Internet Engineering Task Force (IETF)
Request for Comments: 6874
Updates: 3986
Category: Standards Track
ISSN: 2070-1721

B. Carpenter
Univ. of Auckland
S. Cheshire
Apple Inc.
R. Hinden
Check Point

February 2013

Representing IPv6 Zone Identifiers in Address Literals and Uniform Resource Identifiers

Abstract

This document describes how the zone identifier of an IPv6 scoped address, defined as <zone_id> in the IPv6 Scoped Address Architecture (RFC 4007), can be represented in a literal IPv6 address and in a Uniform Resource Identifier that includes such a literal address. It updates the URI Generic Syntax specification (RFC 3986) accordingly.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc6874.

Table of Contents

1. Introduction

The Uniform Resource Identifier (URI) syntax specification [RFC3986] defined how a literal IPv6 address can be represented in the "host" part of a URI. Two months later, the IPv6 Scoped Address Architecture specification [RFC4007] extended the text representation of limited-scope IPv6 addresses such that a zone identifier may be concatenated to a literal address, for purposes described in that specification. Zone identifiers are especially useful in contexts in which literal addresses are typically used, for example, during fault diagnosis, when it may be essential to specify which interface is used for sending to a link-local address. It should be noted that zone identifiers have purely local meaning within the node in which they are defined, often being the same as IPv6 interface names. They are completely meaningless for any other node. Today, they are meaningful only when attached to addresses with less than global scope, but it is possible that other uses might be defined in the future.

The IPv6 Scoped Address Architecture specification [RFC4007] does not specify how zone identifiers are to be represented in URIs. Practical experience has shown that this feature is useful, in particular when using a web browser for debugging with link-local addresses, but because it is undefined, it is not implemented consistently in URI parsers or in browsers.

Some versions of some browsers directly accept the IPv6 Scoped Address syntax [RFC4007] for scoped IPv6 addresses embedded in URIs, i.e., they have been coded to interpret a "%" sign following the literal address as introducing a zone identifier [RFC4007], instead of introducing two hexadecimal characters representing some percent-encoded octet [RFC3986]. Clearly, interpreting the "%" sign as introducing a zone identifier is very convenient for users, although it formally breaches the established URI syntax [RFC3986]. This

document defines an alternative approach that respects and extends the rules of URI syntax, and IPv6 literals in general, to be consistent.

Thus, this document updates the URI syntax specification [RFC3986] by adding syntax to allow a zone identifier to be included in a literal IPv6 address within a URI.

It should be noted that in contexts other than a user interface, a zone identifier is mapped into a numeric zone index or interface number. The MIB textual convention InetZoneIndex [RFC4001] and the socket interface [RFC3493] define this as a 32-bit unsigned integer. The mapping between the human-readable zone identifier string and the numeric value is a host-specific function that varies between operating systems. The present document is concerned only with the human-readable string.

Several alternative solutions were considered while this document was developed. Appendix A briefly describes the various options and their advantages and disadvantages.

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 "Key words for use in RFCs to Indicate Requirement Levels" [RFC2119].

2. Specification

According to IPv6 Scoped Address syntax [RFC4007], a zone identifier is attached to the textual representation of an IPv6 address by concatenating "%" followed by <zone_id>, where <zone_id> is a string identifying the zone of the address. However, the IPv6 Scoped Address Architecture specification gives no precise definition of the character set allowed in <zone_id>. There are no rules or de facto standards for this. For example, the first Ethernet interface in a host might be called %0, %1, %en1, %eth0, or whatever the implementer happened to choose.

In a URI, a literal IPv6 address is always embedded between "[" and "]". This document specifies how a <zone_id> can be appended to the address. According to URI syntax [RFC3986], "%" is always treated as an escape character in a URI, so, according to the established URI syntax [RFC3986] any occurrences of literal "%" symbols in a URI MUST be percent-encoded and represented in the form "%25". Thus, the scoped address fe80::a%en1 would appear in a URI as http://[fe80::a%25en1].

A <zone_id> SHOULD contain only ASCII characters classified as "unreserved" for use in URIs [RFC3986]. This excludes characters such as "]" or even "%" that would complicate parsing. However, the syntax described below does allow such characters to be percent-encoded, for compatibility with existing devices that use them.

If an operating system uses any other characters in zone or interface identifiers that are not in the "unreserved" character set, they MUST be represented using percent encoding [RFC3986].

We now present the necessary formal syntax.

The URI syntax specification [RFC3986] formally defined the IPv6 literal format in ABNF [RFC5234] by the following rule:

```
IP-literal = "[" ( IPv6address / IPvFuture  ) "]"
```

To provide support for a zone identifier, the existing syntax of IPv6address is retained, and a zone identifier may be added optionally to any literal address. This syntax allows flexibility for unknown future uses. The rule quoted above from the previous URI syntax specification [RFC3986] is replaced by three rules:

```
IP-literal = "[" ( IPv6address / IPv6addrz / IPvFuture  ) "]"

ZoneID = 1*( unreserved / pct-encoded )

IPv6addrz = IPv6address "%25" ZoneID
```

This syntax fills the gap that is described at the end of Section 11.7 of the IPv6 Scoped Address Architecture specification [RFC4007].

The established rules for textual representation of IPv6 addresses [RFC5952] SHOULD be applied in producing URIs.

The URI syntax specification [RFC3986] states that URIs have a global scope, but that in some cases their interpretation depends on the end-user’s context. URIs including a ZoneID are to be interpreted only in the context of the host at which they originate, since the ZoneID is of local significance only.

The IPv6 Scoped Address Architecture specification [RFC4007] offers guidance on how the ZoneID affects interface/address selection inside the IPv6 stack. Note that the behaviour of an IPv6 stack, if it is passed a non-null zone index for an address other than link-local, is undefined.

3. Web Browsers

This section discusses how web browsers might handle this syntax extension. Unfortunately, there is no formal distinction between the syntax allowed in a browser's input dialogue box and the syntax allowed in URIs. For this reason, no normative statements are made in this section.

Due to the lack of defined syntax, web browsers have been inconsistent in providing for ZoneIDs. Many have no support, but there are examples of ad hoc support. For example, some versions of Firefox allowed the use of a ZoneID preceded by a bare "%" character, but this feature was removed for consistency with established syntax [RFC3986]. As another example, some versions of Internet Explorer allow use of a ZoneID preceded by a "%" character encoded as "%25", still beyond the syntax allowed by the established rules [RFC3986]. This syntax extension is in fact used internally in the Windows operating system and some of its APIs.

It is desirable for all browsers to recognise a ZoneID preceded by a percent-encoded "%". In the spirit of "be liberal with what you accept", we also suggest that URI parsers accept bare "%" signs when possible (i.e., a "%" not followed by two valid and meaningful hexadecimal characters). This would make it possible for a user to copy and paste a string such as "fe80::a%en1" from the output of a "ping" command and have it work. On the other hand, "%ee1" would need to be manually rewritten to "fe80::a%25ee1" to avoid any risk of misinterpretation.

Such bare "%" signs are for user interface convenience, and need to be turned into properly encoded characters (where "%25" encodes "%") before the URI is used in any protocol or HTML document. However, URIs including a ZoneID have no meaning outside the originating node. It would therefore be highly desirable for a browser to remove the ZoneID from a URI before including that URI in an HTTP request.

The normal diagnostic usage for the ZoneID syntax will cause it to be entered in the browser's input dialogue box. Thus, URIs including a ZoneID are unlikely to be encountered in HTML documents. However, if they do (for example, in a diagnostic script coded in HTML), it would be appropriate to treat them exactly as above.

4. Security Considerations

The security considerations from the URI syntax specification [RFC3986] and the IPv6 Scoped Address Architecture specification [RFC4007] apply. In particular, this URI format creates a specific pathway by which a deceitful zone index might be communicated, as mentioned in the final security consideration of the Scoped Address Architecture specification. It is emphasised that the format is intended only for debugging purposes, but of course this intention does not prevent misuse.

To limit this risk, implementations MUST NOT allow use of this format except for well-defined usages, such as sending to link-local addresses under prefix fe80::/10. At the time of writing, this is the only well-defined usage known.

An HTTP client, proxy, or other intermediary MUST remove any ZoneID attached to an outgoing URI, as it has only local significance at the sending host.

5. Acknowledgements

The lack of this format was first pointed out by Margaret Wasserman some years ago, and more recently by Kerry Lynn. A previous draft document by Martin Duerst and Bill Fenner [LITERAL-ZONE] discussed this topic but was not finalised.

Valuable comments and contributions were made by Karl Auer, Carsten Bormann, Benoit Claise, Stephen Farrell, Brian Haberman, Ted Hardie, Tatuya Jinmei, Yves Lafon, Barry Leiba, Radia Perlman, Tom Petch, Tomoyuki Sahara, Juergen Schoenwaelder, Dave Thaler, Martin Thomson, and Ole Troan.

Brian Carpenter was a visitor at the Computer Laboratory, Cambridge University during part of this work.

6. References

6.1. Normative References

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

[RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, RFC 3986, January 2005.

[RFC4007] Deering, S., Haberman, B., Jinmei, T., Nordmark, E., and B. Zill, "IPv6 Scoped Address Architecture", RFC 4007, March 2005.

[RFC5234] Crocker, D., Ed. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, RFC 5234, January 2008.

[RFC5952] Kawamura, S. and M. Kawashima, "A Recommendation for IPv6 Address Text Representation", RFC 5952, August 2010.

6.2. Informative References

[LITERAL-ZONE] Fenner, B. and M. Duerst, "Formats for IPv6 Scope Zone Identifiers in Literal Address Formats", Work in Progress, October 2005.

[RFC3493] Gilligan, R., Thomson, S., Bound, J., McCann, J., and W. Stevens, "Basic Socket Interface Extensions for IPv6", RFC 3493, February 2003.

[RFC4001] Daniele, M., Haberman, B., Routhier, S., and J. Schoenwaelder, "Textual Conventions for Internet Network Addresses", RFC 4001, February 2005.

Appendix A. Options Considered

The syntax defined above allows a ZoneID to be added to any IPv6 address. The 6man WG discussed and rejected an alternative in which the existing syntax of IPv6address would be extended by an option to add the ZoneID only for the case of link-local addresses. It was felt that the solution presented in this document offers more flexibility for future uses and is more straightforward to implement.

The various syntax options considered are now briefly described.

1. Leave the problem unsolved.

   This would mean that per-interface diagnostics would still have to be performed using ping or ping6:

   ping fe80::a%en1

   Advantage: works today.

   Disadvantage: less convenient than using a browser.

2. Simply use the percent character:

   http://[fe80::a%en1]

   Advantage: allows use of browser; allows cut and paste.

   Disadvantage: invalid syntax under RFC 3986; not acceptable to URI community.

3. Simply use an alternative separator:

   http://[fe80::a-en1]

   Advantage: allows use of browser; simple syntax.

   Disadvantage: Requires all IPv6 address literal parsers and generators to be updated in order to allow simple cut and paste; inconsistent with existing tools and practice.

   Note: The initial proposal for this choice was to use an underscore as the separator, but it was noted that this becomes effectively invisible when a user interface automatically underlines URLs.

4. Simply use the "IPvFuture" syntax left open in RFC 3986:

   http://[v6.fe80::a_en1]

   Advantage: allows use of browser.

   Disadvantage: ugly and redundant; doesn’t allow simple cut and paste.

5. Retain the percent character already specified for introducing zone identifiers for IPv6 Scoped Addresses [RFC4007], and then percent-encode it when it appears in a URI, according to the already-established URI syntax rules [RFC 3986]:

   http://[fe80::a%25en1]

   Advantage: allows use of browser; consistent with general URI syntax.

   Disadvantage: somewhat ugly and confusing; doesn’t allow simple cut and paste.

   This is the option chosen for standardisation.

Authors' Addresses

Brian Carpenter
Department of Computer Science
University of Auckland
PB 92019
Auckland, 1142
New Zealand

EMail: brian.e.carpenter@gmail.com

Stuart Cheshire
Apple Inc.
1 Infinite Loop
Cupertino, CA 95014
United States

EMail: cheshire@apple.com

Robert M. Hinden
Check Point Software Technologies, Inc.
800 Bridge Parkway
Redwood City, CA 94065
United States

EMail: bob.hinden@gmail.com