M3AAWG | Messaging Malware Mobile Anti-Abuse Working Group

Messaging, Malware and Mobile Anti-Abuse Working Group

M^{3}AAWG Border Gateway Protocol (BGP) Flowspec Best Practices

February 2019
URL to reference this document: www.m3aawg.org/flowspec-BP

Executive Summary

This paper is for individuals wanting to learn more about Flowspec and are interested in taking advantage of the numerous opportunities for use that it offers. It is written for network engineers responsible for Network Service Provider (NSP), hosting provider, or enterprise networks. Additionally, it assumes the reader is familiar with Border Gateway Protocol (BGP) routing and other common networking technologies.

I. Introduction

Flow Specification (Flowspec) is a new type of Network Layer Reachability Information (NLRI) for the BGP routing protocol. It is used to apply specific actions on network traffic defined by specific filters to traffic flowing through routers. Flowspec was originally developed to help mitigate Distributed Denial of Service (DDoS) attacks but its use has expanded to numerous other applications.

II. Background

The concept for Flowspec was originally proposed by Pedro Marques (Cisco Systems), Nischal Sheth (Juniper Networks), Robert Raszuk (Cisco Systems), Barry Greene (Juniper Networks), Jared Mauch (NTT America), and Danny McPherson (Arbor Networks). These individuals created the Internet Engineering Task Force (IETF) RFC 5575 which was ratified in August 2009[1]. (https://tools.ietf.org/html/rfc5575). Although, Flowspec was specifically engineered to deal with DDoS attacks it is now also used for additional applications outside of DDoS mitigation.

III. Flowspec Protocol Details

Flowspec rules are typically generated by a controller and advertised to routers via BGP. The controller can be another router that has a Flowspec configuration which is configured to send the rules to other routers via External BGP (eBGP) or Internal BGP (iBGP). For a router to receive Flowspec rules, the router must have enabled either the IPv4 or IPv6 Flowspec address family. Please note that while numerous router vendors support IPv6 Flowspec rules, the IETF draft defining IPv6 Flowspec has not yet been ratified[2] https://tools.ietf.org/html/draft-ietf-idr-flow-spec-v6-08. Flowspec rules can use numerous parameters to match traffic for manipulation. The following is a list of parameters of network traffic that Flowspec can match:

- Src/Dst IP Address/Subnet
- Src/Dst Port (Note: it is also possible to define a range of ports and use greater than/less than notation)
- IP Protocol

- ICMP Type/Code
- TCP Flags (Defined by a bitmask)
- Packet Length
- DSCP Value
- Fragment Bits

The following is a list of actions that can be applied to network traffic that matched a Flowspec rule:

- Drop
- Rate Limit
- Send to a Virtual Routing and Forwarding Instance (VRF)
- Set the Diffserv Code Point (DSCP) value in the packet header
- Traffic sampling (Note that router vendor support for this feature is very limited)

An additional action to set the BGP nexthop IP was proposed in the IETF but the document expired without further action[3] (https://datatracker.ietf.org/doc/draft-simpson-idr-flowspec-redirect/).

Nonetheless, this method of redirection is supported by large router vendors such as Cisco and Juniper. If there is future widespread adoption of this methodology for diversion, this may prove to be the best solution for diverting traffic for further inspection.

IV. ACLs vs Flowspec Rules

a) Why Flowspec instead of ACLs?

Flowspec rules essentially provide the same functionality as Access Lists (ACLs), firewall rules and policy based routing, but Flowspec rules have a major advantage. The advantage is that Flowspec rules are sent out in near real time using BGP and are withdrawn just as quickly. Flowspec rules can be sent out programmatically by a controller to a virtually unlimited number of routers when certain conditions are matched. For example, when a DDoS attack to a victim IP is detected by a platform such as Arbor SP, Flowspec rules can immediately be sent out to peering routers to block the attack traffic to the victim IP. Once the attack traffic has stopped, the controller can then withdraw the Flowspec rules from the routers.

In comparison, with an ACL typically a human would have to receive an alert that a victim IP is under DDoS attack, devise the ACL and then log into each router to apply the ACL. Once the attack is over, the human would then need to again log into all the routers and remove the ACL. The steps of applying and removing the configuration could be scripted but this increases the complexity. With numerous DDoS attacks lasting less than 15 minutes, one can clearly see that there is a distinct advantage of using Flowspec to automatically mitigate the attacks.

b) When to use ACLs as opposed to Flowspec rules?

There is no hard and fast rule of when to use one or the other. A good general rule is that if the filtering rules will be permanent or at least changes are not required very frequently, and the filtering rules need to be in place as soon as the device begins forwarding packets (such as limiting access to a management interface), then ACLs are a good fit. If the filtering rules need to be applied temporarily and in a quick fashion, or if the filtering rules need generated and distributed programmatically, then Flowspec rules could be a good fit.

V. Flowspec Details

Vendor platforms differ, but in most cases when a router receives a Flowspec rule, the router validates the rule to verify that:

- The controller sending the Flowspec rule is also advertising the best match unicast route for the destination IP/prefix.
- There is not a more specific existing unicast route in comparison to the Flowspec destination.

There can be additional validation steps depending on the router platform. Any rule that does not meet these verifications will not be installed by routers that perform validation by default. Typically, routers provide the operator the option to manually disable this validation.

A proposed update to the Flowspec RFC validation section[4] (https://tools.ietf.org/html/rfc5575#section-6) allows for Flowspec rules to be originated from a centralized controller[5] (https://tools.ietf.org/html/draft-ietf-idr-bgp-flowspec-oid-05). If ratified and adopted by router vendors, this would negate the need for routers to offer the operator the choice to manually disable this validation.

Normally, in the default configuration, Flowspec rules received by a router are installed on all interfaces of the device. Most routers have a configuration option to disable Flowspec rules from being applied to specific interfaces. It is recommended that Flowspec be disabled on the management interfaces. Flowspec rules could lock out management access to a network device if the rules are not disabled on the management interfaces. Usually a configuration command is applied by the operator to disable Flowspec on any selected interface.

On some platforms, such as Juniper, multiple interfaces can be assigned to a group and then Flowspec can be disabled on all the interfaces in the group. For example, an operator may use Flowspec rules on the peering interfaces of a peering router to block inbound DDoS traffic from the Internet offnet and onnet, but disable Flowspec on the other internal facing and management interfaces. Flowspec rules are immediately removed by a router when either the Flowspec rule is withdrawn via a BGP update or when the BGP session with the controller is terminated for any reason.

Routers are typically configured as route reflectors to advertise Flowspec rules. A router that is not a route reflector should never re-advertise a Flowspec rule that it has learned from a neighbor even if it has a BGP Flowspec session with that neighbor.

Numerous router platforms can receive and apply Flowspec rules. A non-inclusive list is:

- Cisco Routers running IOS-XR and IOS-XE
- Juniper Networks Routers
- Nokia Networks Routers
- Huawei Routers

Software applications can also act as Flowspec controllers, sending and receiving Flowspec rules. A non-inclusive list is:

- ExaBGP[6] (https://github.com/Exa-Networks/exabgp)
- BIRD[7] (http://bird.network.cz/)

- GoBGP[8] (https://github.com/osrg/gobgp)
- YABGP[9] (https://github.com/smartbgp/yabgp)
- Open Daylight[10] (https://www.opendaylight.org/)
- FastNetMon[11] (https://fastnetmon.com/)
- Arbor SP[12] (https://www.netscout.com/arbor-ddos)
- Deepfield Defender[13] (https://networks.nokia.com/solutions/deepfield-ip-network-analytics-DDoS-protection)
- Radware DefenseFlow[14] (https://www.radware.com/products/defenseflow/)
- Auto-Flowspec Docker Container[15] (https://github.com/racompton/docker-auto-flowspec)
- Bgpflowspectool[16] (https://github.com/Pragma-Innovation/bgpflowspectool)
- Flowspy[17] (https://github.com/grnet/flowspy)
- Fortinet FortiDDoS[18] (https://www.fortinet.com/products/ddos/fortiddos.html)

Most carrier class routers that process Flowspec rules can be manually configured as Flowspec controllers in comparison to the more programmatic options available on the above controllers.

VI. Flowspec Best Practices

There are many ways to limit the negative issues that may occur with the use of Flowspec. One suggested best practice is to tag the Flowspec rules with a specific BGP community. The BGP communities assigned to Flowspec rules can be combined with route policies applied to the BGP session with the peer advertising the Flowspec rules.

The route policy can be used by the router to validate that the proper BGP community is assigned to the Flowspec rule before installing it. If the community is invalid, the Flowspec rule can be rejected. BGP communities can also be used to control Flowspec rule application in a large network. For example, routers in the western half of a network could be configured to only accept Flowspec rules tagged with a BGP community of 65000:1. Routers in the eastern half of the same network could be configured to only accept Flowspec rules with a BGP community of 65000:2. In this way, a Flowspec controller can tag its rules with one tag or the other to define which half of the network the rules get applied to.

Route policies can also be used to validate the source or destination prefix defined in the received Flowspec rule. For example, if a network operator knows that their Flowspec controller will only be advertising rules that contain an IPv4 /32 destination prefix, the route policy can check the prefix length and reject any rules that have a shorter prefix length.

Just like other BGP sessions, MD5 authentication of the BGP session using a shared passphrase between peers can be used to validate the TCP connection between the peers. BGP TTL Security[19] (https://tools.ietf.org/html/rfc5082) can also be used to validate the TTL value that packets should have when they are sent from one BGP peer to another.

There are a limited number of Flowspec rules that a router can handle. The number is normally dependent upon the platform and how much free memory is available on the platform. The restrictions can also be per port, group of ports, linecard or even group of linecards. The best practice is to contact the router vendor to get more information about the restrictions on their specific platform.

Typically, the larger the number of ACL rules and firewall rules that a platform has configured, the lower the number of Flowspec rules it can reliably deal with. For these reasons, it is recommended to configure the maximum prefix setting on the BGP session with the peer that is advertising the Flowspec rules. Usually this setting will include a warning level that will generate an alert on the platform and then a hard limit that will drop the BGP session with the peer if the number of prefixes advertised is exceeded.

For most service provider level platforms, a hard limit of 2000 and a warning of 75 percent of this value is normally a safe setting but it is not recommended to use these values in production without lab testing. Flowspec controllers should incorporate a mechanism to limit the number of Flowspec rules that are advertised to peers as well. The more complex a rule is, the more amount of memory it will consume on the router. For example, a rule that blocks all UDP port 1900 traffic to one /32 IPv4 address will take up less memory in a router than a rule that blocks all traffic matching source port 1024 to 65535 and destination ports 80, 443, 8000, 8080, and 8088 with a TCP flag of SYN set and a DSCP value of 3 and a packet length of 78 bytes.

Some router platforms process Flowspec rules before matching local ACLs and firewall rules or performing netflow sampling. In this case an issue may occur where a DDoS attack is incorrectly detected as stopped as soon as the Flowspec rules are installed on the router because no netflow records are being sent to the DDoS detection solution. It is imperative that lab testing be performed to determine if this preemptive behavior exists on the routers that have been deployed. One possible way to address this issue would be to configure the controller to continue the DDoS mitigation for a fixed amount of time that is longer than the average length of an attack.

VII. Standard Architecture

Normally, a Flowspec controller does not have a BGP session with every router to which it needs to send Flowspec rules. The Flowspec controller usually has an iBGP session to one or more BGP speakers configured as route reflectors and then these route reflectors have an iBGP session to each of the routers that will apply Flowspec rules to network traffic. Also, as stated above, a best practice is to tag Flowspec rules with BGP communities so that the rules can be validated by the routers and to define which routers the rules will be applied to.

VIII. Flowspec Use Cases

The use of Flowspec can be divided into use cases that occur within an operator's network and use cases where Flowspec rules are propagated across network operators ASN boundaries. iBGP is used to distributed Flowspec rules within a single ASN. eBGP can be used when sharing Flowspec rules between operators or between an operator and a BGP customer.

IX. iBGP Flowspec Use Cases

The standard use case for Flowspec rules is to mitigate DDoS attacks. Once a system or a human has identified that a DDoS attack has started and determined the details of the attack, that information can be fed into a Flowspec controller. The controller will then generate and advertise Flowspec rules which can either block, rate limit, or divert the attack traffic. In the case of diversion, traffic is diverted to a separate VRF route target where it can be inspected by a Deep Packet Inspection (DPI) device such as a DDoS mitigation appliance.

An alternate DDoS mitigation use case is to drop traffic from source IPs that are generating attack traffic. Normally, DDoS attacks are sourced from large number of IPs and therefore, there should be some mechanism in place to only block attack traffic from a limited number of sources to prevent routers from becoming overwhelmed with Flowspec rules (see Section VII).

In the same vein, Flowspec can be used as a quick method of blocking malicious traffic of this type across an entire network very quickly. Instances have occurred in the past where the exploitation of a UDP service for DDoS amplification has spread rapidly and caused large outages for various hosts across the Internet. Flowspec can be used to very quickly block this malicious traffic across an entire network with one rule. An example of this is the memcache UDP amplification incident which caused the largest openly reported attack seen to that date. The malicious traffic was identified as being sourced from UDP port 11211 with a packet size of 1424 bytes. A Flowspec rule dropping any traffic matching this description would have mitigated this attack. Note that, if as in the memcache case, the filtering needs to be maintained for a long period, the Flowspec rules should be replaced with more permanent ACL/firewall rules.

Care is needed for some UDP amplification attack types. For example, CHARGEN attacks generate numerous non-first UDP fragments. Blocking all UDP fragments is dangerous as it can lead to blocking valid UDP fragments as well as fragments associated with the attack traffic. UDP amplification traffic may or may not produce fragments; some of them have layer 4 mechanisms to limit the packet size to below the MTU limit so that the reply is returned in multiple unfragmented UDP packets. Large NTP and memcached replies behave this way, but DNS, LDAP and CHARGEN do not, and will therefore cause UDP fragmentation.

Another use case for Flowspec is to divert a specific ISP customer's HTTP, HTTPS and DNS traffic to an alternative path where all traffic is redirected to an environment called a walled garden. A walled garden may display a customized message to a customer no matter what webpage they are attempting to visit. Walled gardens are used by ISPs in the case of bot infections, non-bill pay, and other scenarios. Once the customer has received the message and taken the appropriate action, the Flowspec rule can be withdrawn allowing the customer's traffic to follow the normal path.

An additional use case related to malicious traffic is one where an ISP has a list of malicious IPs and the ports that are being used by bad actors, this list of malicious IPs/ports could perhaps come from a third-party threat intel company. This information can be fed into the Flowspec controller to block or divert inbound or outbound traffic matching these IP/ports to an alternate path using the VRF redirection. This alternate path could employ a DPI appliance, such as an Intrusion Prevention System (IPS) or a Next Generation Firewall (NGFW), to inspect this subset of the outbound traffic to identify and to block malicious traffic – e.g., botnet command and control traffic. An advantage of this system is that the DPI appliances only need to examine a small subset of the total traffic – only traffic that matches the Flowspec redirect rules, instead of all traffic.

X. eBGP Flowspec Use Cases

Flowspec rules are primarily advertised within an organization using iBGP, but they can also be advertised via eBGP to a separate entity such as another ISP with a different Autonomous System Number (ASN).

One use case for eBGP Flowspec rules is similar to the current handling of routes advertised by customers to their upstream ISP with the blackhole community assigned - normally BGP community 666. This is commonly called a Remotely Triggered Blackhole (RTBH). The RTBH is usually used to take an IP address offline when it is the target of a DDoS attack. All the traffic to the blackholed IP is blocked by the upstream ISP before it gets to the victim network. In this way, the victim has sacrificed the network connectivity of the one IP address so the attack traffic does not affect other hosts sharing the same network resources. Using Flowspec rules, the victim can advertise a more specific rule to the upstream ISP to block only the attack traffic coming into the victim's network. The victim's network is not affected by the attack because the attack traffic is blocked upstream and potentially the individual victim IP is still reachable as well.

An ISP customer could potentially advertise a very specific set of filters only allowing traffic they define as valid to their Internet facing hosts, with all other traffic being dropped upstream by the ISP. This would limit the range of attacks that would be received by the customer's network. For example, if the customer only has web servers listening on TCP port 443, then the customer could advertise Flowspec rules permitting TCP port 443, rate limiting ICMP echo requests to 100 kbps, and then drop all other network traffic by the upstream ISP.

Another use case for eBGP Flowspec is DDoS Peering. Like an upstream ISP blocking DDoS traffic for a customer, ISPs that peer with another ISP that is the victim of a DDoS attack can receive Flowspec routes to block or rate limit the attack traffic that is destined to the victim. The victim ISP needs to set up eBGP sessions with its neighbors and add the Flowspec address family to exchange Flowspec rules. Once a DDoS attack starts, the victim ISP identifies what Flowspec rules would mitigate the attack traffic and then advertises them out via eBGP to its neighbors. Once the attack has stopped, the victim ISP can withdraw the Flowspec rules and request that the rules be removed from its neighboring ISPs.

Please note that in the above two cases of a customer advertising Flowspec rules to an upstream ISP, and an ISP advertising Flowspec rules to other ISPs via eBGP, the entity receiving the rules should be very careful about validating the rules before installing them into its network. Please reference Section VII of this paper for more information about some safeguards that can be used to prevent an unintentional outage. For example, an ISP may configure the maximum prefixes option so a customer can only advertise 10 Flowspec rules to the ISP. The ISP can verify that the rule has been tagged with a specific BGP community and specifies a destination /32 or /128 prefix contained within the customer's publicly reachable netblocks. The ISP can drop the advertised Flowspec rule if one or more of these conditions are not met. Also, it may be a good idea to have the Flowspec rules that are advertised by an external entity sent to a user interface, where a human can review the rules before installing them into the network. This step can go a long way to prevent damaging rules from being installed on the network.

XI. Conclusion

This paper was written for network engineers familiar with BGP routing and other common networking technologies who are responsible for NSP, hosting providers or enterprise networks. The concept of BGP Flowspec was originally proposed by the creators of IETF RFC 5575. Originally, Flowspec was specifically engineered to deal with DDoS attacks but it is now also used for additional applications outside of DDoS mitigation. Inherent concerns with the application of Flowspec outside the original concept, can be minimized by the techniques and best practices defined in this paper, thus allowing a successful deployment of Flowspec.

XII. References

1. RFC 5575 (Request for Comments) which was ratified in August 2009. - https://tools.ietf.org/html/rfc5575.

2. Dissemination of Flow Specification Rules for IPv6 draft. - https://tools.ietf.org/html/draft-ietf-idr-flow-spec-v6-08.)

3. BGP Flow-Spec Extended Community for Traffic Redirect to IP Next Hop draft. - https://datatracker.ietf.org/doc/draft-simpson-idr-flowspec-redirect/

4. Dissemination of Flow Specification Rules RFC 5575 - https://tools.ietf.org/html/rfc5575 - section-6

5. Revised Validation Procedure for BGP Flow Specifications draft - (https://tools.ietf.org/html/draft-ietf-idr-bgp-flowspec-oid-05)

6. The BGP Swiss Army Knife of Networking ExaBGP - https://github.com/Exa-Networks/exabgp

7. The BIRD Internet Routing Daemon - http://bird.network.cz/

8. BGP implemented in the Go Programming Language - https://github.com/osrg/gobgp

9. BGP Python Implementation - https://github.com/smartbgp/yabgp

10. Open Daylight - https://www.opendaylight.org/

11. FastNetMon DDoS Detection Tool - https://fastnetmon.com/

12. Arbor DDoS Solutions - https://www.netscout.com/arbor-ddos

13. Deepfield: IP Network Analytics & DDoS Protection - https://networks.nokia.com/solutions/deepfield-ip-network-analytics-DDoS-protection

14. Radware DefenseFlow: Network-Wide DDoS Attack Defense and Centralized Cyber Control - https://www.radware.com/products/defenseflow/

15. Auto-Flowspec Docker Container - https://github.com/racompton/docker-auto-flowspec

16. BGP Flowspec (RFC 5575) tool for DDoS mitigation - https://github.com/Pragma-Innovation/bgpflowspectool

17. GRNET Firewall on Demand platform - https://github.com/grnet/flowspy

18. Fortinet FortiDDoS Advanced DDoS Protection for Enterprise Data Centers - https://www.fortinet.com/products/ddos/fortiddos.html.

19. The Generalized TTL Security Mechanism (GTSM) RFC 5082 -https://tools.ietf.org/html/rfc5082.

As with all documents that we publish, please check the M^{3}AAWG website (www.m3aawg.org) for updates.

M3AAWG124