

Network Working Group
Document: [draft-ietf-lemonade-rfc2192bis-01.txt](#)
Expires: August 2006
Intended category: Standards Track

C. Newman
Sun Microsystems
A. Melnikov
Isode Ltd.
S. H. Maes
Oracle Corporation
February 2006

IMAP URL Scheme

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with [Section 6 of 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>.

A revised version of this draft document will be submitted to the RFC editor as a Proposed Standard for the Internet Community. Discussion and suggestions for improvement are requested, and should be sent to the IMAPEXT Mailing list <ietf-imapext@imc.org>. Distribution of this draft is unlimited.

Copyright Notice

Copyright (C) The Internet Society (2006).

Abstract

IMAP [[IMAP4](#)] is a rich protocol for accessing remote message stores. It provides an ideal mechanism for accessing public mailing list archives as well as private and shared message stores. This document defines a URL scheme for referencing objects on an IMAP server.

1. Conventions used in this document

The key words "MUST", "MUST NOT", "SHOULD", "SHOULD NOT", and "MAY" in this document are to be interpreted as defined in "Key words for use in RFCs to Indicate Requirement Levels" [[KEYWORDS](#)].

2. IMAP scheme

The IMAP URL scheme is used to designate IMAP servers, mailboxes, messages, MIME bodies [[MIME](#)], and search programs on Internet hosts accessible using the IMAP protocol.

The IMAP URL follows the common Internet scheme syntax as defined in [[URI-GEN](#)] <<except that clear text passwords are not permitted>>. If <:port> is omitted, the port defaults to 143.

An IMAP URL takes one of the following forms:

```
imap://<iserver>/
```

```
imap://<iserver>/<enc-list-mailbox>;TYPE=<list-type>
```

```
imap://<iserver>/<enc-mailbox>[uidvalidity][?<enc-search>]
```

```
imap://<iserver>/<enc-mailbox>[uidvalidity]<iuid>  
[isection][ipartial]
```

The first form is used to refer to an IMAP server, the second form refers to a list of mailboxes, the third form refers to the contents of a mailbox or a set of messages resulting from a search, and the final form refers to a specific message or message part, and possibly a byte range in that part. Note that the syntax here is informal. The authoritative formal syntax for IMAP URLs is defined in [section 11](#). The partial specifier semantics conforms to [[IMAP4](#)] partial specifiers.

3. IMAP userinfo component

3.1. IMAP mailbox naming scope

The "enc-user" part of the "iuserinfo" component, if present, denotes mailbox naming scope. If it is absent, the IMAP URL can only reference mailboxes with globally unique names, i.e. mailboxes with names that don't change depending on the user the client authenticated as to the IMAP server. <<AM: Note that not all IMAP implementations support globally unique names>>

For example, a personal mailbox described by the following URL <imap://michael@example.org/INBOX> is most likely be different from a personal mailbox described by <imap://bester@example.org/INBOX>, even though both URLs use the same mailbox name.

<<Chris: whenever possible the shared folder's name for a mailbox should be used to avoid mailbox naming scope issues. AM: There is no standard way of learning a globally unique name for a mailbox. However by looking at the NAMESPACE command result it is possible to see if a mailbox name is globally unique or not.>>

3.2. IMAP User Name and Authentication Mechanism

The userinfo component [[URI-GEN](#)] of an IMAP URI may contains an IMAP user Name (authorization identity, "enc-user") and/or an authentication mechanism. (Note that the "enc-user" also defines a mailbox naming scope as described in [section 3.1](#)). They are used in the "LOGIN" or "AUTHENTICATE" commands after making the connection to the IMAP server.

If no user name or authentication mechanism is supplied, the client must authenticate as anonymous to the server. If the server advertises AUTH=ANONYMOUS IMAP capability, the client MUST use the AUTHENTICATE command with ANONYMOUS [[ANONYMOUS](#)] SASL mechanism. If SASL ANONYMOUS is not available, the user name "anonymous" is used with the "LOGIN" command and the password is supplied as the Internet e-mail address of the end user accessing the resource. <<The latter is to provide for interoperability with existing servers.>>

If the URL doesn't supply a user name, the program interpreting the IMAP URL SHOULD request one from the user (if it is an interactive program) or configuration.

Note that as described in [RFC 3501](#), the LOGIN command MUST NOT be used when the IMAP server advertises the LOGINDISABLED capability.

An authentication mechanism can be expressed by adding ";AUTH=<enc-auth-type>" to the end of the user name. When such an <enc-auth-

type> is indicated, the client SHOULD request appropriate credentials from that mechanism and use the "AUTHENTICATE" command instead of the "LOGIN" command. If no user name is specified, one SHOULD be obtained from the mechanism or requested from the user as appropriate.

The string ";AUTH=*" indicates that the client SHOULD select an appropriate authentication mechanism. It MAY use any mechanism listed in the CAPABILITY command or use an out of band security service resulting in a PREAUTH connection. If no user name is specified and no appropriate authentication mechanisms are available, the client SHOULD fall back to anonymous login as described above. This allows a URL which grants read-write access to authorized users, and read-only anonymous access to other users.

If a user name is included with no authentication mechanism, then ";AUTH=*" is assumed.

Since URLs can easily come from untrusted sources, care must be taken when resolving a URL which requires or requests any sort of authentication. If authentication credentials are supplied to the wrong server, it may compromise the security of the user's account. The program resolving the URL should make sure it meets at least one of the following criteria in this case:

- (1) The URL comes from a trusted source, such as a referral server which the client has validated and trusts according to site policy. Note that user entry of the URL may or may not count as a trusted source, depending on the experience level of the user and site policy.
- (2) Explicit local site policy permits the client to connect to the server in the URL. For example, if the client knows the site domain name, site policy may dictate that any hostname ending in that domain is trusted.
- (3) The user confirms that connecting to that domain name with the specified credentials and/or mechanism is permitted.
- (4) A mechanism is used which validates the server before passing potentially compromising client credentials.
- (5) An authentication mechanism is used which will not reveal information to the server which could be used to compromise future connections.

URLs which do not include a user name must be treated with extra care, since they are more likely to compromise the user's primary account. A URL containing ";AUTH=*" must also be treated with extra care since it might fall back on a weaker security mechanism. Finally, clients are discouraged from using a plain text password as a fallback with ";AUTH=*" unless the connection has strong

Newman & Co

Expires: August 2006

FORMFEED[Page 4]

encryption.

A program interpreting IMAP URLs MAY cache open connections to an IMAP server for later re-use. If a URL contains a user name, only connections authenticated as that user may be re-used. If a URL does not contain a user name or authentication mechanism, then only an anonymous connection may be re-used. If a URL contains an authentication mechanism without a user name, then any non-anonymous connection may be re-used.

Note that if unsafe or reserved characters such as " " or ";" are present in the user name or authentication mechanism, they MUST be encoded as described in [[URI-GEN](#)].

<<URLAUTH overrides the authorization identity>>

<<Include URLAUTH content in this document?>>

[4.](#) IMAP server

An IMAP URL referring to an IMAP server has the following form:

```
imap://<iserver>/
```

A program interpreting this URL would issue the standard set of commands it uses to present a view of the contents of an IMAP server. This is likely to be semantically equivalent to one of the following URLs:

```
imap://<iserver>;TYPE=LIST
imap://<iserver>;TYPE=LSUB
```

See [section 5](#) for details on how such URLs should be processed.

The program interpreting this URL SHOULD use the LSUB form if it supports mailbox subscriptions.

[5.](#) Lists of mailboxes

An IMAP URL referring to a list of mailboxes has the following form:

```
imap://<iserver>/<enc-list-mailbox>;TYPE=<list-type>
```

The <list-type> may be either "LIST" or "LSUB", and is case insensitive. The field "<list-type>" MUST be included.

The <enc-list-mailbox> is any argument suitable for the list-mailbox field of the IMAP [[IMAP4](#)] LIST or LSUB commands. The field <enc-list-mailbox> may be omitted, in which case the program interpreting the IMAP URL may use "*" or "%" as the <enc-list-mailbox>. The program SHOULD use "%" if it supports a hierarchical view, otherwise it SHOULD use "*".

Note that if unsafe or reserved characters such as " " or "%" are present in <enc-list-mailbox> they MUST be encoded as described in [[URI-GEN](#)]. If the character "/" is present in enc-list-mailbox, it SHOULD NOT be encoded.

If the <enc-list-mailbox> field is omitted or is defined to be "*" or "%" the following procedure for listing mailboxes is RECOMMENDED:

The program SHOULD first use the NAMESPACE [[NAMESPACE](#)] command (if supported by the server) to discover available namespaces, for example:

```
C: A001 NAMESPACE
S: * NAMESPACE (("" "/")) NIL ("Public Folders/" "/")
S: A001 OK NAMESPACE command completed
```

After that it should issue a LIST/LSUB command (as described by <list-type>) for each namespace, suppressing redundant list commands if the server lists mailboxes for multiple namespaces as a result of a single list command. For example (continuing the example above), the client first issues <A002 LIST "" *>. If the server returns information for mailboxes under the shared namespace prefix "Public Folders/" the client should omit the subsequent <A003 LIST "Public Folders/" *> command.

If the NAMESPACE command is not supported by the IMAP server, the client SHOULD issue <tag <list-type> "" <enc-list-mailbox>>.

<<Add suggestion to use "LIST %" level by level? Reference \HasChildren and \NoInferiors.>>

6. Lists of messages

An IMAP URL referring to a list of messages has the following form:

```
imap://<iserver>/<enc-mailbox>[uidvalidity][?<enc-search>]
```


The <enc-mailbox> field is used as the argument to the IMAP4 "SELECT" command. Note that if unsafe or reserved characters such as " ", ";", or "?" are present in <enc-mailbox> they MUST be encoded as described in [\[URI-GEN\]](#). If the character "/" is present in enc-mailbox, it SHOULD NOT be encoded.

The [uidvalidity] field is optional. If it is present, it MUST be the argument to the IMAP4 UIDVALIDITY status response at the time the URL was created. This SHOULD be used by the program interpreting the IMAP URL to determine if the URL is stale.

The [?<enc-search>] field is optional. If it is not present, the contents of the mailbox SHOULD be presented by the program interpreting the URL. If it is present, it SHOULD be used as the arguments following an IMAP4 SEARCH command with unsafe characters such as " " (which are likely to be present in the <enc-search>) encoded as described in [\[URI-GEN\]](#).

7. A specific message or message part

An IMAP URL referring to a specific message or message part has the following form:

```
imap://<iserver>/<enc-mailbox>[uidvalidity]<iuid>
[isection][ipartial]
```

The <enc-mailbox> and [uidvalidity] are as defined above.

If [uidvalidity] is present in this form, it SHOULD be used by the program interpreting the URL to determine if the URL is stale.

The <iuid> refers to an IMAP4 message UID, and SHOULD be used as the <set> argument to the IMAP4 "UID FETCH" command.

The [isection] field is optional. If not present, the URL refers to the entire Internet message as returned by the IMAP command "UID FETCH <iuid> BODY.PEEK[]". If present, the URL refers to the object returned by a "UID FETCH <iuid> BODY.PEEK[<section>]" command. The type of the object may be determined with a "UID FETCH <iuid> BODYSTRUCTURE" command and locating the appropriate part in the resulting BODYSTRUCTURE. Note that unsafe characters in [isection] MUST be encoded as described in [\[URI-GEN\]](#).

The [ipartial] field is optional. If present, it effectively appends "<<partial-range>>" to the end of the UID FETCH BODY.PEEK[<section>] command constructed as described in the previous paragraph. In other words it allows the client to request a

byte range of the message/message part.

8. Relative IMAP URLs

Relative IMAP URLs are permitted and are resolved according to the rules defined in [\[URI-GEN\]](#) with one exception. In IMAP URLs, parameters are treated as part of the normal path with respect to relative URL resolution. This is believed to be the behavior of the installed base and is likely to be documented in a future revision of the relative URL specification.

The following observations are also important:

The <iauth> grammar element is considered part of the user name for purposes of resolving relative IMAP URLs. This means that unless a new login/server specification is included in the relative URL, the authentication mechanism is inherited from a base IMAP URL.

URLs always use "/" as the hierarchy delimiter for the purpose of resolving paths in relative URLs. IMAP4 permits the use of any hierarchy delimiter in mailbox names. For this reason, relative mailbox paths will only work if the mailbox uses "/" as the hierarchy delimiter. Relative URLs may be used on mailboxes which use other delimiters, but in that case, the entire mailbox name MUST be specified in the relative URL or inherited as a whole from the base URL.

The base URL for a list of mailboxes or messages which was referred to by an IMAP URL is always the referring IMAP URL itself. The base URL for a message or message part which was referred to by an IMAP URL may be more complicated to determine. The program interpreting the relative URL will have to check the headers of the MIME entity and any enclosing MIME entities in order to locate the "Content-Location" header. This header is used to determine the base URL as defined in section 5 of [\[MHTML\]](#). For example, if the referring IMAP URL contains a "/;SECTION=1.2" parameter, then the MIME headers for [section 1.2](#), for [section 1](#), and for the enclosing message itself SHOULD be checked in that order for "Content-Location" headers.

9. Multinational Considerations

IMAP4 [\[IMAP4\]](#) [section 5.1.3](#) includes a convention for encoding non-US-ASCII characters in IMAP mailbox names. Because this convention is private to IMAP, it is necessary to convert IMAP's encoding to one that can be more easily interpreted by a URL display program.

For this reason, IMAP's modified UTF-7 encoding for mailboxes MUST be converted to UTF-8 [UTF8]. Since 8-bit characters are not permitted in URLs, the UTF-8 characters are encoded as required by the URL specification [URI-GEN], section 2.1. Sample code is included in [Appendix A](#) to demonstrate this conversion.

IMAP usernames are UTF-8 strings and MUST be encoded as required by the URL specification [URI-GEN], section 2.1.

Also note that IMAP SEARCH criteria can contain non US-ASCII characters. 8-bit octets in those strings MUST be encoded as required by the URL specification [URI-GEN], section 2.1.

10. Examples

The following examples demonstrate how an IMAP4 client program might translate various IMAP4 URLs into a series of IMAP4 commands. Commands sent from the client to the server are prefixed with "C:", and responses sent from the server to the client are prefixed with "S:".

The URL:

```
<imap://minbari.example.org/gray-council;UIDVALIDITY=385759045/;
  UID=20/PARTIAL=0.1024>
```

Results in the following client commands:

```
<connect to minbari.example.org, port 143>
C: A001 LOGIN ANONYMOUS sheridan@babylon5.example.org
C: A002 SELECT gray-council
<client verifies the UIDVALIDITY matches>
C: A003 UID FETCH 20 BODY.PEEK[]<0.1024>
```

The URL:

```
<imap://michael@minbari.example.org/users.*;type=list>
```

Results in the following client commands:

```
<client requests password from user>
<connect to minbari.example.org imap server,
  activate strong encryption>
C: A001 LOGIN MICHAEL zipper
C: A002 LIST "" users.*
```


The URL:

```
<imap://psicorp.example.org/~peter/%E6%97%A5%E6%9C%AC%E8%AA%9E/
%E5%8F%B0%E5%8C%97>
```

Results in the following client commands:

```
<connect to psicorp.example.org, port 143>
C: A001 LOGIN ANONYMOUS bester@psycop.psicorp.example.org
C: A002 SELECT ~peter/&ZeVnLIqe-/&U,BTFw-
<commands the client uses for viewing the contents of a mailbox>
```

The URL:

```
<imap:///AUTH=GSSAPI@minbari.example.org/gray-council/;uid=20/
;section=1.2>
```

Results in the following client commands:

```
<connect to minbari.example.org, port 143>
C: A001 AUTHENTICATE GSSAPI
<authentication exchange>
C: A002 SELECT gray-council
C: A003 UID FETCH 20 BODY.PEEK[1.2]
```

If the following relative URL is located in that body part:

```
<;section=1.4>
```

This could result in the following client commands:

```
C: A004 UID FETCH 20 (BODY.PEEK[1.2.MIME]
    BODY.PEEK[1.MIME]
    BODY.PEEK[HEADER.FIELDS (Content-Base Content-Location)])
<Client looks for Content-Base or Content-Location headers in
    result. If no such headers, then it does the following>
C: A005 UID FETCH 20 BODY.PEEK[1.4]
```


The URL:

```
<imap:///;AUTH=*@minbari.example.org/gray%20council?
SUBJECT%20shadows>
```

Could result in the following:

```
<connect to minbari.example.org, port 143>
C: A001 CAPABILITY
S: * CAPABILITY IMAP4rev1 AUTH=DIGEST-MD5
S: A001 OK
C: A002 AUTHENTICATE DIGEST-MD5
<authentication exchange>
S: A002 OK user lennier authenticated
C: A003 SELECT "gray council"
...
C: A004 SEARCH SUBJECT shadows
S: * SEARCH 8 10 13 14 15 16
S: A004 OK SEARCH completed
C: A005 FETCH 8,10,13:16 ALL
...
```

NOTE: In this final example, the client has implementation dependent choices. The authentication mechanism could be anything, including PREAUTH. And the final FETCH command could fetch more or less information about the messages, depending on what it wishes to display to the user.

11. Security Considerations

Security considerations discussed in the IMAP specification [[IMAP4](#)] and the URI specification [[URI-GEN](#)] are relevant. Security considerations related to authenticated URLs are discussed in [section 3](#) of this document.

Many email clients store the plain text password for later use after logging into an IMAP server. Such clients MUST NOT use a stored password in response to an IMAP URL without explicit permission from the user to supply that password to the specified host name.

12. ABNF for IMAP URL scheme

Formal syntax is defined using ABNF [ABNF], extending the ABNF rules in section 9 of [IMAP4]. Elements not defined here can be found in the [ABNF], [IMAP4], [IMAPABNF] or [URI-GEN]. Strings are not case sensitive and free insertion of linear-white-space is not permitted.

```
uchar          = unreserved | pct-encoded

achar          = uchar / "&" / "="
                ;; <<less permissive than "pchar">>

bchar          = achar / ":" / "@" / "/"
                ;; <<less permissive than "pchar">>

enc-auth-type  = 1*achar
                ; encoded version of [IMAP4] "auth-type"

enc-list-mailbox = 1*bchar
                ; encoded version of [IMAP4] "list-mailbox"

enc-mailbox    = 1*bchar
                ; encoded version of [IMAP4] "mailbox"

enc-search     = 1*bchar
                ; encoded version of [IMAPABNF]
                ; "search-program". Note that IMAP4
                ; literals may not be used in
                ; a "search-program", i.e. only
                ; quoted or non-synchronizing
                ; literals (if the server supports
                ; LITERAL+) are allowed.
<<Add text explaining the restriction (code complexity)>>

<<Add example showing use of LITERAL+ with UTF-8 data>>

enc-section    = 1*bchar
                ; encoded version of [IMAP4] "section-spec"

enc-user       = 1*achar
                ; encoded version of [IMAP4] authorization identity
                ; or "userid".

imapurl        = "imap://" iserver "/" [ icommand ]

iauth          = ";AUTH=" ( "*" / enc-auth-type )
```


iccommand = mailboxlist / imessagelist / imessagepart

mailboxlist = [enc-list-mailbox] ";TYPE=" list-type

mailbox-ref = enc-mailbox [uidvalidity]
; <<new production in this draft>>

imessagelist = mailbox-ref ["?" enc-search]
; "enc-search" is [\[URI-GEN\]](#) "query".

imessagepart = mailbox-ref iuid [isection] [ipartial]

ipartial = "/;PARTIAL=" partial-range

isection = "/;SECTION=" enc-section

iserver = [userinfo "@"] host [":" port]
; This is the same as "authority" defined
; in [\[URI-GEN\]](#). See [\[URI-GEN\]](#) for "host"
; and "port" definitions.

iuid = "/;UID=" nz-number
; See [\[IMAP4\]](#) for "nz-number" definition

userinfo = enc-user [iauth] / [enc-user] iauth
; conforms to the generic syntax of
; "userinfo" as defined in [\[URI-GEN\]](#).

list-type = "LIST" / "LSUB"

partial-range = number ["." nz-number]
; partial fetch

uidvalidity = ";UIDVALIDITY=" nz-number
; See [\[IMAP4\]](#) for "nz-number" definition

[13.](#) IANA Considerations

IANA is requested to update "imap" definition in the "Uniform Resource Identifier scheme registry" to point to this document.

The registration teplate (as per <<[draft-hansen-2717bis-2718bis-uri-guidelines-XX.txt](#)>>) is specified in [section 13.1](#) of this document.

13.1. IANA Registration of imap: URI Scheme

This section provides the information required to register the imap: URI scheme.

URI scheme name: imap

Status: permanent

URI scheme syntax:

See [section 12](#) of this RFC.

URI scheme semantics:

The imap: URI scheme is used to designate IMAP servers, mailboxes, messages, MIME bodies [[MIME](#)], and search programs on Internet hosts accessible using the IMAP protocol. <<Discuss configuration storage versa access here?>>

There is no MIME type associated with this URI.

Encoding considerations:

See [Section 9](#) of <<[this RFC]>>.

Applications/protocols that use this URI scheme name:

The imap: URI is intended to be used by applications that might need access to IMAP mailstore. Such applications may include (but not limited to) IMAP-capable web browsers; IMAP clients that wish to access a mailbox, message, or edit a message on the server using <<[CATENATE]>>; <<[SUBMIT]>> clients and servers that are requested to assemble a complete message on submission using <<[BURL]>>.

Interoperability considerations:

Mozilla/Thubderbird use a different imap: scheme internally. <<Is this relevant?>>

Security considerations:

See Security Considerations ([Section 11](#)) of <<[this RFC]>>.

Contact:

Alexey Melnikov <alexey.melnikov@isode.com>

Author/Change controller:

IESG

References:

[This RFC] and [[IMAP4](#)].

14. Normative References

[KEYWORDS] Bradner, "Key words for use in RFCs to Indicate Requirement Levels", [RFC 2119](#), Harvard University, March 1997.

[IMAP4] Crispin, M., "Internet Message Access Protocol - Version 4rev1", [RFC 3501](#), University of Washington, March 2003.

[IMAPABNF] Melnikov, A., and C. Daboo, "Collected extensions to IMAP4 ABNF", work in progress, [draft-melnikov-imap-ext-abnf-XX.txt](#).

[MHTML] Palme, J., Hopmann, A. and N. Shelness, "MIME Encapsulation of Aggregate Documents, such as HTML (MHTML)", [RFC 2557](#), March 1999.

[ABNF] Crocker, Overell, "Augmented BNF for Syntax Specifications: ABNF", [RFC 4234](#), October 2005.

[MIME] Freed, N., Borenstein, N., "Multipurpose Internet Mail Extensions", [RFC 2045](#), Innosoft, First Virtual, November 1996.

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

[UTF-8] Yergeau, F., "UTF-8, a transformation format of ISO 10646", STD 63, [RFC 3629](#), November 2003.

[NAMESPACE] Gahrns, M. and C. Newman, "IMAP4 Namespace", [RFC 2342](#), May 1998.

[ANONYMOUS] K. Zeilenga (Ed.), "The Anonymous SASL Mechanism", work in progress, [draft-ietf-sasl-anon-XX.txt](#) (Obsoletes [RFC 2245](#)).

15. Author's Address

Chris Newman
Sun Microsystems

1050 Lakes Drive
West Covina, CA 91790 USA
EMail: chris.newman@sun.com

Alexey Melnikov (Editor)
Isode Limited
5 Castle Business Village
36 Station Road
Hampton, Middlesex
TW12 2BX, UK
Email: Alexey.Melnikov@isode.com
URI: <http://www.melnikov.ca/>

Stephane H. Maes (Editor)
Oracle Corporation
500 Oracle Parkway
M/S 40p634
Redwood Shores, CA 94065
USA
Phone: +1-650-607-6296
Email: stephane.maes@oracle.com

Appendix A. Sample code

Here is sample C source code to convert between URL paths and IMAP mailbox names, taking into account mapping between IMAP's modified UTF-7 [[IMAP4](#)] and hex-encoded UTF-8 which is more appropriate for URLs. This code has not been rigorously tested nor does it necessarily behave reasonably with invalid input, but it should serve as a useful example. This code just converts the mailbox portion of the URL and does not deal with parameters, query or server components of the URL.

<<Does the code comply with the new URI rules?>>

```
#include <stdio.h>
#include <string.h>

/* hexadecimal lookup table */
static char hex[] = "0123456789ABCDEF";

/* URL unsafe printable characters */
static char urlunsafe[] = " \\#%&+;=<=>?@[\\]^`{|}";

/* UTF7 modified base64 alphabet */
static char base64chars[] =
    "ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789+,";
```



```
#define UNDEFINED 64

/* UTF16 definitions */
#define UTF16MASK    0x03FFUL
#define UTF16SHIFT   10
#define UTF16BASE    0x10000UL
#define UTF16HIGHEST 0xD800UL
#define UTF16HIGHEST 0xDBFFUL
#define UTF16LOSTART 0xDC00UL
#define UTF16LOEND   0xDFFFUL

/* Convert an IMAP mailbox to a URL path
 * dst needs to have roughly 4 times the storage space of src
 * Hex encoding can triple the size of the input
 * UTF-7 can be slightly denser than UTF-8
 * (worst case: 8 octets UTF-7 becomes 9 octets UTF-8)
 */
void MailboxToURL(char *dst, char *src)
{
    unsigned char c, i, bitcount;
    unsigned long ucs4, utf16, bitbuf;
    unsigned char base64[256], utf8[6];

    /* initialize modified base64 decoding table */
    memset(base64, UNDEFINED, sizeof (base64));
    for (i = 0; i < sizeof (base64chars); ++i) {
        base64[base64chars[i]] = i;
    }

    /* loop until end of string */
    while (*src != '\0') {
        c = *src++;
        /* deal with literal characters and &- */
        if (c != '&'amp;' || *src == '-') {
            if (c < ' ' || c > '~' || strchr(urlunsafe, c) != NULL) {
                /* hex encode if necessary */
                dst[0] = '%';
                dst[1] = hex[c >> 4];
                dst[2] = hex[c & 0x0f];
                dst += 3;
            } else {
                /* encode literally */
                *dst++ = c;
            }
            /* skip over the '-' if this is an &- sequence */
            if (c == '&') ++src;
        } else {
            /* convert modified UTF-7 -> UTF-16 -> UCS-4 -> UTF-8 -> HEX */

```



```
bitbuf = 0;
bitcount = 0;
ucs4 = 0;
while ((c = base64[(unsigned char) *src]) != UNDEFINED) {
    ++src;
    bitbuf = (bitbuf << 6) | c;
    bitcount += 6;
    /* enough bits for a UTF-16 character? */
    if (bitcount >= 16) {
        bitcount -= 16;
        utf16 = (bitcount ? bitbuf >> bitcount
                        : bitbuf) & 0xffff;
        /* convert UTF16 to UCS4 */
        if
            (utf16 >= UTF16HIGHESTART && utf16 <= UTF16HIGHEEND) {
            ucs4 = (utf16 - UTF16HIGHESTART) << UTF16SHIFT;
            continue;
        } else if
            (utf16 >= UTF16LOSTART && utf16 <= UTF16LOEND) {
            ucs4 += utf16 - UTF16LOSTART + UTF16BASE;
        } else {
            ucs4 = utf16;
        }
        /* convert UTF-16 range of UCS4 to UTF-8 */
        if (ucs4 <= 0x7fUL) {
            utf8[0] = ucs4;
            i = 1;
        } else if (ucs4 <= 0x7ffUL) {
            utf8[0] = 0xc0 | (ucs4 >> 6);
            utf8[1] = 0x80 | (ucs4 & 0x3f);
            i = 2;
        } else if (ucs4 <= 0xffffUL) {
            utf8[0] = 0xe0 | (ucs4 >> 12);
            utf8[1] = 0x80 | ((ucs4 >> 6) & 0x3f);
            utf8[2] = 0x80 | (ucs4 & 0x3f);
            i = 3;
        } else {
            utf8[0] = 0xf0 | (ucs4 >> 18);
            utf8[1] = 0x80 | ((ucs4 >> 12) & 0x3f);
            utf8[2] = 0x80 | ((ucs4 >> 6) & 0x3f);
            utf8[3] = 0x80 | (ucs4 & 0x3f);
            i = 4;
        }
        /* convert utf8 to hex */
        for (c = 0; c < i; ++c) {
            dst[c] = '%';
            dst[1] = hex[utf8[c] >> 4];
            dst[2] = hex[utf8[c] & 0x0f];
```



```
        dst += 3;
    }
}
/* skip over trailing '-' in modified UTF-7 encoding */
if (*src == '-') ++src;
}
}
/* terminate destination string */
*dst = '\0';
}

/* Convert hex coded UTF-8 URL path to modified UTF-7 IMAP mailbox
 * dst should be about twice the length of src to deal with non-hex
 * coded URLs
 */
void URLtoMailbox(char *dst, char *src)
{
    unsigned int utf8pos, utf8total, i, c, utf7mode, bitstogo, utf16flag;
    unsigned long ucs4, bitbuf;
    unsigned char hextab[256];

    /* initialize hex lookup table */
    memset(hextab, 0, sizeof (hextab));
    for (i = 0; i < sizeof (hex); ++i) {
        hextab[hex[i]] = i;
        if (isupper(hex[i])) hextab[tolower(hex[i])] = i;
    }

    utf7mode = 0;
    utf8total = 0;
    bitstogo = 0;
    while ((c = *src) != '\0') {
        ++src;
        /* undo hex-encoding */
        if (c == '%' && src[0] != '\0' && src[1] != '\0') {
            c = (hextab[src[0]] << 4) | hextab[src[1]];
            src += 2;
        }
        /* normal character? */
        if (c >= ' ' && c <= '~') {
            /* switch out of UTF-7 mode */
            if (utf7mode) {
                if (bitstogo) {
                    *dst++ = base64chars[(bitbuf << (6 - bitstogo)) & 0x3F];
                }
                *dst++ = '-';
                utf7mode = 0;
            }
        }
    }
}
```



```
    }
    *dst++ = c;
    /* encode '&' as '&-' */
    if (c == '&') {
        *dst++ = '-';
    }
    continue;
}
/* switch to UTF-7 mode */
if (!utf7mode) {
    *dst++ = '&';
    utf7mode = 1;
}
/* Encode US-ASCII characters as themselves */
if (c < 0x80) {
    ucs4 = c;
    utf8total = 1;
} else if (utf8total) {
    /* save UTF8 bits into UCS4 */
    ucs4 = (ucs4 << 6) | (c & 0x3FUL);
    if (++utf8pos < utf8total) {
        continue;
    }
} else {
    utf8pos = 1;
    if (c < 0xE0) {
        utf8total = 2;
        ucs4 = c & 0x1F;
    } else if (c < 0xF0) {
        utf8total = 3;
        ucs4 = c & 0x0F;
    } else {
        /* NOTE: can't convert UTF8 sequences longer than 4 */
        utf8total = 4;
        ucs4 = c & 0x03;
    }
    continue;
}
/* loop to split ucs4 into two utf16 chars if necessary */
utf8total = 0;
do {
    if (ucs4 >= UTF16BASE) {
        ucs4 -= UTF16BASE;
        bitbuf = (bitbuf << 16) | ((ucs4 >> UTF16SHIFT)
                                   + UTF16HIGHSTART);
        ucs4 = (ucs4 & UTF16MASK) + UTF16LOSTART;
        utf16flag = 1;
    } else {
```



```
        bitbuf = (bitbuf << 16) | ucs4;
        utf16flag = 0;
    }
    bitstogo += 16;
    /* spew out base64 */
    while (bitstogo >= 6) {
        bitstogo -= 6;
        *dst++ = base64chars[(bitstogo ? (bitbuf >> bitstogo)
                                     : bitbuf)
                              & 0x3F];
    }
    } while (utf16flag);
}
/* if in UTF-7 mode, finish in ASCII */
if (utf7mode) {
    if (bitstogo) {
        *dst++ = base64chars[(bitbuf << (6 - bitstogo)) & 0x3F];
    }
    *dst++ = '-';
}
/* tie off string */
*dst = '\0';
}
```

Appendix B. List of changes since [RFC 2192](#)

Updated boilerplate, list of editor's, etc.
Updated references.
Updated ABNF not to use `_`, to use SP instead of SPACE.
Updated example domains to use example.org.
Fixed ABNF error in "imessagelist" non-terminal.
Updated ABNF, due to changes in [RFC 3501](#), IMAPABNF and [RFC 3986](#).
Renamed "iuserauth" non-terminal to "iuserinfo".
Clarified that the userinfo component describes both authorization identity and mailbox naming scope.
Added "ipartial" specifier that denotes a partial fetch.

Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in [BCP 78](#) and [BCP 79](#).

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Full Copyright Statement

Copyright (C) The Internet Society (2006).

This document is subject to the rights, licenses and restrictions contained in [BCP 78](#), and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.

