Internet   N. Shen
Internet-Draft   C. Pignataro
Intended status: Standards Track   R. Asati
Expires: August 30, 2012   E. Chen
                                                            Cisco Systems
                                                                 A. Atlas
                                                         Juniper Networks
                                                        February 27, 2012

Traceroute and Ping Message Extension
draft-shen-traceroute-ping-ext-04

Abstract

This document specifies extensions to traceroute and ping techniques to convey additional application information to be carried in UDP, TCP and ICMP traceroute probe messages and ICMP echo request and reply messages. The extensions are backward compatible.

Table of Contents

1. Specification of Requirements

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 [RFC2119].

2. Introduction

Traceroute and Ping are two most commonly used tools created since the dawn of the Internet in the diagnosis of network problems. This document proposes the mechanism by which the traceroute probe messages and ICMP echo request/reply messages can be extended to include other user information various applications may want to include; and it can be optionally authenticated by the receiving node(s). These mechanisms are intended for network operators to perform more secured network management and troubleshooting tasks while using traceroute and ping tools. The changes proposed in this document are backward compatible (with the existing traceroute and ping tools) and applicable to both IPv4 and IPv6 networks.

The mechanisms specified in this document apply to to the following traceroute and ping probe protocols: UDP [RFC0768], TCP [RFC0793], and ICMP/ICMPv6 [RFC0792] [RFC4443]. This mechanism also applies to the ICMP/ICMPv6 echo reply messages [RFC0792].

This document defines an extension for traceroute and ping probe messages to optionally include authentication signature object. The intermediate and destination nodes can authenticate the sender of the traceroute or ping packet before providing the requested information in the ICMP response. This document also defines an optional Information-Request Object for the traceroute/ping extension. This Object specifies the types of information the sender expects to be included in the traceroute/ping response (i.e., in the ICMP message elicited by the traceroute/ping packet and generated by the intermediate or destination node or nodes).

Other applications can define their own Trace-Ping objects using this extension.

3. Motivation

The current traceroute or ping has no defined mechanism to include application data on the sender side, or to include application data in the ICMP echo reply on the receiver side. Although the [RFC4884] has defined the multi-part message extension in ICMP, it is applied only to the ICMP type 3, 11 and 12 for traceroute reply messages.

Those mechanisms are not applied to traceroute probe messages or ICMP echo request/reply messages.

For security concerns of traceroute or ping packets, one may employ a rudimentary control mechanism to limit the trusted senders by defining on every router the access control lists specifying source addresses of the traceroute and ping message, such mechanism is deemed configuration intensive, static, and error-prone. Moreover, such mechanism would be susceptible to address spoofing. Additionally, such mechanism does not provide the sender with dynamic control of the different kind of extensions to be requested.

The ICMP reply messages has been extended to support multi-part message inside ICMP [RFC4884] for some ICMP types. Some of the applications [RFC5837] [RFC4950] [I-D.shen-icmp-routing-inst] are designed mainly for internal network troubleshooting by network operators. Network providers may want to limit those applications only to trusted senders of traceroute/ping probes due to security or policy reasons by using this mechanism described in this document.

Other applications, for example the TRILL-OAM [I-D.tissa-trill-oam] can use this scheme to extend their OAM application using ICMP echo request and reply for data center troubleshootings.

4. Trace-Ping Message Extension

This proposed extension is to define a Trace-Ping data structure that starts at a fixed location (i.e. the 64 octet) in the UDP/TCP/ICMP probe packet data field.

4.1. Trace-Ping Extension Structure

The Trace-Ping structure starts in UDP/TCP/ICMP data field location 64th octet, see Section 4.2. It MUST have exactly one Trace-Ping common header followed by zero or more Trace-Ping Objects.

4.1.1. Trace-Ping Common Header

The Common Header is a 8 octets structure has the following format:

```
      0                   1                   2                   3
      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |Version|        Length         |           Checksum            |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |                  Magic-Number (0x54726163)                    |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The fields of the Common Header are defined as follows:

Version: 4 bits. It is defined as 1 in this document.

Length: 12 bits. The total length of the Trace-Ping data structure specifying number of 32-bit words (includes the common header and all the Objects).

Checksum: 16 bits. The one's complement of the one's complement sum of the Trace-Ping data structure, with the checksum field replaced by zero for the purpose of computing the checksum.

Magic Number: 32 bits. It is defined as Hex value of 0x54726163 in this document. This is used mainly for structure identification of this extension version.

4.1.2. **Trace-Ping Object**

Trace-Ping Object have the following format:

```
      0                   1                   2                   3
      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |             Length            |   Class-Num   |     C-Type    |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     .                                                               .
     .                   // (object payload) //                      .
     .                                                               .
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Length: 16 bits. Length of object, measured in octets, including the object header and object payload.

Class-Num: 8 bits. Identifies ICMP Trace-Ping object class.

C-Type: 8 bits. Identifies ICMP Trace-Ping object sub-type.

All the Trace-Ping Objects are optional. This document defines two Trace-Ping Objects below.

4.1.2.1. Trace-Ping Authentication Object

This Object carries the HMAC authentication related information. It verifies both the data integrity and the authenticity of the entire message. This Object has the following format:

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |             Length            |   Class-Num   |     C-Type    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |     Auth Type                 |    Key ID     | Auth Data Len |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   |                     Auth Data (Variable)                      |
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+ ...
```

Length: Variable, in octets.

Class-Num: IANA allocation from ICMP Trace-Ping extension registry.

C-Type: 1

Auth Type: 16 bits. The following values are proposed:

* Type=0 signifies no authentication.

* Type=1 signifies simple password based authentication.

* Type=2 signifies Cryptographic authentication.

Please note that the above type values are in line with IANA allocated values for other protocols (e.g., OSPF).

Key ID: 8 bits. This allows multiple secret keys to be active simultaneously. Using Key IDs makes the key rollover convenient. Each secret key must be associated with the hash algorithm. This may be done through provisioning on each node.

Auth Data Len: 8 bits. This specifies the length of the authentication data (and allows for the support of current and future authentication schemes).

Auth Data: Variable length. This field carries the result (e.g., HMAC code) of the HMAC algorithm applied over the entire traceroute/ping IP/IPv6 packet. When the Auth data is calculated, the shared key is stored in this field, and the checksum fields in the IP header, UDP/TCP/ICMP header and Trace-Ping common header are set to zero. The result of the algorithm is placed in the Auth Key field. The following lists algorithms that could be commonly supported:

* HMAC-MD5

* HMAC-SHA1

* HMAC-SHA2 variants (e.g., 224, 256, 384, 512, etc.)

At least HMAC-MD5 and HMAC-SHA1 algorithms should be supported on all the nodes compliant with this specification.

4.1.2.2. Trace-Ping Information-Request Object

This Information-Request Object is defined using a bitmap of 32-bits field to represent an array of attributes. The attribute information can be referenced in [RFC4950] [RFC5837] [I-D.shen-icmp-routing-inst]. If detailed information needs to be specified, new objects will have to be defined and it is outside the scope of this document.

The Information-Request Object has the following format:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Length            |   Class-Num   |     C-Type    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Info Request                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Length: 8

Class-Num: IANA allocation, the same Class-Num value as in Section 4.1.2.1.

C-Type: 2

Info-Req: 32 bits. This bitflag field lists the request items the probe sender is interested. The bit number ranges from the right most bit to the left most bit. Currently defined as the following:

Bit Number	Information Item
0	MPLS label related attributes
1	Interface related attributes
2	IP/IPv6 address related attributes
3	Routing Instance related attributes
4	Nexthop(s) related attributes

4.2. Trace-Ping Extension Offset

The Trace-Ping Extension data structure starts at the fixed location of 64th octet inside UDP, TCP and ICMP data field. The first 64 octets data is not defined and can be used by the probe packet sender.

5. Trace-Ping Port Number

The Trace-Ping Port SHOULD be used for the UDP destination port, TCP destination port or the ICMP echo request Identifier field with this Trace-Ping extension. If the implementation uses the port field for the packet sequence purposes, then the sequence information can be written in the private space in the first 64 octets of the data field in probe packets.

When the UDP or TCP, either in Traceroute or Ping operation, packet reaches the destination, the host or router will return the ICMP DESTINATION UNREACHABLE message back to the sender.

6. Scaling Considerations on Internet

Although this extension allows new features easily being developed on top of the existing and popular Traceroute and Ping applications, it does create challenges on the Internet as how to distinguish the regular Traceroute and Ping packets from the new feature usages without incurring rather substantial resource overhead. Steps need to be taken on both implementation and operational sides.

6.1. Implementation and Operation Considerations

Implementation of this extension SHOULD use configuration knobs to enable the new features on the device and leave the standard behavior of Traceroute and Ping treatment if the explicit configuration for

this extension is not present.

The probe sender SHOULD use the Trace-Ping Port in their UDP and TCP Traceroute or Ping packets when using this extension; and the probe sender SHOULD use the Trace-Ping Port in the Identifier of ICMP echo request packet. This will allow the receiver side to easily identify the new features the network wants to support.

Implementation SHOULD allow filters or access-list mechanism to be attached to this extension configurations. For example, the checking or verifying the existence of this extension in the probe packets is only performed when the probe packet is sourced from certain network prefix range. Different features using this extension MAY have different filters or access-lists.

Although this extension allows Traceroute and Ping packets to be rate-limited just as the regular packets, the implementation SHOULD apply special rate-limit if the feature is configured. This special rate-limit SHOULD be configurable due to the nature of the features, the device resource consumption of the features and the handling of DoS attacks. The default special rate-limit SHOULD not exceeds the rate-limit of regular Traceroute and Ping operations on the device.

On the prober packet or the sender side, implementation SHOULD allow specifying the requested information, thus only a subset of the regular objects need to be included in the replying ICMP packets when the receiver is configured to support this feature.

7. Security Considerations

This extension enhances the security of traceroute and ping operation in a backwards-compatible fashion. The mechanism allows the receiver to verify the sender of the traceroute/ping packet such that certain sensitive application, interface and network related information can be supplied in the internal network or across trusted networks.

The use of Cryptographic authentication (i.e., an Auth Type value of 2) allows for a strong authentication mechanism since the keys cannot be discerned by intercepting the packets. The proposed Keyed authentication does not prevent replay attacks. However, in the case of replay attacks, since the packet source IP/IPv6 address of the traceroute/ping probe can not be changed, there is no easy way for the attacker to retrieve the ICMP messages.

A router needs to protect against purposefully-bogus Traceroute packets with extensions that fail the authentication, as a high rate of messages can require significant processing time. [RFC1812]

specifies how rate-limiting is applied to the generation of ICMP messages, and this rate-limiting deters the threat when applied before checking the Authentication. Additionally, when using Cryptographic authentication, the HMAC includes the source IP address, which means the HMAC will not validate if the traceroute/ ping packet is sent over a NAT.

8. IANA Considerations

The IANA is requested to assign a well-known port number, Trace-Ping Port, for the UDP and TCP of this Trace-Ping extensions.

The IANA is also requested to allocate the same Trace-Ping Port to be used for the Identifier in the ICMP Echo Request with this Trace-Ping extensions.

The Trace-Ping Extension contains Trace-Ping Objects. IANA is requested to assign a new Class-Num for the Trace-Ping extension, and a sub-registry under Trace-Ping extension to include c-types. This document has defined c-type 1 and 2 for authentication and information-request objects. c-types 3-0xF6 are allocated through Expert Review [RFC5226]. C-types 0xF7 to 0xFF are reserved for private use.

IANA should also establish a registry for Trace-Ping Info-Request Bits under the information-request sub-registry. This document defines bits 0 - 5 in Section 4.1.2.2. Bits 6-29 are allocated through Expert Review. Bits 30 - 31 are reserved for private use.

9. Acknowledgements

Many thanks to Dan Wing, Tony Li, and Tissa Senevirathne for their insightful comments and valuable suggestions regarding this document. Many thanks to Ron Bonica, Thomas Narten, Jared Mauch, Warren Kumari, Wes George, Shane Amante, S. Moonesamy who have made operational and design comments and suggestions in particular to the scaling issues on the Internet.

10. References

10.1. Normative References

[RFC0768] Postel, J., "User Datagram Protocol", STD 6, RFC 768, August 1980.

[RFC0792] Postel, J., "Internet Control Message Protocol", STD 5, RFC 792, September 1981.

[RFC0793] Postel, J., "Transmission Control Protocol", STD 7, RFC 793, September 1981.

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

10.2. Informative References

[I-D.shen-icmp-routing-inst] Shen, N. and E. Chen, "ICMP Extensions for Routing Instances", draft-shen-icmp-routing-inst-00 (work in progress), November 2006.

[I-D.tissa-trill-oam] Senevirathne, T., Dutt, D., Manral, V., and S. Aldrin, "ICMP based OAM Solution for TRILL", draft-tissa-trill-oam-03 (work in progress), January 2012.

[RFC1812] Baker, F., "Requirements for IP Version 4 Routers", RFC 1812, June 1995.

[RFC4443] Conta, A., Deering, S., and M. Gupta, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", RFC 4443, March 2006.

[RFC4884] Bonica, R., Gan, D., Tappan, D., and C. Pignataro, "Extended ICMP to Support Multi-Part Messages", RFC 4884, April 2007.

[RFC4950] Bonica, R., Gan, D., Tappan, D., and C. Pignataro, "ICMP Extensions for Multiprotocol Label Switching", RFC 4950, August 2007.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

[RFC5837] Atlas, A., Bonica, R., Pignataro, C., Shen, N., and JR. Rivers, "Extending ICMP for Interface and Next-Hop Identification", RFC 5837, April 2010.

Authors' Addresses

Naiming Shen
Cisco Systems
225 West Tasman Drive
San Jose, CA 95134
USA

Email: naiming@cisco.com

Carlos Pignataro
Cisco Systems
7200 Kit Creek Road
Research Triangle Park, NC 27709
USA

Email: cpignata@cisco.com

Rajiv Asati
Cisco Systems
7025 Kit Creek Road
Research Triangle Park, NC 27709
USA

Email: rajiva@cisco.com

Enke Chen
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
USA

Email: enkechen@cisco.com

Alia K. Atlas
Juniper Networks
10 Technology Park Drive
Westford, MA 01886
USA

Email: akatlas@juniper.net