

TOMORROW starts here.



Cisco *live!*

BGP Optimising the Foundational SDN Technology

BRKSPG-2641

Oliver Boehmer
Cisco AS Solutions Architect

Agenda

- Some words about SDN

- BGP-Assisted SDN Use-case
 1. WAN Orchestration – BGP-LS
 2. Flow Steering/Security Policies – BGP-FS
 3. Peering Diagnostics – BMP



Introduction to SDN



The network paradigm as we know it...



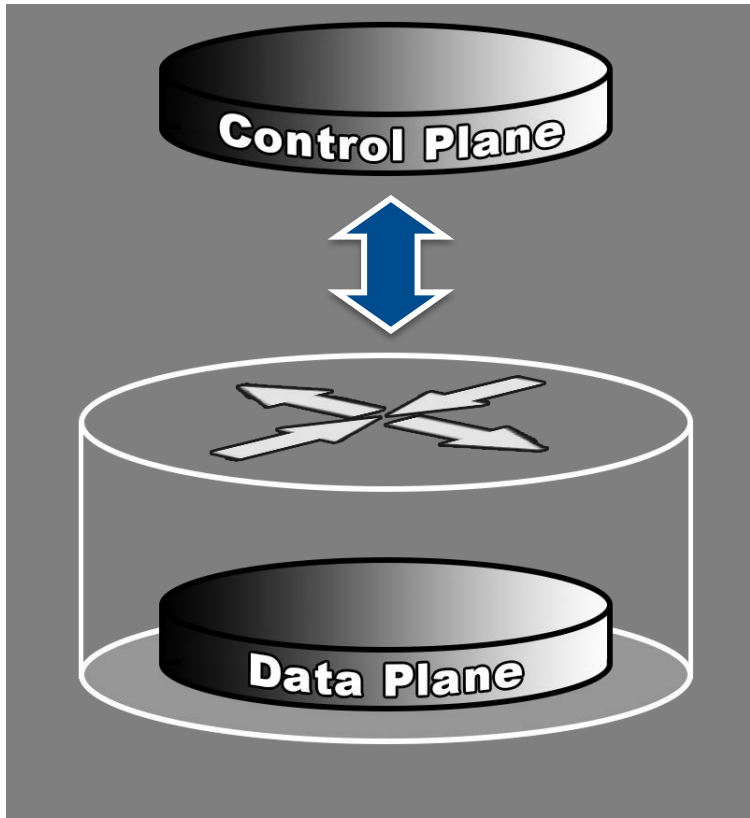
Control and Data Plane resides within Physical Device



What is SDN?

(per Wikipedia definition)

Software defined networking (SDN) is an approach to building computer networks that separates and abstracts elements of these systems



In other words...

In the SDN paradigm, not all processing happens inside the same device

A Better Definition

SDN Definition

Centralisation of control of the network via the

Separation of control logic to off-device compute, that

Enables **automation and orchestration** of network services via

Open **programmatic** interfaces


SDN Benefits

Efficiency: optimise existing applications, services, and infrastructure

Scale: rapidly grow existing applications and services

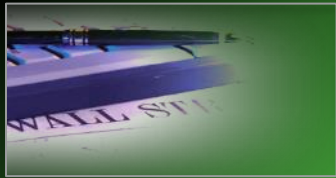
Innovation: create and deliver new types of applications and services and business models

In Lament's Terms



SDN? It's all about
speed!!

Different Customers, Different Pain Points



Research/ Academia

- Experimental OpenFlow/SDN components for production networks

Network
“Slicing”



Massively Scalable Data Centre

- Customise with Programmatic APIs to provide deep insight into network traffic

Network Flow
Management



Cloud

- Automated provisioning and programmable overlay, OpenStack

Scalable
Multi-Tenancy



Service Providers

- Policy-based control and analytics to optimise and monetise service delivery

Agile Service
Delivery
Transport Efficiency



Enterprise

- Virtual workloads, VDI, Orchestration of security profiles

Private Cloud
Automation

**Diverse Programmability Requirements Across Segments
Most Requirements are for Automation & Programmability**

Cisco's SDN Vision

**Program for
Optimised
Experience**

Policy & Intent

Applications

**Network
Intelligence,
Guidance**

**Harvest
Network
Intelligence**

**Services
Orchestration**

Analytics

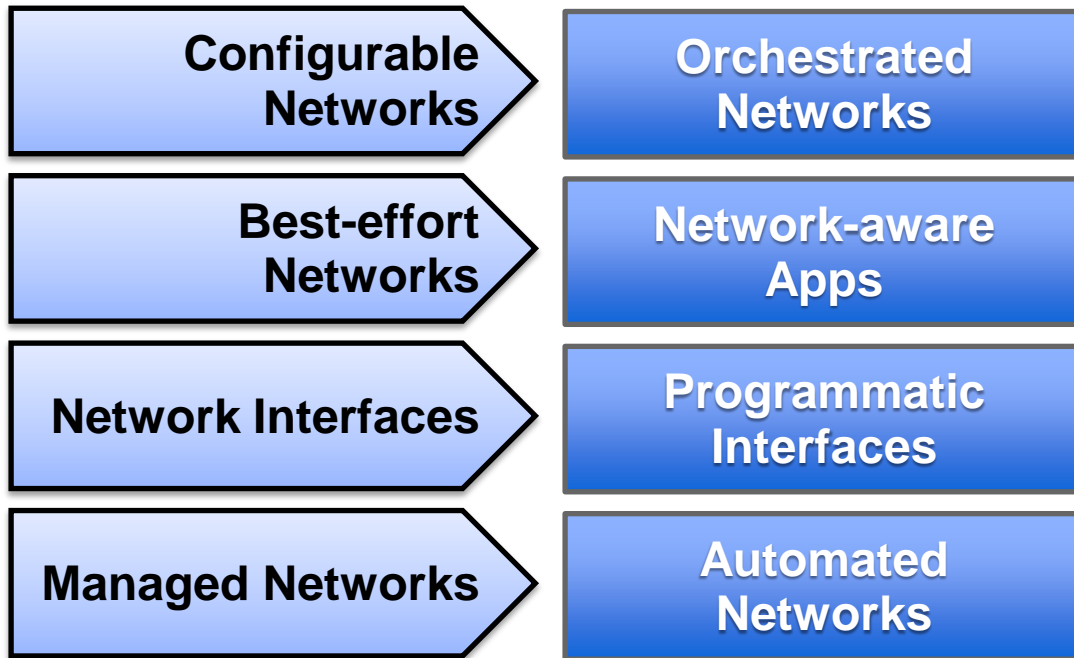
Programmability

Network

**Stats, State &
Events**

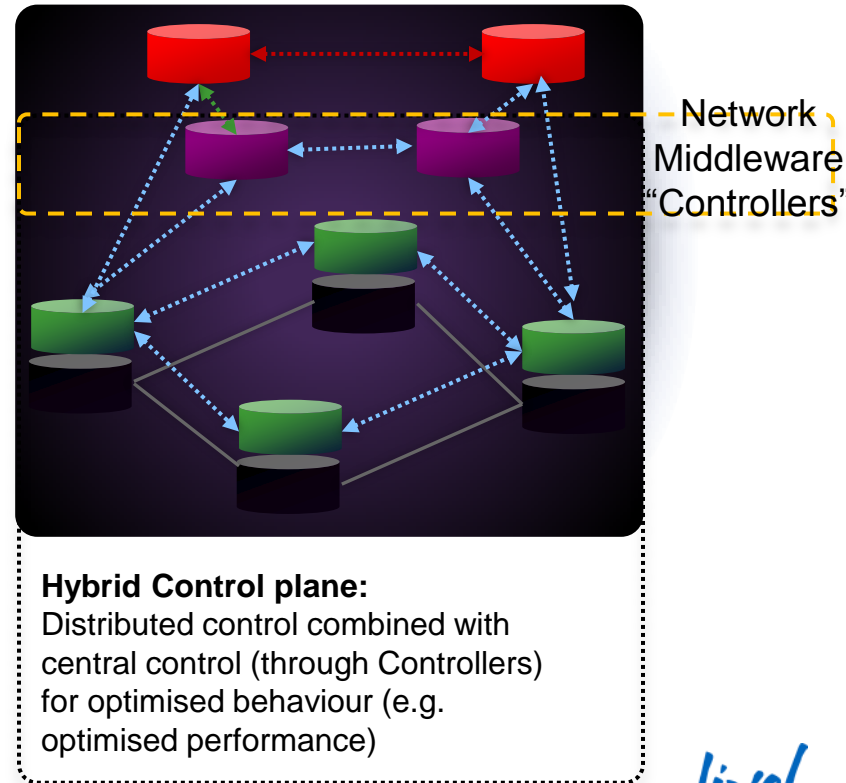
Towards A New Area In Networking

Make everything go faster, easier and more agile



SDN Hybrid Approach

- 20+ Years investment in Distributed Control Planes—capex, skills and expertise—
by both vendors and customers
- Distributed Control Planes designed to survive battlefield conditions with the possibility of multiple failures
- Leave the distributed control plane in place for “normal” traffic, use SDN for traffic that needs special handling (routing, bandwidth reservation etc.)
- In the event of an SDN Controller failure, you still have a network that works, maybe not as optimally





About BGP

Why is BGP Successful?

Extensible

- Multi-protocols, AFs
- Incremental
- NLRI, PA, Community
- Capability Negotiation
- Flexible Policy
- Many Services !!

Simple and Scalable

- Structured (Route Reflector)
- Divide and Conquer (Confederation)
- Low protocol overhead
- Simple FSM
- Simple Messages

HA and Secure

- Run over TCP
- NSR
- PIC, Add-Path
- MD5 authentication
- RPKI validation

“Driven by Pragmatism”, “Not perfect, but good enough”

-- Yakov Rekhter

Control-plane Evolution

Most of services are moving towards BGP

Service/transport	200x and before	2013 and future
IDR (Peering)	BGP	BGP (IPv6)
SP L3VPN	BGP	BGP + FRR + Scalability
SP Multicast VPN	PIM	BGP Multicast VPN
DDOS mitigation	CLI	BGP flowspec
Network Monitoring	SNMP	BGP monitoring protocol
Security	Filters	BGP Sec (RPKI), DDoS Mitigation
Proximity		BGP connected app API
SP-L3VPN-DC		BGP Inter-AS, VPN4DC
Business & CE L2VPN	LDP	BGP PW Sign (VPLS)
DC Interconnect L2VPN		BGP MAC Sign (EVPN)
MPLS transport	LDP	BGP+Label (Unified MPLS)
Data Centre	OSPF/ISIS	BGP + Multipath
Massive Scale DMVPN	NHRP / EIGRP	BGP + Path Diversity
Campus/Ent L3VPN	BGP	BGP



Use Case #1: WAN Orchestration

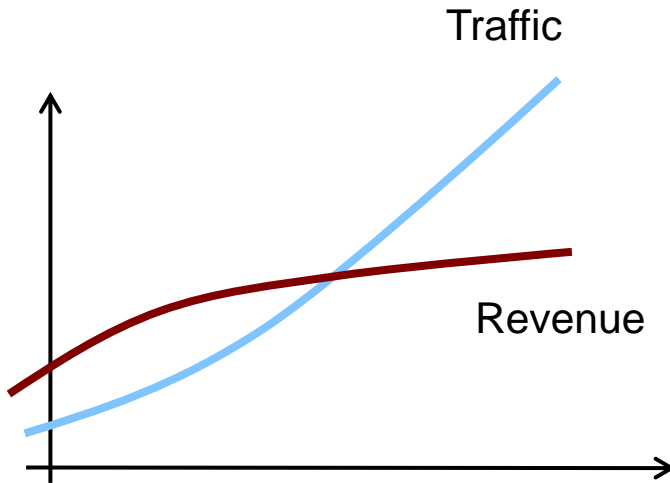
“.. not sure why folks keep talking about SDN as a datacenter technology - the value is in the WAN..”

Vijay Gill

<https://twitter.com/vgill/status/227539039979446272>



The SP Challenge

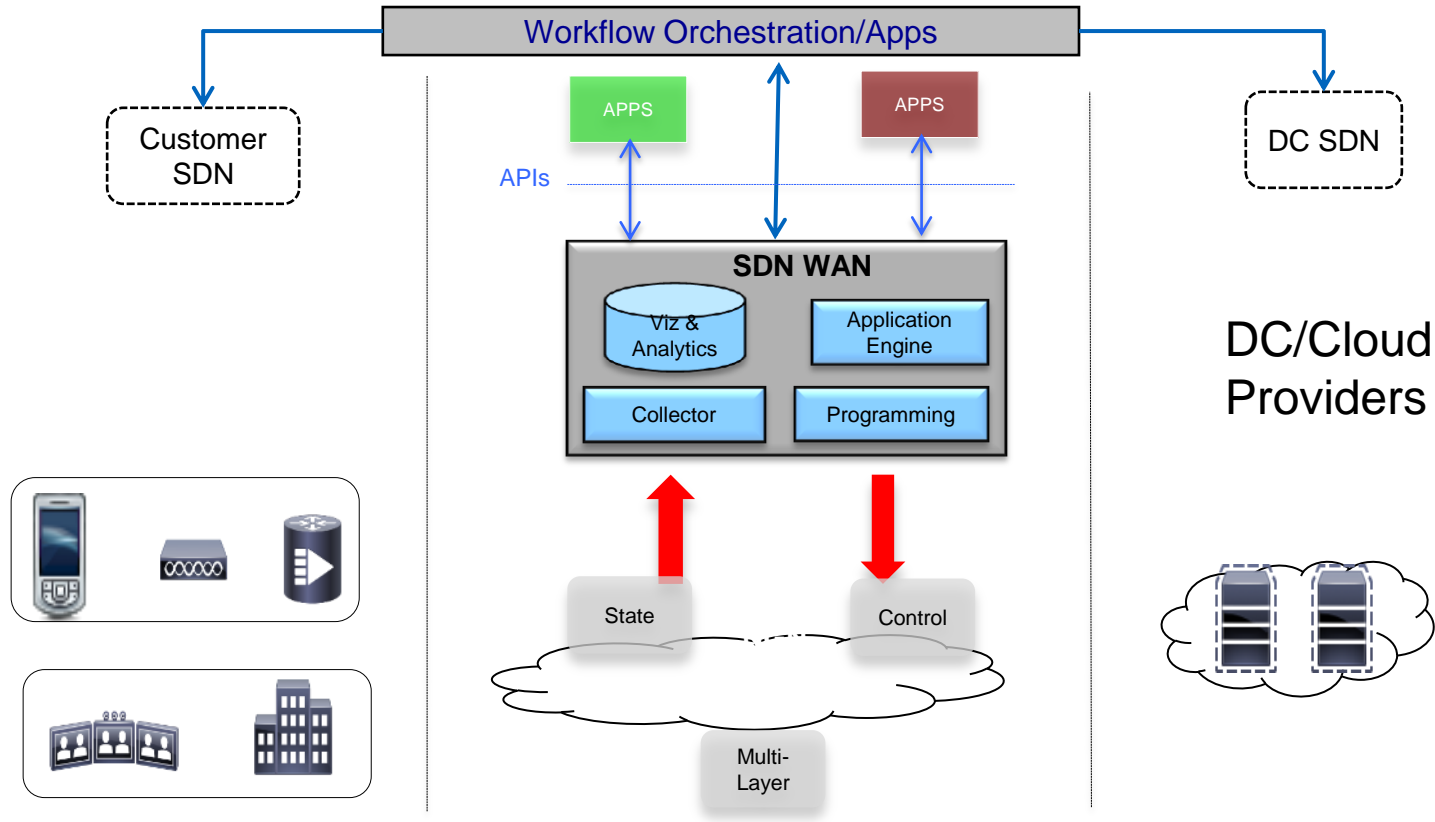


- Traffic continues to increase, while revenue declines
- On top of SPs' minds:
 - Increase efficiency of existing assets
 - Create new revenue opportunities, and be faster at it
- SDN efforts in SP attempt to help with the above!

Netting out the Challenges

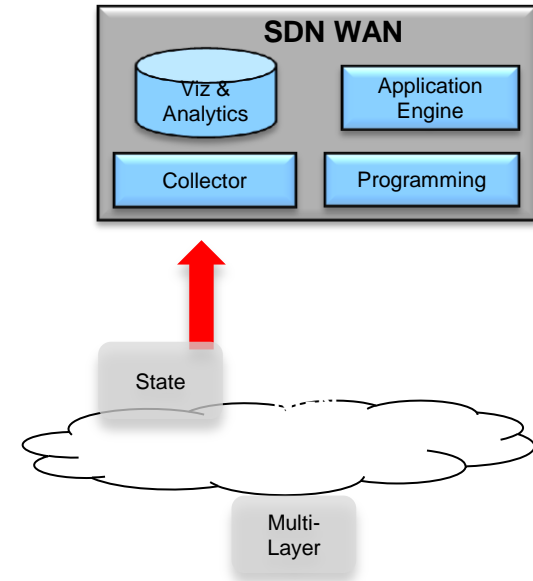
- Make it easier to operate – Simplify!!
- Run the network hotter!
- Act and re-act faster
 - To changing network conditions – adapt MPLS-TE or Metrics, or even logical topology
 - Provision a desired service
- Make \$\$
 - Doing more with the same or less
 - Introduce “on-demand”, “scheduling”, “instant”, “premium”, “secure”, “backup”, etc. choices to the services portfolio

SDN WAN Orchestration End-to-End



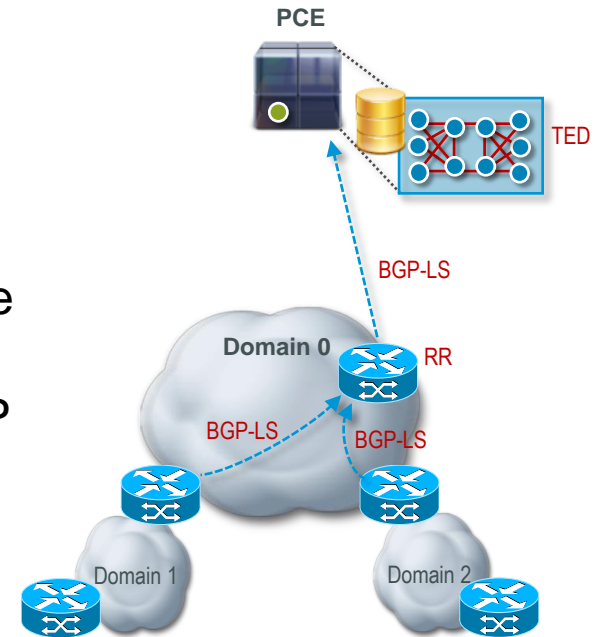
Gathering up-to-date WAN Network State

- To do its job SDN WAN Controller requires up-to-date network visibility information, primarily about
 - Load/Capacity
 - ➔ SNMP, NetFlow
 - Topology
 - ➔ IGP (OSPF/ISIS) information, direct link/passive, or better: **BGP**



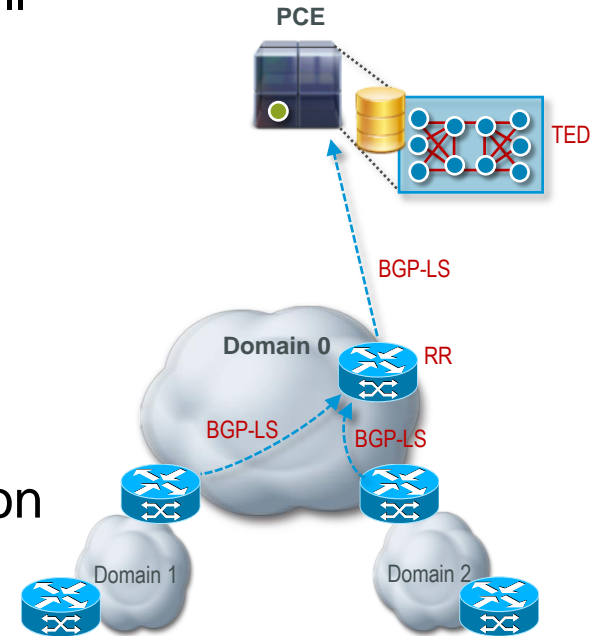
High Level Perspective of BGP-LinkState (BGP-LS)

- BGP may be used to advertise link state and link state TE database of a network (BGP-LS)
- Provides a familiar operational model to easily aggregate topology information across domains
- New link-state address family
- Support for distribution of OSPF and IS-IS link state databases
- Topology information distributed from IGP into BGP (only if changed)
- Support introduced in IOS XR 5.1.1



BGP-LS for Topology Distribution

- One or more BGP speaker per routing area will translate LSDB/TE into Network Layer Reachability Information (NLRI) extensions
- Classical BGP operations and rules apply
 - Selection algorithm
 - Route Reflection / propagation
 - Attributes
- BGP allows multi-hop sessions and hence a much more flexible way to distribute information
 - I.e.: no need to have layer-3 adjacencies

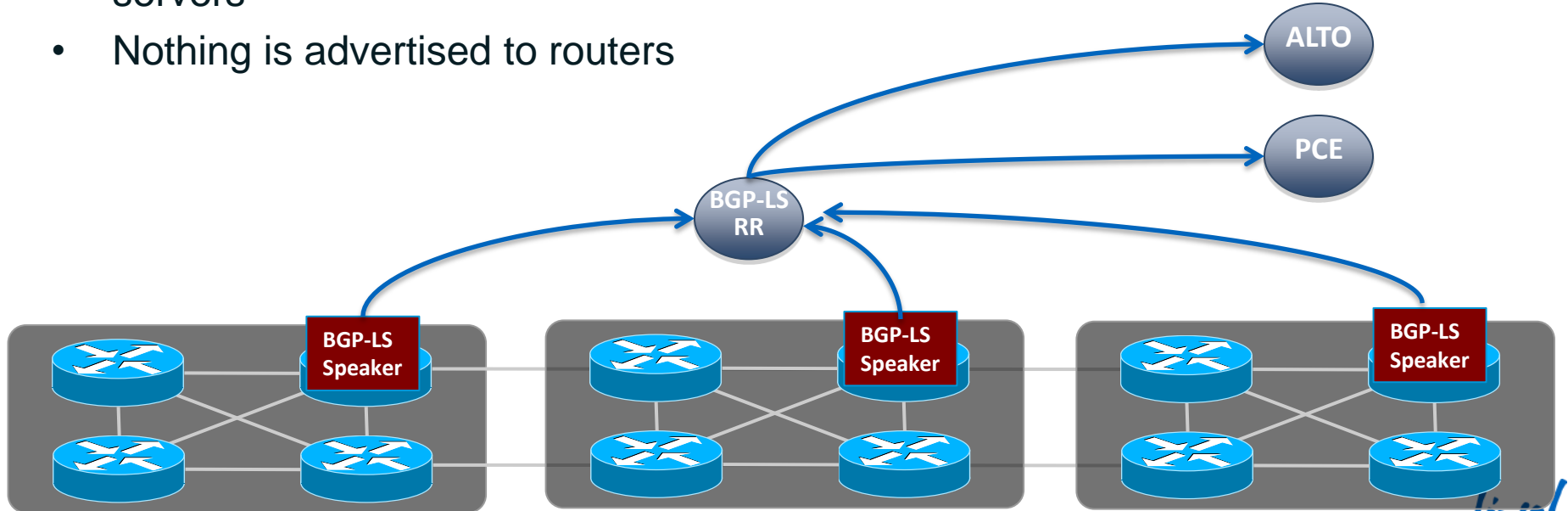


BGP-LS for Topology Distribution

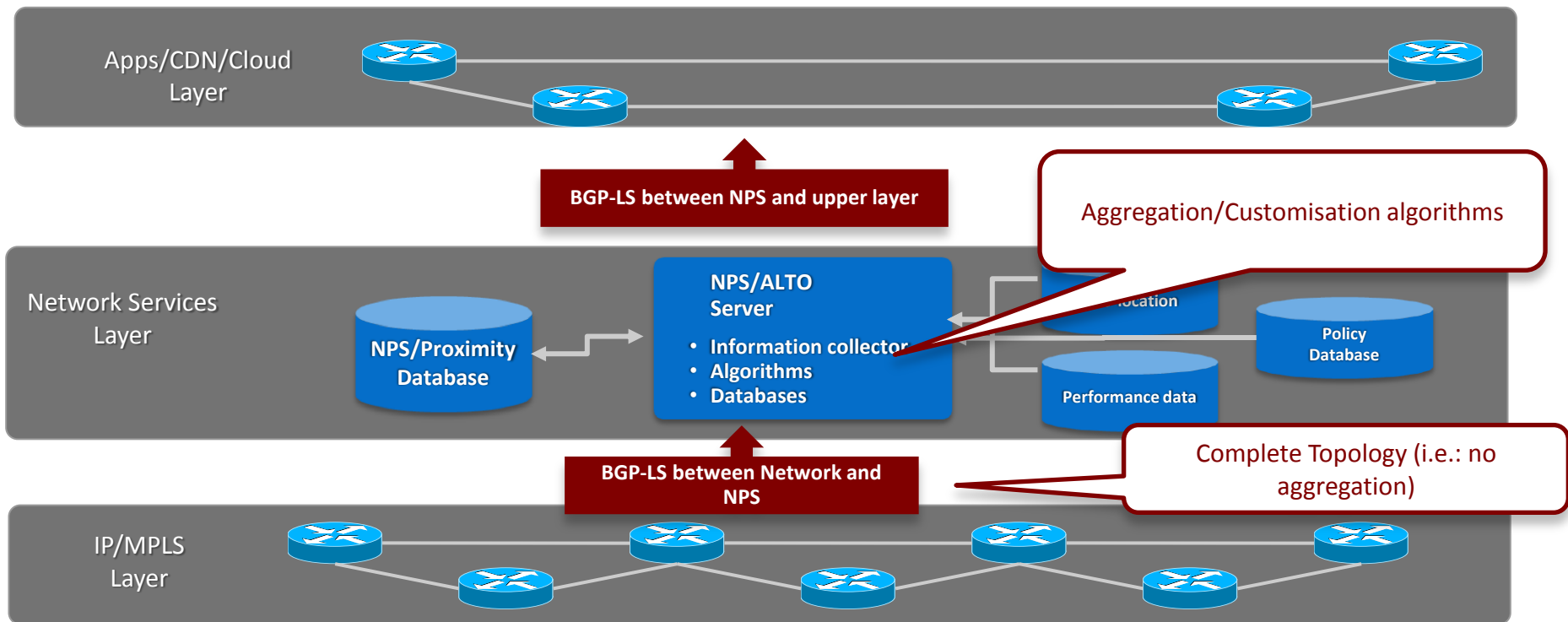
- New BGP NLRI for:
 - Link and Node descriptors
 - Draft tends to minimise new encoding format
 - Replicate what available in ISIS and OSPF encodings
- NLRI TLVs allow Link-State & TE Database encoding
 - With all attributes
- However, any form of topology (real, virtualised) can be encoded
 - Links/Nodes can be aggregated: only advertise big pipes
 - Links/Nodes can be hidden: only advertise what consumer needs
- The scheme allows maximum flexibility in order to deliver topology

BGP-LS for Topology Distribution

- One or two routers per area redistribute IGP topology into BGP-LS NLRIs
- BGP-LS NLRI are sent to BGP-LS RR that reflects them to ALTO and PCE servers
- Nothing is advertised to routers



BGP-LS: Network Guidance Use Case



BGP Link State Configuration – Cisco IOS XR 5.1.1

```
router isis DEFAULT
  is-type level-2-only
  net 49.0000.1720.1625.5001.00
  distribute bgp-ls level 2
  address-family ipv4 unicast
    metric-style wide
  mpls traffic-eng level-2-only
  mpls traffic-eng router-id Loopback0
  !
[...]
```

```
router bgp 65172
  address-family link-state link-state
  !
  neighbor 172.31.0.1
    description Controller
    remote-as 65172
    update-source Loopback0
    address-family link-state link-state
  !
  !
```



Distribute level-2 link state database into BGP-LS

Enable link-state addresses and specify BGP-LS peer

BGP Link State Prefixes

- BGP-LS prefix string has the following general format

`[NLRI-Type] [Area] [Protocol-ID] [Local node descriptor] [Remote node descriptor] [Attributes]/prefix-length`

- Node descriptors and attributes consists of potentially multiple TLVs
- Node descriptors and attributes are shown as

`[X[TLV1] [TLV2] ...]`

– Where X identifies object (e.g. local node, remote node, link, etc.)

- TLVs are shown in the format

`[yVALUE]`

– Where y identifies field type (e.g. AS number, interface address, etc.)

BGP Link State Verification – Cisco IOS XR 5.1.1

```
RP/0/RSP0/CPU0:asr9000-pe1#sh bgp link-state link-state
[...]
```

```
Status codes: s suppressed, d damped, h history, * valid, > best
```

```
          i - internal, r RIB-failure, S stale, N Nexthop-discard
```

```
Origin codes: i - IGP, e - EGP, ? - incomplete
```

```
Prefix codes: E link, V node, T IP reachable route, u/U unknown
```

```
          I Identifier, N local node, R remote node, L link, P prefix
```

```
          L1/L2 ISIS level-1/level-2, O OSPF, D direct, S static
```

```
          a area-ID, l link-ID, t topology-ID, s ISO-ID,
```

```
          c confed-ID/ASN, b bgp-identifier, r router-ID,
```

```
          i if-address, n nbr-address, o OSPF Route-type, p IP-prefix
```

```
          d designated router address
```

```
Network                  Next Hop                  Metric LocPrf Weight Path
```

```
*> [V] [L2] [I0x1] [N[c65172] [b172.16.255.1] [s1720.1625.5001.00]]/328
```

```
                  0.0.0.0
```

```
                  0 1
```

```
*> [E] [L2] [I0x1] [N[c65172] [b172.16.255.1] [s1720.1625.5001.00]] [R[c65172]
```

```
[b172.16.255.1] [s1720.1625.5002.00]] [L[i172.16.0.1] [n172.16.0.0]]/696
```

```
                  0.0.0.0
```

```
                  0 1
```

```
:
```

Prefix codes

Node

Link

Summary

- WAN orchestration provides significant value to customers in terms of
 - Operational simplification
 - Network flexibility
 - Revenue opportunities
- BGP-LS is important technology component for network topology/state collection, hand-in-hand with other protocols (PCE/BGP-LS) to program state into the underlying network



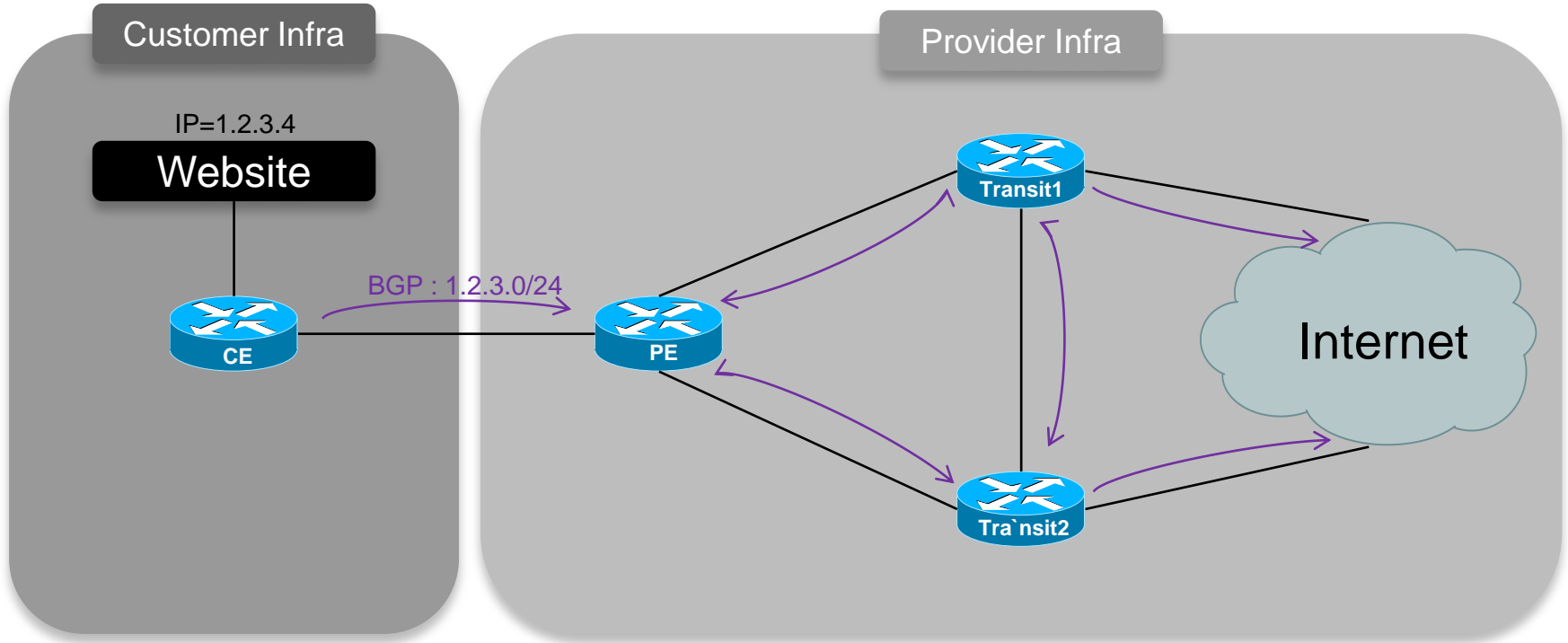
Use Case #2: Controlling Flows via BGP

Introduction

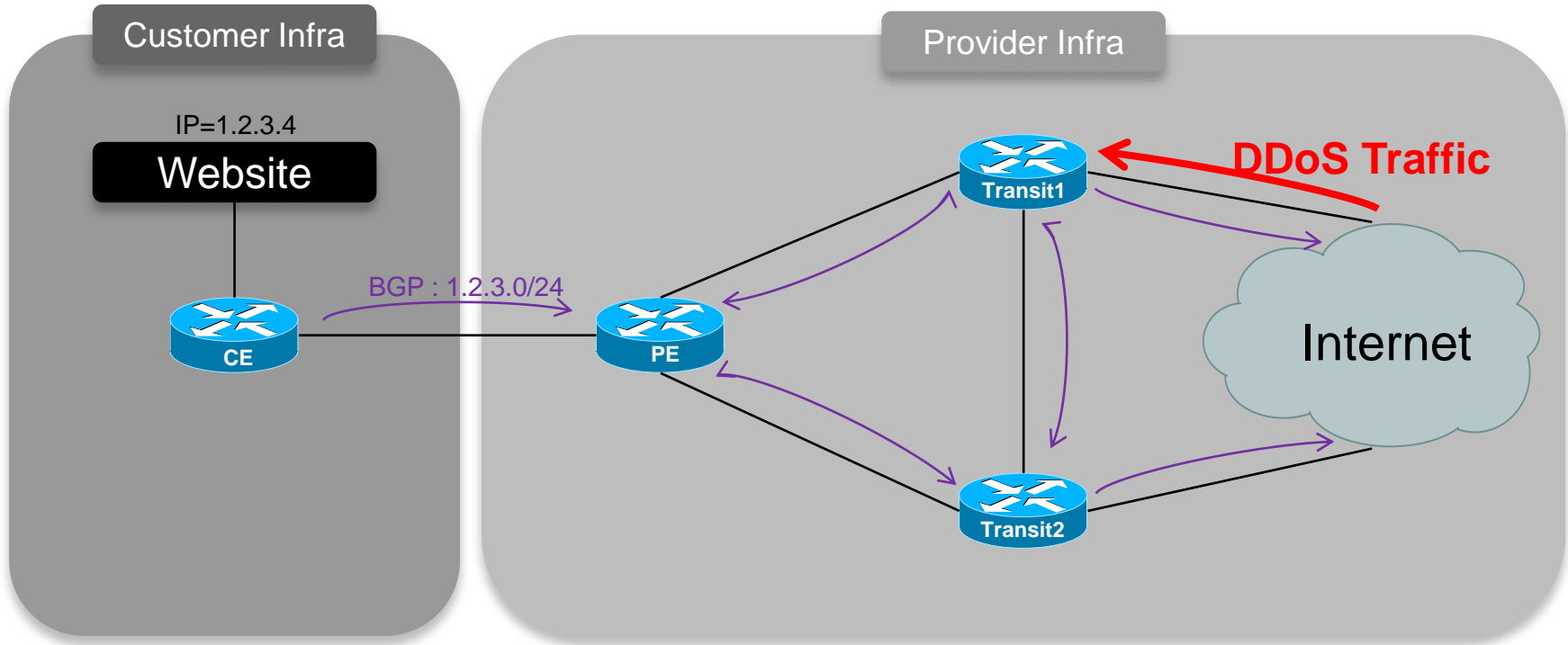
- BGP (like any other routing protocol) influences destination-based routing
- BGP routing information can be injected from a central place (“route server”)
- Why not use it for more than just giving a destination address to route packets to?

- “Flow Specification Rules”
 - Application aware Filtering/redirect/mirroring
 - Dynamic and adaptive technology
 - Simple to configure

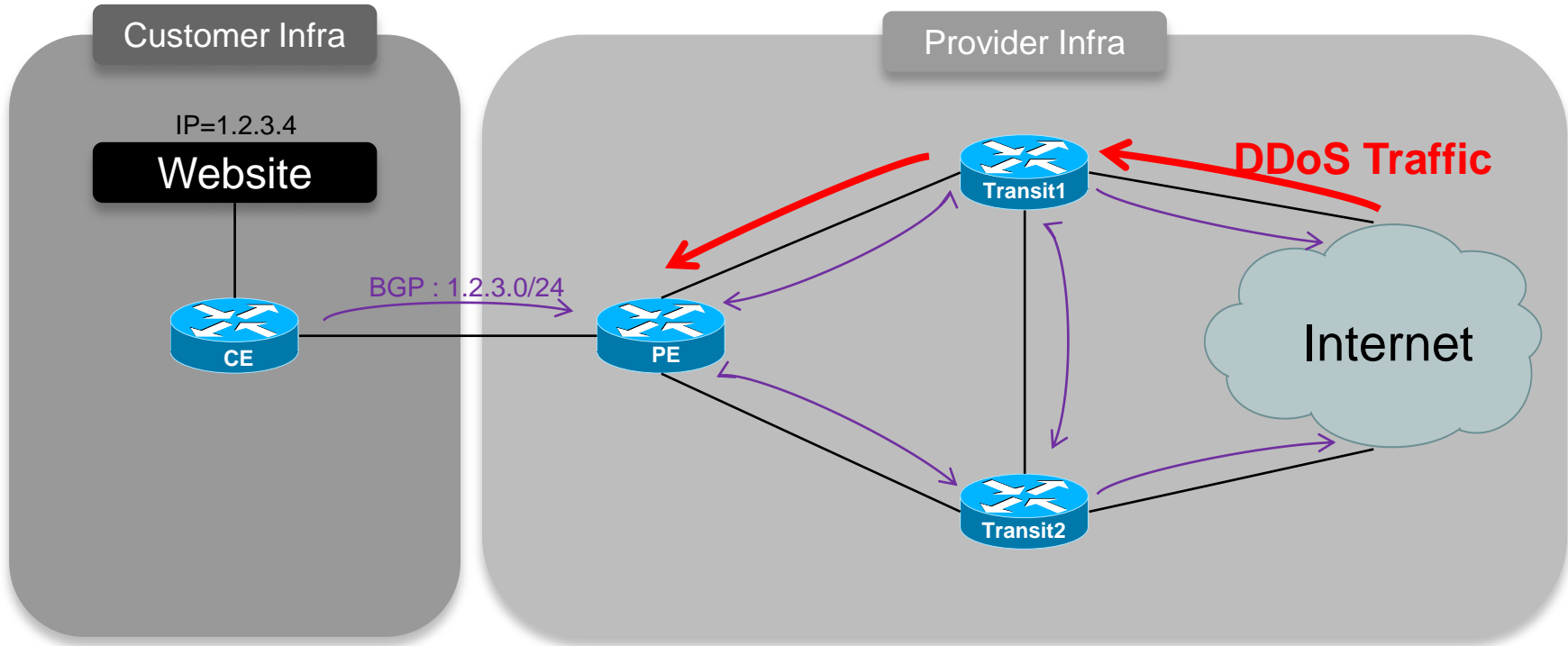
An Example: Denial of Service Mitigation



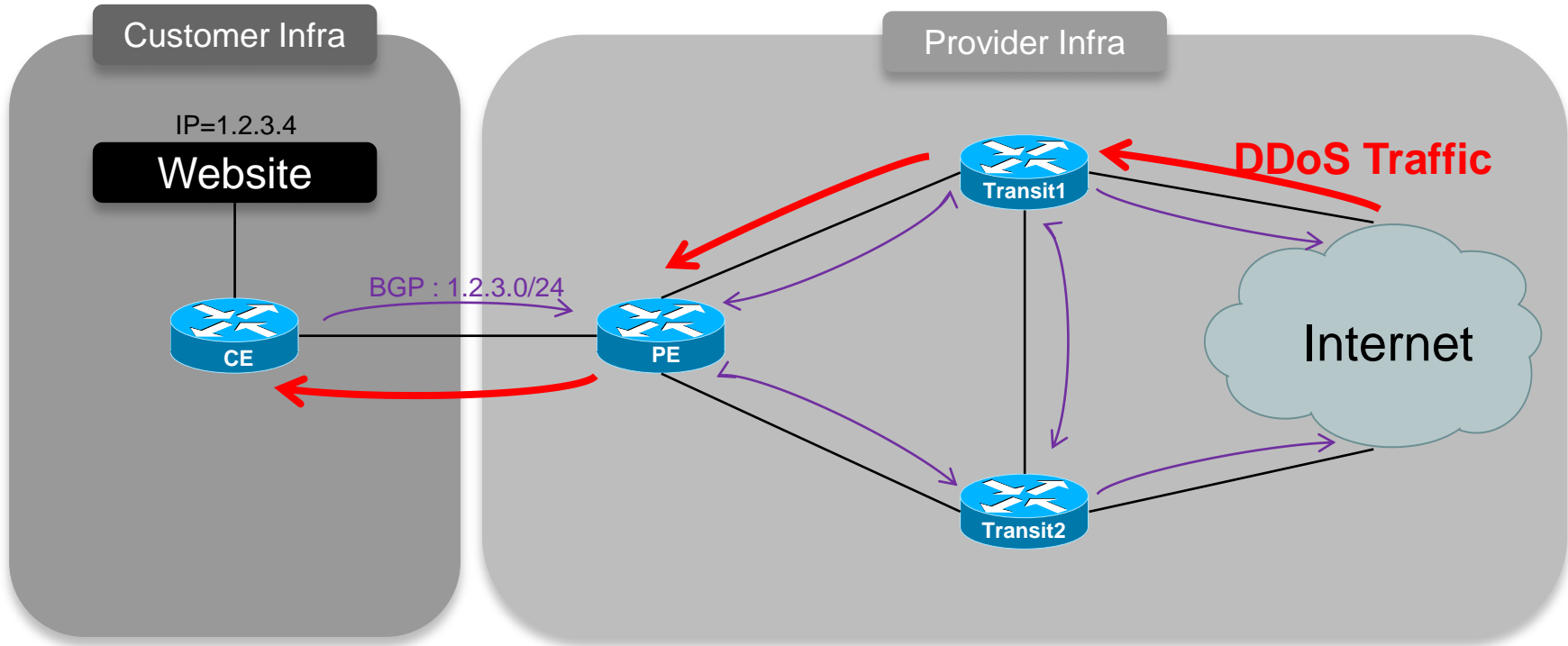
An Example: Denial of Service Mitigation



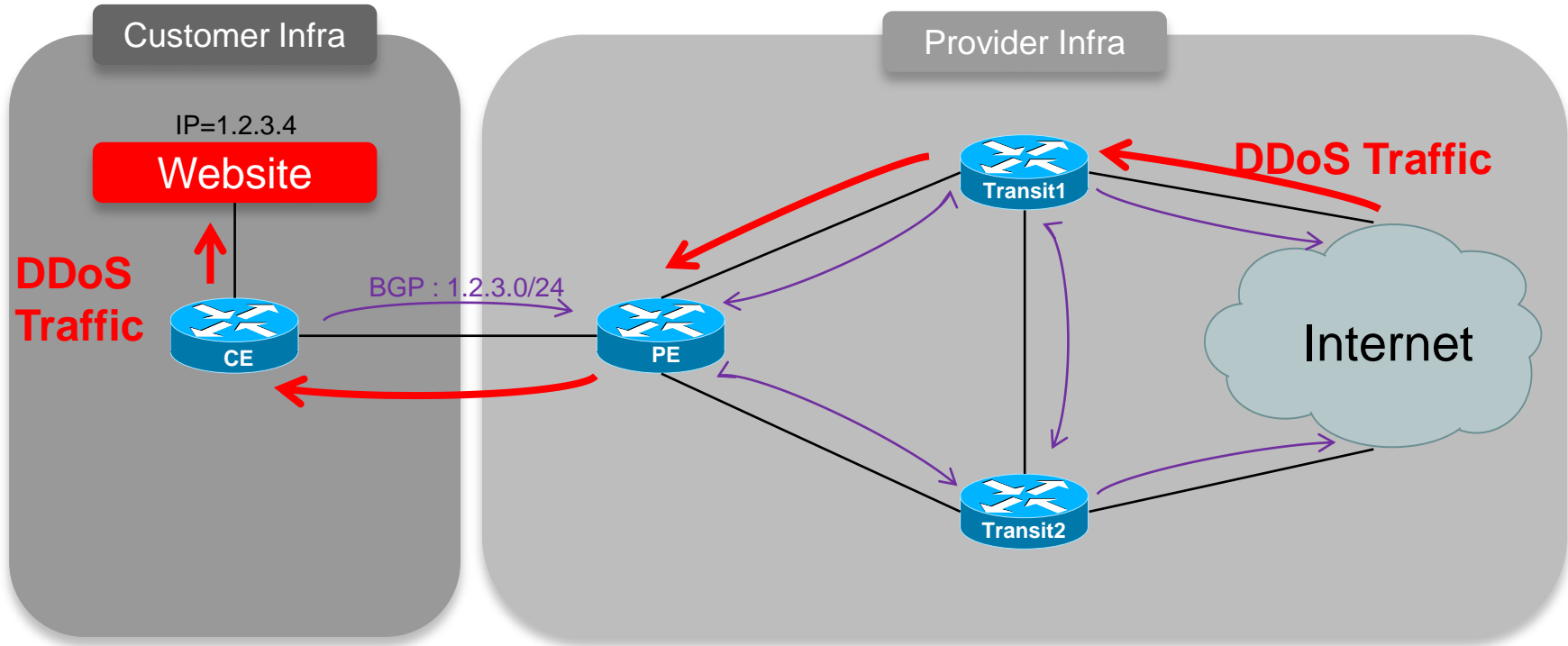
An Example: Denial of Service Mitigation



An Example: Denial of Service Mitigation

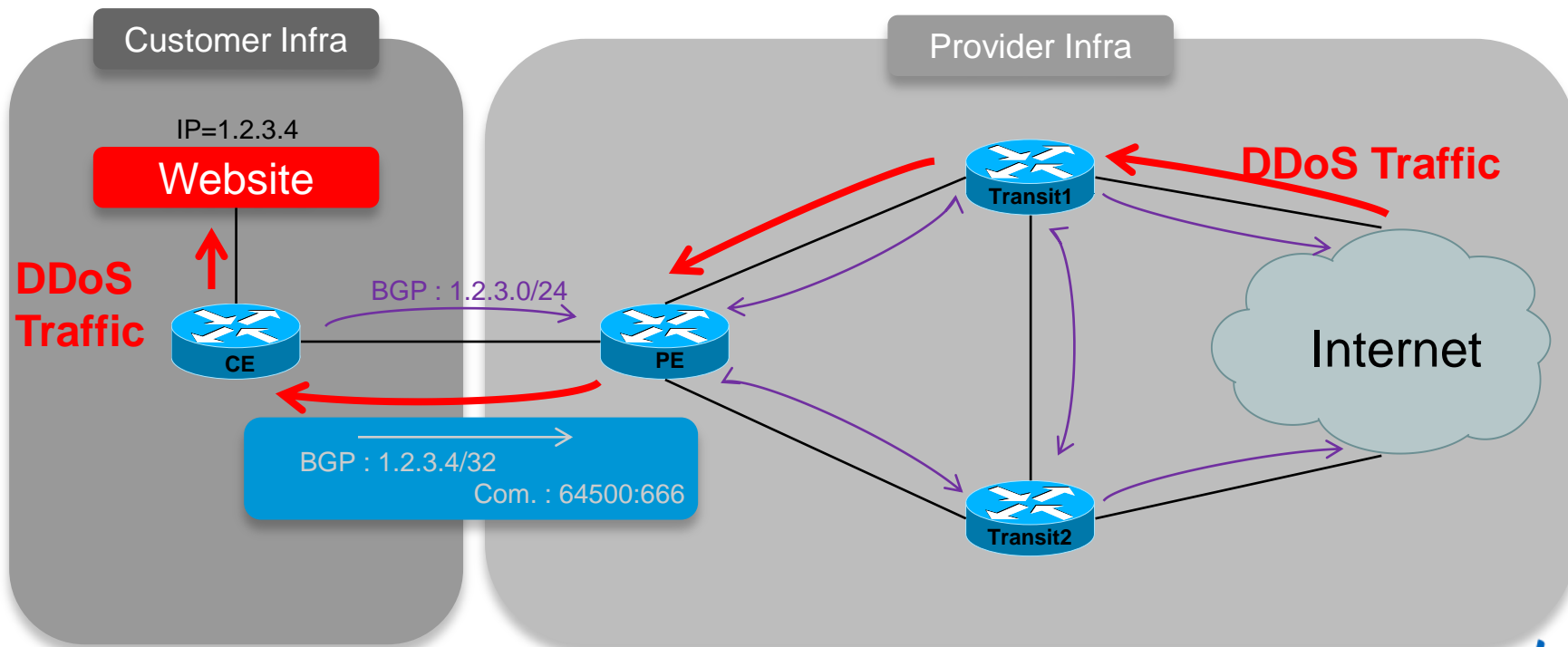


An Example: Denial of Service Mitigation



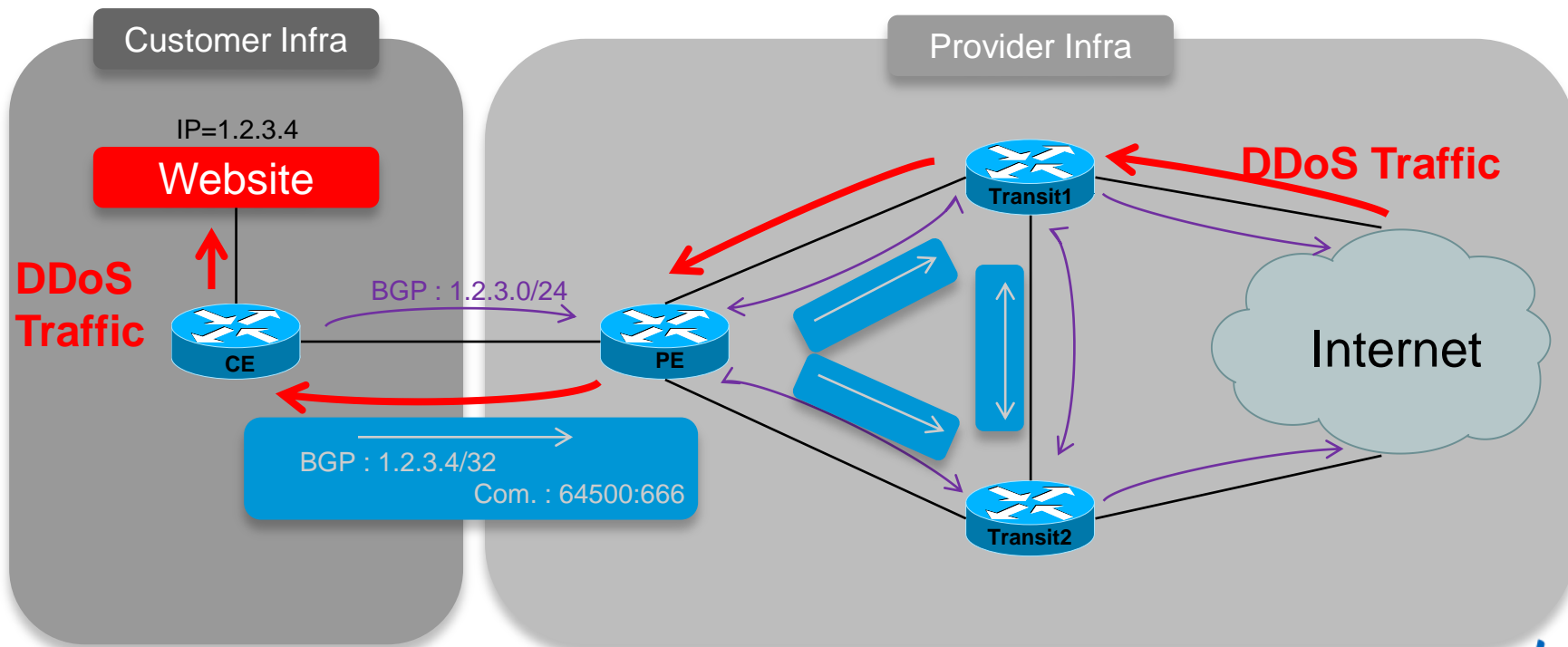
Solution: Remotely Triggered Black Hole

It is time to use the blackhole community given by the provider (i.e. 64500:666)



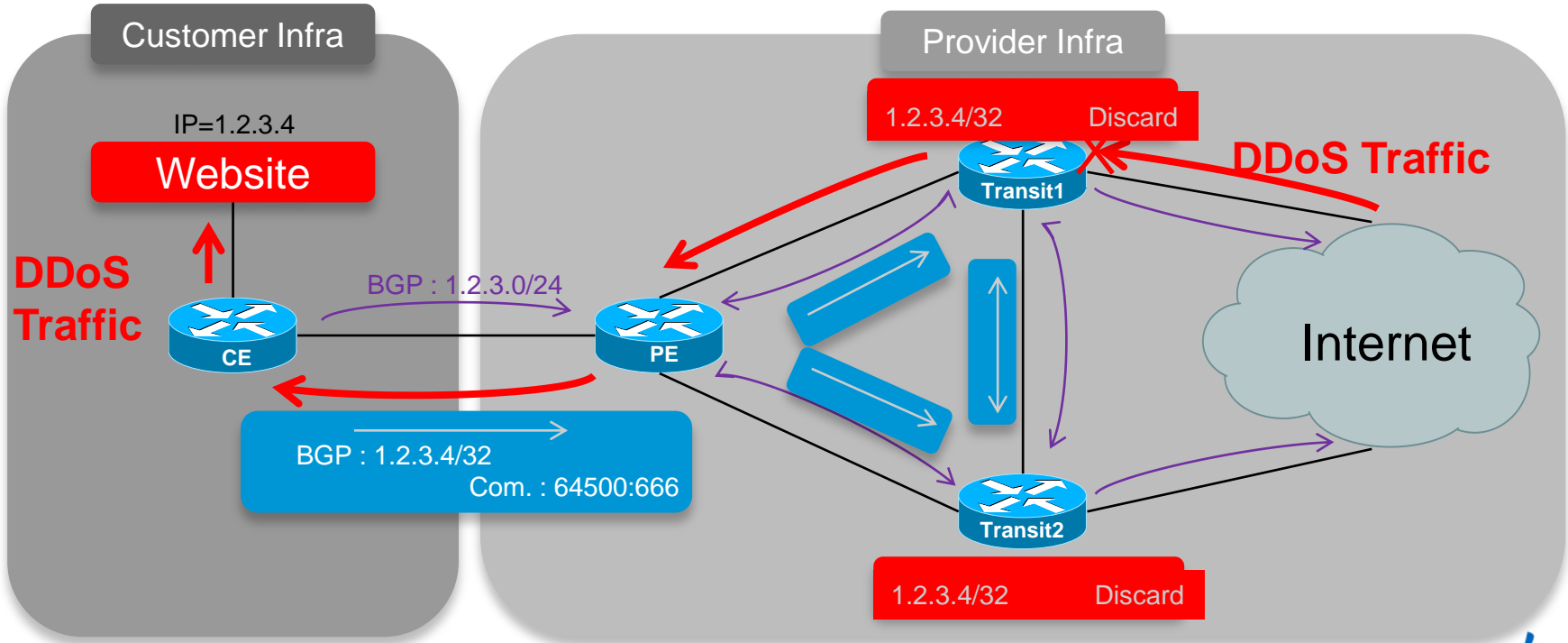
Solution: Remotely Triggered Black Hole

It is time to use the blackhole community given by the provider (i.e. 64500:666)



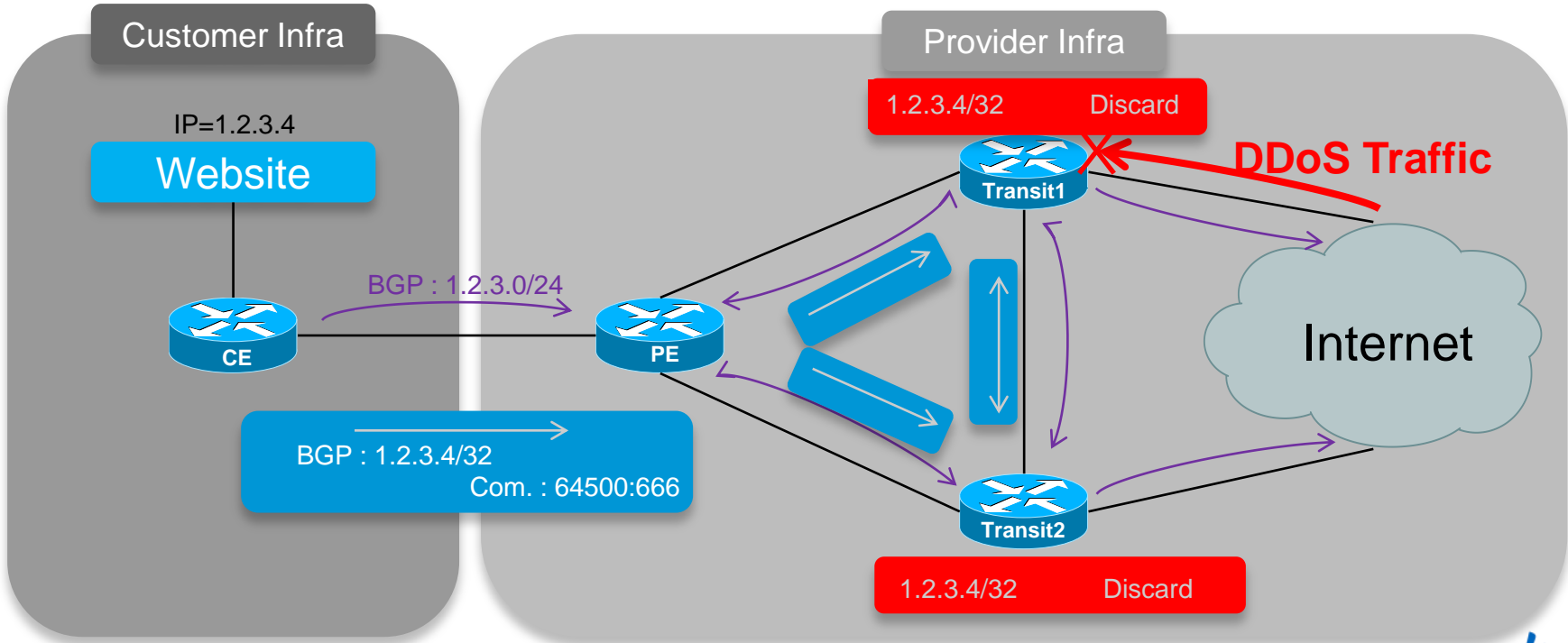
Solution: Remotely Triggered Black Hole

All prefixes with blackhole community get assigned a special nexthop which recurses to Null0



Solution: Remotely Triggered Black Hole

All prefixes with blackhole community get assigned a special nexthop which recurses to Null0



Solution: Remotely Triggered Black Hole

- Great, I have my server responding again!
 - No more DDoS traffic on my network
 - **But** no more traffic at all on my website....
- Well, maybe it was not the solution I was looking for....

Alternative Solution: Policy Based Routing

- Identification of DDoS traffic: based around a conditions regarding MATCH statements
 - Source/Destination address
 - Protocol
 - Packet size
 - Etc...
- Actions upon DDoS traffic
 - Discard
 - Logging
 - Rate-Limiting
 - Redirection
 - Etc...
- Doesn't this sound like a great solution?

Alternative Solution: Policy Based Routing

- Good solution for
 - Done with hardware acceleration even on carrier grade routers
 - Can provide surgical precision of match statements and actions to impose
- But...
 - Customer need to call my provider
 - Customer need the provider to accept and run this filter on each of their backbone/edge routers
 - Customer need to call the provider and remove the rule after!
- Reality: It won't happen...

BGP FlowSpec as a Better Alternative

- Comparison with the other solutions
 - Makes static PBR a dynamic solution!
 - Allows to propagate PBR rules
 - Existing control plane communication channel is used
- How?
 - By using your existing MP-BGP infrastructure

Dissemination of Flow Specification Rules

(RFC5575)

- Why use BGP?
 - Simple to extend by adding new reachability information
 - Network-wide loop-free point-to-multipoint path is already setup
 - Already used for all kinds of technology (IPv4, IPv6, VPN, Multicast, Labels, etc...)
 - Inter-domain support
 - Networking engineers and operations perfectly understand BGP

Dissemination of Flow Specification Rules (RFC5575)

New NLRI defined (AFI=1, SAFI=133)

- | | |
|---------------------------|-------------------|
| 1. Destination IP Address | 7. ICMP Type |
| 2. Source IP Address | 8. ICMP Code |
| 3. IP Protocol | 9. TCP Flags |
| 4. Port | 10. Packet length |
| 5. Destination port | 11. DSCP |
| 6. Source Port | 12. Fragment |

```
+-----+
| Address Family Identifier (2 octets)
+-----+
| Subsequent Address Family Identifier (1 octet)
+-----+
| Length of Next Hop Network Address (1 octet)
+-----+
| Network Address of Next Hop (variable)
+-----+
| Reserved (1 octet)
+-----+
| Network Layer Reachability Information (variable)
+-----+
```

The MP_REACH_NLRI – RFC 4760

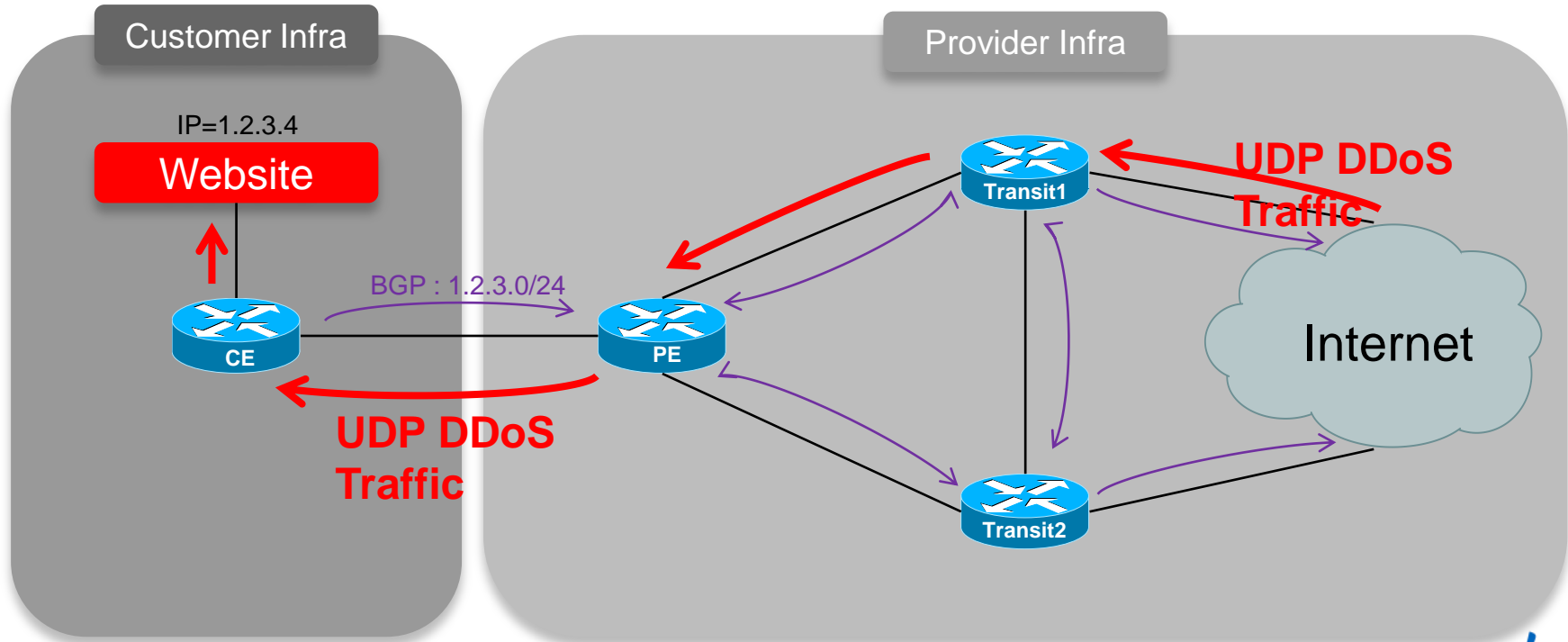
Notice from the RFC: “Flow specification components must follow strict type ordering. A given component type may or may not be present in the specification, but if present, it MUST precede any component of higher numeric type value.”

BGP Flowspec Traffic Actions

Action	Description
Traffic-Rate	Ability to police flow to a given amount
Traffic-Marking	Rewrite DSCP value
Redirect VRF	Redirect to a VRF (using route-target) Ex: “cleaning” traffic
Redirect NH	Redirect to an alternate next-hop
Traffic-Action	Drop/Discard or Sample (not yet implemented)

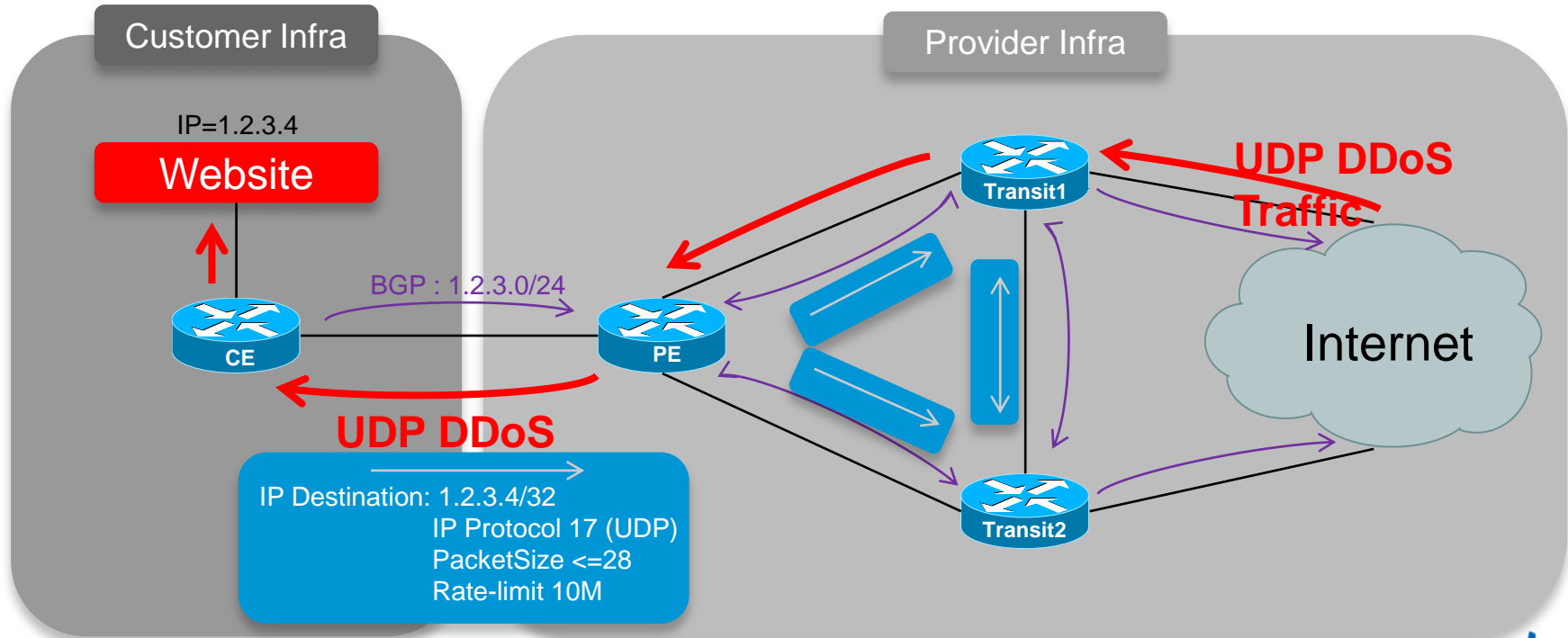
DDoS Mitigation using BGP FlowSpec

Let's do this better now with the new BGP FlowSpec functionality

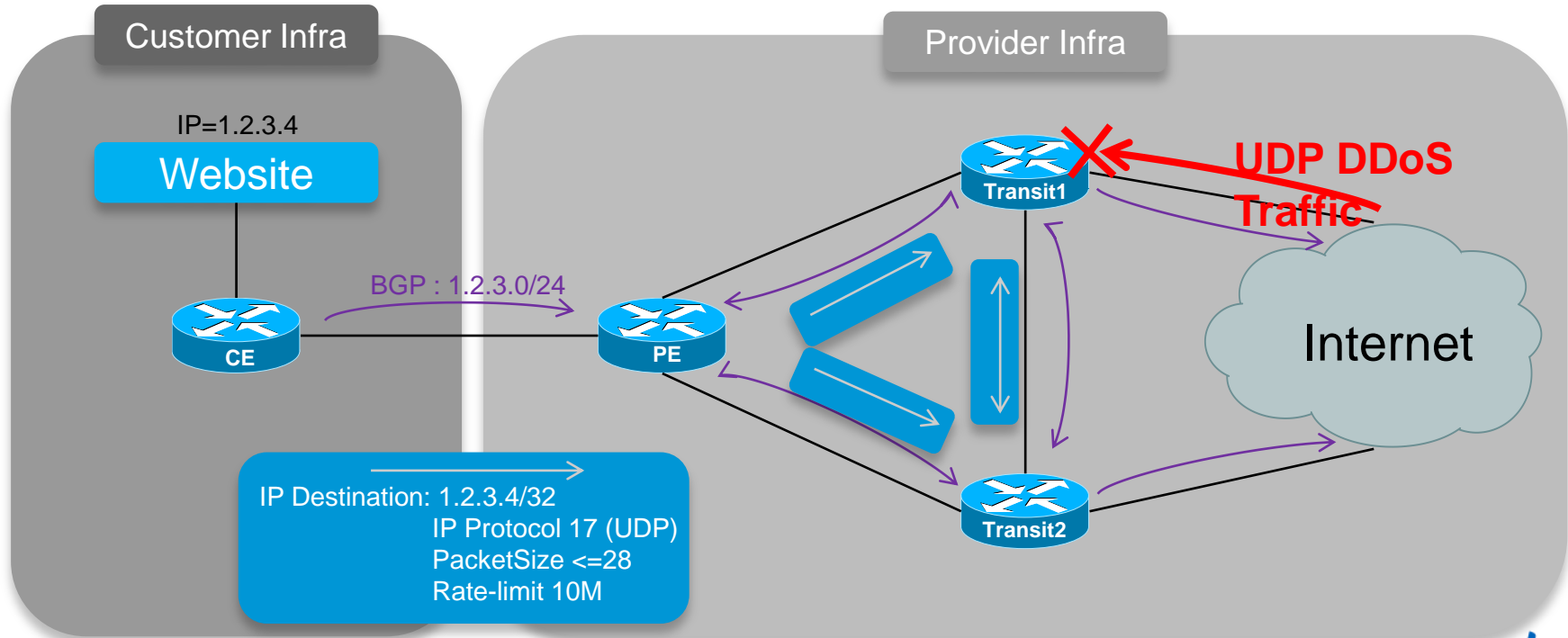


DDoS Mitigation using BGP FlowSpec

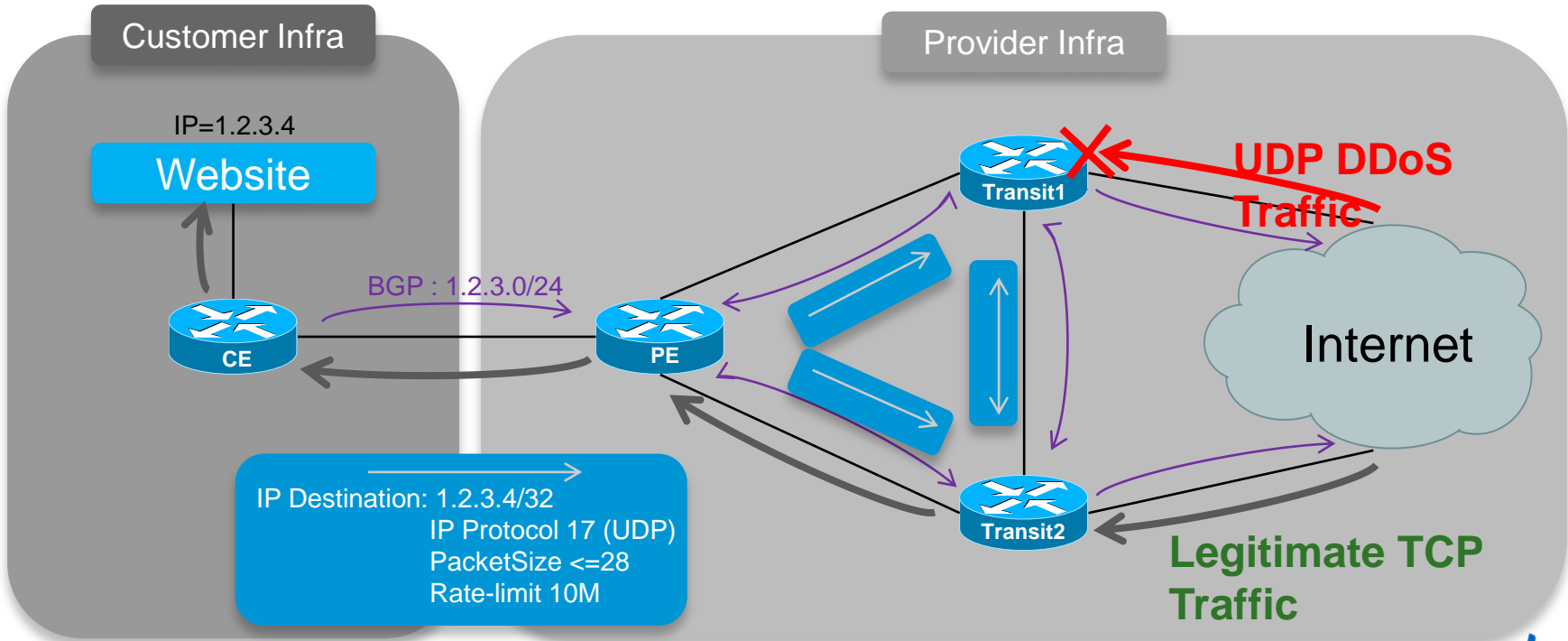
Customer advertises the web server's address with granular flow information



DDoS Mitigation using BGP FlowSpec



DDoS Mitigation using BGP FlowSpec

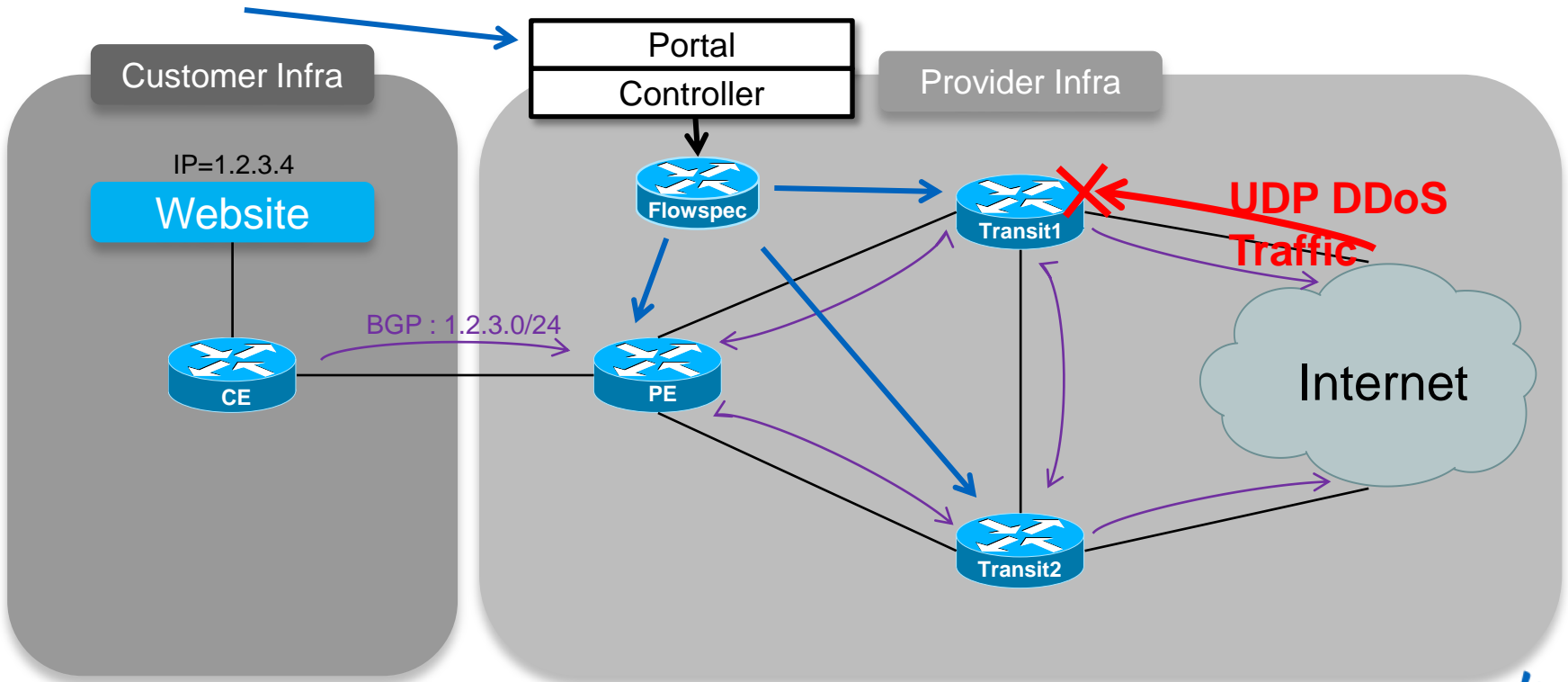


Real Life Architecture

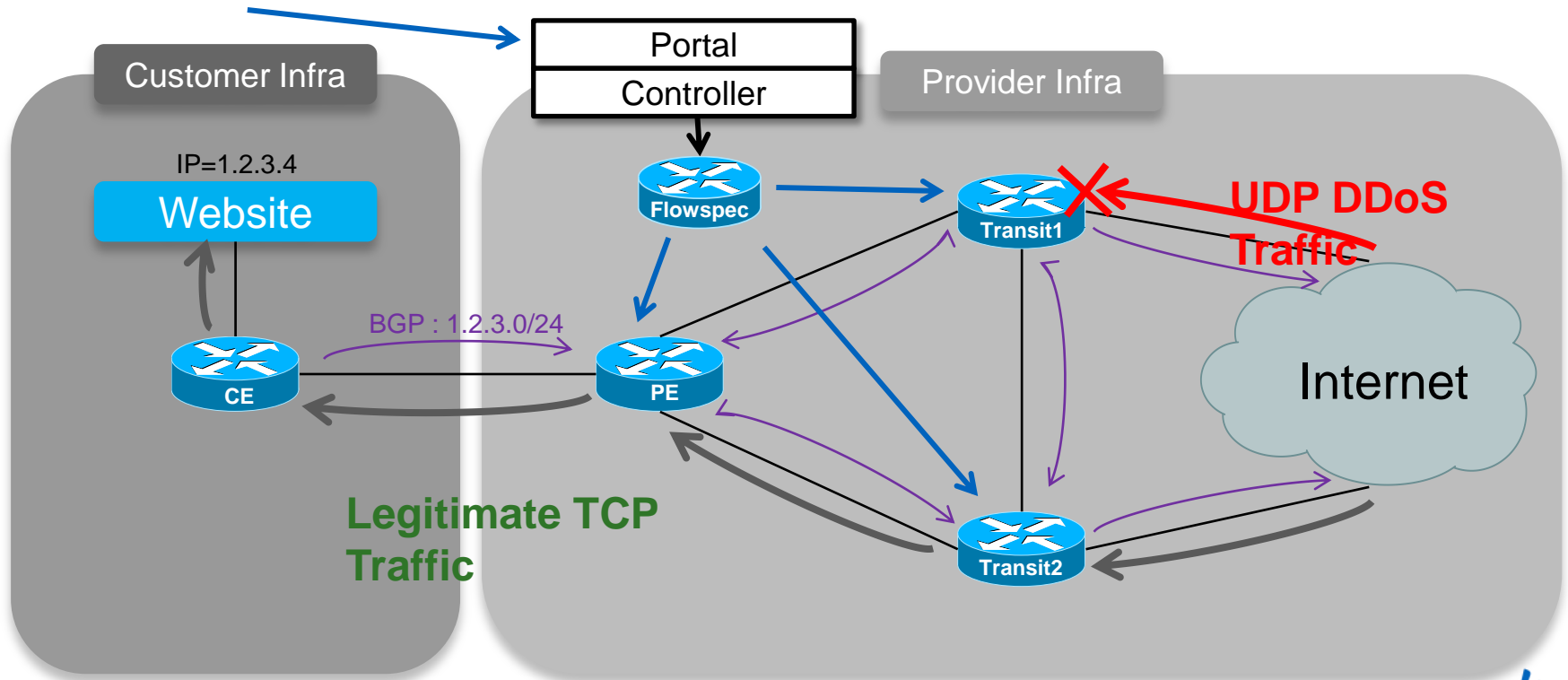
- In reality this architecture is not deployed
 - Service Provider DO NOT trust the Customer (at least not that much ;-)
 - It requires new BGP AFI/SAFI combination to be deployed between Customer and Service provider
 - Both these result in Flowspec not commonly being deployed between Customer and SP

- What is done instead?
 - SP utilise a central Flowspec speaker(s)
 - Have it BGP meshed within the Service Provider routers
 - Only the central Flowspec speaker is allowed to distribute Flowspec rules
 - Central Flowspec speaker is considered “trusted” by the network (no-validate)
 - Central Flowspec speaker is managed by the service provider

Central FlowSpec Speaker



Central FlowSpec Speaker





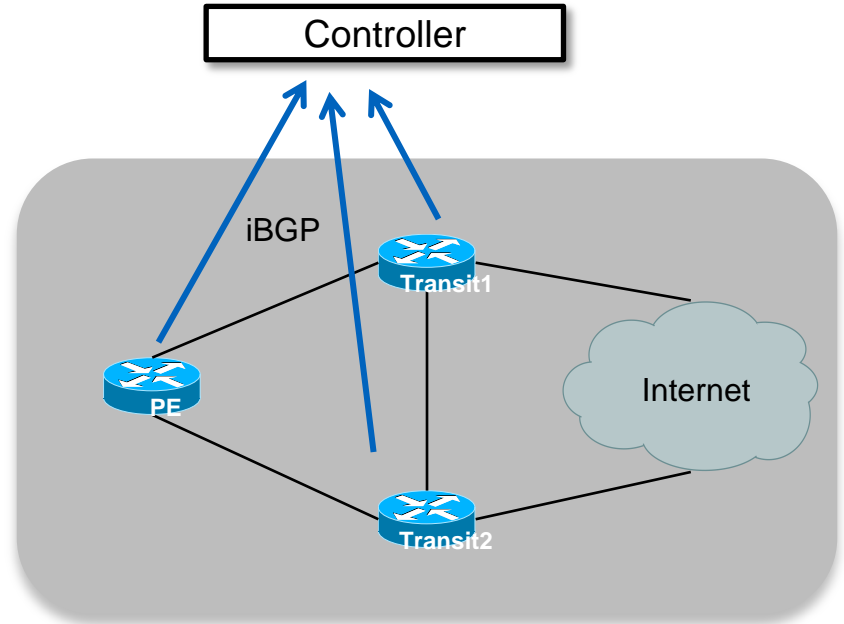
Use Case #3: Routing Visibility

Optimising Routing Towards the Internet

- When your network is multi-homed to multiple SPs, balancing the traffic across the potential exit points can become a cumbersome task:
 1. Baseline the situation
 2. Tweak BGP attributes (MED, local preference, AS-path) to shift traffic to other exits
 3. Watch the result
 4. If not happy, go back to 2
- How about letting software do this for you?
- It knows the topology (via BGP-LS, see earlier)
- It knows the traffic/matrix (via NetFlow, LSP stats, interface load)
- It misses information about the BGP routing table and its attributes

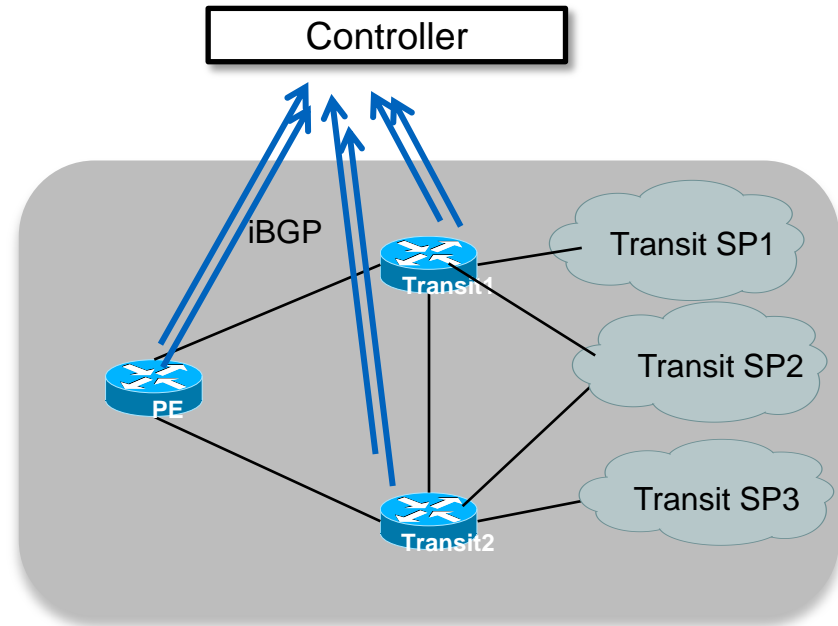
Achieving Routing Visibility

- As a routing protocol, it can also be used to update the controller with granular routing information
- Easy.
- Really?



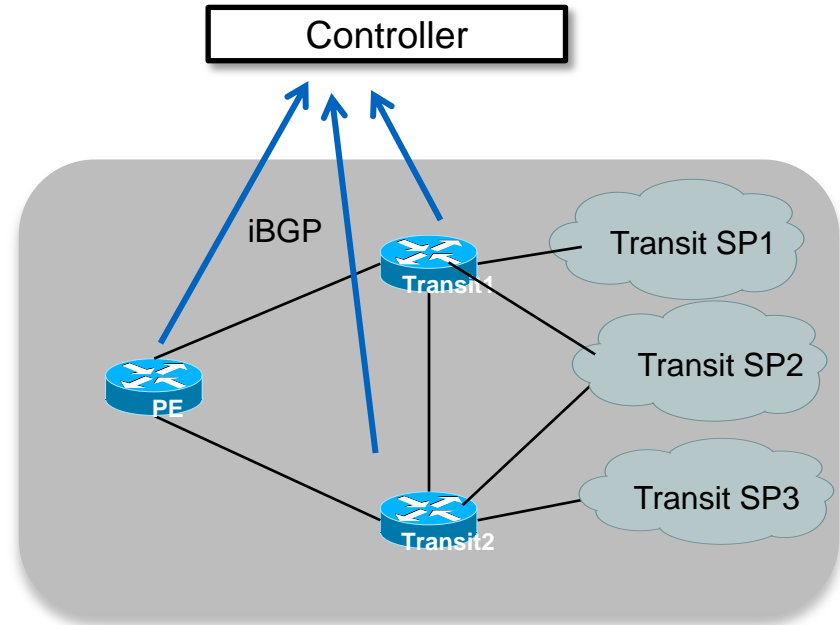
Routing Visibility – Add-Path

- BGP selects one best path and advertises it to its peers
- But if I have multiple neighbours advertising the same prefix, the controller should know about all the paths
- Solution: **BGP Add-Path**
 - Selects Best Path, but also sends one or more additional paths
 - New protocol capability, needs to be enabled



Routing Visibility

- Ok, now the controller has all the information, and can do its “magic”
- It changes BGP routing policy (route-maps/RPL) on the devices, modifying BGP attributes, etc.
- But now we might have modified the attributes which were originally sent to us by the SP
- But we might want to know about the original attributes when the next optimisation run is due?

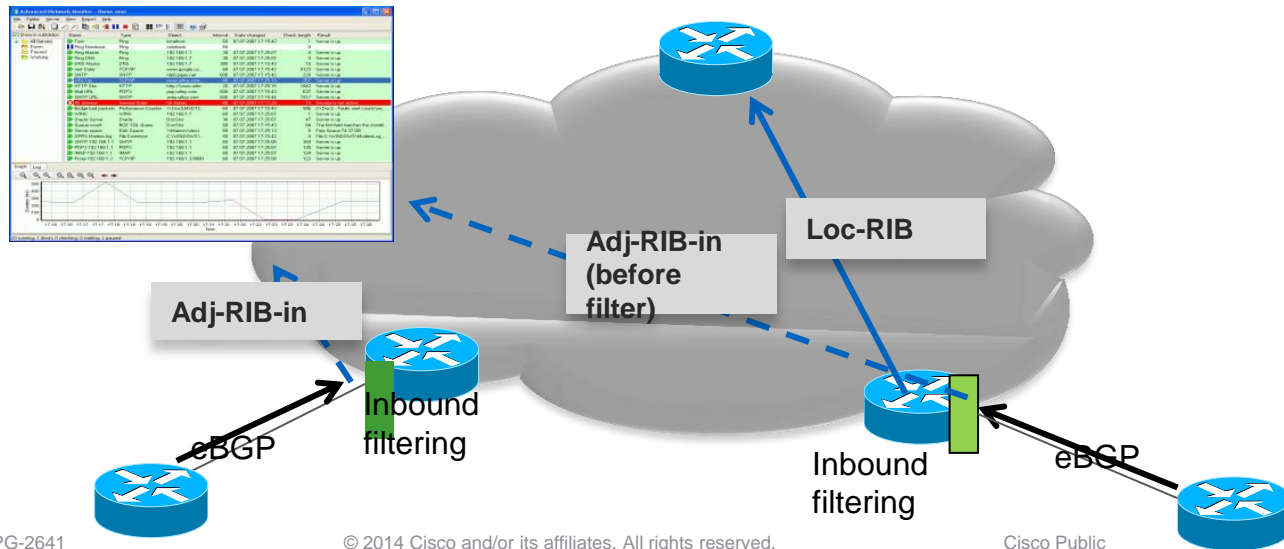


BGP RIBs

- BGP speaker maintains multiple Routing Tables:
- **Adj-RIB-in** (per neighbour)
 - These are the updates as received by the peer
 - Incoming route policy is applied, attributes are changed
 - Updates which are dropped by the incoming route-policy are discarded, to save on memory
 - “soft-reconfiguration inbound” keeps them, paths flagged with “received-only” in “show bgp ...”
- **Loc-RIB** (or Local RIB)
 - BGP calculates best path among eligible paths in Adj-RIB in and places them into Loc-RIB
 - provides a view of all entries kept by the BGP router to forward traffic

BGP Monitor Protocol

- We saw one case where we want to know exactly what the neighbour sent us (original attributes)
- For troubleshooting/monitoring, a record of prefixes received by neighbours (even those we configured to ignore) can be valuable tool

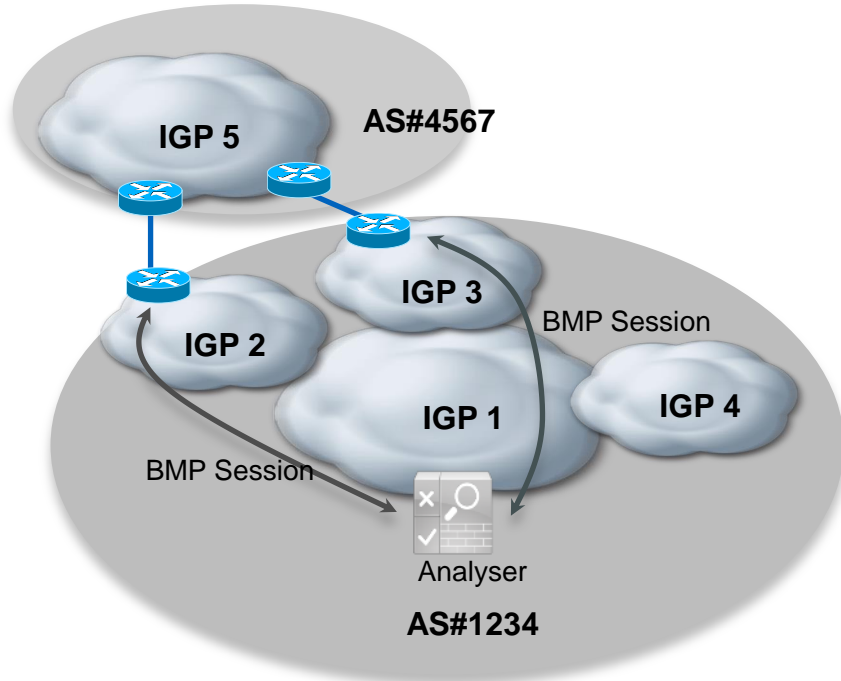


What is BMP?

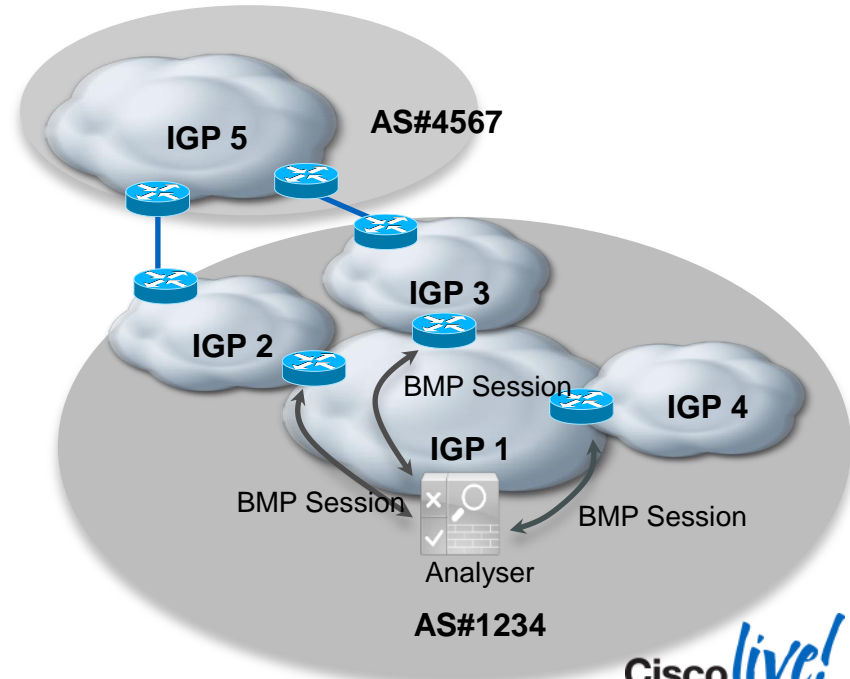
- BMP is intended to be used for monitoring BGP sessions
- BMP is intended to provide a more convenient interface for obtaining route views
- Design goals
 - Simplicity
 - Easy to use
 - Minimal service affecting
- BMP is not impacting the routing decision process and is only used to provide monitoring information
- BMP provides access to the Adj-RIB-In of a BGP peer on an ongoing basis and provides a periodic dump of statistical information. A monitoring station can use this for further analysis
- <http://tools.ietf.org/html/draft-ietf-grow-bmp-07>

Deployment Models

- Deployment Model 1
 - Peering diagnostics and analytics



- Deployment Model 2
 - Internal diagnostics and analytics

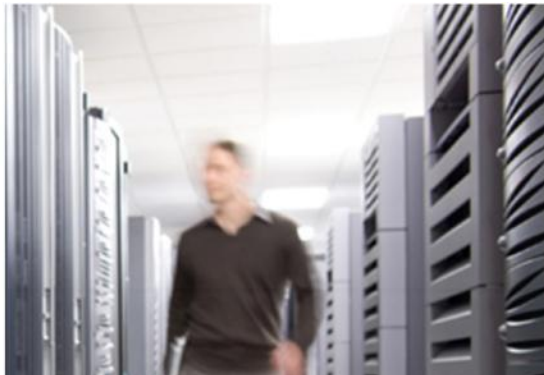


Configuration

```
router bgp <asn>
  neighbor <ip-address> BMP monitor all / server 1 server 2 ...

  bmp server <1-32>
    activate
    address <ipv4/6 address> port-number <num>
    update-source <interface>
    description <string>
    failure-retry-delay <seconds>
    flapping-delay <seconds>
    initial-delay <seconds>
    set ip dscp value <1-7>
    stats-reporting-period <seconds>

  bmp buffer-size <megabytes>
  bmp initial-refresh {delay <seconds> | skip }
```



Wrapping Up

Summary

- SDN enhances the way we're doing networking, automates tasks, introduces new possibilities through open APIs
- SDN is much more than OpenFlow, has many aspects for many different use cases
- SDN can co-exist with traditional networking protocols, it even leverages them
- BGP provides a couple of essential tools in the toolbox for topology and routing distribution and flow control
- We hope you will make use of them to make your network infrastructure more agile and cost-effective



Q & A

Complete Your Online Session Evaluation

Give us your feedback and receive a Cisco Live 2014 Polo Shirt!

Complete your Overall Event Survey and 5 Session Evaluations.

- Directly from your mobile device on the Cisco Live Mobile App
- By visiting the Cisco Live Mobile Site www.ciscoliveaustralia.com/mobile
- Visit any Cisco Live Internet Station located throughout the venue

Polo Shirts can be collected in the World of Solutions on Friday 21 March 12:00pm - 2:00pm



Learn online with Cisco Live!

Visit us online after the conference for full access to session videos and presentations.

www.CiscoLiveAPAC.com



CISCO™