- Verify prefix count after refresh is still below the maximum-prefix limit. If not, continue to suppress.
- If all is well, throw out suppressed prefixes and perform BGP best path decisions on prefixes received.

How would it work? (cont'd)



Potential problems?

- Resource constraints to the router by performing this action?
  - Additional memory to hold BGP adj-in to compare against
  - Timestamps on BGP routes received (most vendors already do that, right?)
- What if a peer is oscillating a leak? Perhaps enable a safety mechanism to shutdown a peer when they've cycled through max-prefix several times in a set time period?
- What if the leak is slow/gradual and not a sudden burst?

Additional maximum-prefix gripes

- Allow maximum-prefix knobs to permit pre-policy and post-policy:
  - Cisco – Post-Policy
  - Juniper – Pre-Policy
- Perhaps maximum-prefix **outbound**?

  (Suggested by Eric Bell years ago)
- Fix Cisco as-path regex bug (CSCse92685)

  (Regex does not work 100% of the time)

Other knobs that would be nice

- Need to provide a method to signal/communicate to a BGP peer for various reasons (drafts have been submitted over the past few years):
  - Planned maintenance – provide NOC contact information, expected duration
  - Administratively shutdown (use your imagination)
  - You are approaching the configured maximum-prefix limit – send a message to the peer to have them update this
  - Maximum-prefix exceeded – shutting session down
  - Pay your bills

Thank You

Tom Scholl **tom.scholl@att.com**