ARBOR NETWORKS

Mitigation using BGP flow spec

Ferran Orsola
forsola@arbor.net

BGP flow Spec

POSITIONING

ARBOR NETWORKS

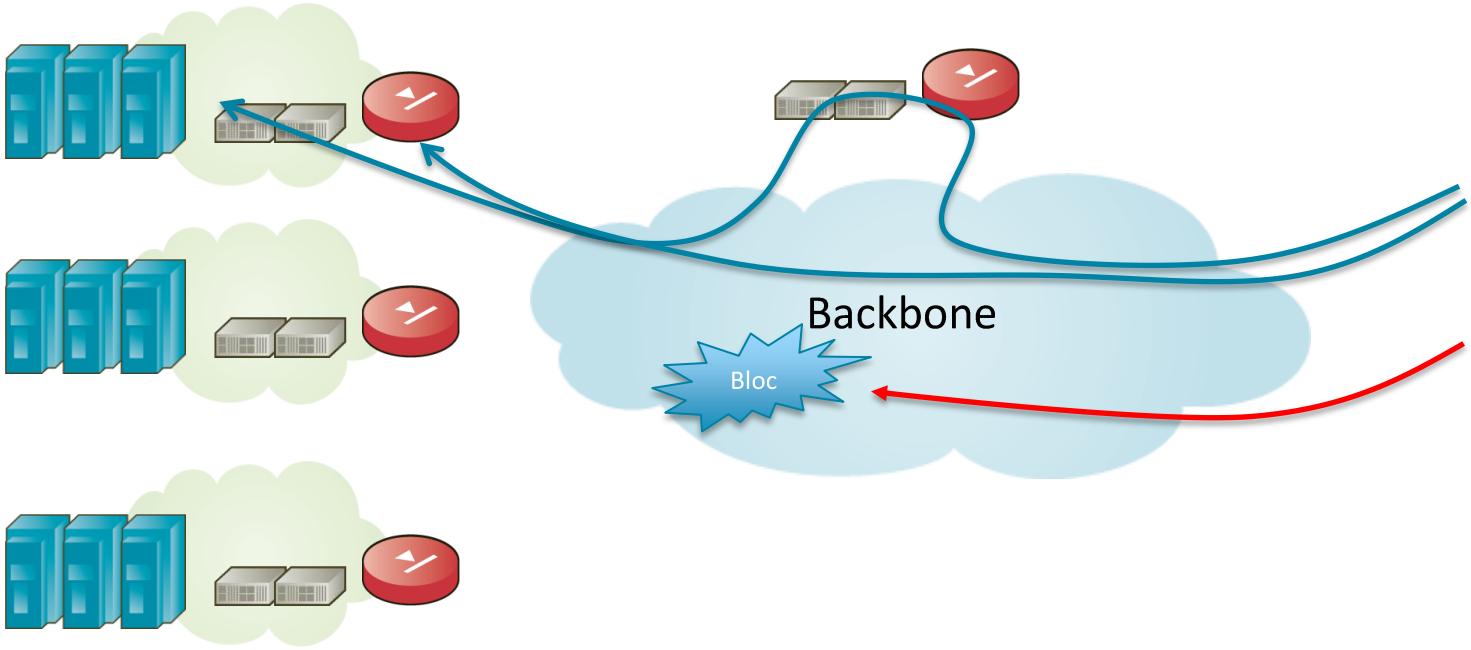
BGP flow spec in few words

- BGP flow spec is a MP BGP capability just like IPv4 or IP VPN, IPv6, …
  - Negotiated during BGP session establishment,
  - Address Family Identifier (AFI) Subsequent Address Family Identifier (SAFI)
    - AFI 1 (IPv4) / SAFI 133 (flow spec) and 1 (IPv4) /134 (IPv4 VPN)
  - Dedicated NLRI : Network Layer Reachability Information
    - Opaque key transported by MP BGP and managed by control plan application layer
  - Allows to specify flow information via BGP NLRI
  - Allows to define action associated to that flow
    - Traffic rate in bytes per seconds (0 means black hole)
    - Traffic action : start stop filter, apply sampling
    - Redirect : redirect traffic to a IP VPN (Route Target)
    - Traffic Marking
- RFC 5575
  - Standard track 2009
- Side remark: could required a session reset (Juniper)

ARBOR NETWORKS

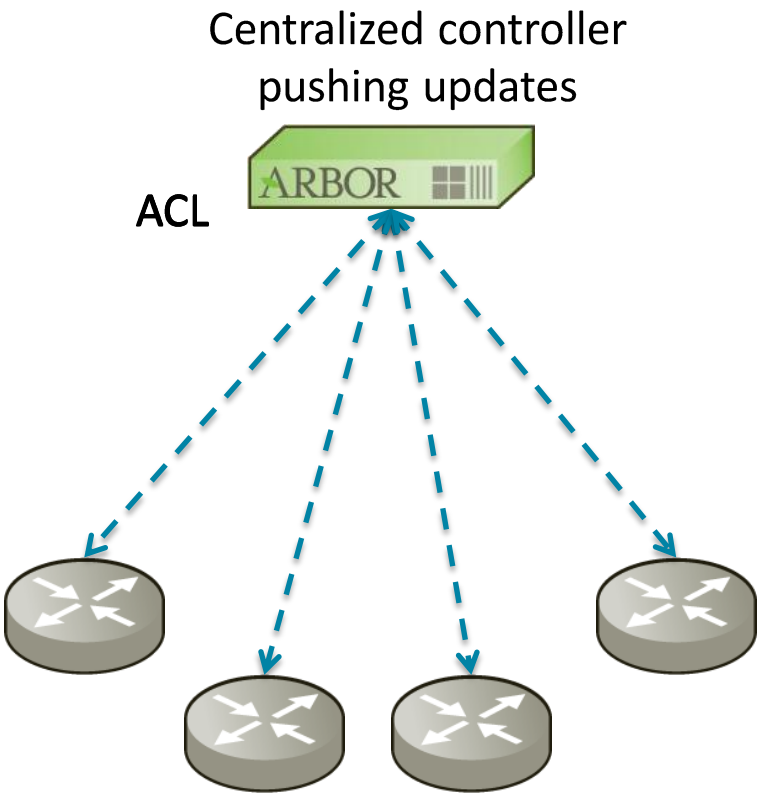
BGP flow spec: why ?

- Interact with the network in order to modify its behavior, optimize QoS, optimize application aware infrastructure
  - Modify the way a traffic behave in a network: QoS, Rate limit, forwarding …



ARBOR NETWORKS

Quick description about Flow spec



- Simple way to understand BGP Flow Spec:
  - Provision ACL/PBR dynamically via MP-BGP
- For this purpose you need to
  - Identify a traffic (Flow)
  - Ask the router to apply an action to this traffic
- Just like with ACL/PBR you can
  - Drop, Rate limit, mark, re-mark and redirect, …
- We are more or less doing SDN for real wo Open Flow

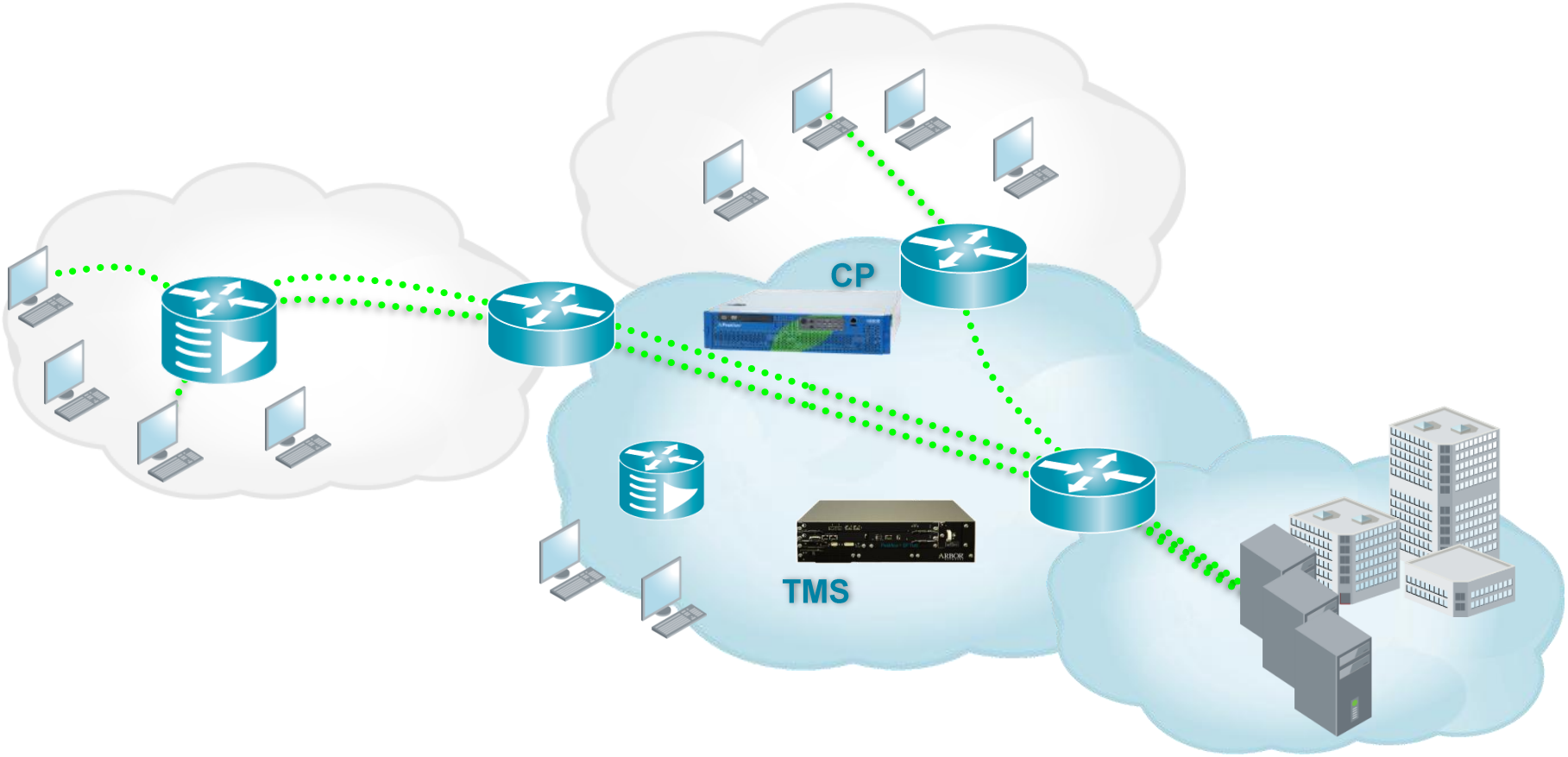
ARBOR NETWORKS

BGP Flow Spec

PEAKFLOW IMPLEMENTATION

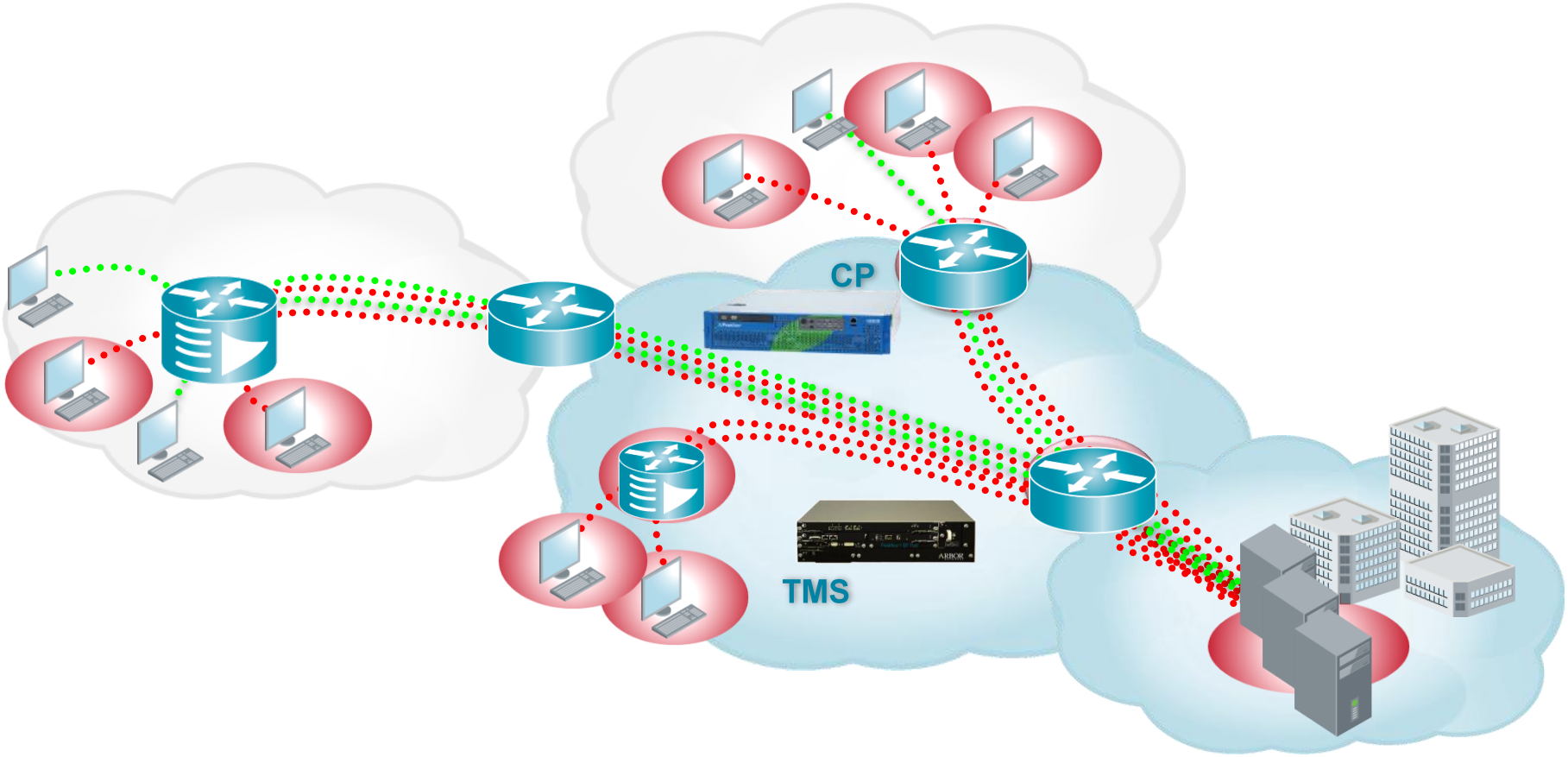
ARBOR NETWORKS

Traditional DDoS - Mitigation



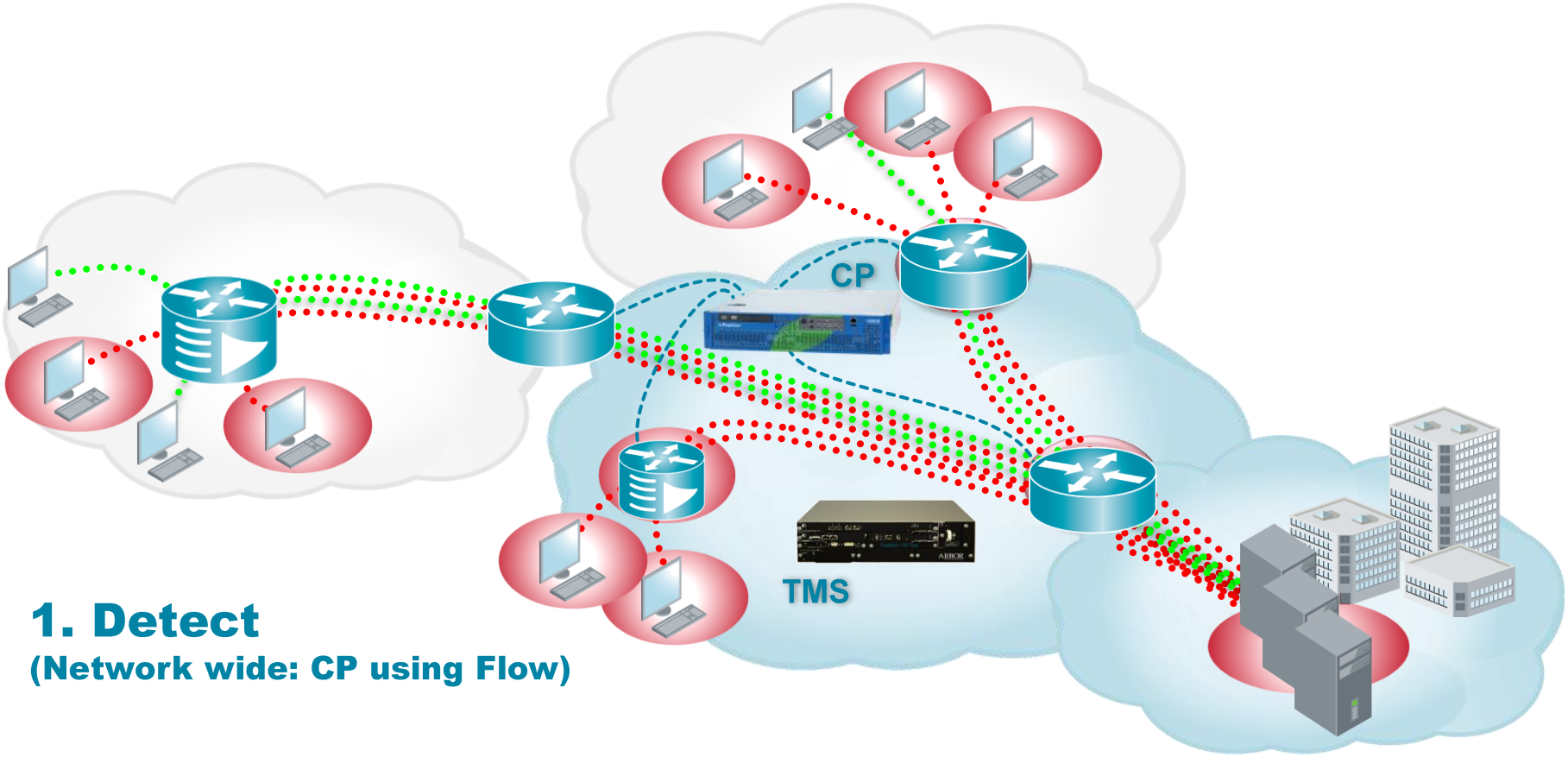
ARBOR NETWORKS

DDoS - Mitigation



ARBOR NETWORKS

DDoS - Mitigation

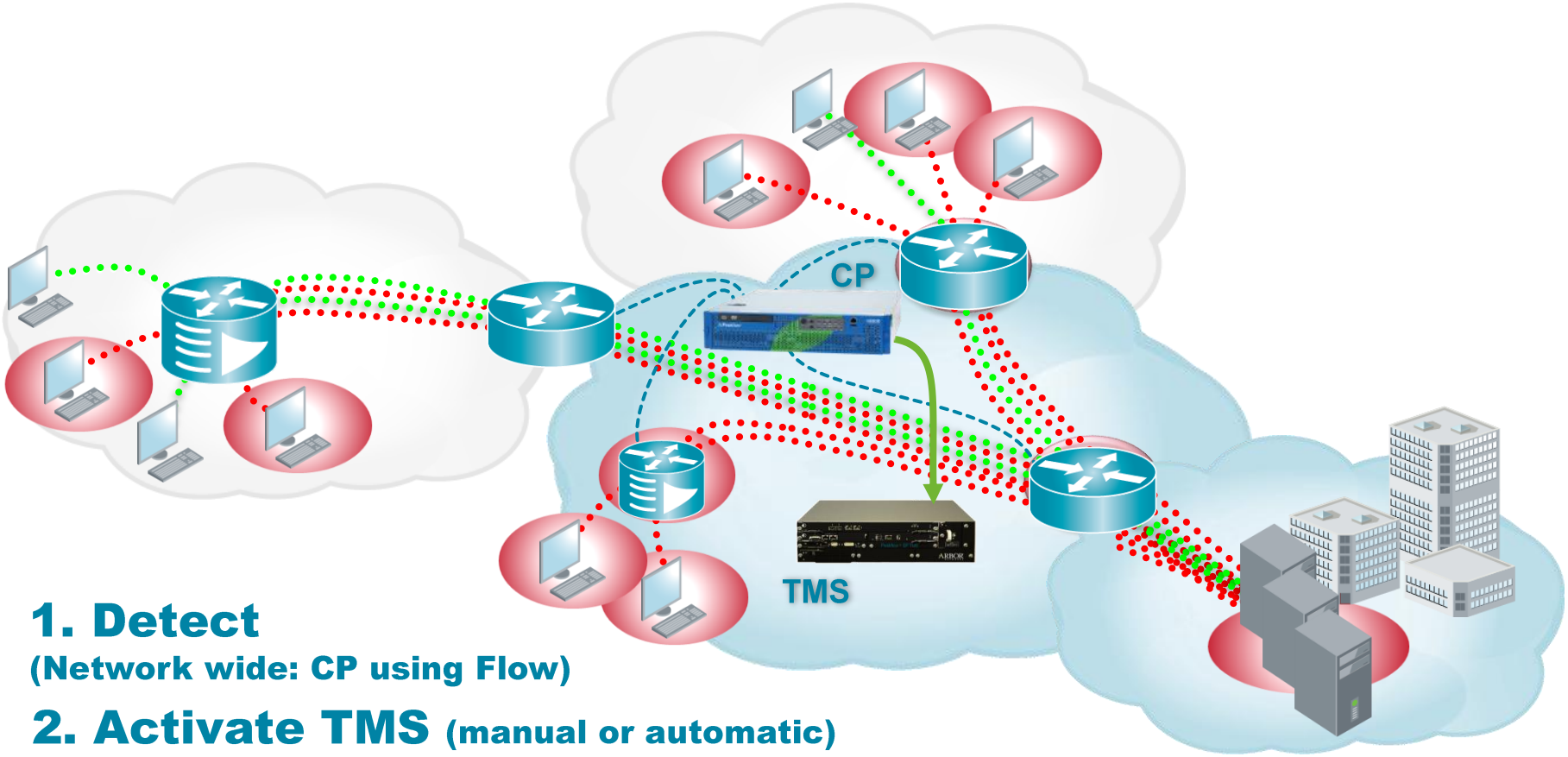


1. Detect

(Network wide: CP using Flow)

ARBOR NETWORKS

DDoS - Mitigation

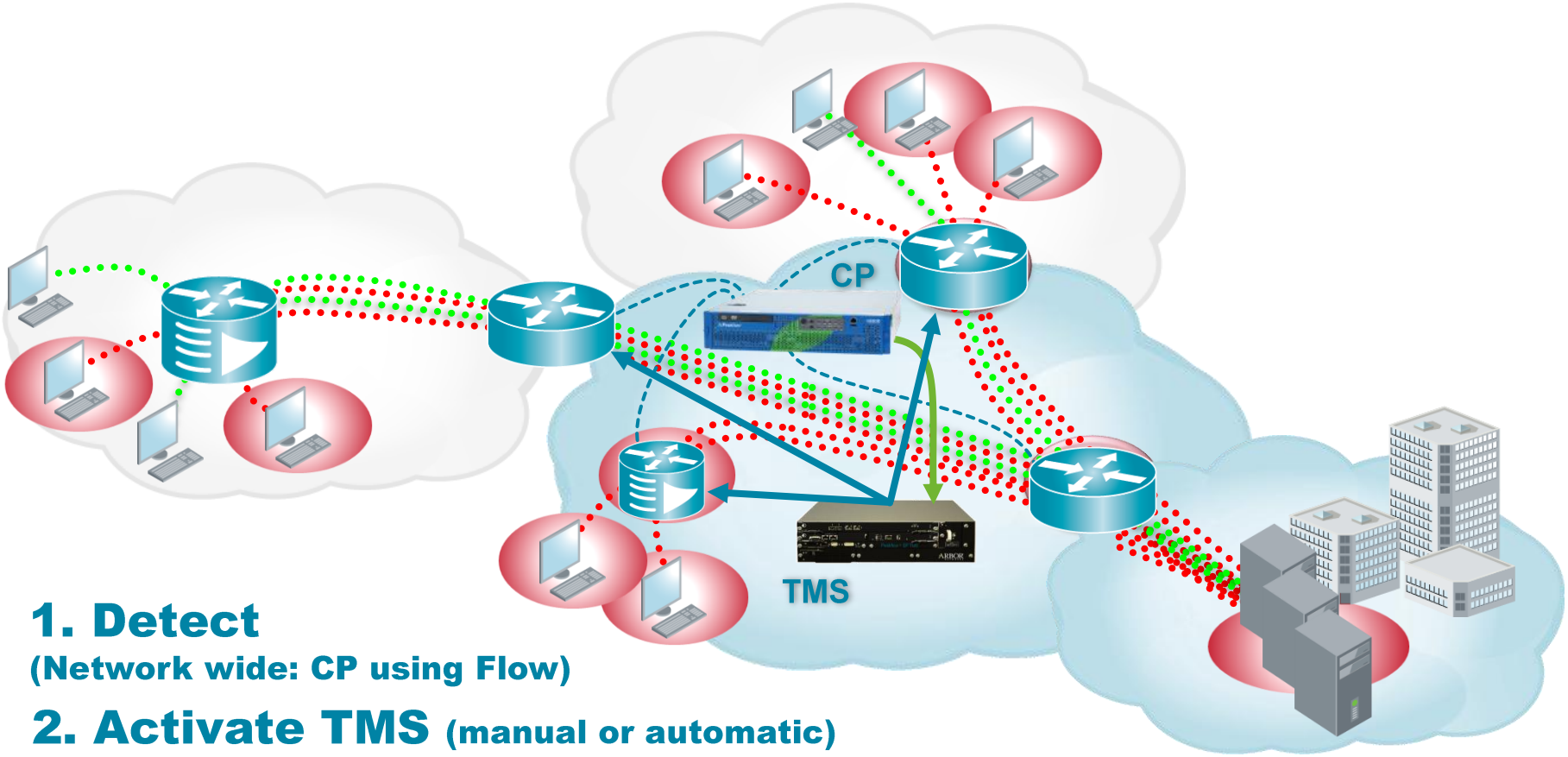


1. Detect
(Network wide: CP using Flow)

2. Activate TMS (manual or automatic)

ARBOR NETWORKS

DDoS - Mitigation



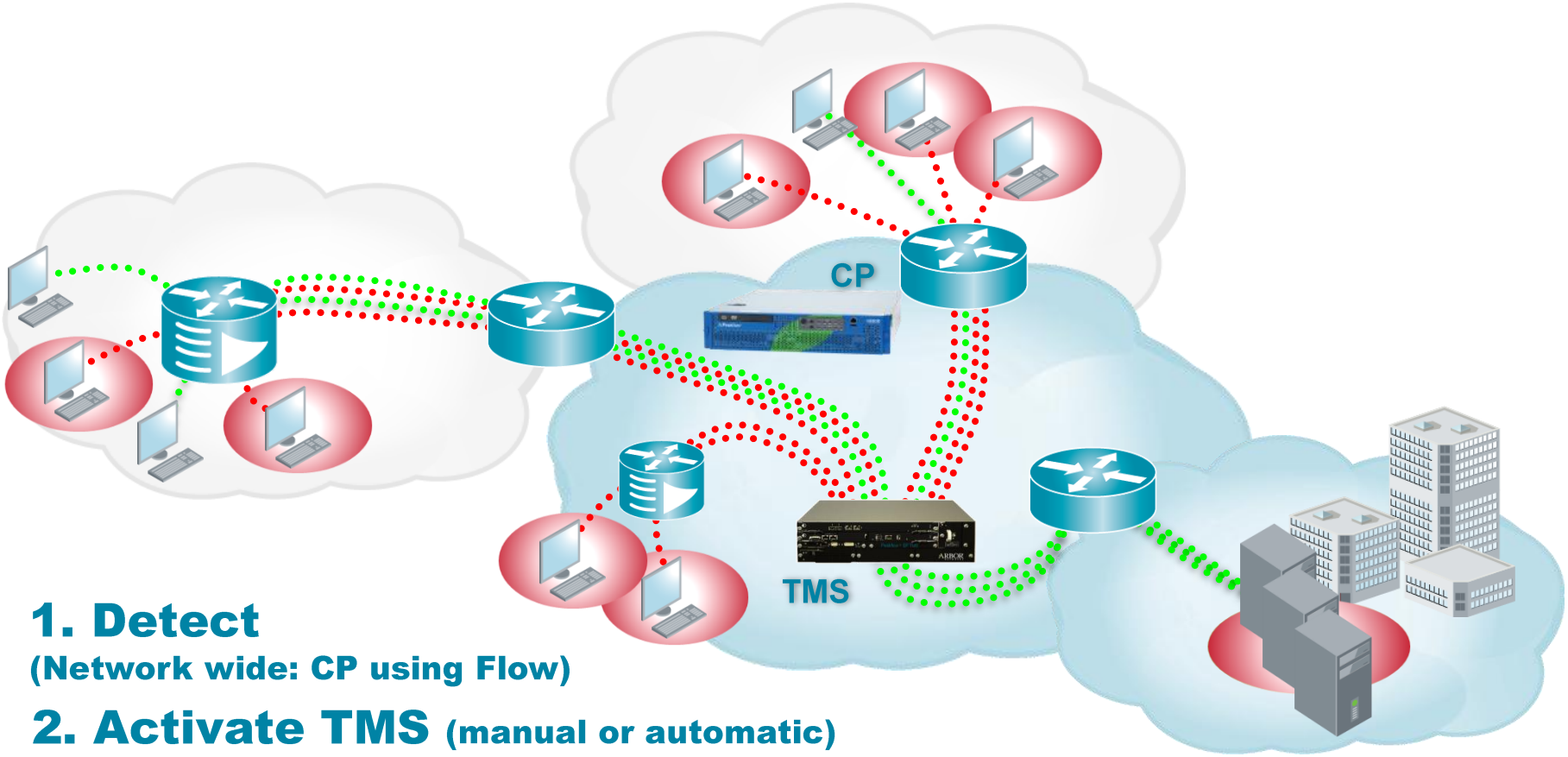
1. Detect
(Network wide: CP using Flow)

2. Activate TMS **(manual or automatic)**

3. Divert Traffic **(Network wide: BGP OFF-Ramp announcement)**

ARBOR NETWORKS

DDoS - Mitigation



1. Detect
(Network wide: CP using Flow)

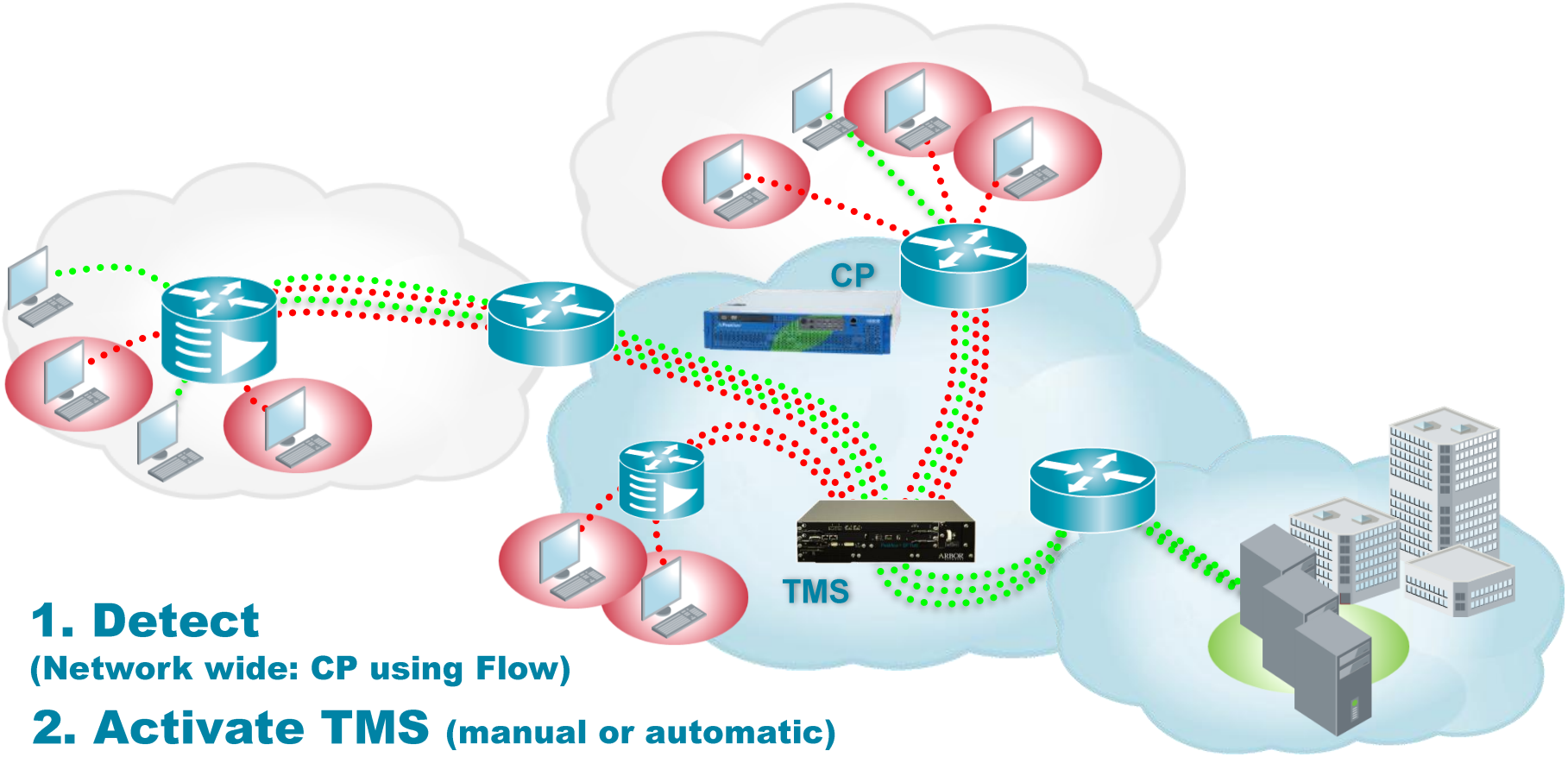
2. Activate TMS (manual or automatic)

3. Divert Traffic (Network wide: BGP OFF-Ramp announcement)

4. Clean the Traffic and forward the legitimate
(Network wide: using ON-Ramp Technique [e.g. MPLS, GRE, VLAN, ...])

ARBOR NETWORKS

DDoS - Mitigation



1. Detect
(Network wide: CP using Flow)

2. Activate TMS (manual or automatic)

3. Divert Traffic (Network wide: BGP OFF-Ramp announcement)

4. Clean the Traffic and forward the legitimate
(Network wide: using ON-Ramp Technique [e.g. MPLS, GRE, VLAN, ...])

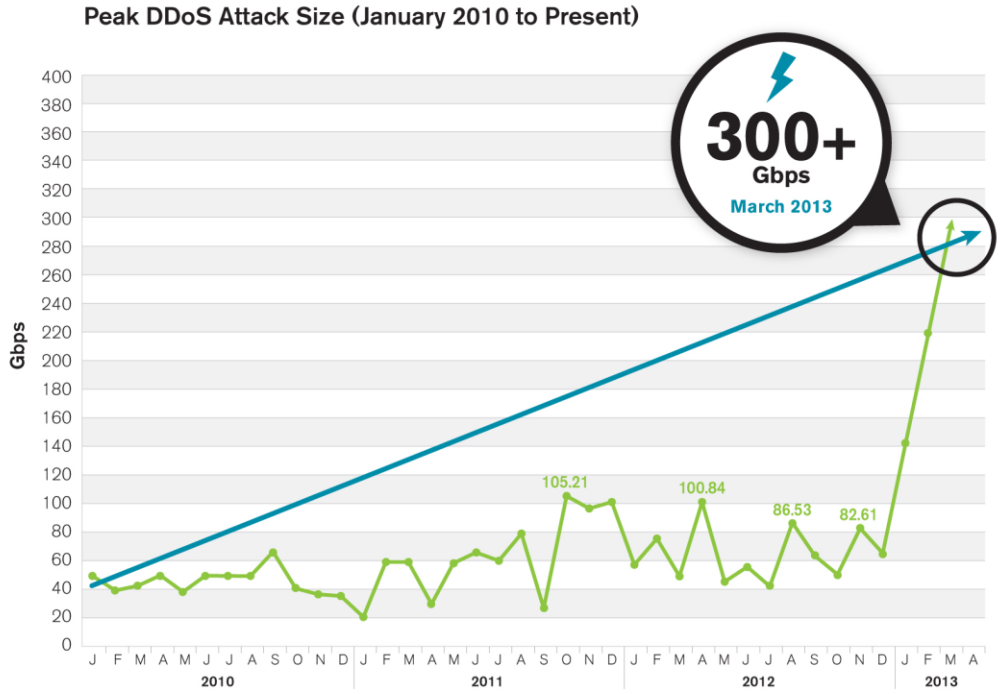
5. Protected

ARBOR NETWORKS

Why FlowSpec is important for Arbor?

1. It is one of our core competitive advantages
2. It is one of our “core” technologies that we should promote
3. It is one of the best responses to “hockey stick era”

D. McPherson
Arbor Networks
August 2009



Source: Arbor Networks, Inc.

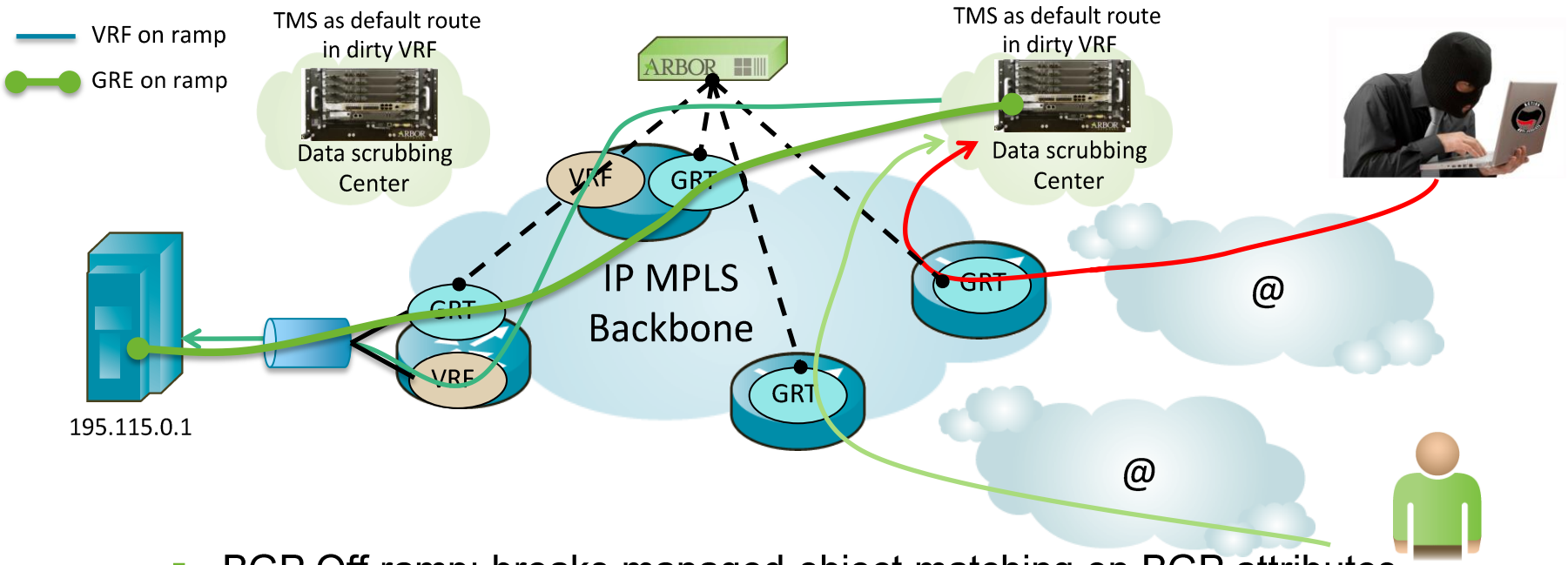
ARBOR NETWORKS

BGP flow spec
Implementation milestones

- With Peakflow SP supports BGP flow spec action rate limit and black hole
  - Allows similar behavior RTBH and S/RTBH: source IP / destination IP (simple L3 blackhole)
  - Allows traffic drop a flow using matching on TCP/UDP ports, ICMP type, ICMP code, TCP flags, packet lengths, DSCP, Fragment, …
- With Peakflow SP/TMS supports BGP flow spec action redirect
  - Allows to redirect IP packets matching a flow to an IP VPN for off ramp purposes

ARBOR NETWORKS

BGP off ramp: today's constraints

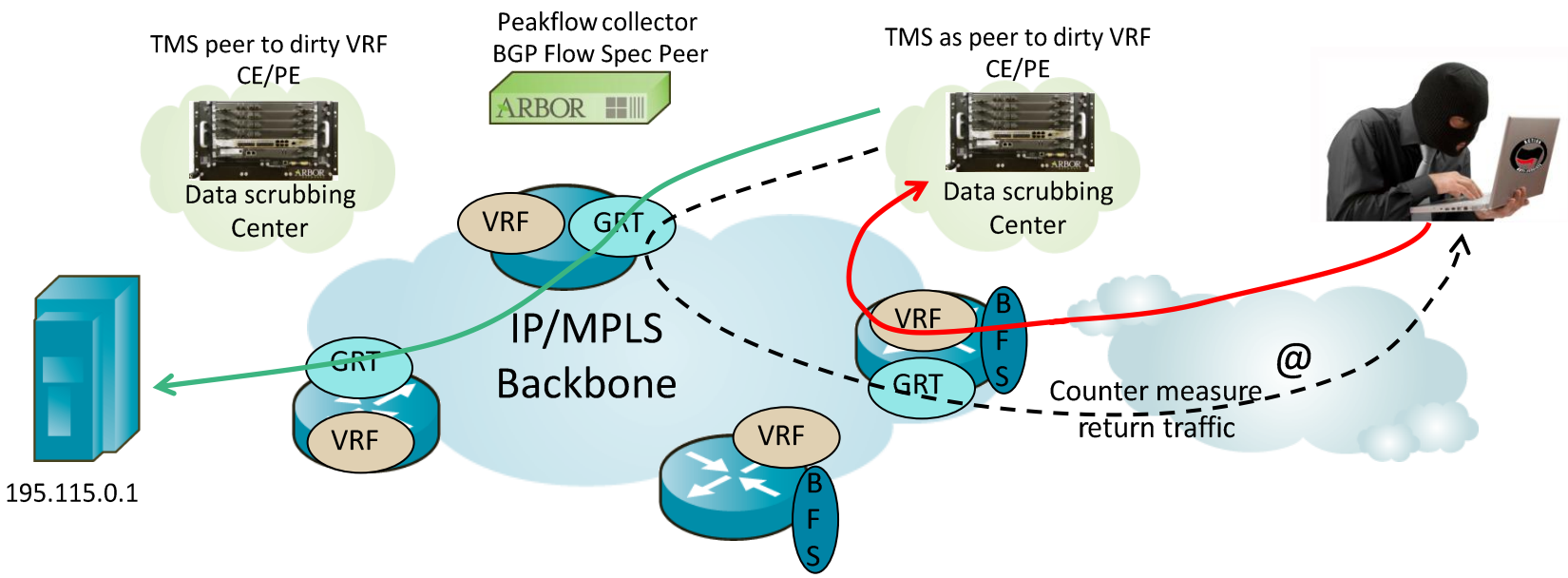


- BGP Off ramp: breaks managed object matching on BGP attributes which makes issues with alerting and auto-mitigation impossible
  - Off ramp update doesn't preserve AS path nor communities
- VRF on ramp:
  - Manual leaking and stating routing for each and every protected prefix
  - Always challenging to use the same interface/vlan for GRP and on ramp traffic on the server side
- GRE tunnel
  - Scalability issues for GRE termination, service card issue (Juniper/ALU),
  - GRE proliferation issue when we have several TMS's
  - Static route and GRE manual provisioning
  - GRE troubleshooting: no real OAM, keep alive not always easy to use

ARBOR NETWORKS

BGP diversion using BGP flow spec

- BGP flow spec can be applied on a predefined set of routers
  - Typically peering edge routers



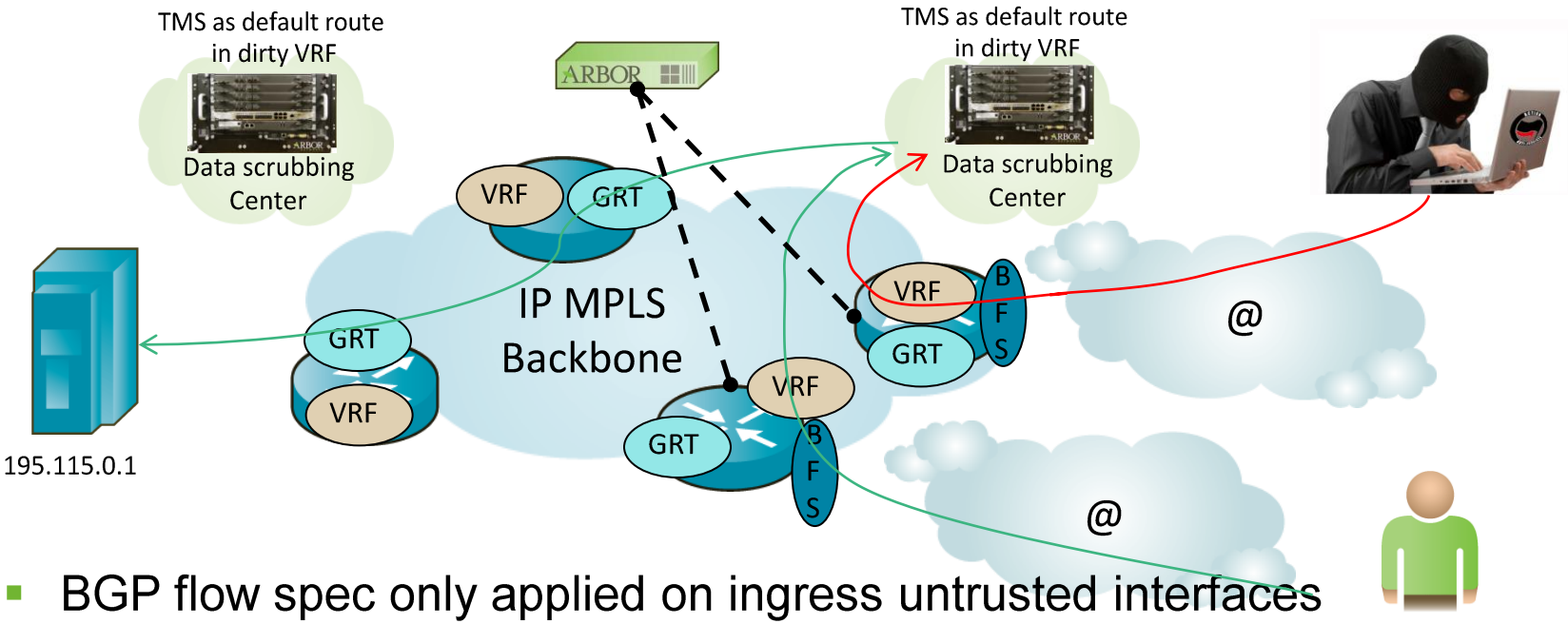
Main pro's against today's approach

- Can be automatically provisioned without any manual configuration and for whatever IP being under attack
  - no manual configuration like route leaking, static route configuration, …
  - We do not impact global routing table for the return path of the clean traffic
- We are really surgical : only diversion of specific flows
  - We can select traffic based on source/dest IP, TCP/UDP ports

ARBOR NETWORKS

Re-Injection with BGP flow spec

- You just have nothing to do as you didn't impact routing table for diversion



- BGP flow spec only applied on ingress untrusted interfaces (ALU)
- BGP flow spec only on peering edge (Juniper)
  - Make sure that protected customer or server is not attached to a peering router

ARBOR NETWORKS

Caveats and limitations

- Flow Spec availability
  - ALU 7750:
    - R9 and above
    - Full support and flexibility to enable flow spec on a per interface basis
    - IPv4 and IPv6
  - Juniper:
    - JunOS R7.0 (may be earlier version) and above
    - Full support but flow spec rules applied to all router interface
  - Cisco
    - CRS models with version 5.2.2 IOS-XR version
- FlowSpec works differently in three main Vendors: Cisco, Juniper and ALU.
- BGP flow spec IPv6 limeted support in Router Vendors

ARBOR NETWORKS

BGP Flow Spec

WORKING WITH FLOWSPEC

ARBOR NETWORKS

Does it work ?

Test Description	Juniper Result	Alcatel Result	Cisco Result
Blackhole a flow by Source IP	PASSED	PASSED	PASSED
Blackhole a flow by Destination IP	PASSED	PASSED	PASSED
Blackhole a flow by Source Prefix	PASSED	PASSED	PASSED
Blackhole a flow by Destination Prefix	PASSED	PASSED	PASSED
Blackhole a flow by Destination IP and Protocol Number UDP (17)	PASSED	PASSED	PASSED
Blackhole a flow by Destination IP and Protocol Number TCP (6)	PASSED	PASSED	PASSED
Blackhole a flow by Destination IP and Protocol Number ICMP (1)	PASSED	PASSED	PASSED
Backhole a flow by src/dst IP and src/dst Port	PASSED	PASSED	PASSED
Blackhole a flow by destination IP and fragment	PASSED	PASSED	PARTIAL FAILED
Blackhole a flow by destination IP, Protocol Number UDP(17) and Packet Size (range of size)	PASSED	FAILED	PASSED
Blackhole a flow by destination IP, Protocol Number UDP(17) and Packet Sice (fixed size)	PASSED	FAILED	PASSED
Blackhole a flow by Source IP, Protocol Number TCP(6) and TCP Sync Flag	PARTIAL FAILED	PASSED	PASSED
Rate Limiting Flow	PARTIAL FAILED	FAILED	PASSED
Redirect a flow to a specific VRF	PARTIAL FAILED	PASSED	PASSED

ARBOR NETWORKS

FlowSpec tests in major ISP

Platform	Current limitations	Major steps
Juniper MX	- FlowSpec is applied to all interfaces	Working on next phase: -Redirect to IP next-hop -Increase flow route scaling
ALU 7950	- 512 flow routes - Can't rate-limit>0 (can drop only)	No information
Cisco ASR9K	- Partial support of Fragmentation flags - Typhoon/Tomahawk line card needed. - Trident Card line not supported.	Test done using 5.2.0.
Huawei NE5000E		NE5000E supports FlowSpec with V8R3. Probably there should be a custom build for NE40E?

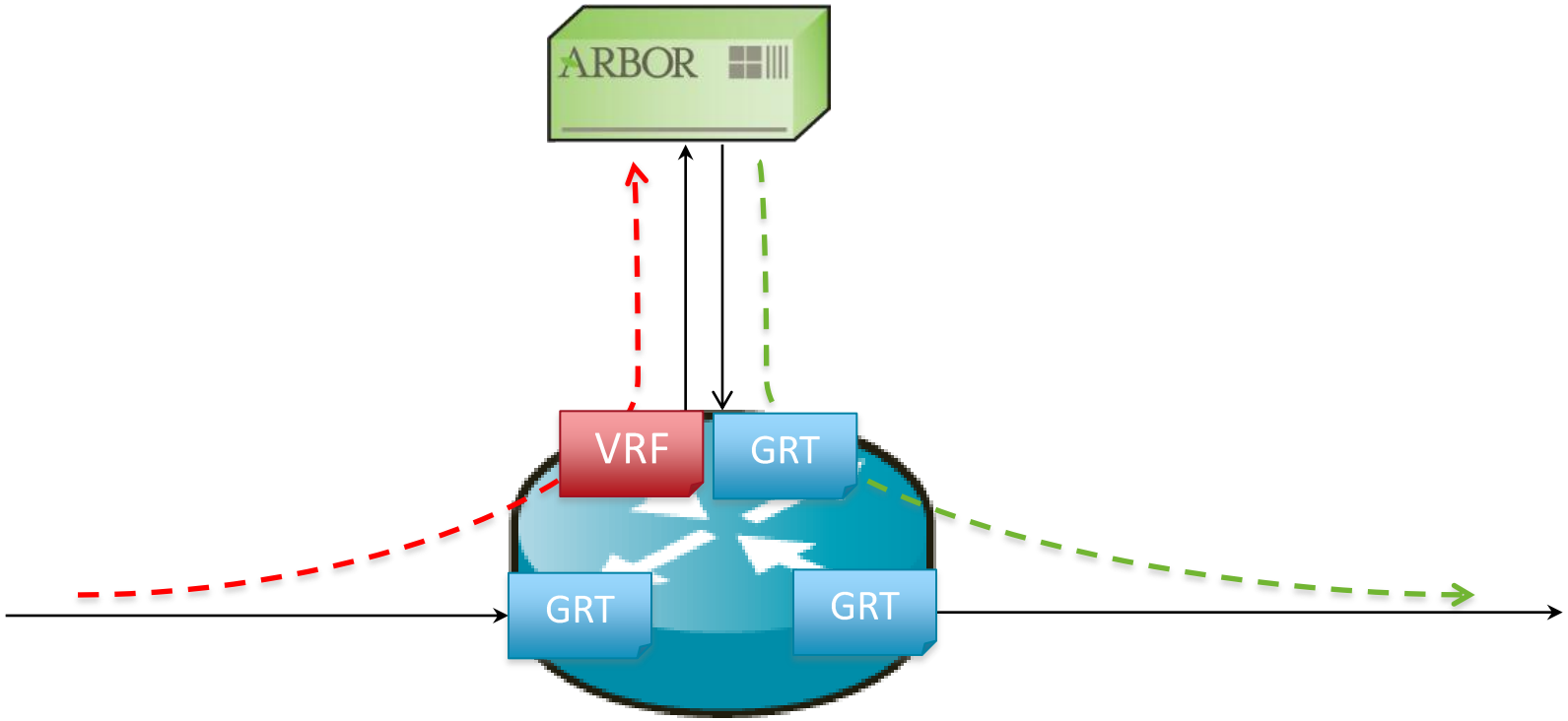
ARBOR NETWORKS

BGP Flow Spec

REDIRECTING TRAFFIC WITH FLOWSPEC

ARBOR NETWORKS

FlowSpec redirection: no issues with

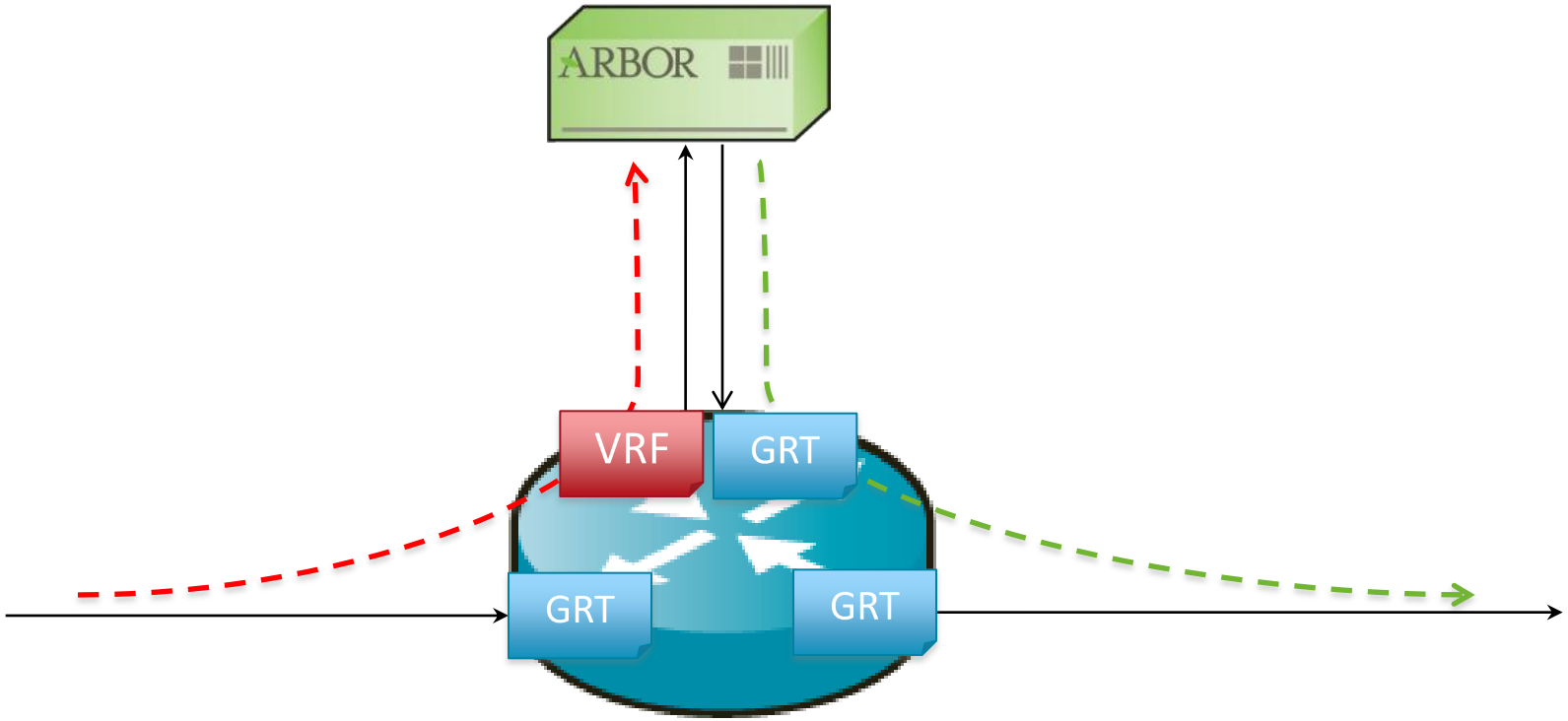


ALU's FlowSpec works as OPT-Out per interface

-By default ALL interfaces accept FlowSpec.

-FlowSpec can be disabled in specific Interfaces.

FlowSpec redirection: no issues with

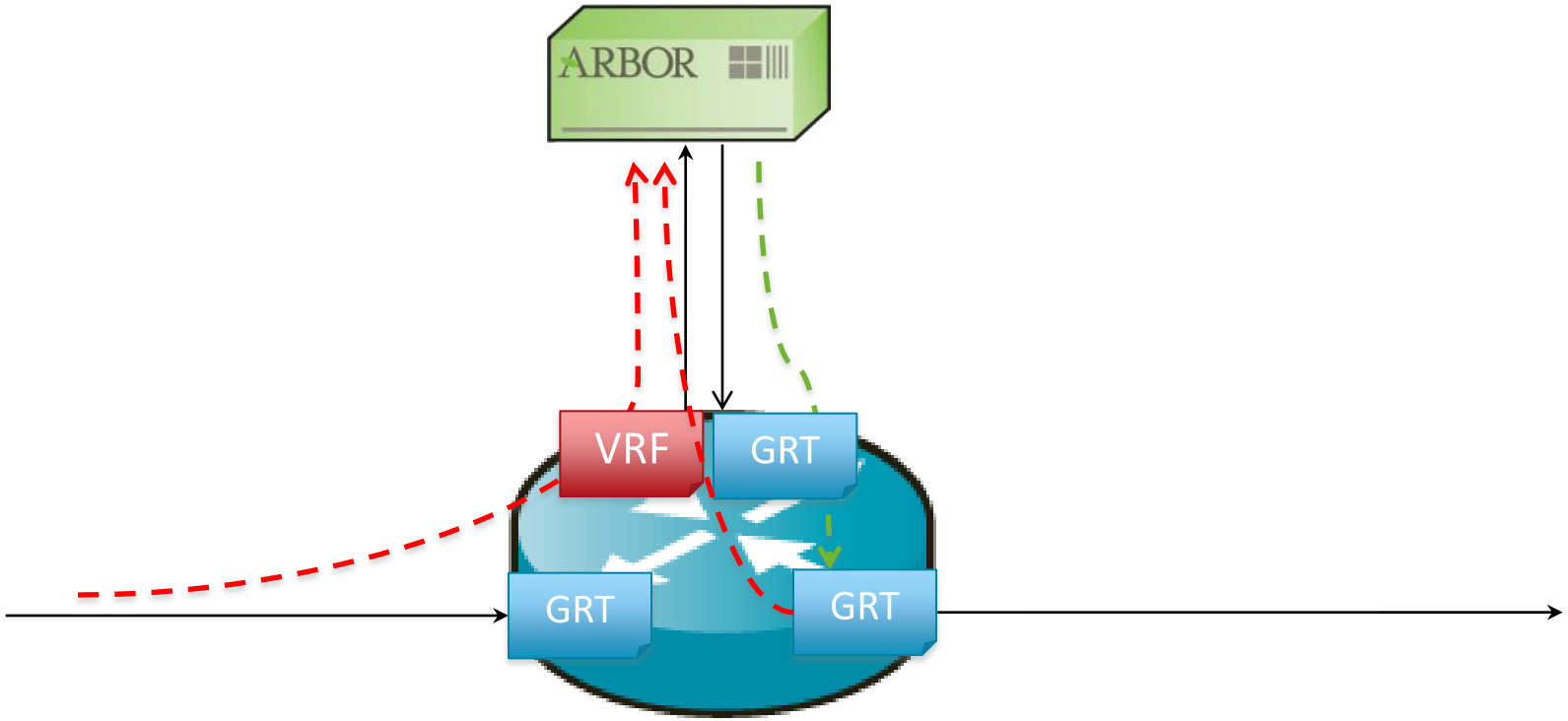


Cisco's FlowSpec works as OPT-In per interface

-By default non interfaces accept FlowSpec.

-FlowSpec must be enabled in specific Interfaces.

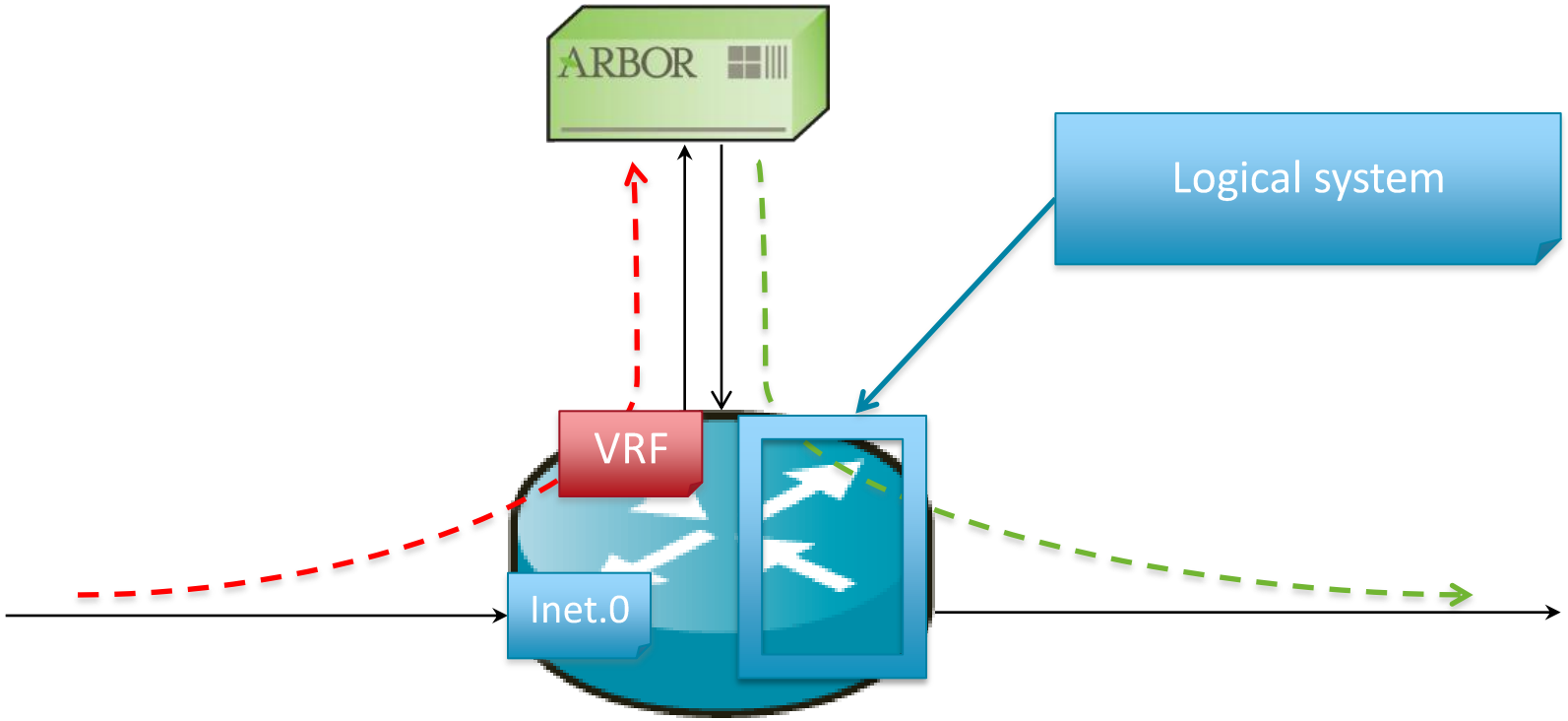
FlowSpec redirection: issues with



Junos FlowSpec works in ALL Interfaces

- If On-ramp traffic goes to another interface in GRT a loop is created for this flow.

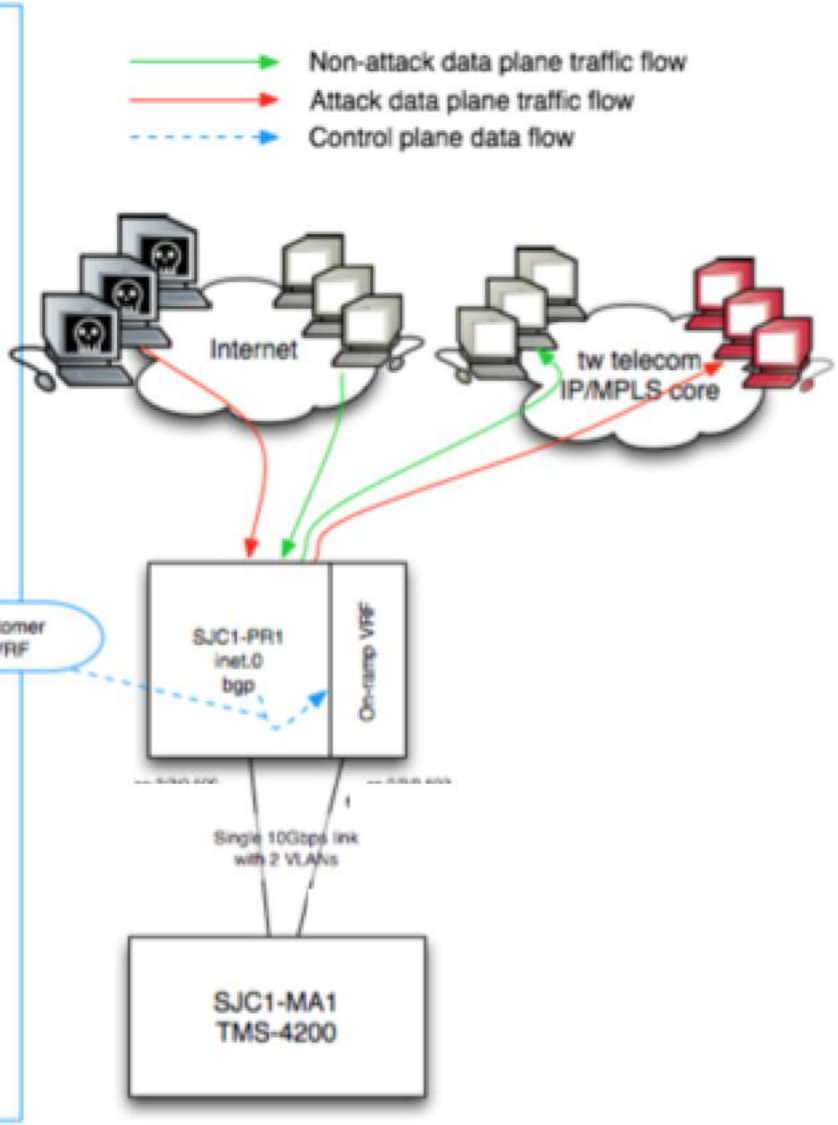
FlowSpec redirection: workaround #1



Create logical-systems for clean traffic. Not very well perceived by customers because of complexity: we need BGP session between LS and master, we need LT interfaces etc…

FlowSpec redirection: workaround #2

```
routing-options {
    rib-groups {
        TMS_ONRAMP_RIB {
            import-rib [ inet.0 :    _TMS_ONRAMP_VRF.inet.0 ];
            import-policy TWTC_TMS_ONRAMP_RIB_IMPORT;
        }
    }
}
protocols {
    bgp {
        family inet {
            unicast {
                rib-group     C_TMS_ONRAMP_RIB;
            }
        }
    }
}
policy-options {
   policy-statement     :_TMS_ONRAMP_RIB_IMPORT {
        term TMS_BH {
            from {
                protocol bgp;
                community TMS_BLACKHOLE;
            }
            then reject;
        }
        term ACCEPT_CUSTOMER {
            from {
                protocol bgp;
                community [ ROUTING_BGPCUST ROUTING_STATCUST
ROUTING_CITYAGG ];
            }
            then accept;
        }
        term REJECT {
            then reject;
        }
    }
}
routing-instance {
    TWTC_    ONRAMP_VRF {
        instance-type vrf;
        interface xe-2/2/0.502;
        vrf-target target:4___:113000;
    }
}
```



FlowSpec for offramp, VRF for onramp. Use of Rib-Groups to leak the traffic from VRF to GRT with no Lookup.

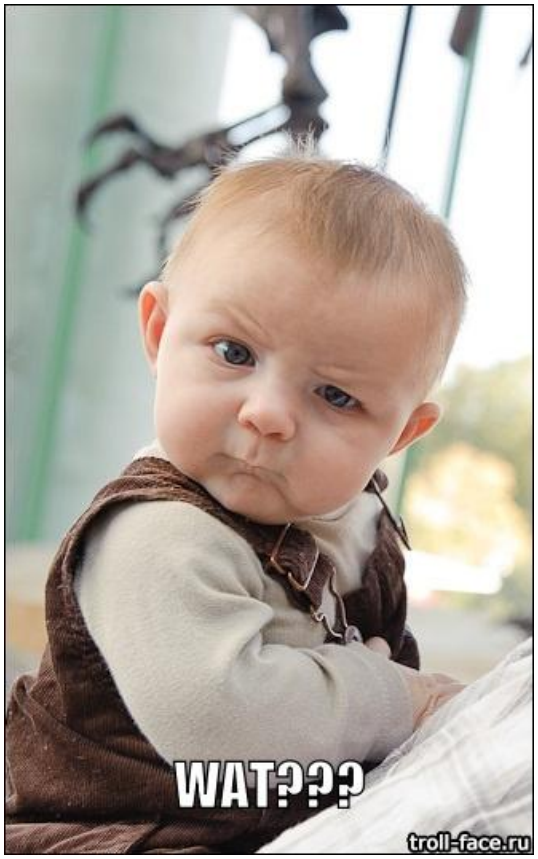
BGP Flow Spec

BLOCKING TRAFFIC WITH FLOWSPEC

ARBOR NETWORKS

Mitigating DNS

Generic messaging: use ACLs / FlowSpec, but do not block UDP/53 completely.



Customer: so what EXACTLY you propose me to do?

ARBOR NETWORKS

What you should NOT do

Block traffic from UDP/53 completely

1. It drops legitimate DNS replies
2. It does not drop fragments as non-initial fragments do not contain UDP header
   - Normally amplified responses are 3K-4K bytes long
   - Initial fragment is 1500 bytes
   - Followed by 2 or 3 fragments
   - Blocking UDP/53 you miss 50-70% of attack traffic

A better than best practice of DNS mitigation

1. Run FlowSpec to drop initial DNS fragments
2. Run BGP redirect to divert non-initial fragments to TMS
3. Let Invalid packets take care of non-initial fragments:

"A fragment considered to be incomplete if TMS tries to get all fragments of the datagram during one second and fails. In this case all these are considered to be incomplete"

Invalid packets CM details: https://wiki.arbor.net/arbor/kirill

ARBOR NETWORKS

Things to keep in mind

1. Requires TMS capacity around 50-70% of attack size
2. Test fragmentation bitmask before using:

About Fragmentation Bitmask Menus

Identifiers for fragmentation bitmask menus

The following table contains common traffic identifiers for bitmask menus:

Bitmask Value	Purpose
0	Do not fragment
1	Is a fragment
2	First fragment
3	Last fragment

How does this bitmask work?

ARBOR NETWORKS

How does fragmentation bitmask work?

Fragmentation bitmask is used to check these fields:

- Don't fragment (DF)
- More fragments (MF)
- Fragmentation offset

My wild guess on how bitmask matches IP header fields:

Bitmask	Description	DF	MF	Offset
0	Do not fragment	1	0	0
1	Is a fragment	0	any	! 0
2	First fragment	0	1	0
3	Last fragment	0	0	! 0

ARBOR NETWORKS

What Junos expects?

Tested in 11.4R9.4.

Bitmask	Description	DF	MF	Offset
1	Do not fragment	1	0	0
2	Is a fragment	0	any	! 0
4	First fragment	0	1	0
8	Last fragment	0	0	! 0

Had no chance to test it on different images, but I guess it is the same. And it seems to match RFC wording:

```
Uses bitmask operand format defined above.

         0   1   2   3   4   5   6   7
       +---+---+---+---+---+---+---+---+
       |   Reserved    |LF |FF |IsF|DF |
       +---+---+---+---+---+---+---+---+

Bitmask values:

+  Bit 7 - Don't fragment (DF)

+  Bit 6 - Is a fragment (IsF)

+  Bit 5 - First fragment (FF)

+  Bit 4 - Last fragment (LF)
```

Oooops…

Oct 29 13:47:21.040 2013 M7-AR5 /kernel: %KERN-6: pid 1997 (dfwd), uid 0: exited on signal 6 (core dumped)
Oct 29 13:47:21.108 2013 M7-AR5 dfwd[47379]: %DAEMON-4: (LOG_INFO) CH_NET_SERV_KNOB_STATE read 0, firewall chassis state NORMAL (All FPC)
Oct 29 13:47:35.878 2013 M7-AR5 /kernel: %KERN-6: pid 47379 (dfwd), uid 0: exited on signal 6 (core dumped)
Oct 29 13:47:36.230 2013 M7-AR5 mgd[47515]: %INTERACT-6-UI_CHILD_START: Starting child '/usr/sbin/dfwd'
Oct 29 13:47:36.243 2013 M7-AR5 mgd[47515]: %INTERACT-6-UI_CHILD_STATUS: Cleanup child '/usr/sbin/dfwd', PID 47519, status 0
Oct 29 13:47:40.905 2013 M7-AR5 dfwd[47493]: %DAEMON-4: (LOG_INFO) CH_NET_SERV_KNOB_STATE read 0, firewall chassis state NORMAL (All FPC)

Congrats! DFWD is cored and tries to restart with no success.

Isn't it one more reason why Inter-ISP FlowSpec is not that popular?

DFWD: dynamic firewall daemon

Mitigating DNS amplification without TMS

You don’t want to divert a volumetric attack to TMS if you have no available TMS resource.

Try this then:

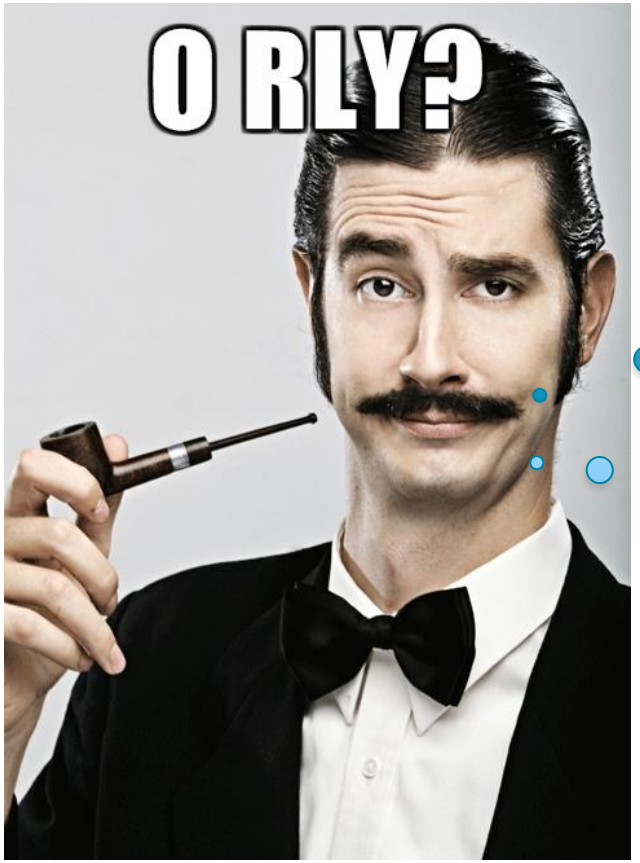
☐ **Drop initial DNS fragments**	**Dst:** 1.1.0.1/32 **Protocols:** 17 **Src Ports:** 53 **Fragment:** 4
☐ **Drop non-initial UDP fragments**	**Dst:** 1.1.0.0/32 **Protocols:** 17 **Fragment:** 2

ARBOR NETWORKS

Mitigating NTP attacks

Ah, that is easy: time synch with NTP utilized packets that 76 bytes long. This simple FCAP should stop NTP attack:

Drop proto udp and src port 123 and not (bpp 76)



Have you thought about fragments?

Have you tested your router on packet size match?

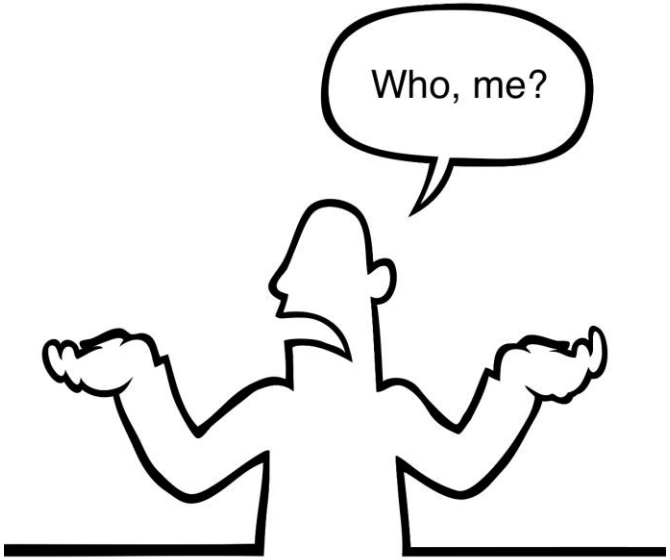
ARBOR NETWORKS

PR # 968125

Summary: BGP flowspec routes with packet-length/icmp-code/icmp-type matching rules take no effect on the firewall filter of the received router.

Fixed in: 13.3R3, 14.1R2, 14.2R1

You might think you JunOS version is not affected?



PR # 968125: who is affected?

Installed Platforms: ACX,EX-Series,M-Series,MX-Series,PTX-Series,T-Series.

Software Versions: 10.4R14,10.4R15,10.4R16,11.4R10,11.4R10-S,11.4R10-S1,11.4R10-S2,11.4R11,11.4R11-S,11.4R11-S1,11.4R11-S2,11.4R12,11.4R8,11.4R8-S,11.4R8-S1,11.4R8-S1.1,11.4R8-S2,11.4R9,11.4R9-S,11.4R9-S1,11.4R9-S2,12.1R10,12.1R5-S,12.1R5-S1,12.1R5-S1.0,12.1R5-S2,12.1R5-S2.2,12.1R5-S3,12.1R5-S3.1,12.1R6,12.1R6-S,12.1R6-S1,12.1R6-S1.1,12.1R6-S2,12.1R6.5,12.1R7,12.1R7-S,12.1R7-S1,12.1R7-S3,12.1R8,12.1R8-S,12.1R8-S1,12.1R8-S2,12.1R8-S2.1,12.1R8-S3,12.1R9,12.1R9-S,12.1R9-S1,12.2R4,12.2R4-S,12.2R4-S1,12.2R4-S2,12.2R4-S3,12.2R4.5,12.2R5,12.2R5-S,12.2R5-S1,12.2R5-S2,12.2R5-S3,12.2R6,12.2R6-S,12.2R6-S1,12.2R6-S2,12.2R7,12.2R7-S,12.2R7-S1,12.2R8,12.2R8-S,12.2R8-S1,12.3R2,12.3R2-S,12.3R2-S1,12.3R2-S1.1,12.3R2-S2,12.3R2-S3,12.3R2-S3.1,12.3R2-S4,12.3R2-S4.2,12.3R2-S5,12.3R2-S6,12.3R2-S7,12.3R2-S7.1,12.3R2-S8,12.3R2.5,12.3R3,12.3R3-S,12.3R3-S1,12.3R3-S2,12.3R3-S2.1,12.3R3-S3,12.3R3-S3.1,12.3R3-S3.2,12.3R3-S4,12.3R3-S4.1,12.3R3-S4.2,12.3R3-S4.3,12.3R3-S5,12.3R3-S5.3,12.3R3-S6,12.3R3-S6.1,12.3R3-S7,12.3R3-S8,12.3R3-S8.1,12.3R3-S9,12.3R4,12.3R4-S,12.3R4-S1,12.3R4-S2,12.3R4-S3,12.3R4-S3.1,12.3R4-S4,12.3R4-S5,12.3R5,12.3R5-S,12.3R5-S1,12.3R5-S1.1,12.3R5-S2,12.3R5-S2.1,12.3R5-S3,12.3R5-S3.1,12.3R5-S4,12.3R6,12.3R6-S,12.3R6-S1,12.3R6-S2,12.3R6-S3,12.3R7,13.1R1,13.1R1.6,13.1R2,13.1R3,13.1R3-S,13.1R3-S1,13.1R4,13.2R1,13.2R1-S,13.2R1-S1,13.2R2,13.2R2-S,13.2R2-S1,13.2R2-S2,13.2R2-S2.1,13.2R2-S3,13.2R2-S4,13.2R2-S5,13.2R2-S5.1,13.2R2-S5.2,13.2R3,13.2R3-S,13.2R3-S1,13.2R3-S2,13.2R3-S3,13.2R3-S3.1,13.2R3-S3.2,13.2R3-S4,13.2R3-S4.1,13.2R4,13.2R4-S,13.2R4-S1,13.2R4-S1.1,13.2R4-S2,13.2R5,13.3R1,13.3R1-S,13.3R1-S1,13.3R1-S1.1,13.3R2,13.3R2-S,13.3R2-S1,14.1R1

Questions???

ARBOR NETWORKS

ARBOR NETWORKS

Thank You