Internet Engineering Task Force (IETF)
Request for Comments: 6783
Obsoletes: 5983
Category: Informational
ISSN: 2070-1721

J. Levine
Taughannock Networks
R. Gellens
Qualcomm Incorporated

November 2012

Mailing Lists and Non-ASCII Addresses

Abstract

This document describes considerations for mailing lists with the introduction of non-ASCII UTF-8 email addresses. It outlines some possible scenarios for handling lists with mixtures of non-ASCII and traditional addresses but does not specify protocol changes or offer implementation or deployment advice.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

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). Not all documents approved by the IESG are a candidate for any level of Internet Standard; see 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/rfc6783.

Table of Contents

1. Introduction

This document describes considerations for mailing lists with the introduction of non-ASCII UTF-8 email addresses. The usage of such addresses is described in [RFC6530].

Mailing lists are an important part of email usage and collaborative communications. The introduction of internationalized email addresses affects mailing lists in three main areas: (1) transport (receiving and sending messages); (2) message headers of received and retransmitted messages; and (3) mailing list operational policies.

A mailing list is a mechanism that distributes a message to multiple recipients when the originator sends it to a single address. An agent, usually software rather than a person, at that single address receives the message and then causes the message to be redistributed to a list of recipients. This agent usually sets the envelope return address (henceforth called the "bounce address") of the redistributed message to a different address from that of the original message. Using a different bounce address directs error and other automatically generated messages to an error-handling address associated with the mailing list. This sends error and other automatic messages to the list agent, which can often do something useful with them, rather than to the original sender, who typically doesn't control the list and hence can't do anything about them.

In most cases, the mailing list agent redistributes a received message to its subscribers as a new message, that is, conceptually it uses message submission [RFC6409] (as did the sender of the original message). The exception, where the mailing list is not managed by a

separate agent that receives and redistributes messages in separate transactions but is implemented by an expansion step within an SMTP transaction where one local address expands to multiple local or non-local addresses, is not addressed by this document.

1.1. Mailing List Header Additions and Modifications

Some list agents alter message header fields, while others do not. A number of standardized list-related header fields have been defined, and many lists add one or more of these headers. Separate from these standardized list-specific header fields, and despite a history of interoperability problems from doing so, some lists alter or add header fields in an attempt to control where replies are sent. Such lists typically add or replace the "Reply-To" field, and some add or replace the "Sender" field. Some lists alter or replace other fields, including "From".

Among these list-specific header fields are those specified in RFCs 2369 [RFC2369] and 2919 [RFC2919]. For more information, see Section 3.

1.2. Non-ASCII Email Addresses

While the mail transport protocol is the same for regular email recipients and mailing list recipients, list agents have special considerations with non-ASCII email addresses because they retransmit messages composed by other agents to potentially many recipients.

There are considerations for non-ASCII email addresses in the envelope as well as in header fields of redistributed messages. In particular, a message with non-ASCII addresses in the headers or envelope cannot be sent to non-SMTPUTF8 recipients.

With mailing lists, there are two different types of considerations: first, the purely technical ones involving message handling, error cases, and the like, and second, those that arise from the fact that humans use mailing lists to communicate. As an example of the first, list agents might choose to reject all messages from non-ASCII addresses if they are unprepared to handle SMTPUTF8 mail. As an example of the second, a user who sends a message to a list often is unaware of the list membership. In particular, the user often doesn’t know if the members are SMTPUTF8 mail users or not, and often neither the original sender nor the recipients personally know each other. As a consequence of this, remedies that may be readily available for one-to-one communication might not be appropriate when dealing with mailing lists. For example, if a user sends a message that is undeliverable, normally the telephone, instant messaging, or other forms of communication are available to obtain a working

address. With mailing lists, the users may not have any recourse. Of course, with mailing lists, the original sender usually does not know which list members successfully received a message or if it was undeliverable to some.

Conceptually, a mailing list's internationalization can be divided into three capabilities. First, does the list have a non-ASCII submission address? Second, does the list agent accept subscriptions for addresses containing non-ASCII characters? And third, does the list agent accept messages that require SMTPUTF8 capabilities?

If a list has subscribers with ASCII addresses, those subscribers might or might not be able to accept SMTPUTF8 messages.

2. Scenarios Involving Mailing Lists

Generally (and exclusively within the scope of this document), an original message is sent to a mailing list as a completely separate and independent transaction from the list agent sending the retransmitted message to one or more list recipients. In both cases, the message might be addressed only to the list address or might have recipients in addition to the list. Furthermore, the list agent might choose to send the retransmitted message to each list recipient in a separate message submission transaction or might choose to include multiple recipients per transaction. Often, list agents are constructed to work in cooperation with, rather than include the functionality of, a message submission server; hence, the list transmits to a single submission server one copy of the retransmitted message. The submission server then decides which recipients to include in which transaction.

2.1. Fully SMTPUTF8 Lists

Some lists may wish to be fully SMTPUTF8. That is, all subscribers are expected to be able to receive SMTPUTF8 mail. For list hygiene reasons, such a list would probably want to prevent subscriptions from addresses that are unable to receive SMTPUTF8 mail. If a putative subscriber has a non-ASCII address, it must be able to receive SMTPUTF8 mail, but there is no way to tell whether a subscriber with an ASCII address can receive SMTPUTF8 mail short of sending an SMTPUTF8 probe or confirmation message and somehow finding out whether it was delivered, e.g., if the user clicked a link in the confirmation message.

2.2. Mixed SMTPUTF8 and ASCII Lists

Other lists may wish to handle a mixture of SMTPUTF8 and ASCII subscribers, either as a transitional measure as subscribers upgrade to SMTPUTF8-capable mail software or as an ongoing feature. While it is not possible in general to downgrade SMTPUTF8 mail to ASCII mail, list software might divide the recipients into two sets, SMTPUTF8 and ASCII recipients, and create a downgraded version of SMTPUTF8 list messages to send to ASCII recipients. See Sections 3.2 and 3.3.

To determine which set an address belongs in, list software might make the conservative assumption that ASCII addresses get ASCII messages, it might try to probe the address with an SMTPUTF8 test message, or it might let the subscriber set the message format manually, similar to the way that some lists now let subscribers choose between plain text and HTML mail, or individual messages and a daily digest.

To determine whether a message needs to be downgraded for ASCII recipients, list software might assume that any message received via an SMTPUTF8 SMTP session is an SMTPUTF8 message or might examine the headers and body of the message to see whether it needs SMTPUTF8 treatment. Depending on the interface between the list software and the Mail Transfer Agent (MTA) and Mail Delivery Agent (MDA) that handle incoming messages, it may not be able to tell the type of session for incoming messages.

2.3. SMTP Issues

Mailing list software usually changes the envelope addresses on each message. The bounce address is set to an address that will return bounces to the list agent, and the recipient addresses are set to the subscribers of the list. For some lists, all messages to a list get the same bounce address. For others, list software may create a bounce address per recipient or a unique bounce address per message per recipient, bounce management techniques known as Variable Envelope Return Paths or VERP [VERP].

The bounce address for a list typically includes the name of the list, so a list with a non-ASCII name will have a non-ASCII bounce address. Given the unknown paths that bounce messages might take, list software might instead use an ASCII bounce address to make it more likely that bounces can be delivered back to the list agent. Similarly, a VERP address for each subscriber typically embeds a version of the subscriber’s address so the VERP bounce address for a non-ASCII subscriber address will be a non-ASCII address. For the same reason, the list software might use ASCII bounce addresses that encode the recipient’s identity in some other way.

3. List Headers

List agents typically add list-specific headers to each message before resending the message to list recipients.

3.1. SMTPUTF8 List Headers

The list headers in RFCs 2369 [RFC2369] and 2919 [RFC2919] were all specified before SMTPUTF8 mail existed, and their definitions do not address where non-ASCII characters might appear. These include, for example:

```
List-Id: List Header Mailing List
   <list-header.example.com>
List-Help:
   <mailto:list@example.com?subject=help>
List-Unsubscribe:
   <mailto:list@example.com?subject=unsubscribe>
List-Subscribe:
   <mailto:list@example.com?subject=subscribe>
List-Post:
   <mailto:list@example.com>
List-Owner:
   <mailto:listmom@example.com> (Contact Person for Help)
List-Archive:
   <mailto:archive@example.com?subject=index%20list>
```

As described in [RFC2369], "[t]he contents of the list header fields mostly consist of angle-bracket ('<', '>') enclosed URLs, with internal whitespace being ignored". [RFC2919] specifies that "[t]he list identifier will, in most cases, appear like a host name in a domain of the list owner". Since these headers were defined in the context of ASCII mail, these headers permit only ASCII text, including in the URLs.

The most commonly used URI schemes in List-* headers tend to be http and mailto [RFC6068], although they sometimes include https and ftp and, in principle, can contain any valid URI.

Even if a scheme permits an internationalized form, it should use a pure ASCII form of the URI described in [RFC3986]. Future work may extend these header fields or define replacements to directly support unencoded non-ASCII outside the ASCII repertoire in these and other header fields, but in the absence of such extension or replacement, non-ASCII characters can only be included by encoding them as ASCII.

The encoding technique specified in [RFC3986] is to use a pair of hex digits preceded by a percent sign, but percent signs have been used informally in mail addresses to do source routing. Although few mail systems still permit source routing, a lot of mail software still forbids or escapes characters formerly used for source routing, which can lead to unfortunate interactions with percent-encoded URIs or any URI that includes one of those characters. If a program interpreting a mailto: URI knew that the Mail User Agent (MUA) in use were able to handle non-ASCII data, the program could pass the URI in unencoded non-ASCII, avoiding problems with misinterpreted percent signs, but at this point, there is no standard or even informal way for MUAs to signal SMTPUTF8 capabilities. Also, note that whether internationalized domain names should be percent-encoded or appear in A-label form [RFC5890] depends on the context in which they occur.

The List-ID header field uniquely identifies a list. The intent is that the value of this header remain constant, even if the machine or system used to operate or host the list changes. This header field is often used in various filters and tests, such as client-side filters, Sieve filters [RFC5228], and so forth. If the definition of a List-ID header field were to be extended to allow non-ASCII text, filters and tests might not properly compare encoded and unencoded versions of a non-ASCII value. In addition to these comparison considerations, it is generally desirable that this header field contain something meaningful that users can type in. However, ASCII encodings of non-ASCII characters are unlikely to be meaningful to users or easy for them to accurately type.

3.2. Downgrading List Headers

If list software prepares a downgraded version of an SMTPUTF8 message, all the List-* headers must be downgraded. In particular, if a List-* header contains a non-ASCII mailto (even encoded in ASCII), it may be advisable to edit the header to remove the non-ASCII address or replace it with an equivalent ASCII address if one is known to the list software. Otherwise, a client might run into trouble if the decoded mailto results in a non-ASCII address. If a header that contains a mailto URL is downgraded by percent encoding, some mail software may misinterpret the percent signs as attempted source routing.

When downgrading list headers, it may not be possible to produce a downgraded version that is satisfactorily equivalent to the original header. In particular, if a non-ASCII List-ID is downgraded to an ASCII version, software and humans at recipient systems will typically not be able to tell that both refer to the same list.

If lists permit mail with multiple MIME parts, some MIME headers in SMTPUTF8 messages may include non-ASCII characters in file names and other descriptive text strings. Downgrading these strings may lose the sense of the names, break references from other MIME parts (such as HTML IMG references to embedded images), and otherwise damage the mail.

3.3. Subscribers' Addresses in Downgraded Headers

List software typically leaves the original submitter's address in the From: line, both so that recipients can tell who wrote the message and so that they have a choice of responding to the list or directly to the submitter. If a submitter has a non-ASCII address, there is no way to downgrade the From: header and preserve the address so that ASCII recipients can respond to it, since non-SMTPUTF8 mail systems can't send mail to non-ASCII addresses.

Possible work-arounds (none implemented that we know of) might include allowing subscribers with non-ASCII addresses to register an alternate ASCII address with the list software, having the list software itself create ASCII forwarding addresses, or just putting the list's address in the From: line and losing the ability to respond directly to the submitter.

4. Security Considerations

None beyond what mailing list agents do now.

5. References

5.1. Normative References

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

[RFC6068] Duerst, M., Masinter, L., and J. Zawinski, "The 'mailto' URI Scheme", RFC 6068, October 2010.

[RFC6409] Gellens, R. and J. Klensin, "Message Submission for Mail", STD 72, RFC 6409, November 2011.

[RFC6530] Klensin, J. and Y. Ko, "Overview and Framework for Internationalized Email", RFC 6530, February 2012.

5.2. Informative References

[RFC2369] Neufeld, G. and J. Baer, "The Use of URLs as Meta-Syntax for Core Mail List Commands and their Transport through Message Header Fields", RFC 2369, July 1998.

[RFC2919] Chandhok, R. and G. Wenger, "List-Id: A Structured Field and Namespace for the Identification of Mailing Lists", RFC 2919, March 2001.

[RFC5228] Guenther, P. and T. Showalter, "Sieve: An Email Filtering Language", RFC 5228, January 2008.

[RFC5890] Klensin, J., "Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework", RFC 5890, August 2010.

[VERP] Bernstein, D., "Variable Envelope Return Paths", February 1997, <http://cr.yp.to/proto/verp.txt>.

Authors' Addresses

John Levine
Taughannock Networks
PO Box 727
Trumansburg, NY 14886
US

Phone: +1 831 480 2300
EMail: standards@taugh.com
URI: http://jl.ly

Randall Gellens
Qualcomm Incorporated
5775 Morehouse Drive
San Diego, CA 92121
US

EMail: rg+ietf@qti.qualcomm.com