Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: May 9, 2019

A. Khare, Ed.
J. Scudder
Juniper Networks, Inc.
L. Jalil
M. Gallagher
Verizon

November 5, 2018

BGP FlowSpec Payload Matching
draft-khare-idr-bgp-flowspec-payload-match-02

Abstract

The rise in frequency, volume, and pernicious effects of DDoS attacks has elevated them from fare for the specialist to generalist press. Numerous reports detail the taxonomy of DDoS types, the varying motivations of their attackers, as well as the resulting business and reputation loss of their targets.

BGP FlowSpec (RFC 5575, "Dissemination of Flow Specification Rules") can be used to rapidly disseminate filters that thwart attacks, being particularly effective against the volumetric type. Operators can use existing FlowSpec components to match on pre-defined packet header fields. However recent enhancements to forwarding plane filter implementations allow matches at arbitary locations within the packet header and, to some extent, the payload. This capability can be used to detect highly amplified attacks, whose attack signature remains relatively constant.

We define a new FlowSpec component, "Flexible Match Conditions", with similar matching semantics to those of existing components. This component will allow the operator to define bounded match conditions using offsets and bitmasks.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any

time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on May 9, 2019.

Table of Contents

1. Introduction 3
1.1. Requirements Language 3
2. Motivation 3
2.1. Volumetric attacks 3
2.2. Tunneled traffic 4
2.3. Non-IP traffic 4
3. Details 4
3.1. Flexible Match Conditions 5
3.1.1. Operator 5
3.1.2. Value 6
3.1.2.1. String Comparison 6
3.1.2.2. Numeric Range Comparison 7
3.1.3. Example 7
3.2. Error Handling 8
3.3. Security Considerations 8
3.4. IANA Considerations 8
4. Acknowledgements 9
5. References 9
5.1. Normative References 9
5.2. Informative References 9
5.3. URIs 10
Authors' Addresses 10

1. Introduction

BGP FlowSpec [RFC5575] can be used to rapidly disseminate filters that thwart attacks, being particularly effective against the volumetric type. Operators can use existing FlowSpec components to match on pre-defined packet header fields. However recent enhancements to forwarding plane filter implementations allow matches at arbitary locations within the packet header and, to some extent, the payload. This capability can be used to detect highly amplified attacks whose attack signature remains relatively constant, or the burgeoning variety of tunneled traffic.

We define a new FlowSpec component, "Flexible Match Conditions", with similar matching semantics to those of existing components. This component will allow the operator to define bounded match conditions using offsets and bitmasks.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119] .

2. Motivation

BGP FlowSpec couples both the advertisement of NLRI-specific match conditions, as well as the forwarding instance to which the filter is attached. This makes sense since BGP FlowSpec advertisements are most commonly generated, or at least verified, by human operators. The operator finds it intuitive to configure match conditions as human-readable values, native to each address family.

It is much friendlier, for instance, to define a filter that matches a source address of 192.168.1.1/32, than it is to work with the equivalent binary representation of that IPv4 address. Further, it is easier to use field names such as 'IPv4 source address' as part of the match condition, than it is to demarc that field using byte and bit offsets.

However, there are a number of use cases that benefit from the latter, more machine-readable approach.

2.1. Volumetric attacks

Launching a DDoS is easier and more cost-effective than ever. The will to attack matters more than wherewithal. Those with the inclination can initiate one from the comfort of their homes [1], or

even buy DDoS-as-a-Service [2], complete with 24x7 support and flexible payment plans.

Despite their effectiveness, such attacks are easily thwarted - once identified. The challenge lies in fishing out a generally unvarying attack signature from a data stream. Machine analysis may prove superior here, given the size of input involved. The resulting pattern may not lie within a well-defined field; even if it happens to, it may be a more straight-forward workflow to have machine analysis result in a machine-readable filter.

2.2. Tunneled traffic

Tunnels continue to proliferate due to the benefits they provide. They can help reduce state in the underlay network. Tunnels allow bypassing routing decisions of the transit network. Traffic that is tunneled is often done so to obscure or secure. Common tunnel types include IPsec [RFC4301], Generic Routing Encapsulation (GRE) [RFC2890], Virtual eXtensible Local Area Network (VXLAN) [RFC7348], GPRS Tunneling Protocol (GTP) [3GPP.29.281], et al.

By definition, transit nodes that are not the endpoints of the tunnel hold no attendant control or management plane state. These very qualities make it challenging to filter tunneled traffic at non-endpoints. Often though, the forwarding hardware at these transit-only nodes is capable of reading the byte stream that comprises the protocol being tunneled. Despite this capability, it is usually infeasible to filter based on the content of this passenger protocol's header since BGP FlowSpec does not provide the operator a way to address arbitrary locations within a packet.

2.3. Non-IP traffic

Not all traffic is forwarded as IP packets. Layer 2 services abound, including flavors of BGP-signaled Ethernet VPNs such as BGP-EVPN, BGP-VPLS, FEC 129 VPWS (LDP-signaled VPWS with BGP Auto-Discovery).

Ongoing efforts such as [I-D.ietf-idr-flowspec-l2vpn] offer one approach, which is to add layer 2 fields as additional match conditions. This may suffice if a filter needs to be applied only to layer 2, or only to layer 3 header fields.

3. Details

3.1. Flexible Match Conditions

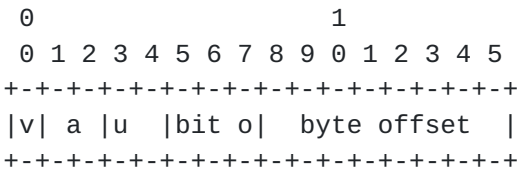
We define a new FlowSpec component, Type TBD, named "Flexible Match Conditions".

Encoding: <type (1 octet), op, value>

It contains a single {operator, value} tuple that is used to match packets according to the rules given below.

3.1.1. Operator

The operator field is encoded as:



v - Type of value being matched, string comparison (Section 3.1.2.1) if this bit is set, and numeric range (Section 3.1.2.2) if unset.

a - Anchor. A 2-bit unsigned integer whose value indicates where in the packet the match should start. To avoid ambiguity with tunneled packets, the match SHOULD be anchored at the outermost header. An example is given below (Section 3.1.3).

Value	Symbolic Name	Match start
0	d	Layer 2 (d)ata-link layer Ethernet header
1	i	Layer 3 (I)Pv4/IPv6 header
2	t	Layer 4 TCP/UDP (t)ransport header
3	p	Layer 4-specific (p)rotocol-specific payload

Anchor Field Values

u - Reserved. MUST be set to 0. MUST be ignored on receipt.

bit offset - A 3-bit unsigned integer indicating how many bits to ignore, following the byte offset.

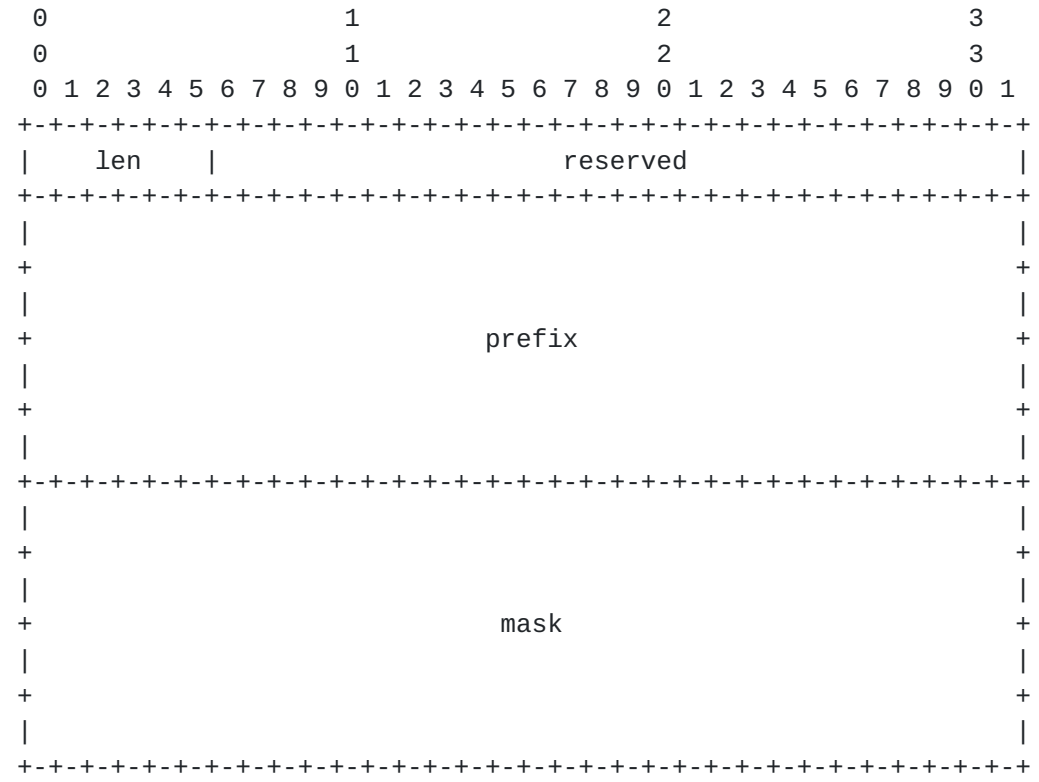
byte offset - An 8-bit unsigned integer indicating how many bytes to ignore, after the match start as determined by the first selected anchor bit.

3.1.2. Value

The operator field indicates where to start matching; by contrast, the value operand indicates what to match and where to stop matching. The value operand MUST be of the type indicated by the 'v' bit, as signaled in the operator. As a result it can take on one of two forms - string vs. numeric range comparison.

The length of the numeric range is constant. It uses two 64-bit fields. A string comparison uses two 128-bit fields. Its length field indicates the extent of how much of the prefix and mask fields to use in the AND operation. This is deemed sufficient for stateless inspection and practical for efficient hardware forwarding plane implementations.

3.1.2.1. String Comparison



len - Indicates the number of corresponding bits in the prefix and mask fields to read. This length field is interpreted as

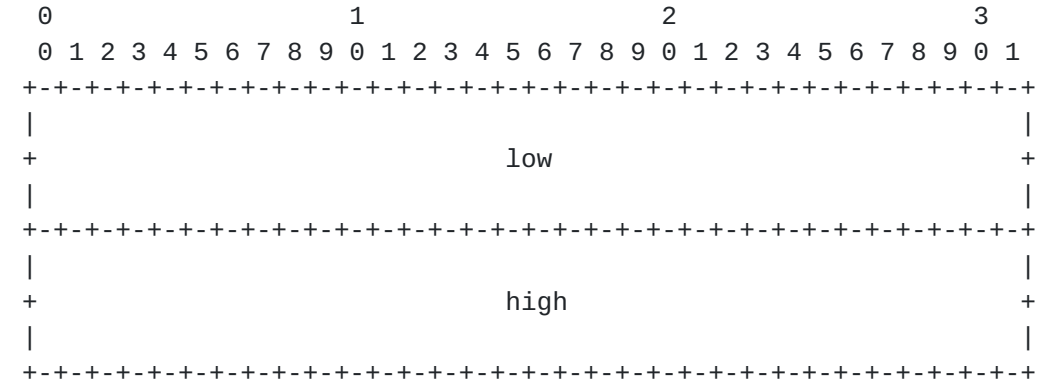
(len + 1 << 1). This allows even unsigned values ranging from 2-128.

prefix - Provides a bit string to be matched. The prefix and mask fields are bitwise AND'ed to create a resulting pattern. The number of bits used in the AND operation are indicated by the preceding length field.

mask - Paired with the prefix field to create a bit string match. An unset bit is treated as a 'do not care' bit in the corresponding position in the prefix field. When a bit is set in the mask, the value of the bit in the corresponding location in the prefix field must match exactly.

Implementations MUST only extract the number of bits from the prefix and mask fields as indicated by the preceding length field.

3.1.2.2. Numeric Range Comparison



low - The low value of the desired inclusive numeric range. This value MUST be numerically lower than the high value.

high - The high value of the desired inclusive numeric range. This value MUST be numerically higher than the low value.

3.1.3. Example

As an example, consider that the canonical Virtual eXtensible Local Area Network (VXLAN) [RFC7348] packet has the following headers:

- Outer Ethernet Header: Source MAC address of the originating VXLAN Tunnel End Point (VTEP).

- Outer IPv4/IPv6 Header: Source IP address of the originating VXLAN Tunnel End Point (VTEP).

- Outer UDP Header: Random source port used to generate entropy for load balancing, and destined to the IANA-assigned VXLAN port 4789.

- VXLAN Header: Used to identify a specific VXLAN overlay network.

- Inner Ethernet Header and payload: Original MAC frame being encapsulated.

The following table outlines where the match would start based on the anchor setting:

Anchor value	Match start
d	Outer Ethernet Header
i	Outer IPv4/IPv6 Header
t	Outer UDP Header
p	VXLAN Header

3.2. Error Handling

Malicious, misbehaving, or misunderstanding implementations could advertise semantically incorrect values. Care must be taken to minimize fallout from attempting to parse such data. Any well-behaved implementation SHOULD verify that the minimum packet length undergoing a match equals (match start header length + byte offset + bit offset + value length).

3.3. Security Considerations

This document introduces no additional security considerations beyond those already covered in [RFC5575] .

3.4. IANA Considerations

IANA is requested to assign a type from the First Come First Served range of the "Flow Spec Component Types" registry:

Type Value	Name	Reference
TBD	Flexible Match Conditions	this document

4. Acknowledgements

Thanks to Rafal Jan Szarecki, Sudipto Nandi, Ron Bonica, and Jeff Haas for their valuable comments and suggestions on this document.

5. References

5.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC5575] Marques, P., Sheth, N., Raszuk, R., Greene, B., Mauch, J., and D. McPherson, "Dissemination of Flow Specification Rules", RFC 5575, DOI 10.17487/RFC5575, August 2009, <https://www.rfc-editor.org/info/rfc5575>.

5.2. Informative References

[3GPP.29.281] 3GPP, "General Packet Radio System (GPRS) Tunnelling Protocol User Plane (GTPv1-U)", 3GPP TS 29.281 10.3.0, September 2011.

[I-D.ietf-idr-flowspec-l2vpn] Weiguo, H., liangqiandeng, l., Uttaro, J., Litkowski, S., and S. Zhuang, "Dissemination of Flow Specification Rules for L2 VPN", draft-ietf-idr-flowspec-l2vpn-08 (work in progress), July 2018.

[RFC2890] Dommety, G., "Key and Sequence Number Extensions to GRE", RFC 2890, DOI 10.17487/RFC2890, September 2000, <https://www.rfc-editor.org/info/rfc2890>.

[RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, DOI 10.17487/RFC4301, December 2005, <https://www.rfc-editor.org/info/rfc4301>.

[RFC7348] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC 7348, DOI 10.17487/RFC7348, August 2014, <https://www.rfc-editor.org/info/rfc7348>.

5.3. URIs

[1] https://github.com/649/Memcrashed-DDoS-Exploit

[2] https://www.facebook.com/PutinStresser/photos/a.1687498801469198/2024483917770683/?type=3

Authors' Addresses

Anurag Khare (editor)
Juniper Networks, Inc.
2251 Corporate Park Drive
Herndon, Virginia 20171
US

Email: anuragk@juniper.net

John Scudder
Juniper Networks, Inc.
1133 Innovation Way
Sunnyvale, CA 94089
US

Email: jgs@juniper.net

Luay Jalil
Verizon

Email: luay.jalil@one.verizon.com

Michael Gallagher
Verizon

Email: michael.gallagher@verizon.com