Workgroup: 6MAN
Internet-Draft: draft-ietf-6man-rfc6874bis-09
Obsoletes: 6874 (if approved)
Updates: 3986, 3987 (if approved)
Published: 2 July 2023
Intended Status: Standards Track
Expires: 3 January 2024

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

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 and Internationalized Resource Identifier specifications (RFC 3986, RFC 3987) accordingly, and obsoletes RFC 6874.

Discussion Venue

This note is to be removed before publishing as an RFC.

Discussion of this document takes place on the 6MAN mailing list (ipv6@ietf.org), which is archived at https://mailarchive.ietf.org/arch/browse/ipv6/.

Status of This Memo

Copyright Notice

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. 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, usually being the same as IPv6 interface names. They are completely meaningless for any other node. Today, they are meaningful only when attached to link-local addresses, but it is possible that other uses might be defined in the future.

The IPv6 Scoped Address Architecture specification does not specify how zone identifiers are to be represented in URIs. Practical experience has shown that this feature is necessary in various use cases, including the following:

1. A web browser may be used for simple debugging actions involving link-local addresses on a host with more than one active link interface. For example, the existence of a device may today be checked via "ping fe80::1234%eth0" but not via "https://[fe80::1234%eth0]".

2. A web browser must sometimes be used to configure or reconfigure a device which only has a link-local address and whose only configuration tool is a web server, again in a host with more than one active link interface. For example, a typical home router may today be configured via "http://192.168.178.1" but not via "http://[fe80::1%eth0]".

3. The Apple and open-source CUPS printing mechanism [CUPS] [OP-CUPS] uses an HTTP-based protocol [RFC3510][RFC7472] to establish link-local relationships, so requires the specification of the relevant interface.

4. The Microsoft Web Services for Devices (WSD) virtual printer port mechanism can generate an IPv6 link-local URL such as "http://[fe80::823b:f9ff:fe7b:d9dc%10]:80/WebServices/Device" in which the zone identifier is present, but is not recognized by any current browser.

5. The National Marine Electronics Association (NMEA) has recently defined its "OneNet Marine IPv6 Ethernet Networking Standard" [ONE-NET], which includes a specific requirement for device configuration via a browser using link-local addresses. Such requirements have already spawned a hack to work around the current limitation [LL-HACK].

For these use cases, it is highly desirable that a complete IPv6 link-local address can be cut and pasted from one context (such as the output from a system command) to another (such as a browser dialogue box). Since such addresses may include quite long hexadecimal strings, any solution except cut-and-paste is highly error prone.

The use cases listed above apply to relatively simple actions on end systems. The zone identifiers that can be used are limited by the character set allowed in URIs. In particular, upper case letters and most non-alphanumeric characters are intrinsically problematic in the host part of a URI. This is not an issue on typical end systems, which generally use lower case alphanumeric interface names, but it is likely to arise, for example, in network infrastructure devices. These may have large numbers of interfaces, which are commonly named for network management purposes in styles such as "Ethernet1/0/1" or "ge-0/0/0.0", reflecting the hardware structure and depending on the manufacturer. Generally speaking, such names are handled by various network management mechanisms and specialized commands, and do not need to be included in URIs. Nevertheless, we describe below how an interface name containing non-conforming characters can be replaced by a numeric value in case it is needed in a URI.

For avoidance of doubt, devices whose network stack does not support the RFC 4007 model of a readable Zone ID plus a numeric index are out of scope for this document.

As IPv6 deployment becomes widespread, the lack of a solution for handling complete link-local addresses in web browsers is becoming an acute problem for increasing numbers of operational and support personnel. It will become critical as IPv6-only networks, with no native IPv4 support, appear. For example, the NMEA use case mentioned above is an immediate requirement. This is the principal reason for documenting this requirement and its solution now.

It should be noted that whereas some operating systems and network APIs support a default zone identifier as recommended by the IPv6 scoped address architecture [RFC4007], others do not, and for them an appropriate URI syntax is particularly important.

In the past, some browser versions directly accepted the IPv6 Scoped Address syntax for scoped IPv6 addresses embedded in URIs, i.e., they were coded to interpret a "%" sign following the literal address as introducing a zone identifier, instead of introducing two hexadecimal characters representing some percent-encoded octet as explained in Section 2.1 of [RFC3986]. Clearly, interpreting the "%" sign as introducing a zone identifier is very convenient for users, although it is not supported by the URI syntax in RFC 3986 or the Internationalized Resource Identifier (IRI) syntax in [RFC3987]. Therefore, this document updates RFC 3986 and RFC 3987 by adding syntax to allow a zone identifier to be included in a literal IPv6 address within a URI.

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. (However, note that interface numbers are limited to positive signed 32-bit integers (see InterfaceIndex defined in [RFC2863] and if-index defined in [RFC8343]) while the zone index allows for unsigned 32-bit integers.)

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 that is typically displayed in an operating system's user interface. However, in most operating systems it is possible to use the underlying interface number, represented as a decimal integer, as an equivalent to the human-readable string. This is recommended by Section 11.2 of RFC 4007, but not required. This provides a solution for cases where the assigned zone identifier uses characters not allowed in a URI. The user must find the interface number corresponding to the displayed interface name. For example, on Linux, a user can determine interface numbers by issuing the command "ip link show" and then use "fe80::1%5" instead of "fe80::1%Ethernet+0+1", if the interface number happens to be 5. In such operating systems, the decimal integer can be used in a URI in place of the zone identifier, although this does not allow cut-and-paste of the human-readable identifier.

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

This document obsoletes its predecessor [RFC6874] by greatly simplifying its recommendations and requirements for URI parsers. Its effect on the formal URI syntax [RFC3986] is different from that of RFC 6874.

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.

2. Issues with Implementing RFC 6874

Several issues prevented RFC 6874 being implemented in browsers:

1. There was some disagreement with requiring percent-encoding of the "%" sign preceding a zone identifier. This requirement is dropped in the present document.

2. The requirement to delete any zone identifier before emitting a URI from the host in an HTTP message was considered both too complex to implement and in violation of normal HTTP practice [RFC9110], although required by Section 11.2 of RFC 4007. This requirement has been dropped from the present document.

3. The suggestion to pragmatically allow a bare "%" sign when this would be unambiguous was considered both too complex to

implement and confusing for users. This suggestion has been dropped from the present document since it is now irrelevant.

3. Specification

According to the 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, %25, %en1, %eth0, or whatever the implementer happened to choose.

This lack of precision leads to two specific difficulties when set against the general rules for the host subcomponent of a URI [RFC3986]:

1. The URI host component is case-insensitive. RFC 4007 implies case sensitivity.

2. The URI host component must be composed from a specific character set. RFC 4007 simply requires an ASCII string.

The syntax specified below clarifies these two items.

In a URI, a literal IPv6 address is always embedded between "[" and "]". This document specifies how a zone identifier can be appended to the address. The URI syntax defined by RFC 3986 does not allow the presence of a percent ("%") character within an IPv6 address literal. For this reason, it is backwards compatible to allow the use of "%" within an IPv6 address literal as a delimiter only, such that the scoped address "fe80::abcd%en1" would appear in a URI as "http://[fe80::abcd%en1]" or "https://[fe80::abcd%en1]".

This use of "%" as a delimiter applies only within an IPv6 address literal, and is irrelevant to and exempt from the percent-encoding mechanism of RFC 3986.

A zone identifier used in a URI MUST contain only ASCII characters classified as "unreserved" for use in URIs by RFC 3986. This excludes characters such as "/", "]" or even "%" that would complicate parsing. For the avoidance of doubt, note that a zone identifier consisting of "25" or starting with "25" is valid and is used in some operating systems. A parser MUST NOT apply percent decoding to the IPv6 address literal in a URI, including cases such as "http://[fe80::abcd%25]" and "http://[fe80::abcd%25xy]".

If an operating system uses any characters in zone or interface identifiers that are not in the "unreserved" character set, identifiers including them cannot be used in a URI.

Section 6.2.2.1 of RFC 3986 states unambiguously that "the scheme and host are case-insensitive and therefore should be normalized to lowercase". Therefore, even if an operating system supports case-sensitive zone or interface identifiers, such identifiers including upper case letters cannot be used in the host component of a URI, because they will be incorrectly converted to lower case.

We now present the corresponding formal syntax.

The URI syntax specification in RFC 3986 formally defines 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 RFC 3986 is replaced by four rules:

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

ZoneID = 1*( lc-unreserved )

lc-unreserved = %x61-7A / DIGIT / "-" / "." / "_" / "~"

IPv6addrz = IPv6address "%" ZoneID
```

Note that this change restricts the character set left open by RFC 4007, and because of the lower case issue it restricts the "unreserved" character set of RFC 3986.

This ABNF change also applies to [RFC3987].

This syntax fills the gap that is described at the end of Section 11.7 of the IPv6 Scoped Address Architecture specification [RFC4007]. It replaces and obsoletes the syntax in Section 2 of [RFC6874].

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

RFC 3986 states that URIs have a global scope, but that in some cases their interpretation depends on the end-user's context. URIs including a zone identifier are an example of this, since the zone identifier is of local significance only. Such a zone identifier cannot be correctly interpreted outside the host to which it applies, so it must be treated as an opaque string.

When defining zone identifiers compatible with RFC 4007, it is RECOMMENDED to use only lower case letters, digits, and the symbols "-", ".", "_" or "~", in order to also be compatible with URI syntax. In case this recommendation is not adopted, an implementation SHOULD follow the recommendation in Section 11.2 of RFC 4007 to support numeric identifiers.

RFC 4007 offers guidance on how the zone identifier 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.

In cases where the RFC 6874 encoding is currently used between specific software components rather than between a browser and a web server, such usage MAY continue indefinitely.

4. Scope and Deployment

A URI (or IRI) using this format has no meaning outside the scope of the individual host that originates it and of the specific layer 2 link concerned. It may in fact be delivered in an HTTP message to a server that does not support this format and which will reject the message as invalid. For the diagnostic use cases concerned, this is of no importance: an HTTP error response will serve the diagnostic purpose of establishing that the link and remote host are operational. The other use cases shown above are only meaningful if the remote host also accepts this format; otherwise they will fail with an HTTP error response. As a result, this format can be deployed progressively as required, with no wider consequences.

It is worth noting that there is nothing new about a URI that refers to a local resource. URIs referring to local domains under ".local" are normal. Any URI such as "https://169.254.0.1" (link-local IPv4, [RFC3927]), "https://10.1.1.1" (private IPv4, [RFC1918]), or "https://[fd63:45eb:cd14:0:80b2:5c79:62ae:d341]" (IPv6 unique local address, [RFC4193]) refers to a local resource and has no meaning off the link or outside the local domain. In operating systems with support for a default zone identifier, URLs such as "https://[fe80::2e3a:12cd:fea4:dde7]" already work as expected. Deployment of support for link-local IPv6 addresses with zone identifiers introduces no new principle compared to these currently operational examples.

There has been considerable concern about potential security concerns caused by locally scoped URIs. A recent W3C Community Group draft report [PNA-REP] provides background on the issue of cross-origin resource sharing (CORS), a mechanism which "allows a server to indicate any origins (domain, scheme, or port) other than its own from which a browser should permit loading resources." This mechanism was originally devised for the case of private IPv4 addresses, but has been expanded to cover other cases, explicitly including link-local IPv6 addresses. Addresses are sorted into three scopes: loopback, local and public. It could be argued that link-local addresses which include a zone identifier should be treated on the same basis as a loopback address, since they are meaningless outside the originating host (see Section 11.2 of [RFC4007]). In any case, link-local addresses can clearly be handled by the CORS mechanism, regardless of the presence or absence of a zone identifier. To respect the general prohibition on transmitting zone identifiers in Section 11.2 of RFC 4007, CORS can ensure that they are not processed by the receiving node.

5. URI Parsers

This section discusses how URI (or IRI) parsers, such as those embedded in web browsers, might handle this syntax extension.

In practice, although parsers respect the established syntax, many are coded pragmatically rather than being formally syntax-driven. Typically, IP address literals are handled by an explicit code path. Parsers have been inconsistent in providing for zone identifiers. Most have no support, but there have been examples of ad hoc support. For example, some versions of Firefox allowed the use of a zone identifier preceded by a bare "%" character, but this feature was removed for consistency with the established syntax of RFC 3986. As another example, some versions of Internet Explorer allowed use of a zone identifier preceded by a "%" character encoded as "%25", still beyond the syntax allowed by the established rules. This syntax extension is in fact used internally in the Windows operating system and some of its APIs.

URI parsers SHOULD accept a zone identifier according to the syntax defined in Section 3, rather than treating the URI as invalid as they do today. An IPv6 address literal never contains percent-encodings. In terms of Section 2.4 of [RFC3986], the "%" character preceding a zone identifier is acting as a delimiter, not as data. Any code handling percent-encoding or percent-decoding must be aware of this.

While the ABNF syntax defined above is consistent, there are many existing URI parsers that apply percent decoding liberally (including within IPv6 literals) regardless of the ABNF, so the probability of practical and operational problems is claimed to be very high, especially during the period when some parsers have been updated and others have not. For example, the URI "http://

[fe80::cd%21]" might be incorrectly decoded as "http://[fe80::cd!]", which will fail. However, as discussed in the first paragraph of Section 4, errors of this type will not prevent progressive deployment of the new syntax on devices that need it.

As noted above, a zone identifier included in a URI has no meaning outside the originating HTTP client node. This has two consequences:

1. In some use cases, such as CUPS, the host address embedded in the URI will be reflected back to the client, using exactly the representation of the zone identifier that the client sent. Otherwise, the zone identifier is of no value to the server.

2. A URI parser which is not running in the originating host cannot verify the validity of the zone identifier, since that is only possible on the originating host. It can only verify that it conforms to the ABNF.

The various use cases for the zone identifier syntax will usually require it to be entered in a browser's input dialogue box. However, URIs including a zone identifier might occur in HTML documents. For example, a diagnostic script in an HTML page might be tailored for a particular host. Because of such usage, it is appropriate for browsers to treat such URIs in the same way whether they are entered in the dialogue box or encountered in an HTML document.

6. 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.

However, this format is only meaningful for link-local addresses under prefix fe80::/10. It is not necessary for web browsers to verify this, or to validate the zone identifier, because the operating system will do so when the address is passed to the socket API, and return an error code if the zone identifier is invalid. This is in addition to the protection offered by CORS when a zone identifier is transmitted to another device, as discussed in Section 4.

A zone identifier in a URI will be revealed to the recipient of an HTTP message containing it (typically in the "Host" field [RFC9110]). A server that receives a zone identifier in an HTTP message or otherwise SHOULD NOT make use of it, for validation of authority or any other purpose, since it has no meaning outside the originating host. Existing practice for controlling cross-origin resource sharing applies, as discussed above Section 4.

Visibility of the zone identifier to a server is anyway a minor security concern, since the information revealed is of local significance only and will be exploitable only if both the client host and the server have both already been compromised.

Unfortunately there is no formal limit on the length of the zone identifier string in RFC 4007. An implementation SHOULD apply a reasonable length limit when generating a URI, in order to minimize the risk of a buffer overrun. For example, a limit to 16 ASCII characters would correspond to the existing limit on Linux interface names.

An implementation SHOULD NOT include ASCII NULL characters in a zone identifier string as this could cause inconsistencies in subsequent string processing.

It is conceivable that this format could be misused to remotely probe a local network configuration or to fingerprint a host. In particular, a script included in an HTML web page could originate HTTP messages intended to determine if a particular link-local address is valid, for example to discover and misuse the address of the first-hop router. However, such attacks are already possible, by probing IPv4 addresses, routeable IPv6 addresses or link-local addresses without a zone identifier. Indeed, with a zone identifier present, the attacker's job is harder because they must also guess the zone identifier itself; the zone identifier increases the search space compared to guessing only the interface identifier. Zone identifiers vary widely between operating systems; in some cases they are easily guessed small integers or conventional names such as "eth0" but in other cases they contain arbitrary characters derived from MAC addresses. In any case, an attacker must discover them before probing any link-local addresses. This argues against the recommendation of [RFC4007] to support a default zone identifier. Nevertheless, the principal defence against scanning attacks remains the 64 bit size of the IPv6 interface identifier [RFC7707].

In the case that a zone identifier contains the hexadecimal MAC address of a network interface, it will be revealed to the HTTP recipient and to any observer on the link. Since the MAC address will also be visible in the underlying layer 2 frame, this is not a new exposure. Nevertheless, this method of naming interfaces might be considered to be a privacy issue.

It should be noted that if a node uses an interface identifier in the outdated Modified EUI format [RFC4291] for its link-local address, the search space for an attacker is very significantly reduced, as discussed in Section 4.1.1.1 of [RFC7707]. The resultant recommendations of [RFC8064] apply to all nodes, including routers, since they ensure that the search space for an attacker is of size 2**64, which is impractically large.

Nevertheless, even a Modified EUI link-local address is significantly harder to guess than typical IPv4 addresses for devices such as home routers, which are often included in published documentation.

7. IANA Considerations

This document makes no request of IANA.

8. References

8.1. 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>.

[RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, RFC 3986, DOI 10.17487/RFC3986, January 2005, <https://www.rfc-editor.org/info/rfc3986>.

[RFC3987] Duerst, M. and M. Suignard, "Internationalized Resource Identifiers (IRIs)", RFC 3987, DOI 10.17487/RFC3987, January 2005, <https://www.rfc-editor.org/info/rfc3987>.

[RFC4007] Deering, S., Haberman, B., Jinmei, T., Nordmark, E., and B. Zill, "IPv6 Scoped Address Architecture", RFC 4007, DOI 10.17487/RFC4007, March 2005, <https://www.rfc-editor.org/info/rfc4007>.

[RFC5234] Crocker, D., Ed. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, RFC 5234, DOI 10.17487/RFC5234, January 2008, <https://www.rfc-editor.org/info/rfc5234>.

[RFC5952] Kawamura, S. and M. Kawashima, "A Recommendation for IPv6 Address Text Representation", RFC 5952, DOI 10.17487/RFC5952, August 2010, <https://www.rfc-editor.org/info/rfc5952>.

[RFC8064] Gont, F., Cooper, A., Thaler, D., and W. Liu, "Recommendation on Stable IPv6 Interface Identifiers", RFC 8064, DOI 10.17487/RFC8064, February 2017, <https://www.rfc-editor.org/info/rfc8064>.

[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/info/rfc8174>.

8.2. Informative References

[CUPS] Apple, "Apple CUPS", 2022, <https://www.cups.org/>.

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

[LL-HACK] Jin, P., "Snippets: IPv6 link local connect hack", 2021, <https://website.peterjin.org/wiki/Snippets:IPv6_link_local_connect_hack>.

[ONE-NET] NMEA, "The OneNet Standard for IP Networking of Marine Electronic Devices", 2023, <https://www.nmea.org/nmea-onenet.html>.

[OP-CUPS] Sweet, M., "OpenPrinting CUPS", 2022, <https://openprinting.github.io/cups/>.

[PNA-REP] Rigoudy, T., Ed., "Private Network Access", 2023, <https://wicg.github.io/private-network-access/>.

[RFC1918] Rekhter, Y., Moskowitz, B., Karrenberg, D., de Groot, G. J., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, DOI 10.17487/RFC1918, February 1996, <https://www.rfc-editor.org/info/rfc1918>.

[RFC2863] McCloghrie, K. and F. Kastenholz, "The Interfaces Group MIB", RFC 2863, DOI 10.17487/RFC2863, June 2000, <https://www.rfc-editor.org/info/rfc2863>.

[RFC3493] Gilligan, R., Thomson, S., Bound, J., McCann, J., and W. Stevens, "Basic Socket Interface Extensions for IPv6", RFC 3493, DOI 10.17487/RFC3493, February 2003, <https://www.rfc-editor.org/info/rfc3493>.

[RFC3510] Herriot, R. and I. McDonald, "Internet Printing Protocol/1.1: IPP URL Scheme", RFC 3510, DOI 10.17487/RFC3510, April 2003, <https://www.rfc-editor.org/info/rfc3510>.

[RFC3927] Cheshire, S., Aboba, B., and E. Guttman, "Dynamic Configuration of IPv4 Link-Local Addresses", RFC 3927, DOI 10.17487/RFC3927, May 2005, <https://www.rfc-editor.org/info/rfc3927>.

[RFC4001] Daniele, M., Haberman, B., Routhier, S., and J. Schoenwaelder, "Textual Conventions for Internet Network Addresses", RFC 4001, DOI 10.17487/RFC4001, February 2005, <https://www.rfc-editor.org/info/rfc4001>.

[RFC4193] Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses", RFC 4193, DOI 10.17487/RFC4193, October 2005, <https://www.rfc-editor.org/info/rfc4193>.

[RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, DOI 10.17487/RFC4291, February 2006, <https://www.rfc-editor.org/info/rfc4291>.

[RFC6874] Carpenter, B., Cheshire, S., and R. Hinden, "Representing IPv6 Zone Identifiers in Address Literals and Uniform Resource Identifiers", RFC 6874, DOI 10.17487/RFC6874, February 2013, <https://www.rfc-editor.org/info/rfc6874>.

[RFC7472] McDonald, I. and M. Sweet, "Internet Printing Protocol (IPP) over HTTPS Transport Binding and the 'ipps' URI Scheme", RFC 7472, DOI 10.17487/RFC7472, March 2015, <https://www.rfc-editor.org/info/rfc7472>.

[RFC7707] Gont, F. and T. Chown, "Network Reconnaissance in IPv6 Networks", RFC 7707, DOI 10.17487/RFC7707, March 2016, <https://www.rfc-editor.org/info/rfc7707>.

[RFC8343] Bjorklund, M., "A YANG Data Model for Interface Management", RFC 8343, DOI 10.17487/RFC8343, March 2018, <https://www.rfc-editor.org/info/rfc8343>.

[RFC9110] Fielding, R., Ed., Nottingham, M., Ed., and J. Reschke, Ed., "HTTP Semantics", STD 97, RFC 9110, DOI 10.17487/RFC9110, June 2022, <https://www.rfc-editor.org/info/rfc9110>.

Appendix A. Options Considered

The syntax defined above allows a zone identifier 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 zone identifier 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::abcd%en1
   ```

   Advantage: works today.

   Disadvantage: less convenient than using a browser. Leaves use cases unsatisfied.

2. Simply use the percent character:

   ```
   http://[fe80::abcd%en1]
   ```

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

   Disadvantage: requires code changes to all URI parsers, some of which differ in their interpretation of the percent-encoding rules.

   This is the option chosen for standardisation.

3. Use an alternative separator:

   ```
   http://[fe80::abcd-en1]
   ```

   Advantage: allows use of browser; simple syntax.

   Disadvantages: requires code changes to all URI parsers; requires manual editing during 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 may become invisible or unclear when a user interface automatically underlines URLs.

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

   ```
   http://[v6.fe80::abcd-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 [RFC3986]:

   ```
   http://[fe80::abcd%25en1]
   ```

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

   Disadvantages: somewhat ugly and confusing; requires manual editing during cut and paste; requires code changes to all URI parsers, some of which differ in their interpretation of the percent-encoding rules.

Appendix B. Change log

This section is to be removed before publishing as an RFC.

*draft-ietf-6man-rfc6874bis-09, 2023-07-02:

-Noted scope is limited to RFC 4007 model.

-Updated W3C reference.

*draft-ietf-6man-rfc6874bis-08, 2023-04-06:

-Noted minor inconsistency with RFC 4007.

*draft-ietf-6man-rfc6874bis-07, 2023-04-12:

-Clarified character set restrictions and the applicability of numeric identifiers as a work-around.

-Updated ABNF to require lower case, reorganized text as a result.

-Expanded text on handling of zone ID at server.

-Other nits.

*draft-ietf-6man-rfc6874bis-06, 2023-04-07:

-Noted potential exposure of MAC addresses in zone IDs.

-Expanded detail on lower-case normalization.

-Added specific use case examples.

-Added NMEA use case.

- Clearly explained cut-and-paste requirement.

- Indicated that network infrastructure devices are out of scope.

- Noted the work-around using interface numbers.

- Mentioned .local as another case of locally significant URIs.

- Added discussion of CORS.

- Update descriptions of rejected alternatives

- Noted parsing fragility re % sign.

- Other IESG review nits.

*draft-ietf-6man-rfc6874bis-05, 2022-11-07:

- Noted lower case issue.

*draft-ietf-6man-rfc6874bis-04, 2022-10-19:

- should accept -> SHOULD.

- Suggested maximum length of zone ID.

*draft-ietf-6man-rfc6874bis-03, 2022-09-30:

- Strengthened motivation for publishing this requirement now.

- Removed unnecessary sentence about browsers.

- Noted that zone ID will be revealed to HTTP server.

- Noted that servers should make no use of received zone IDs.

- Noted that zone IDs have no length limit.

- Added section on scope and deployment, specifically noting that URIs with local scope are nothing new.

- Other Last Call clarifications and nits.

*draft-ietf-6man-rfc6874bis-02, 2022-07-05:

- Improve discussion of URLs in HTML documents

- Discuss scripting attack and Modified EUI IIDs

- Several editorial clarifications

- Some nits fixed

*draft-ietf-6man-rfc6874bis-01, 2022-04-07:

- Extended use cases

- Clarified relationship with RFC3986 language

- Allow for legacy use of RFC6874 format

- Augmented security considerations

- Editorial and reference improvements

*draft-ietf-6man-rfc6874bis-00, 2022-03-19:

- WG adoption

- Clarified security considerations

*draft-carpenter-6man-rfc6874bis-03, 2022-02-08:

- Changed to bare % signs.

- Added IRIs, RFC3987

- Editorial fixes

*draft-carpenter-6man-rfc6874bis-02, 2021-18-12:

- Give details of open issues

- Update authorship

- Editorial fixes

*draft-carpenter-6man-rfc6874bis-01, 2021-07-11:

- Added section on issues with RFC6874

- Removed suggested heuristic for bare % signs

- Editorial fixes

*draft-carpenter-6man-rfc6874bis-00, 2021-07-05:

- Initial version

Appendix C. Acknowledgements

The lack of this format was first pointed out by Margaret Wasserman and later by Kerry Lynn. A previous draft document by Bill Fenner and Martin Dürst [LITERAL-ZONE] discussed this topic but was not finalised. Michael Sweet and Andrew Cady explained some of the difficulties caused by RFC 6874. The ABNF syntax proposed above was drafted by Andrew Cady.

Valuable comments and contributions were made by Karl Auer, Carlos Bernardos, Carsten Bormann, Benoit Claise, Martin Dürst, David Farmer, Stephen Farrell, Brian Haberman, Ted Hardie, Philip Homburg, Tatuya Jinmei, Leif Johansson, Nate Karstens, Yves Lafon, Barry Leiba, Ted Lemon, Ben Maddison, Radia Perlman, Tom Petch, Michael Richardson, Tomoyuki Sahara, Jürgen Schönwälder, Nico Schottelius, Dave Thaler, Martin Thomson, Philipp S. Tiesel, Ole Troan, Éric Vyncke, Shang Ye, several IESG members, and others.

Authors' Addresses

Brian Carpenter
School 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 of America

Email: cheshire@apple.com

Robert M. Hinden
Check Point Software
959 Skyway Road
San Carlos, CA 94070
United States of America

Email: bob.hinden@gmail.com