

Network Working Group
Internet-Draft
Updates: [959](#) (if approved)
Intended status: Standards Track
Expires: October 1, 2012

J. Klensin
March 30, 2012

FTP TYPE Extension for Internationalized Text
draft-klensin-ftpext-typeu-00

Abstract

The traditional FTP protocol includes a TYPE command to specify the data representation. That command has values for ASCII and EBCDIC text, plus binary ("IMAGE") transmission. As the Internet becomes more international, there is a growing requirement to be able to transmit textual data, encoded in Unicode, in a way that is independent of the coding and line representation forms of particular operating systems. This memo specifies a new FTP representation TYPE value for Unicode data.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

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."

This Internet-Draft will expire on October 1, 2012.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect

Internet-Draft

FTP Unicode TYPE

March 2012

to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Internet-Draft

FTP Unicode TYPE

March 2012

Table of Contents

| | | |
|-----------------------------|---|--------------------|
| 1. | Introduction | 4 |
| 1.1. | Context and Overview | 4 |
| 1.2. | Summary of History of Internationalization of FTP | 4 |
| 1.3. | History of the TYPE Command | 4 |
| 1.4. | Terminology | 6 |
| 1.5. | Discussion List | 6 |
| 2. | Specification | 6 |
| 2.1. | Existing TYPES | 6 |
| 2.2. | Unicode TYPE | 7 |
| 2.3. | Data Structure | 7 |
| 2.4. | Feature Negotiation | 7 |
| 3. | Net-Unicode Format for FTP | 8 |
| 4. | Acknowledgments | 8 |
| 5. | IANA Considerations | 8 |
| 6. | Security Considerations | 8 |
| 7. | References | 9 |
| 7.1. | Normative References | 9 |
| 7.2. | Informative References | 9 |
| Appendix A. | Change Log | 10 |
| A.1. | New Version and File Name: draft-ietf-ftpext2-typeu-00 | 10 |
| A.2. | Version -01 | 10 |
| A.3. | Version -02 | 10 |
| A.4. | Version -03 | 11 |
| A.5. | New Version and File Name: draft-klensin-ftpext-typeu-00 | 11 |
| | Author's Address | 11 |

1. Introduction

1.1. Context and Overview

The traditional FTP protocol, as documented in [RFC 959](#) [[RFC0959](#)], includes a TYPE command to specify the data representation. That command was originally specified as having values for ASCII and EBCDIC text, plus binary ("IMAGE") transmission. The Host Requirements specification [[RFC1123](#)] made other changes to FTP, but did not alter the TYPE command or the environment for which it provided.

As the Internet becomes more international, there is a growing requirement to be able to transmit textual data, encoded in Unicode [[Unicode](#)], in a way that is independent of the coding and line representation forms of particular operating systems. This memo specifies a new FTP TYPE value for Unicode data.

1.2. Summary of History of Internationalization of FTP

[RFC 2640](#) [[RFC2640](#)] is described as providing internationalization of FTP, but only addresses the use of FTP in internationalized (non-ASCII or extended ASCII [[ASCII](#)]) file systems. Its facilities were slightly enhanced in a more general extensions specification [[RFC3659](#)], which builds on a more general FTP extension mechanism [[RFC2389](#)]. The specification in this document addresses the transfer of non-ASCII text files only, building on the TYPE command of the original FTP specification [[RFC0959](#)].

1.3. History of the TYPE Command

[[Note in Draft: AppsAWG: please decide whether this subsection should be included in the final version as informative or dropped as surplus text that doesn't contribute to an implementer understanding of what should be done.]]

When the FTP protocol was first defined in 1971 [[RFC0114](#)], hosts on the ARPANET were extremely diverse. ASCII and EBCDIC were both in active use, as were several completely different character encodings, and ASCII was encoded in a variety of different forms inside different systems (TENEX/TOPS-20, Multics, Unix on 16 and then 32 bit architectures, and the original IBM ASCII all used different encodings. In mid-1972, the late John McCarthy described some aspects of the issues [[RFC0373](#)]. Within a relatively short period of time, it was understood that expecting every system to adapt to the formats of every other system -- a fairly large n-squared problem -- was crazy. At least for text, the solution was to expect all FTP-supporting hosts to convert between their local formats and a

network-standard ASCII encoding and, optionally, to also identify, and permit, EBCDIC files to be transferred in canonical form. The TYPE command was incorporated into FTP to support client specification of those forms for on-the-wire transfer and also to support a pair of TYPEs to support transferring data in forms that were likely to be operating system and hardware specific (see [Section 2.1](#) for more details).

Because of the need to handle these different text character sets and encoding forms without that n-squared problem, TYPE was very commonly used unless it was known that the sending and receiving systems were homogeneous. Several arrangements for single-line FTP commands did not make explicit provision for TYPE specifications, but they tended to make exactly that homogeneity assumption.

By the late 1980s, the ARPANET was converging toward a single basic host system architecture. Almost all significant computer systems used 32 bit architectures or felt an obligation to be able to simulate them. EBCDIC had fallen into disuse on the network. ASCII, encoded right-justified in eight bits with a leading zero, had become pervasive. An Image transfer among diverse systems might well encounter differences with line termination or, occasionally, record

structures rather than stream ones (both of which TYPE A would have smoothed out), but the character encodings were almost certain to be the same. So, with allowances for those line termination problems -- which have been a large issue in many cases -- Image ("binary") and ASCII transfers were almost equivalent and the TYPE command became less-used. Some client FTP implementations also adopted an "automatic" mode in which they tried to determine heuristically, based on either file names or content inspection, whether the relevant file consisted of ASCII characters or binary information and to send the appropriate TYPE command without user intervention. Because there were usually only two choices in practice, they often (but not always) got it right.

However, migration to Unicode has reintroduced many of the old issues. When Unicode is used inside a system, it can be used with several different encodings (e.g., UTF-8 and several variations on UTF-16 (possibly with surrogate pairs), different assumptions about normalization (see "Terminology for Use in Internationalization" [[i18n-terms](#)] for more discussion) and even new variations on line termination conventions. When those files are transferred to another system with Image type, the result may be completely uninterpretable on the target system. This specification extends to non-ASCII character transfers the early concept of having a very small number of common/ canonical network transfer formats for characters, having systems able to convert to or from them. By doing so, it avoids a Unicode version of the n-squared problems and the general confusion

that led to the definition of TYPE.

[1.4.](#) Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [[RFC2119](#)].

This document assumes that the reader is familiar with the terminology of [RFC 959](#). Those terms, especially reply, server-FTP process, user-FTP process, server-PI, user-PI, logical byte size, and user, if used here, are used in the same way. For the convenience of contemporary readers, the terms "client" and "server" are used interchangeably with the historic terms "user-FTP process" and "server-FTP process". The document also assumes the terminology and

changes in the updates to FTP specified in [RFC 1123](#) and [RFC 2389](#) [[RFC2389](#)].

[1.5.](#) Discussion List

[[anchor5: RFC Editor: please remove this section before publication.]]

This proposal is being discussed on the IETF FTP Extensions mailing list at ftpeft@ietf.org. Subscription and other information is available from <https://www.ietf.org/mailman/listinfo/ftpeft>.

[2.](#) Specification

[2.1.](#) Existing TYPEs

The FTP TYPE command, described in [[RFC0959](#)] accepts four possible first argument values, as described below. Note that the descriptions in this subsection are provided for the reader's convenience; the definitions in [RFC 959](#) remain normative.

- A The data are expected to be in, and are transformed by the server if needed to, an ASCII [[ASCII](#)] data stream conforming to the "NVT" specification (See [RFC 959](#) [[RFC0959](#)] and [Appendix B of RFC 5198](#) [[RFC5198](#)] for more information).
- E The data are expected to be in, and are transformed by the server if needed to, an EBCDIC data stream as specified in [RFC 959](#).

- I The data are transferred in "image" form, i.e., exactly as they appear in the server. Because it is the only TYPE form in which true binary data can be transferred, TYPE I is often referred to as "binary" or "binary transfer".
- L The data are transmitted in logical bytes of a size specified in an additional argument. See [RFC 959](#).

Any of these four argument variations to TYPE except "TYPE A" (with non-print format) MAY be rejected by the server-FTP process with a 504 response code if it does not support that type and the necessary conversions.

[2.2.](#) Unicode TYPE

The client-PI MAY transmit TYPE U to the server-PI as an alternative to other TYPE commands and arguments. If it does, the server MAY return reply-code 504, indicating that the TYPE U feature is not supported (unchanged from [RFC 959](#)) or MUST respond to any data retrieval request (e.g., RETR) by sending the data in a stream conformant to the Net-Unicode format specified in [Section 3](#). Similarly, if the client-PI sends TYPE U and the server accepts it, the client MUST send any data streams in that format while the option is in effect. No second parameter is used or permitted for TYPE U.

[2.3.](#) Data Structure

The default and only permitted data structure for TYPE U is "file structure". Use of the STRU command SHOULD be avoided. If is used, its argument MUST be "F".

[2.4.](#) Feature Negotiation

[RFC 2389](#) [[RFC2389](#)] specifies a feature negotiation mechanism for new extensions to FTP. Since the TYPE command is a required part of the base FTP specification, the client-PI is not required to issue the FEAT command prior to issuing TYPE U. However, it MAY do so and Server-FTP implementations that include TYPE U SHOULD support FEAT as described below. If the FEAT command is transmitted from the client-PI to the server-PI, and this extension and FEAT are supported, the response MUST include a TYPE line that lists all TYPE values supported by the server (including the required ones). For example, if an FTP-server supports all of TYPEs A, E, I, and U, the FEAT response line would contain each of the possible arguments separated by semicolons, e.g.,

This specification does not change either [RFC 959](#) or [RFC 2389](#). In particular, no FEAT response line is required for TYPE unless this, or some other, extension to TYPE is supported by the FTP-server.

[3.](#) Net-Unicode Format for FTP

This section specifies a profile of Net-Unicode [[RFC5198](#)] for use with FTP TYPE U.

Unicode characters must be transmitted in UTF-8 [[RFC3629](#)] as specified for Net-Unicode. Because FTP is used in data transmission, the characters and sequences that are discouraged in Section 2 of [RFC 5198](#) are permitted to be transported by FTP. However, line-ending sequences MUST conform to the CRLF convention specified there. Consistent with Paragraph 4 of that Section, strings SHOULD be normalized before transmission if at all possible.

The implicit logical byte size for this transmission type is eight bits.

[4.](#) Acknowledgments

This document draws heavily on [RFC 959](#); appreciation is expressed to its authors and to the authors of [RFC 2398](#). The work of Mark P. Peterson and Douglas J. Papenthien on other FTP extensions finally motivated production of this document in 2008 after a long delay; that contribution is appreciated as well. Specific useful comments on this draft or its immediate predecessors were provided by the late and much-lamented Mike Padlipsky and by Mykyta Yevstifeyev.

[5.](#) IANA Considerations

When this specification is approved, IANA is requested to add an additional table to the FTP Extensions Registry established by [RFC 5797](#) [[RFC5797](#)]. That table should be titled "TYPE command arguments" and should include "A (m) [RFC 959](#)", "E (o) [RFC 959](#)", "I (o) [RFC 959](#)", "L (o) [RFC 959](#)", and "U (o) RFCNNNN".

[6.](#) Security Considerations

This specification makes no substantive change to the FTP command stream (the argument to the standard TYPE command is changed). It

only alters the presentation of data in the data stream. Consequently, it should have no negative security implications that are not already present in the earlier FTP specifications described in [Section 1](#) and in the Net-Unicode specification [[RFC5198](#)]. By specifying an exact canonical form for the identification and transfer of Unicode strings, it may eliminate some problems that might be encountered when such strings are transmitted without identification or without restrictions (e.g., using TYPE I to obtain a "binary" transfer).

[7.](#) References

[7.1.](#) Normative References

- [ASCII] American National Standards Institute (formerly United States of America Standards Institute), "USA Code for Information Interchange", ANSI X3.4-1968, 1968.
- ANSI X3.4-1968 has been replaced by newer versions with slight modifications, but the 1968 version remains definitive for the Internet.
- [RFC0959] Postel, J. and J. Reynolds, "File Transfer Protocol", STD 9, [RFC 959](#), October 1985.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC2389] Hethmon, P. and R. Elz, "Feature negotiation mechanism for the File Transfer Protocol", [RFC 2389](#), August 1998.
- [RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO 10646", STD 63, [RFC 3629](#), November 2003.
- [RFC5198] Klensin, J. and M. Padlipsky, "Unicode Format for Network Interchange", [RFC 5198](#), March 2008.
- [Unicode] The Unicode Consortium. The Unicode Standard, Version 6.0.0, defined by:, "The Unicode Standard, Version 6.0.0", (Mountain View, CA: The Unicode Consortium, 2011. ISBN 978-1-936213-01-6)., <<http://www.unicode.org/versions/Unicode6.0.0/>>.

[7.2.](#) Informative References

[RFC0114] Bhushan, A., "File Transfer Protocol", [RFC 114](#), April 1971.

Klensin

Expires October 1, 2012

[Page 9]

Internet-Draft

FTP Unicode TYPE

March 2012

[RFC0373] McCarthy, J., "Arbitrary Character Sets", [RFC 373](#), July 1972.

[RFC1123] Braden, R., "Requirements for Internet Hosts - Application and Support", STD 3, [RFC 1123](#), October 1989.

[RFC2640] Curtin, B., "Internationalization of the File Transfer Protocol", [RFC 2640](#), July 1999.

[RFC3659] Hethmon, P., "Extensions to FTP", [RFC 3659](#), March 2007.

[RFC5797] Klensin, J. and A. Hoenes, "FTP Command and Extension Registry", [RFC 5797](#), March 2010.

[i18n-terms]

Hoffman, P. and J. Klensin, "Terminology Used in Internationalization in the IETF", June 2011, <<https://datatracker.ietf.org/doc/draft-ietf-appsawg-rfc3536bis/>>.

[Appendix A](#). Change Log

[[anchor13: RFC Editor: Please remove this section]]

[A.1](#). New Version and File Name: [draft-ietf-ftpext2-typeu-00](#)

This version of the document is a slight update to [draft-klensin-ftp-typeu-00](#), posted in July 2008). It includes some updated references to work completed in the interim, information about the FTPEXT2 WG, a new Security Considerations section (omitted from the prior draft), and a few other minor corrections.

[A.2](#). Version -01

- o Corrected a typographical error in the -00 change log entry and made a cosmetic change to that section.
- o Added additional metadata.

- o Added a new introductory subsection ([Section 1.3](#)) to clarify the relationship of this spec to FTP's development and some other ongoing discussions in the IETF.

[A.3.](#) Version -02

- o Changed title per suggestion from Mykyta Yevstifeyev

Klensin

Expires October 1, 2012

[Page 10]

Internet-Draft

FTP Unicode TYPE

March 2012

- o Removed reference to ABNF since it turned out to be possible to write the document without it.
- o Rewrote the IANA Considerations to specify a table for TYPE argument values.
- o Made a number of other relatively minor corrections and clarifications.
- o Updated Unicode reference to 6.0.
- o Moved this section to an appendix for easier handling later.

[A.4.](#) Version -03

- o Draft reissued to reactivate it.
- o Many small editorial changes and clarifications with no substantive change to the specification itself.

[A.5.](#) New Version and File Name: [draft-klensin-ftpext-typeu-00](#)

This version of the document is identical, except for the date, file name and updated mailing list information, to [draft-ietf-ftpext2-typeu-03](#), posted on 12 July 2012. This return the draft to individual submission status after the shutdown of the FTPEXT2 WG.

Author's Address

John C Klensin

1770 Massachusetts Ave, Ste 322
Cambridge, MA 02140
USA

Phone: +1 617 245 1457
Email: john+ietf@jck.com