Internet Engineering Task Force
I2RS working group
Internet Draft
Category: Informational

N. Bitar
Verizon
G. Heron
L. Fang
Microsoft
R. Krishnan
Brocade Communications
N. Leymann
Deutshe Telekom
H. Shah
Ciena
S. Chakrabarti
W. Haddad
Ericsson

Expires: August 2014 February 14, 2014

Interface to the Routing System (I2RS) for Service Chaining: Use Cases and Requirements

draft-bitar-i2rs-service-chaining-01

Abstract

Service chaining is the concept of applying an ordered set of services to a packet or a flow. Services in the chain may include network services such as load-balancing, firewalling, intrusion prevention, and routing among others. Criteria for applying a service chain to a packet or flow can be based on packet/flow attributes that span the OSI layers (e.g., physical port, Ethernet MAC header information, IP header information, transport, and application layer information). This document describes use cases and I2RS (Information to the Rousting System) requirements for the discovery and maintenance of services topology and resources. It also describes use cases and I2RS requirements for controlling the forwarding of a packet/flow along a service chain based on packet/flow attributes.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on August 14, 2014.

Copyright Notice

Conventions used in this document

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].

Table of Contents

1. Introduction

Several networking scenarios involve applying a set of services to a packet or flow. For instance, when a host in a protected zone initiates a session to a server outside the zone, the session may be directed to a chain of a Wide Area Network (WAN) application acceleration service, a network address and port translation (NAPT) service, and a firewall. On the server side, another set of services may also be applied. Such a sequence of services applied to a packet or flow is referred to as a service chain. Services in the chain may include deep packet inspection (DPI), load-balancing, firewalling, intrusion prevention, and routing among others.

Criteria for applying a service chain to a packet or flow can be based on packet/flow attributes that span the OSI layers. Such attributes may include the physical/virtual port on which the packet arrives, Ethernet MAC header information (e.g., VLAN ID), IP header information (e.g., source IP address), transport header information (e.g., TCP destination port number), and application layer information among others.

The transition from one service to the next in a service chain may be conditioned on the output of the current service, or may be non-conditional (pre-determined). A new mechanism, to be defined, may also enrich the packet transition in a service chain by passing service-specific information and/or information pertaining to preceding services in the chain along with the packet being processed. This type of mechanism and its influence are outside the scope of this document. In addition, this version of the document addresses the simple use case of pre-determined service chains applied to non-dropped packets with no additional information from preceding services. The service path for a packet/flow may be established via a management plane or routing, and may be enforced in the data plane via different mechanisms, as discussed in this document.

Services in a chain can be co-located on one system and/or physically separated across systems. In either case, a service may be running in its own virtualized system space or natively on the hosting system.

This document describes use cases and I2RS [i2rs-prob] requirements for the discovery and maintenance of services topology and resources. It also describes use cases and I2RS requirements for controlling the forwarding of a packet/flow along a service chain based on packet/flow attributes.

2. Abbreviations and Definitions

2.1. Abbreviations

2.2. Definitions

3. Service Chaining Use Cases and Requirements

A service chain is an ordered set of services applied to a packet or flow. It is often the case that when a flow in a bidirectional session is assigned to a service chain, the reverse flow of the same session is required to traverse the same chain in the reverse order. Assigning a flow to a service chain is often defined at an abstract level. Mapping a service chain to a network requires knowledge of the available services, their locations and available resources so that services are properly engineered on the services infrastructure. This section describes requirements and applicability for such information, and for directing traffic through a service chain.

3.1. Services topology

In order to establish a service chain that applies to a packet/flow, it is important to have a topology of the service nodes. A service node can be a service running natively within a system (e.g., a service card or a service engine in a router), a virtual machine (VM) hosted on a server, a VM hosted on a service engine within a system (e.g., a service card in a router), or a dedicated standalone service hardware appliance. In addition, a service node may be dedicated to a customer (e.g., an IPVPN customer), globally shared across customers or a customer set of VPNs, or available to be assigned in whole or in part to a customer or a set customer VPNs. A customer and tenant are used synonymously in this document. How a service node is created is outside the scope of this document. Resources on a service node that are not assigned to a customer context (e.g., VRF) will be logically referred to as a non-assigned service node with free available resources. A service node that can be shared in a global context will be referred to as a global service node. It should be noted, that once a service node is bound to a context, then it is only available for a virtual network (VN) associated with that context.

Different service node types may have information specific to the service(s) they provide. A service node information model needs to describe information common (generic) to all service node types and extensible to be sub-classed so that the service

specific information can be represented. The common information is:

. Service node address: A service node must have a unique address in a service topology. A service node identifier address can be:

- An IP address when feasible. Such a service node can be a VM, a services engine within a system, or a hardware appliance.

- The tuple (service node IP address, hosting system IP address). This applies when there is need to identify the system hosting the service node or when the service node IP address is only reachable within the hosting system.

- The tuple (hosting system IP-address, system internal identifier for the service engine). This applies when the service engine is not IP addressable and is within a system. A potential system internal identifier for a service engine may be (system_slot_number.subslot_number.engine_number).

. For each service node, the following information is required:

- Supported service type (e.g., NAT, FW). A node may support multiple service types.

- Number of virtual contexts (tenants) that can be supported. This parameter will indicate the maximum number of contexts that can be created on the service node.

- Number of virtual contexts (e.g., VRFs) available.

- Supported context type (e.g., VRF).

- Customer ID if the service node is dedicated to a customer. This indicates who can use this service node.

- List of supported (customer ID, virtual contexts). Note that one context per customer is a degenerate case. This will be the global context for a given customer on a service node.

For each service node, virtual context and service type, the following information may be specified, depending on the service resource requirement. That is, some of the information listed here may not be relevant for some services.

- Service bandwidth capacity
- Supported Packet rate (packets per second)
- Supported Bandwidth (e.g., in kbps)
- IP Forwarding Information Base size per address family
- Routing Information Base size
- MAC Forwarding database size
- Number of 64-bit statistics counters for policy-based accounting
- Number of supported Access lists (ACLs) per type (e.g., number of bits per ACL, and ACL type if applicable)
- Number of supported flows for services that require it (e.g., Firewall, NAT, stateful load-balancing, Deep Packet Inspection (DPI)) per flow type (i.e., fields identifying a flow) or flow identification key size. For systems that allow flexible memory usage across flow types and/or key sizes, it is sufficient to track available memory allocated for flows.

In addition to the services topology, it is important to have a view of the Virtual Network (VN) topology (VNT) and access points to which a services topology applies. The topology of such a VN could be relatively static, but it may also be dynamic, especially in a cloud environment where compute, storage, applications and associated networks may be created and removed over a short time scale. The description of a VN topology encompassing the access points is important in order to enable installation of policies for service chaining at the right access points, instantiate the services if needed, and perform the necessary monitoring as described in later sections. VN topology information requirements are described in [i2rs-topology-reqts], but they need to be augmented with the following information:

. Access ports (systems and ports) per VN. A port may be physical or logical on a physical port.

. Addresses reachable on an access port.

3.2. Monitoring Information

Service chaining requires the ability to monitor the state of each service node, including liveliness and resource utilization. If a service node failure is detected, an action may be taken to create another service node and steer traffic to it. If a service node is hitting a resource utilization threshold, traffic may be directed to other service nodes, and/or additional service nodes may be created.

The following is a set of parameters that needs be monitored per service node per virtual context, and per service type as applicable. It should be noted that some services may not require all the parameters listed here to be monitored.

. Bandwidth utilization (e.g., in kbps)

. Packet rate utilization (packets per second)

. Bandwidth utilization per CoS (e.g., in kbps)

. Packet rate utilization per Cos

. Memory utilization and available memory

. RIB utilization per address family

. FIB utilization per address family

. Flow resource utilization per flow type

. CPU utilization as applicable

. Available storage

The following is a set of parameters that needs to be monitored globally per physical system (e.g., host server) providing services or hosting service nodes. Note that some parameters may not be needed for some services:

. Bandwidth utilization (e.g., in kbps)

- Packet rate utilization (packets per second)
- Bandwidth utilization per Class of Service (CoS)
- Packet rate per CoS (packets per second)
- Memory utilization and available memory
- RIB utilization and available RIB memory if applicable per address family
- FIB utilization and available FIB entries if applicable per address family
- Flow resource utilization per flow type if applicable
- CPU utilization if applicable
- Power utilization
- Available storage

Such information needs to be maintained on the distributed system hosting a service node, and/or service node as applicable. In addition, a mechanism to monitor the liveliness of a service node must be available. For some use cases, liveliness and resource utilization information needs to be accessible to a management/control plane that provides for creation of service nodes and orchestration of service chains. Some of this information may also be maintained in the management/orchestration system and validated with the distributed system where the services are instantiated. For some other use cases, a service node and/or hosting system may need to be programmed to update a management system with that information periodically or when a configured high watermark or low watermark is reached for a parameter. Thus, the interface to the service nodes and/or hosting systems must provide a mechanism that enables a management/control system to pull resource utilization information from these nodes and systems, and for these nodes and system to send updates on resource utilization to a designated system.

3.3. Traffic Redirection, Forwarding and Service Chaining

In a service chain, it is important to be able to direct traffic from one service node to another. Some solutions may provide this capability via dynamic routing, data-plane based

policy-based routing, source based routing or a combination. Traffic redirection to a service chain requires the ability to program the routing system with a classification rule that identifies a packet/flow and an associated action that directs the corresponding packet(s) to the first node in the service chain. The focus in this section is on a hop-by-hop policy-based routing (PBR) and source based service routing. At the redirection point, classification rules MUST support the following information that encompasses Layer1-7 information, any of which may be wild-carded or left unspecified for a particular case:

- Port
- VLAN/VLAN stack
- MAC source address
- MAC destination address
- Host/subnet Source IP address
- Host/subnet Destination IP address
- IP version
- IP protocol
- Source port/port-range
- Destination port/port-range
- Optionally, application-layer information such as key words in a URI, content type or user agent

As a result of the classification, an action will need to be specified to direct the matching packet to a service node, or to perform other action(s). The following actions MUST be supported:

- Forward to a specified Outgoing port (physical or logical):
  - VLAN ID
  - IP/GRE tunnel

- o RSVP-TE tunnel
- o Pseudowire (PW)
- o Other types of tunneling protocols

. Steer the packet to a VRF

. Mirror packet:

- o To an IP destination
- o To a port
- o over a VLAN
- o over an IP/GRE tunnel
- o over an RSVP-TE tunnel
- o over a Pseudowire
- o over other types of tunneling

. Route. This could be the default behavior at the tail end of a chain or the result of no match.

. Route the packet to a specific system that is multiple IP hops away (Layer 3 policy based routing). The destination system IP address must be specified along with the tunneling type. The action must result in encapsulating the packet to the destination. At the destination, a policy must be installed to apply a service in a specific context to the arriving packet, or direct the traffic to a local service node.

. Insert a source route header in the transmitted packet that identifies the nodes along the service path. The service route may be composed of IPv4 routes, IPv6 routes and/or a stack of MPLS labels. The source route may capitalize on existing mechanism or new mechanisms that are outside the scope of this document. At the destination, a policy must be installed to apply a service in a specific context to the arriving packet, or direct the traffic to a local service node.

. Insert a source route+service header that identifies the service path and the service type to be applied at each node. This will require the definition of a new data plane header that carries such information.

The number of classification rules and associated actions, as well as the rate of programmability/removal of these rules will be highly application dependent. When the service chain is based on static policy (e.g., applied to a port, a source subnet, a VN), these rules will be programmed on a system at the rate of provisioning. When the attributes of the policies are relatively static (e.g., applied to a fixed port in fixed wireline access), the rate of provisioning on the forwarding system could be low, on the order of few hundred per day. When the attributes are more dynamic, such as in a mobile environment on a system handling a large number of users, that rate could be much higher. In a cloud environment where tenant systems may be spun up and removed on a relatively short time scale this rate could be on the order of few hundreds to thousands a minute at a DC GW for instance. In all cases, if the state is not kept in a persistent storage on the forwarding system(s), system reboot actions will trigger the need for a high provisioning rate, on the order at few thousands per second. When policies are triggered by data-plane, the rate of policy provisioning will be on the order of flow rates and removal will be dependent on the flow duration. These rates will be highly dependent on the applications as well, but at a system that is handling a large number of flows, the protocol used in provisioning must be very efficient to handle a very large number of flows.

4. Service Chaining via BGP-based Redirection

BGP-based steering of a traffic flow to a first service point may be required in certain cases. In this case, a router hosting a service node or connected to a service node will advertise a flow specification that causes a system that receives the advertisement to redirect a packet or mirror a copy of the packet that matches the flow specification to the advertising route [BGP-flowspec]. When the advertising router supports the i2rs BGP service, ann I2RS interface to the router can provision the router with the appropriate BGP policy as well as install on that router a forwarding policy that directs the packet when received to the appropriate service node. Such BGP advertisements can be chained to effect the chaining of multiple services.

5. Operational Considerations

6. IANA Considerations

There is IANA action required by this document.

7. Security Considerations

Service chaining imposes several security issues that must be addressed. First, the control system that installs policies on the routing system must be trusted by that system. An untrusted control system may install policies that hijack traffic, cause denial of service, or mirror traffic to an untrusted entity for eavesdropping. Thus the communication channel between a control system and routing system must be authenticated, and may be encrypted. In addition, when services are being offered to multiple VPN customers with overlapping IP addresses, it is important that the customer privacy is maintained when applying a service chain to a customer packet/flow. Thus, the ability to identify the context in which a service needs to be applied is important. In addition, policies must be installed in the appropriate context. Finally, congesting a service node can result in packet drops that may effectively result in a denial of service. Thus, obtaining information about the performance of a service node is important to detect overload conditions and take corrective action.

8. Acknowledgements

The authors thank David McDysan and Alia Atlas for their comments.

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[i2rs-prob] Atlas, A., Nadeau, T., and Ward, D., "Interface to the Routing System Problem Statement", draft-ietf-i2rs-problem-statement-00, August 2013. Work in progress.

[i2rs-topology-reqts] Medved, J., et al., "Topology API Requirements", draft-medved-i2rs-topology-requirements, February 2013. Work in progress.

[BGP-flowspec] Uttaro, J., et al., "BGP Flow-Spec Extended Community for Traffic Redirect to IP Next Hop", draft-simpson-idr-flowspec-redirect-02, November 2012. Work in Progress.

9.2. Informative References

Authors' Addresses

Nabil Bitar
Verizon
60 Sylvan Rd.
Waltham, MA 02145
EMail: nabil.n.bitar@verizon.com

Giles Heron
Cisco Systems
EMail: giheron@cisco.com

Luyuan Fang
Microsoft
EMail: luyuanf@gmail.com

Ram Krishnan
Brocade Communications
San Jose, CA 95134
EMail: ramk@brocade.com

Nicolai Leymann
Deutsche Telekom
Winterfeldtstrasse 21-27
10781 Berlin
Germany
EMail: n.leymann@telekom.de

Himanshu Shah
Ciena
EMail: hshah@ciena.com

Samita Chakrabatri
Ericsson
EMail: samita.chakrabarti@ericsson.com

Wassim Haddad

Ericsson
EMail: wassim.haddad@ericsson.com