BGP Flow Specification for Route Servers at IXPs

NANOG 73 – 26th of June 2018
Benedikt Rudolph
Research & Development, DE-CIX

DE CIX

Motivation

Why BGP Flow Specification?

malicious traffic

analysis

redirection

ensure QoS

➔Expose router packet-processing features in BGP

- Actions based on n-tuple matching

➔Easy to automate (like RTBH)

- Leverage existing BGP control plane to distribute flow information

➔Leverages BGP best-practices and policy controls

- NEXT_HOP validation against unicast routing

BGP Flow Specification (FlowSpec)

➔RFC 5575 for IPv4 [4], IPv6 (draft) [5]

➔Complement routing information with traffic flow specification

- Take advantage of router ACL capabilities

➔FlowSpec rules encoded as BGP NLRI (address family)

- n-tuple of matching criteria (src/dst prefix, protocol, port, …)

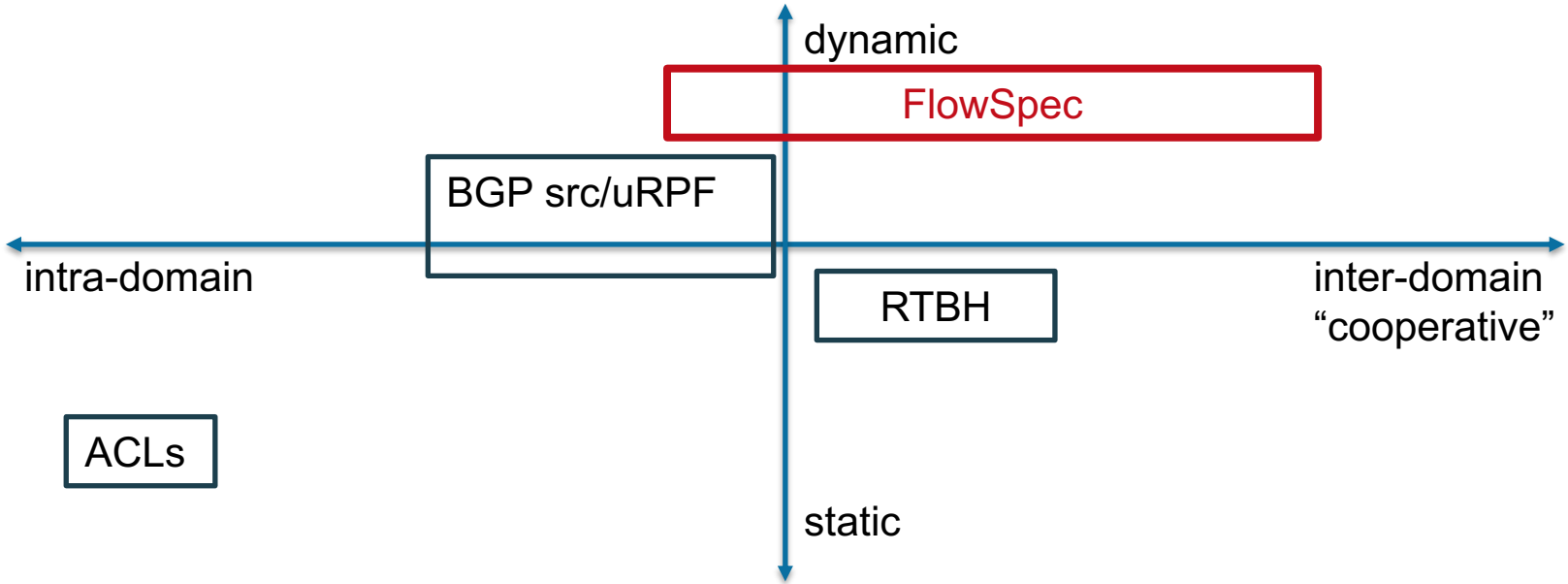
➔FlowSpec rules complement unicast routing (no interference)

➔Extended communities to specify action: discard, limit, sample, redirect

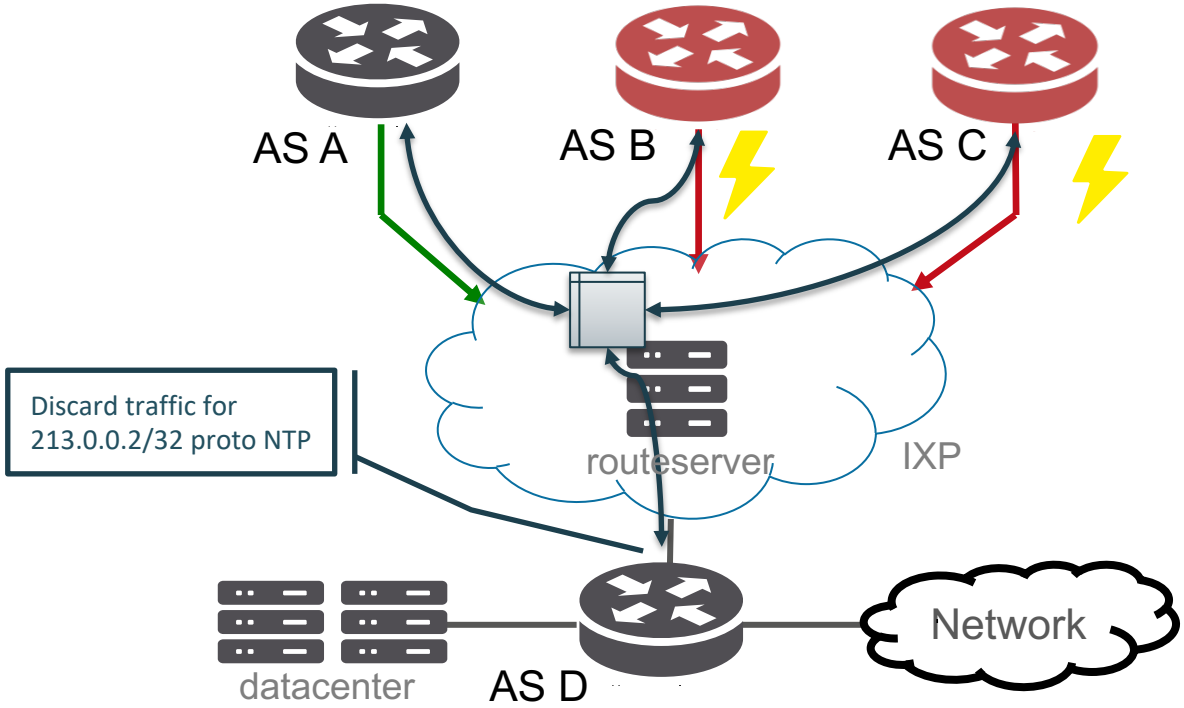
➔Validation: accept filter when advertised by next-hop for destination prefix

DDoS Mitigation Techniques

Network operators use several techniques to mitigate DDoS attacks:



DDoS Mitigation with FlowSpec at an IXP



BGP FlowSpec Performance

Network Working Group
Request for Comments: 5575
Category: Standards Track

P. Marques
Cisco Systems
N. Sheth
Juniper Networks
R. Raszuk
Cisco Systems
B. Greene
Juniper Networks
J. Mauch
NTT America
D. McPherson
Arbor Networks
August 2009

Dissemination of Flow Specification Rules

Abstract

This document defines a new Border Gateway Protocol Network Layer Reachability Information (BGP NLRI) encoding format that can be used to distribute traffic flow specifications. This allows the routing system to propagate information regarding more specific components of the traffic aggregate defined by an IP destination prefix.

Additionally, it defines two applications of that encoding format: one that can be used to automate inter-domain coordination of traffic filtering, such as what is required in order to mitigate (distributed) denial-of-service attacks, and a second application to provide traffic filtering in the context of a BGP/MPLS VPN service.

The information is carried via the BGP, thereby reusing protocol algorithms, operational experience, and administrative processes such as inter-provider peering agreements.

- RFC warns about possible impacts to performance
- Encourages **careful** use of automation

Requirements Analysis

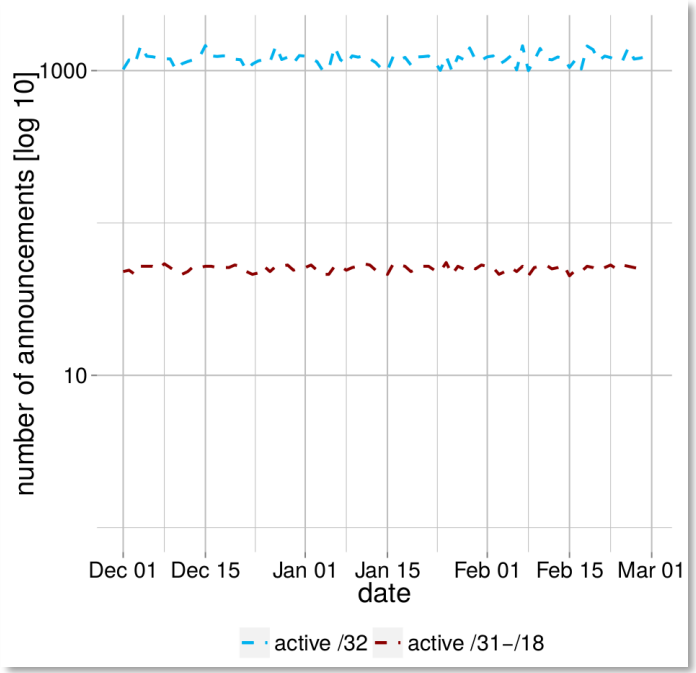
➔ What # of flowspec rules to expect per neighbor?

➔ Update frequency of flowspec rules?

➔ Use RTBH data as basis for flowspec test scenario

- **RTBH usage**: *automated* or *manual*? – hint for kind of use

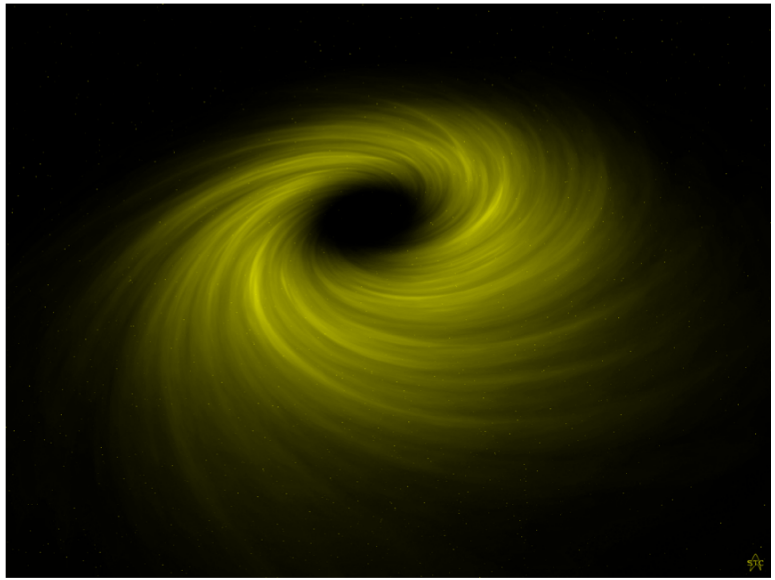
Blackholing Usage – Active Announcements



- ➔About 23,000 announcements collected in three months time
- ➔Stable number of active /32 blackholes (~1200)
- ➔Also, stable number of less specifics /31 - /18 (~50)
- ➔Results confirmed by 2017 data
  - ~1750 blackholes (95 %entile)

source: [1] and NANOG 67

Blackholing Usage – Prefixes



→ Mainly /32 announcements (97%)

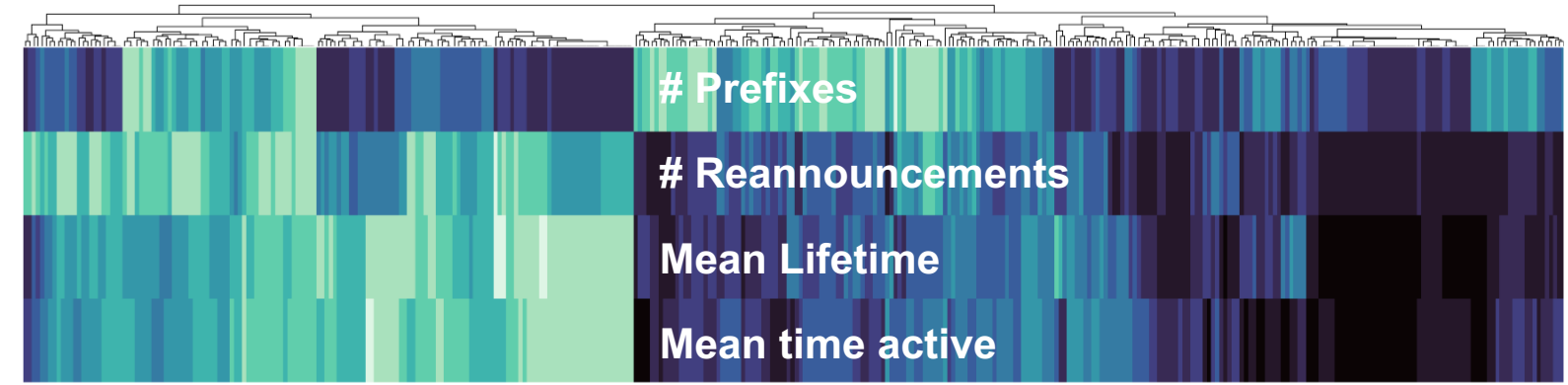
- Majority are short-lived (~50% <= 3 hours)

→ Confirmed by 2017 data

- /32s at 98%
- 95 percent of ASNs have <5 prefixes
- At max 421 prefixes per ASN

source: [1] and NANOG 67

RTBH Activity Clustering (Route Server)



~70%: High amount of prefixes or high prefix activity.
→ **High degree of automation?**

~30%: Low amount of prefixes and low prefix activity.
→ **Low degree of automation?**

FlowSpec Capability “in the wild” (1)

Steps for successful BGP FlowSpec deployment:

1. Vendor / firmware support (multiprotocol BGP + FlowSpec capability)
2. Administrative policy (needs to be enabled)
3. Negotiation during BGP protocol handshake

➔ Analyze BGP OPEN Messages

```
Optional Parameter: Capability
    Parameter Type: Capability (2)
    Parameter Length: 6
    Capability: Multiprotocol extensions capability
        Type: Multiprotocol extensions capability (1)
        Length: 4
        AFI: IPv4 (1)
        Reserved: 00
        SAFI: Flow Spec Filter (133)
```

Flowspec Capability “In the wild” (2)

➔Results:

- NYC-based router with 81 BGP sessions **=> 0**
- DE-CIX’s FRA route servers connected to 964 peers **=> 0**

➔Even if flowspec is supported it is not enabled “by accident”

- Careful use (if used at all)
  - ➔ No test data
  - ➔ Need to generate test data

BIRD for BGP Route Servers

➔ BIRD v2.0.x is a refactoring with new features:

- Multiprotocol BGP (IPv4 and IPv6 integration)
- FlowSpec support (among others)

➔ Expressive filter-language e.g. for route validation

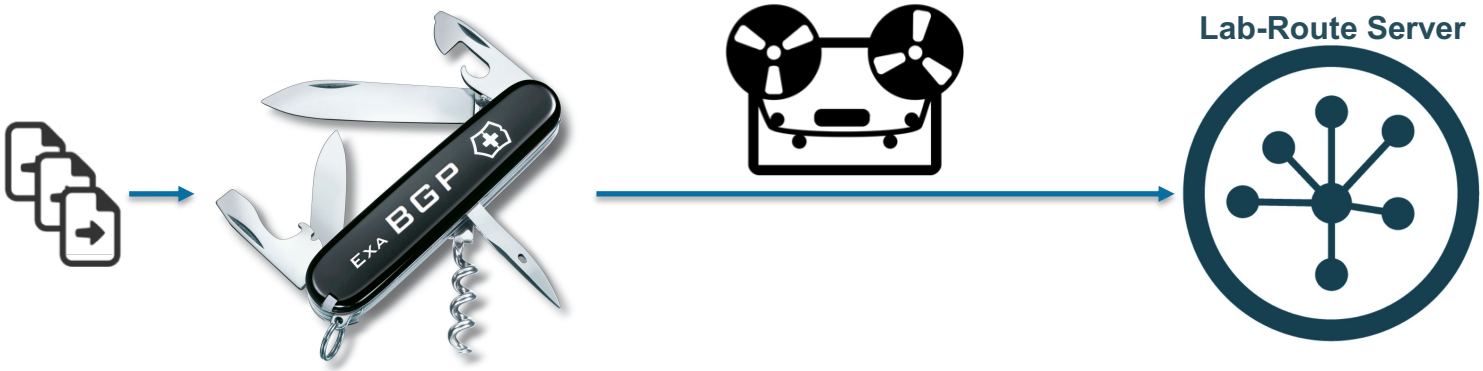
➔ Limitations for flow4/6 routes:

- No access to flow-fields (e.g. src, dst)
- Common BGP fields accessible

```
flow4 { dst 255.255.255.255/32; src 172.31
   Type: BGP univ
   BGP.origin: IGP
   BGP.as_path: 553
   BGP.local_pref: 100
   BGP.ext_community: (generic, 0x8006000
```

➔ Basic reflection of FlowSpec routes

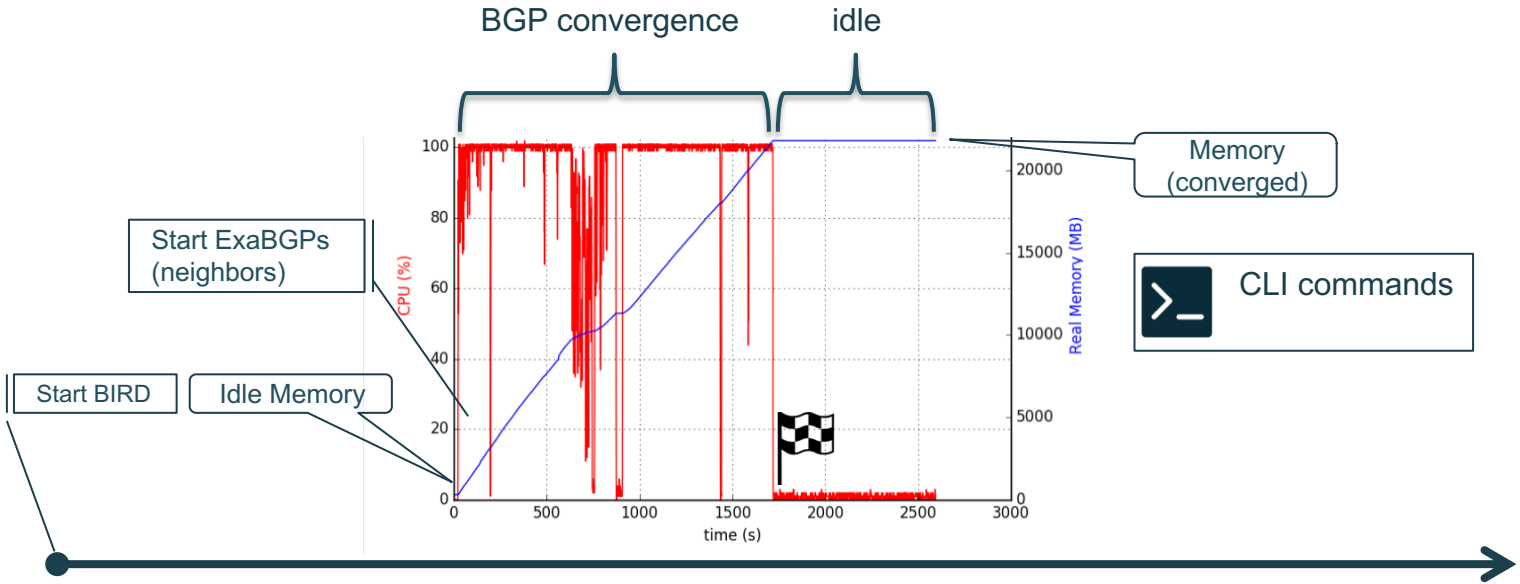
Test Setup for Peering LAN



➔Peers: emulated with ExaBGP (version 3.4.26)

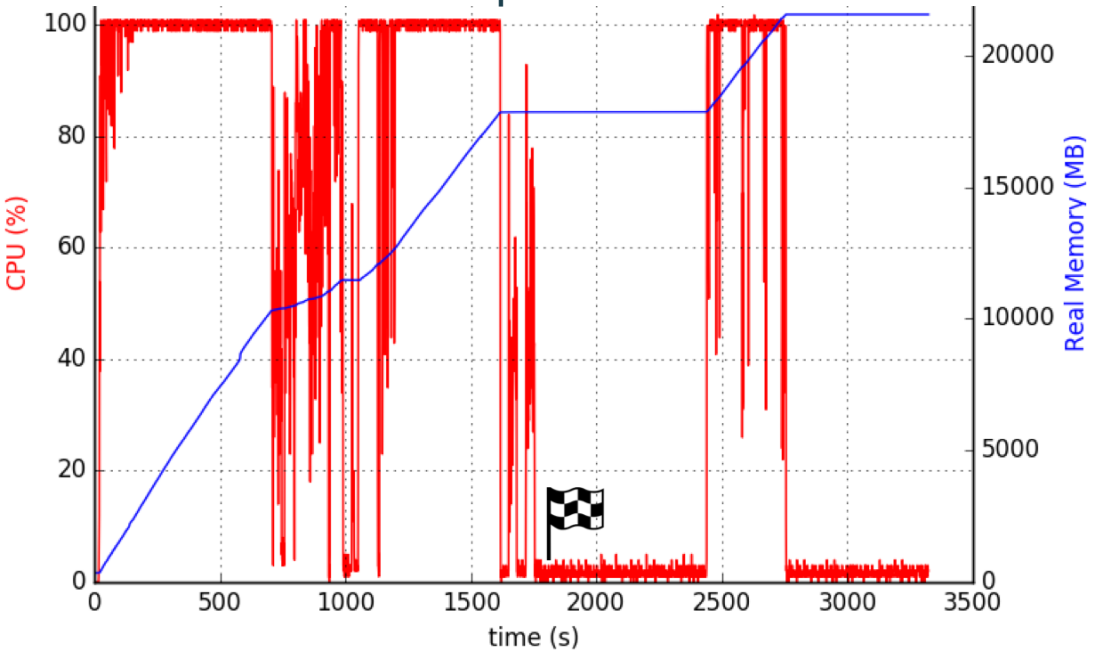
- Non-filtered / filtered routes (to exercise prefix filters)
- 742 neighbors, ~316.000 routes for ~182.000 networks
- dedicated server

Scenario "Route Server Reboot"



Scenario “Route Server Reboot”

➔BIRD v2.0.2 “flowspec”



➔ Convergence: 1736 s

➔ 1 neighbor delayed

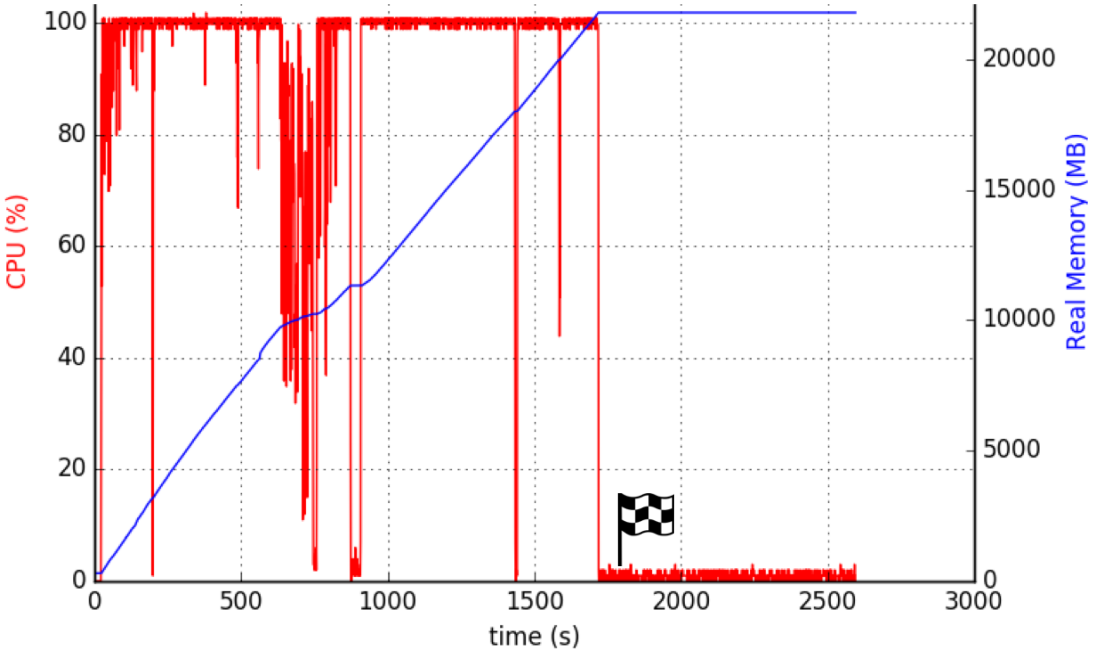
➔ spiky CPU pattern (at times)

➔ 347 MiB initial mem

➔21,04 GiB max. mem

Scenario “Route Server Reboot”

➔BIRD v1.6.4 “stable”



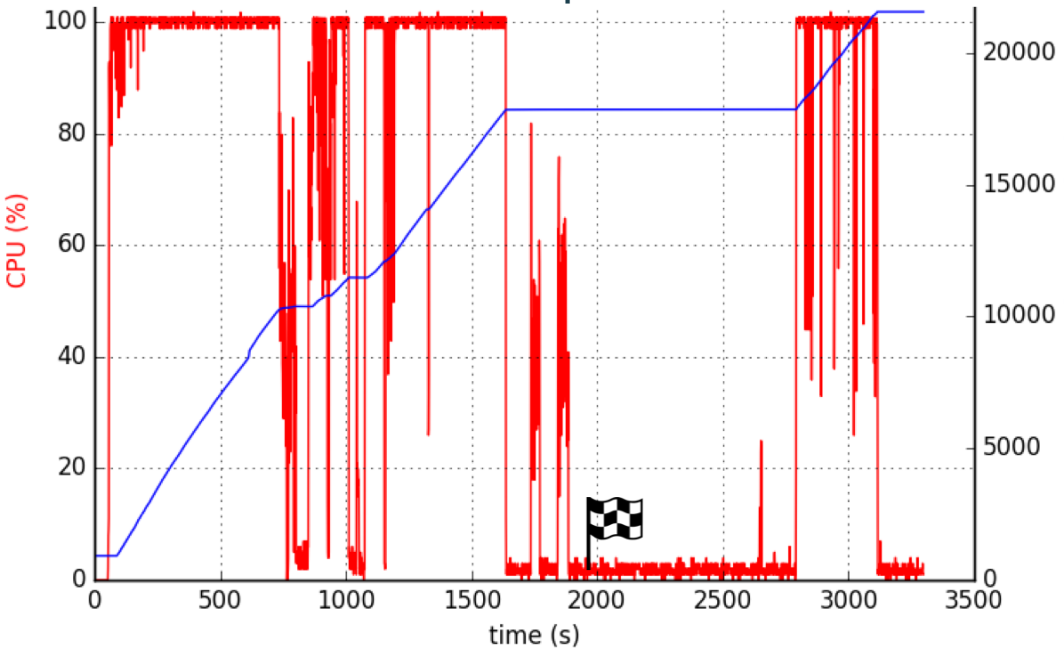
➔ Convergence: 1698 s

➔ 301 MiB initial mem

➔ 21 GiB max. mem

Scenario “Route Server Reboot”

➔BIRD v2.0.2 “no flowspec”



➔ Convergence: 1832 s

➔ 1 neighbor delayed

➔ spiky CPU pattern (at times)

➔ 347 MiB initial mem

➔21,04 GiB max. mem

Scenario “Route Server reboot” Summary

➔BIRD v2.0.2 “no flowspec”

```
$ birdc show memory
```

BIRD memory usage	idle	convergent
Routing tables:	1743 kB	20 GB
Route attributes:	78 kB	91 MB
Protocols:	3514 kB	10 MB
Total:	345 MB	21 GB

➔ + 44 MB idle memory compared to BIRD 1

➔ a little bit slower

➔ *What is the effect of adding flowspec to the existing setup?*

Scenario “Route Server reboot” Summary

➔BIRD v2.0.2 “flowspec”

```
$ birdc show memory
```

BIRD memory usage	idle	convergent
Routing tables:	+5409 kB	20 GB
Route attributes:	78 kB	91 MB
Protocols:	3514 kB	+ 3 MB
Total:	345 MB	21 GB

➔ more mem for routing tables

➔ more mem for protocols

➔Little more memory in total

➔*How does it interact with neighbors?*

Scenario “FlowSpec single update”

➔ Distribution of one FlowSpec (IPv4) rule to 742 neighbors

➔ Method: capture traffic at the route server

➔ **Announce**: 19 ms (incl. processing)

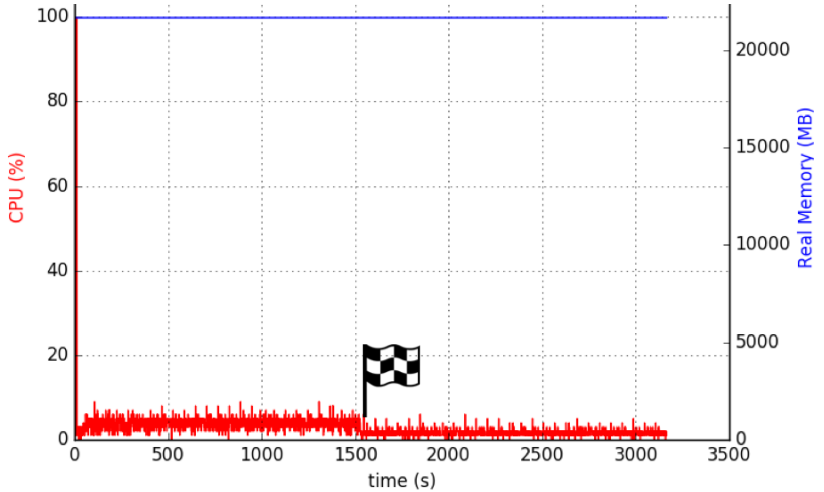
- Propagation to all neighbors: 17,4 ms
- Total time: 36,4 ms

➔ **Withdraw**: 1,1 ms (incl. processing)

- Propagation to all neighbors: 14,3 ms
- Total time: 15,4 ms

Scenario "FlowSpec bulk-update" – I

- ➔ **Announce**
  2288 IPv4 rules to 741 neighbors
- ➔ 2288 rules: extrapolated growth
- ➔ Method: observe BIRD process
- ➔ Initial announcement: 1469 s
- ➔ Bottleneck: exabgpcli + bash
- ➔ Memory: no change



Scenario "FlowSpec bulk-update" – II

➔ **Withdraw**

2288 IPv4 rules to 741 neighbors

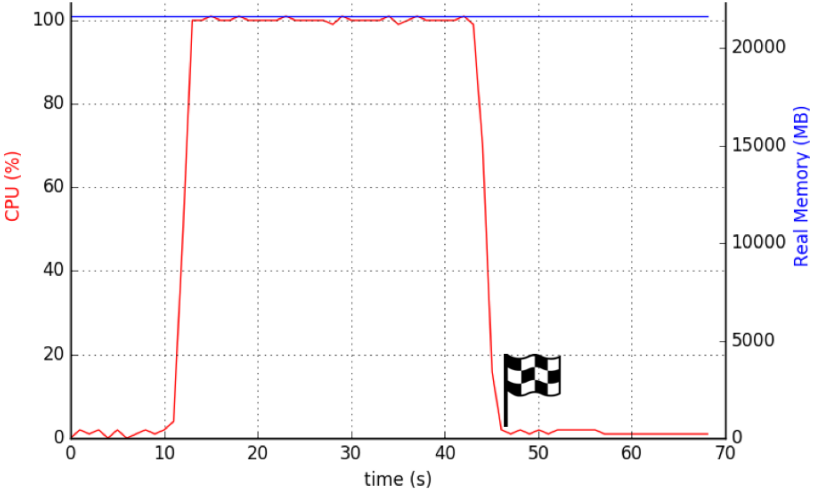
➔ Method: observe BIRD process

➔ **Withdraw**: 34 s

```
$ birdc disable protocol <proto_ID>
```

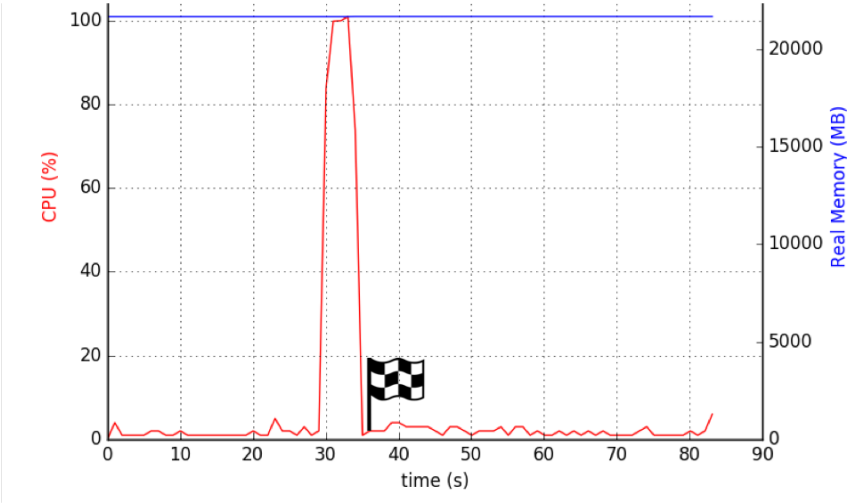
➔ BIRD process at 100% CPU

➔ Memory: no change



Scenario "FlowSpec bulk-update" – III

- **Re-Announce**
  2288 IPv4 rules to 741 neighbors
- Method: observe BIRD process
- **Re-Announce**: 6 s
- BIRD process at 100% CPU
- Fast (RIB to RIB transfer)
- Memory increase 8444 kB



Conclusions

➔ BIRD v2 and FlowSpec seem ready for deployment at IXPs

➔ Tests in worst-case conditions => 100% FlowSpec adoption unrealistic

➔ Scalability looks good

➔ Lots of interesting applications possible

Sources

- [1] Christoph Dietzel, Anja Feldmann, Thomas King: **Blackholing at IXPs: On the Effectiveness of DDoS Mitigation in the Wild**; In: Thomas Karagiannis, Xenofontas Dimitropoulos: Passive and Active Measurement, 17th International Conference, PAM 2016, Heraklion, Greece, March 31 - April 1, 2016. Proceedings
- [2] The BIRD Internet Routing Daemon, by CZ.NIC http://bird.network.cz
- [3] ExaBGP, by Thomas-Mangin https://github.com/Exa-Networks/exabgp
- [4] RFC5575, Dissemination of Flow Specification Rules, https://www.rfc-editor.org/rfc/rfc5575.txt
- [5] Dissemination of Flow Specification Rules for IPv6 (Internet-Draft) https://www.ietf.org/archive/id/draft-ietf-idr-flow-spec-v6-09.txt

Comments? Questions?

DE-CIX Management GmbH | Lindleystr. 12 | 60314 Frankfurt | Germany
Phone +49 69 1730 902 0 | rnd@de-cix.net | www.de-cix.net