scone
Internet-Draft
Intended status: Standards Track
Expires: 30 October 2025

D. Wing
Cloud Software Group
T. Reddy
Nokia
S. Rajagopalan
Cloud Software Group
L. Contreras
Telefonica
28 April 2025

Throughput Advice Object for SCONE
draft-brw-scone-throughput-advice-blob-03

Abstract

Traffic exchanged over a network may be subject to rate-limit policies for various operational reasons. This document specifies a generic object (called, Throughput Advice) that can be used by mechanisms for hosts to dynamically discover these network rate-limit policies. This information is then passed to applications that might adjust their behaviors accordingly.

The design of the throughput advice object is independent of the discovery channel (protocol, API, etc.).

Table of Contents

1. Introduction

Connectivity services are provided by networks to customers via dedicated terminating points, such as customer edges (CEs) or User Equipment (UE). To facilitate data transfer via the provider network, it is assumed that appropriate setup is provisioned over the links that connect customer terminating points and a provider network (usually via a Provider Edge (PE)), successfully allowing data exchange over these links. The required setup is referred to in this document as network attachments, while the underlying link is referred to as "bearers".

The bearer can be a physical or logical link that connects a customer device to a provider network. A bearer can be a wireless or wired link. The same or multiple bearer technologies can be used to establish the bearer (e.g., WLAN or cellular) to graft customer terminating points to a network.

Figure 1 shows an example of a network that connects CEs and hosts (UE, for example). These CEs are servicing other (internal) hosts. The identification of these hosts is hidden from the network. The policies enforced at the network for a network attachment are per-subscriber, not per-host. Typically, if a CE is provided with a /56 IPv6 prefix, policies are enforced in the network on that /56 not the individual /64s that will be used by internal hosts. A customer terminating point may be serviced with one (e.g., UE#1, CE#1, and CE#3) or multiple network attachments (e.g., CE#2). For the sake of simplicity, Figure 1 does not show the interconnection with other networks or multi-homed CEs.

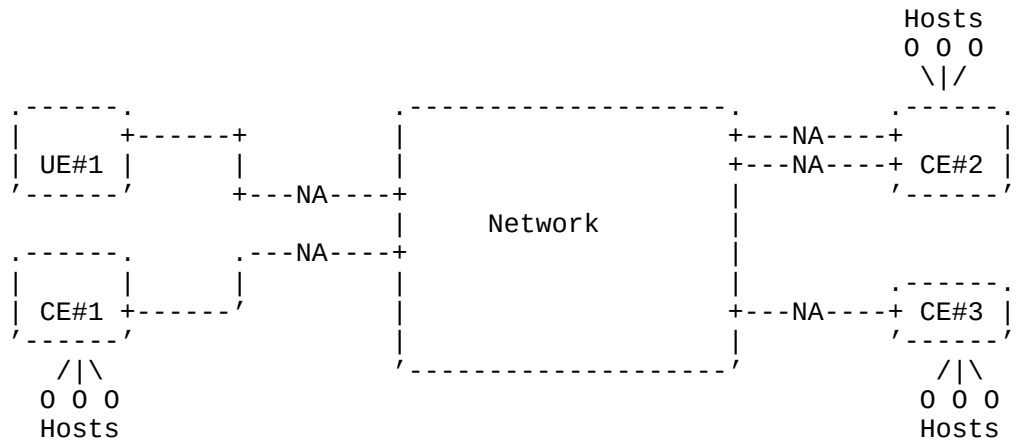

Figure 1: Sample Network Attachments

Customer terminating points are provided with a set of information (e.g., IP address/prefix) to successfully be able to send and receive traffic over a network attachment. The required set of parameters to provision a network attachment is a function of the connectivity service offering. For example, a very limited set of parameters is required for mass-market service offering while a more elaborated set is required for Enterprise services. A comprehensive list of provisioning parameters that are available on the PE-side of a network attachment is specified in [I-D.ietf-opsawg-ntw-attachment-circuit].

As discussed, e.g., in Section 4.2 of [RFC7567], packet dropping by network devices occurs mainly to protect the network (e.g., congestion-unresponsive flows) and also to ensure fairness over a shared link. These policies may be intentional policies (e.g., enforced as part of the activation of the network attachment and typically agreed upon service subscription) or be reactive policies (e.g., enforced temporarily to manage an overload or during a DDoS attack mitigation). Rate-limits are usually configured in (ingress) nodes. These rate-limits can be shared with customers when subscribing to a connectivity service (e.g., "A YANG Data Model for Layer 2 Virtual Private Network (L2VPN) Service Delivery" [RFC8466]).

Section 5 defines a set of parameters that can be used by networks to share the rate-limit policies applied on a network attachment: Throughput Advice. The set of parameters are independent of the address family.

This document does not assume nor preclude any specific signaling protocol to share the throughput advices. These parameters are independent of the channel that is used by hosts to discover such policies.

Whether host-to-network, network-to-host, or both policies are included in throughput advice is deployment specific. All these combinations are supported in this document.

Also, one or more throughput advice instances may be returned for a given traffic direction. Examples of such instances are discussed in Section 6.

As one can infer from the name, a throughput advice is advisory in nature. The advice is provided solely as a hint.

In order to ease mapping with specific signaling mechanisms, allow for future extensions, and ensure consistent use of the advice, a new IANA registry is created in Section 8.

2. What’s Out?

This document does not make any assumption about:

* The type of network (fixed, cellular, etc.) that terminates a network attachment.

* The services or applications that are delivered over a network attachment. Whether one or multiple services are bound to the same network attachment is deployment specific.

* How the throughput advice is computed/set.

* The protocol machinery for validating, refreshing, detecting stale, and flushing out received advices.

* How applications running over a host can learn the bitrates associated with a network attachment. Typically, this can be achieved by invoking a dedicated API. However, the exact details of the API(s) is OS-specific and, thus, out of scope of this document.

3. Conventions and Definitions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

This document makes use of the following term:

Rate-limit: Used as a generic term to refer to a policy to restrict the maximum bitrate over a network attachment.

It can be used with or without any traffic classification.

A rate-limit can involve limiting the rate and/or burst size.

4. Sample Deployment Cases

Some deployment use cases for throughput advice discovery are provided below:

Adaptive Application Behavior: Discovery of intentional policy applied on network attachments when such information is not made available during the service activation or when network upgrades are performed. Adaptive applications will use the information to adjust their behavior.

Concretely, applications are supposed to have access to all throughput advice instances and would, thus, adjust their behavior as a function of the parameters indicated in a throughput policy.

Likewise, a host with multiple network attachments may use the

discovered throughput advice instances over each network attachment to decide how to distribute its flows over these network attachments (prefer a network attachment to place an application session, migrate connection, etc.). That's said, this document does not make any recommendation about how a receiving host uses the discovered policy.

The throughput advice can feed mechanisms such as Section 4.4.2 of [RFC7661] or Section 7.8 of [RFC9002] to control the maximum burst size.

Network Assisted Offload: A network may advertize a throughput advice when it is overloaded, including when it is under attack. The rate-limit policy is basically a reactive policy that is meant to adjust the behavior of connected hosts to better control the load during these exceptional events (issue with RAN resources, for example).

The mechanism can also be used to enrich the tools that are already available to better handle attack traffic close to the source [RFC9066].

Better Local Services: A user may configure policies on the CE such as securing some resources to a specific internal host used, e.g., for gaming or video streaming. The CE can use the throughput advice to share these rate-limit policies to connected hosts to adjust their forwarding behavior. Controlling the load at the source will allow to partition the resources between connected hosts.

5. Throughput Advice Object

5.1. Throughput Parameters

The throughput advice parameters leverage existing technologies for configuring policies in provider networks. Appendix A provides a brief overview of how inbound policies are enforced in ingress network nodes. The reader may refer to [RFC2697], [RFC2698], and [RFC4115] for examples of how various combinations of Committed Information Rate (CIR), Committed Burst Size (CBS), Excess Information Rate (EIR), Excess Burst Size (EBS), Peak Information Rate (PIR), and Peak Burst Size (PBS) are used for policing. Typically:

* A Single-Rate, Two-Color Marker (1r2c) uses CIR and CBS.

* A Single-Rate, Three-Color Marker (1r3c) [RFC2697] uses CIR, CBS, and EBS.

* A Dual-Rate, Three-Color Marker (2r3c) [RFC2698] uses CIR, CBS, PIR, and PBS.

* 2r3c when implemented with [RFC4115] uses CIR, CBS, EIR, and EBS. This mode allows for a better handling of in-profile traffic (refer to Section 1 of [RFC4115] for more details).

An implementation example of these variants (and others) can be found at [VPP].

This version of the document uses the common denominator of all these policies: CIR/CBS.

5.2. Overall Object Structure

A throughput advice object may include multiple throughput advices (referred to as "throughput advice instances"), each covering a specific match criteria. Each of these instances adheres to the structure defined in Section 5.3.

Throughput advice objects are bound to the network interface over which the advice was received.

The throughput advice object is described in CDDL [RFC8610] format shown in Figure 2. This format is meant to ease mapping with encoding specifics of a given discovery channel that supplies the throughput advice.

```
; Provides information about the rate-limit policy that is
; enforced for a network attachment.
; One or more throughput instances can be present in an advice.

throughput-advice =  [+ throughput-instance]

throughput-instance =  {
  ? instance-flags => flags,
  ? traffic-category => category,
  throughput => rate-limit
}

; Indicates scope, traffic direction, and reliability type.
; Default value for scope is per subscriber policy.
; Default value for direction is network-to-host direction.
; Default value for reliability is false (i.e., the policy is
; applicable to both reliable and unreliable traffic).
; If any of these parameters is not present, this is equivalent
; to enclosing the parameter with its default value.

flags =  {
  ? scope: &scope-values .default subscriber,
  ? direction: &direction-values .default n2h,
  ? reliability: &reliability-values .default any
}

scope-values = (subscriber: 0, host: 1, flow: 2)
direction-values = (n2h: 0, h2n: 1, bidir: 2)
reliability-values = (any: 0, reliable: 1, unreliable: 2)

; Indicates traffic category to which the policy is bound.
; If the value is set to 0, this means that the policy is
; enforced for all traffic.

category =  {
  ? tc: uint .default 0
}

; Indicates the rate and burst limits.
; Only CIR/CBS are mandatory to include.

rate-limit = {
  cir: uint,          ; Mbps
  cbs: uint .gt 0,    ; bytes
}
```

Figure 2: Throughput Advice Object Format in CDDL

5.3. Throughput Advice Instance Attributes

This section defines the set of attributes that are included in a throughput advice instance:

Instance Flags (IF): These flags are used to express some generic properties of the applicability of the instance. The following flags are defined:

S (Scope): Indicates the granularity of enforcing policies.

This parameter specifies whether the policy is a per-subscriber, per-host, or per-flow policy.

D (Direction): Indicates the direction on which to apply the enclosed policy.

When set to "00b", this flag indicates that this policy is for network-to-host direction.

When set to "01b", this flag indicates that this policy is for host-to-network direction.

When set to "10b", this flag indicates that this policy is for both network-to-host and host-to-network directions.

R (Reliability): Indicates the reliability type of traffic on which to apply the enclosed policy.

For example, reliable could map to Queue-Building (QB) and unreliable could map to Non-Queue-Building (NQB). One of the ways for application to make reliability markings visible is by following, e.g., the considerations in Section 4 of [I-D.ietf-tsvwg-nqb].

When set to "00b", this flag indicates that this policy is for both reliable and unreliable traffic.

When set to "01b", this flag indicates that this policy is for unreliable traffic.

When set to "10b", this flag indicates that this policy is for reliable traffic.

No meaning is associated with setting the field to "11b". Such value MUST be silently ignored by the receiver.

U: Unassigned flags. See Section 8.2.

TC (Traffic Category): Specifies a traffic category to which this policy applies.

The following values are supported:

* "0": All traffic. This is the default value.

* 1-63: Unassigned values. See Section 8.3.

Committed Information Rate (CIR) (Mbps): An average rate that specifies the maximum number of bits that a network can send (or receive) during one second over a network attachment.

The CIR value MUST be greater than or equal to 0.

If set to 0 (or a very low value), this indicates to the host that alternate paths (if any) should be preferred over this one.

This parameter is mandatory.

Committed Burst Size (CBS) (bytes): Specifies the maximum burst size that can be transmitted at CIR.

MUST be greater than zero.

This parameter is mandatory.

6. Examples

For the sake of illustration, Figure 3 exemplifies the content of a throughput advice using JSON notations. The advice includes one rate-limit instance that covers network-to-host traffic direction and is applicable to all traffic destined to any host of a subscriber.

```
{
    "throughput-advice": [
        {
            "direction": 0,
            "scope": 0,
            "tc": 0,
            "cir": 50,
            "cbs": 10000
        }
    ]
}
```

Figure 3: A JSON Example

The advice conveyed in Figure 4 is similar to the advice in Figure 3. The only difference is that default values are not explicitly signaled in Figure 4.

```
{
    "throughput-advice": [
        {
            "cir": 50,
            "cbs": 10000
        }
    ]
}
```

Figure 4: A JSON Example with Default Values Not Explicitly Signaled

Figure 5 shows the example of an advice that encloses two instances, each for one traffic direction.

```
{
    "throughput-advice": [
        {
            "direction": 0,
            "cir": 50,
            "cbs": 10000
        },
        {
            "direction": 1,
            "cir": 30,
            "cbs": 8000
        }
    ]
}
```

Figure 5: A JSON Example with Both Traffic Directions

If both directions are covered by the same rate-limit policy, then the advice can be supplied as shown in Figure 6

```
{
    "throughput-advice": [
        {
            "direction": 2,
            "cir": 50,
            "cbs": 10000
        }
    ]
}
```

Figure 6: A JSON Example with Single Bidir Rate-Limit Policy

7. Security Considerations

As discussed in Section 4, sharing a throughput advice helps networks mitigate overloads, particularly during periods of high traffic volume.

An attacker who has the ability to change the throughput advice objects exchanged over a network attachment may:

Decrease the bitrate value: This may lower the perceived QoS if the host aggressively lowers its transmission rate.

Increase the bitrate value: The network attachment will be overloaded, but still the rate-limit at the network will discard excess traffic.

Delete or remove the advice: This is equivalent to deployments where the advice is not shared.

8. IANA Considerations

8.1. Rate-Limit Policy Objects Registry Group

This document requests IANA to create a new registry group entitled "SCONE Rate-Limit Policy Objects".

8.2. Instance Flags Registry

This document requests IANA to create a new registry entitled "Instance flags" under the "SCONE Rate-Limit Policy Objects" registry group (Section 8.1).

The initial values of this registry is provided in Table 1.

Bit Position	Label	Description	Reference
1	S	Scope	This-Document
2-3	D	Direction	This-Document
4-5	R	Reliability	This-Document
6-8		Unassigned	

Table 1: Instance Flags

The allocation policy of this new registry is "IETF Review" (Section 4.8 of [RFC8126]).

8.3. Traffic Category Registry

This document requests IANA to create a new registry entitled "Traffic Category Types" under the "SCONE Rate-Limit Policy Objects" registry group (Section 8.1).

The initial values of this registry is provided in Table 2.

Value	Description	Reference
0	All traffic	This-Document
1-63	Unassigned	

Table 2: Traffic Category Values

The allocation policy of this new registry is "IETF Review" (Section 4.8 of [RFC8126]).

8.4. Rate Parameters Registry

This document requests IANA to create a new registry entitled "Rate Parameters" under the "SCONE Rate-Limit Policy Objects" registry group (Section 8.1).

The initial values of this registry is provided in Table 3.

Parameter	Description	Mandatory (Y/N)	Reference
cir	Committed Information Rate (CIR)	Y	This-Document
cbs	Committed Burst Size (CBS)	Y	This-Document

Table 3: Initial Rate Parameters Values Values

The allocation policy of this new registry is "IETF Review" (Section 4.8 of [RFC8126]).

9. References

9.1. 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/rfc/rfc2119>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/rfc/rfc8126>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/rfc/rfc8174>.

[RFC8610] Birkholz, H., Vigano, C., and C. Bormann, "Concise Data Definition Language (CDDL): A Notational Convention to Express Concise Binary Object Representation (CBOR) and JSON Data Structures", RFC 8610, DOI 10.17487/RFC8610, June 2019, <https://www.rfc-editor.org/rfc/rfc8610>.

9.2. Informative References

[I-D.ietf-opsawg-ntw-attachment-circuit]
Boucadair, M., Roberts, R., de Dios, O. G., Barguil, S., and B. Wu, "A Network YANG Data Model for Attachment Circuits", Work in Progress, Internet-Draft, draft-ietf-opsawg-ntw-attachment-circuit-16, 23 January 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-opsawg-ntw-attachment-circuit-16>.

[I-D.ietf-teas-5g-ns-ip-mpls]
Szarkowicz, K. G., Roberts, R., Lucek, J., Boucadair, M., and L. M. Contreras, "A Realization of Network Slices for 5G Networks Using Current IP/MPLS Technologies", Work in Progress, Internet-Draft, draft-ietf-teas-5g-ns-ip-mpls-18, 3 April 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-teas-5g-ns-ip-mpls-18>.

[I-D.ietf-tsvwg-nqb]
White, G., Fossati, T., and R. Geib, "A Non-Queue-Building Per-Hop Behavior (NQB PHB) for Differentiated Services", Work in Progress, Internet-Draft, draft-ietf-tsvwg-nqb-27, 8 November 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-tsvwg-nqb-27>.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, DOI 10.17487/RFC2475, December 1998, <https://www.rfc-editor.org/rfc/rfc2475>.

[RFC2697] Heinanen, J. and R. Guerin, "A Single Rate Three Color Marker", RFC 2697, DOI 10.17487/RFC2697, September 1999, <https://www.rfc-editor.org/rfc/rfc2697>.

[RFC2698] Heinanen, J. and R. Guerin, "A Two Rate Three Color Marker", RFC 2698, DOI 10.17487/RFC2698, September 1999, <https://www.rfc-editor.org/rfc/rfc2698>.

[RFC4115] Aboul-Magd, O. and S. Rabie, "A Differentiated Service Two-Rate, Three-Color Marker with Efficient Handling of in-Profile Traffic", RFC 4115, DOI 10.17487/RFC4115, July 2005, <https://www.rfc-editor.org/rfc/rfc4115>.

[RFC7567] Baker, F., Ed. and G. Fairhurst, Ed., "IETF Recommendations Regarding Active Queue Management", BCP 197, RFC 7567, DOI 10.17487/RFC7567, July 2015, <https://www.rfc-editor.org/rfc/rfc7567>.

[RFC7661] Fairhurst, G., Sathiaseelan, A., and R. Secchi, "Updating TCP to Support Rate-Limited Traffic", RFC 7661, DOI 10.17487/RFC7661, October 2015, <https://www.rfc-editor.org/rfc/rfc7661>.

[RFC8466] Wen, B., Fioccola, G., Ed., Xie, C., and L. Jalil, "A YANG Data Model for Layer 2 Virtual Private Network (L2VPN) Service Delivery", RFC 8466, DOI 10.17487/RFC8466, October 2018, <https://www.rfc-editor.org/rfc/rfc8466>.

[RFC9002] Iyengar, J., Ed. and I. Swett, Ed., "QUIC Loss Detection and Congestion Control", RFC 9002, DOI 10.17487/RFC9002, May 2021, <https://www.rfc-editor.org/rfc/rfc9002>.

[RFC9066] Reddy.K, T., Boucadair, M., Ed., and J. Shallow, "Distributed Denial-of-Service Open Threat Signaling (DOTS) Signal Channel Call Home", RFC 9066, DOI 10.17487/RFC9066, December 2021, <https://www.rfc-editor.org/rfc/rfc9066>.

[VPP] Vector Packet Processor (VPP), "Policing", <https://s3-docs.fd.io/vpp/23.06/developer/corefeatures/policer.html>.

Appendix A. Overview of Network Rate-Limit Policies

As discussed, for example in [I-D.ietf-teas-5g-ns-ip-mpls], a provider network’s inbound policy can be implemented using one of following options:

* 1r2c (single-rate two-color) rate limiter

  This is the most basic rate limiter, described in Section 2.3 of [RFC2475]. It meters at an ingress interface a traffic stream and marks its packets as in-profile (below CIR being enforced) or out-of-profile (above CIR being enforced). In-profile packets are accepted and forwarded. Out-of profile packets are either dropped right at the ingress node (hard rate limiting), or remarked (with different MPLS TC or DSCP TN markings) to signify ’this packet should be dropped in the first place, if there is a congestion’ (soft rate limiting), depending on the business policy of the provider network. In the second case, while packets above CIR are forwarded at an ingress node, they are subject to being dropped during any congestion event at any place in the provider network.

* 2r3c (two-rate three-color) rate limiter

This was initially defined in [RFC2698], and its improved version in [RFC4115]. The traffic is assigned to one of the these three categories:

- Green, for traffic under CIR
- Yellow, for traffic between CIR and PIR
- Red, for traffic above PIR

An inbound 2r3c meter implemented with [RFC4115], compared to [RFC2698], is more 'customer friendly' as it doesn't impose outbound peak-rate shaping requirements on customer edge (CE) devices or hosts. 2r3c meters in general give greater flexibility for provider network edge enforcement regarding accepting the traffic (green), de-prioritizing and potentially dropping the traffic on transit during congestion (yellow), or hard dropping the traffic (red).

Acknowledgments

Thanks to Eduard Vasilenko for the comments.

Authors' Addresses

Dan Wing
Cloud Software Group Holdings, Inc.
United States of America
Email: danwing@gmail.com

Tirumaleswar Reddy
Nokia
India
Email: kondtir@gmail.com

Sridharan Rajagopalan
Cloud Software Group Holdings, Inc.
United States of America
Email: sridharan.girish@gmail.com

Luis M. Contreras
Telefonica
Spain
Email: luismiguel.contrerasmurillo@telefonica.com