FTPEXT Working Group
Internet Draft
Expiration Date: August 1999

R. Elz
University of Melbourne

P. Hethmon
Hethmon Brothers

February 1999

Extensions to FTP

draft-ietf-ftpext-mlst-06.txt

Status of this Memo

Abstract

In order to overcome the problems caused by the undefined format of the current FTP LIST command output, a new command is needed to transfer standardized listing information from Server-FTP to Client-FTP. Commands to enable this are defined in this document.

In order to allow consenting clients and servers to interact more freely, a quite basic, and optional, virtual file store structure is defined.

This proposal also extends the FTP protocol to allow character sets other than US-ASCII[1] by allowing the transmission of 8-bit characters and the recommended use of UTF-8[2] encoding.

Much implemented, but long undocumented, mechanisms to permit restarts of interrupted data transfers in STREAM mode, are also included here.

Lastly, the HOST command has been added to allow a style of "virtual site" to be constructed.

New in this version of this document: Servers now permitted to refuse to allow connections with no HOST command if they have good reason (ie: MUST NOT changed to SHOULD NOT). Make it clear that REST is only defined to complete a partially completed transfer, and that other uses lead to undefined results. Make it clear(er) that type=cdir names are allowed only when the type fact is included in the output. Make it clear(er) that facts are only included when relevant, and that the order of their inclusion may be random. Make it clear that a server is not required to list every file that might be in a directory listed (it may hide "hidden" files if it desired). Allow OPTS MLST to be refused before authentication. Several MLST related examples have been added. Created IANA registries of OS specific MLST facts and filetypes. This paragraph will be deleted from the final version of this document.

Table of Contents

1. Introduction

This document amends the File Transfer Protocol (FTP) [3]. Five new commands are added: "SIZE", "HOST", "MDTM", "MLST", and "MLSD". The existing command "REST" is modified. Of those, the "SIZE" and "MDTM" commands, and the modifications to "REST" have been in wide use for many years. The others are new.

These commands allow a client to restart an interrupted transfer in transfer modes not previously supported in any documented way, to support the notion of virtual hosts, and to obtain a directory listing in a machine friendly, predictable, format.

An optional structure for the server's file store (NVFS) is also defined, allowing servers that support such a structure to convey that information to clients in a standard way, thus allowing clients more certainty in constructing and interpreting path names.

2. Document Conventions

This document makes use of the document conventions defined in BCP14 [4]. That provides the interpretation of capitalized imperative words like MUST, SHOULD, etc.

This document also uses notation defined in STD 9 [3]. In particular, the terms "reply", "user", "NVFS", "file", "pathname", "FTP commands", "DTP", "user-FTP process", "user-PI", "user-DTP", "server-FTP process", "server-PI", "server-DTP", "mode", "type", "NVT", "control connection", "data connection", and "ASCII", are all used here as defined there.

Syntax required is defined using the Augmented BNF defined in [5]. Some general ABNF definitions are required throughout the document,

those will be defined later in this section. At first reading, it may be wise to simply recall that these definitions exist here, and skip to the next section.

2.1. Basic Tokens

This document imports the core definitions given in Appendix A of [5]. There definitions will be found for basic ABNF elements like ALPHA, DIGIT, SP, etc. To that, the following terms are added for use in this document.

```
TCHAR          = VCHAR / SP / HTAB    ; visible plus white space
RCHAR          = ALPHA / DIGIT / "," / "." / ":" / "!" /
                 "@" / "#" / "$" / "%" / "^" /
                 "&" / "(" / ")" / "-" / "_" /
                 "+" / "?" / "/" / "\" / "'" /
                 DQUOTE   ; <"> -- double quote character (%x22)
```

The VCHAR (from [5]), TCHAR, and RCHAR types give basic character types from varying sub-sets of the ASCII character set for use in various commands and responses.

```
token          = 1*RCHAR
```

A "token" is a string whose precise meaning depends upon the context in which it is used. In some cases it will be a value from a set of possible values maintained elsewhere. In others it might be a string invented by one party to an FTP conversation from whatever sources it finds relevant.

Note that in ABNF, string literals are case insensitive. That convention is preserved in this document, and implies that FTP commands added by this specification have names that can be represented in any case. That is, "MDTM" is the same as "mdtm", "Mdtm" and "MdTm" etc. However note that ALPHA, in particular, is case sensitive. That implies that a "token" is a case sensitive value. That implication is correct.

2.2. Pathnames

Various FTP commands take pathnames as arguments, or return pathnames in responses. When the MLST command is supported, as indicated in the response to the FEAT command [6], pathnames are to be transferred in one of the following two formats.

```
pathname       = utf-8-name / raw
utf-8-name     = <a UTF-8 encoded Unicode string>
raw            = <any string not being a valid UTF-8 encoding>
```

Which format is used is at the option of the user-PI or server-PI sending the pathname. UTF-8 encodings [2] contain enough internal structure that it is always, in practice, possible to determine whether a UTF-8 or raw encoding has been used, in those cases where it matters. While it is useful for the user-PI to be able to correctly display a pathname received from the server-PI to the user, it is far more important for the user-PI to be able to retain and retransmit the identical pathname when required. Implementations are advised against converting a UTF-8 pathname to a local encoding, and then attempting to invert the encoding later. Note that ASCII is a subset of UTF-8.

Unless otherwise specified, the pathname is terminated by the CRLF that terminates the FTP command, or by the CRLF that ends a reply. Any trailing spaces preceding that CRLF form part of the name. Exactly one space will precede the pathname and serve as a separator from the preceding syntax element. Any additional spaces form part of the pathname. See [7] for a fuller explanation of the character encoding issues. All implementations supporting MLST MUST support [7].

Implementations should also beware that the control connection uses Telnet NVT conventions [8], and that the Telnet IAC character, if part of a pathname sent over the control connection, MUST be correctly escaped as defined by the Telnet protocol.

Implementors should also be aware that although Telnet NVT conventions are used over the control connections, Telnet option negotiation MUST NOT be attempted. See section 4.1.2.12 of [9].

2.2.1. Pathname Syntax

Except where TVFS is supported (see section 7) this specification imposes no syntax upon pathnames. Nor does it restrict the the character set from which pathnames are created. This does not imply that the NVFS is required to make sense of all possible pathnames. Server-PIs may restrict the syntax of valid pathnames in their NVFS in any manner appropriate to their implementation or underlying filesystem. Similarly, a server-PI may parse the pathname, and assign meaning to the components detected.

2.2.2. Wildcarding

For the commands defined in this specification, all pathnames are to be treated literally. That is, for a pathname given as a parameter to a command, the file whose name is identical to the pathname given is implied. No characters from the pathname may be treated as special or "magic", thus no pattern matching (other than for exact equality) between the pathname given and the files present in the NVFS of the Server-FTP is permitted.

Clients that desire some form of pattern matching functionality must obtain a listing of the relevant directory, or directories, and implement their own filename selection procedures.

2.3. Times

The syntax of a time value is:

```
time-val       = 14DIGIT [ "." 1*DIGIT ]
```

The leading, mandatory, fourteen digits are to be interpreted as, in order from the leftmost, four digits giving the year, with a range of 1000-9999, two digits giving the month of the year, with a range of 01-12, two digits giving the day of the month, with a range of 01-31, two digits giving the hour of the day, with a range of 00-23, two digits giving minutes past the hour, with a range of 00-59, and finally, two digits giving seconds past the minute, with a range of 00-60 (with 60 being used only at a leap second). Years in the tenth century, and earlier, cannot be expressed. This is not considered a serious defect of the protocol.

The optional digits, which are preceded by a period, give decimal fractions of a second. These may be given to whatever precision is appropriate to the circumstance, however implementations MUST NOT add precision to time-vals where that precision does not exist in the underlying value being transmitted.

Symbolically, a time-val may be viewed as

```
YYYYMMDDHHMMSS.sss
```

The "." and subsequent digits ("sss") are optional.

Time values are always represented in UTC (GMT), and in the Gregorian calendar regardless of what calendar may have been in use at the date and time indicated at the location of the server-PI.

The technical differences between GMT, TAI, UTC, UT1, UT2, etc, are not considered here. A server-FTP process should always use the same time reference, so the times it returns will be consistent. Clients are not expected to be time synchronized with the server, so the possible difference in times that might be reported by the different time standards is not considered important.

2.4. Server Replies

Section 4.2 of [3] defines the format and meaning of replies by the server-PI to FTP commands from the user-PI. Those reply conventions are used here without change.

```
error-response = error-code SP *TCHAR CRLF
error-code     = ("4" / "5") 2DIGIT
```

Implementors should note that the ABNF syntax (which was not used in [3]) used in this document, and other FTP related documents, sometimes shows replies using the one line format. Unless otherwise explicitly stated, that is not intended to imply that multi-line responses are not permitted. Implementors should assume that, unless stated to the contrary, any reply to any FTP command (including QUIT) may be of the multiline format described in [3].

Throughout this document, replies will be identified by the three digit code that is their first element. Thus the term "500 reply" means a reply from the server-PI using the three digit code "500".

3. File Modification Time (MDTM)

The FTP command, MODIFICATION TIME (MDTM), can be used to determine when a file in the server NVFS was last modified. This command has existed in many FTP servers for many years, as an adjunct to the REST command for STREAM mode, thus is widely available. However, where supported, the "mtime" fact which can be provided in the result from the new MLST command is recommended as a superior alternative.

When attempting to restart a RETRieve, if the User FTP makes use of the MDTM command, or "mtime" fact, it can check and see if the modification time of the source file is more recent than the modification time of the partially transferred file. If it is, then most likely the source file has changed and it would be unsafe to restart in the middle of the file transfer.

When attempting to restart a STORe, the User FTP can use the MDTM command to discover the modification time of the partially transferred file. If it is older than the modification time of the file that is about to be STORed, then most likely the source file has

changed and it would be unsafe to restart in the middle of the file transfer.

Note that using MLST (described below) where available, can provide this information, and much more, thus giving an even better indication that a file has changed, and that restarting a transfer would not give valid results.

Note that this is applicable to any RESTart attempt, regardless of the mode of the file transfer.

3.1. Syntax

The syntax for the MDTM command is:

```
mdtm          = "MdTm" SP pathname CRLF
```

As with all FTP commands, the "MDTM" command label is interpreted in a case insensitive manner.

The "pathname" specifies an object in the NVFS which may be the object of a RETR command. Attempts to query the modification time of files that are unable to be retrieved generate undefined responses.

The server-PI will respond to the MDTM command with a 213 reply giving the last modification time of the file whose pathname was supplied, or a 550 reply if the file does not exist, the modification time is unavailable, or some other error has occurred.

```
mdtm-response = "213" SP time-val CRLF /
                error-response
```

3.2. Error responses

Where the command is correctly parsed, but the modification time is not available, either because the pathname identifies no existing entity, or because the information is not available for the entity named, then a 550 reply should be sent. Where the command cannot be correctly parsed, a 500 or 501 reply should be sent, as specified in [3].

3.3. FEAT response for MDTM

When replying to the FEAT command [6], a FTP server process that supports the MDTM command MUST include a line containing the single word "MDTM". This MAY be sent in upper or lower case, or a mixture of both (it is case insensitive) but SHOULD be transmitted in upper case only. That is, the response SHOULD be

```
C> Feat
S> 211- <any descriptive text>
S>  ...
S>  MDTM
S>  ...
S> 211 End
```

The ellipses indicate placeholders where other features may be included, and are not required. The one space indentation of the feature lines is mandatory [6].

3.4. MDTM Examples

If we assume the existence of three files, A B and C, and a directory D, and no other files at all, then the MTDM command may behave as indicated. The "C>" lines are commands from user-PI to server-PI, the "S>" lines are server-PI replies.

```
C> MDTM A
S> 213 19980615100045.014
C> MDTM B
S> 213 19980615100045.014
C> MDTM C
S> 213 19980705132316
C> MDTM D
S> 550 D is not retrievable
C> MDTM E
S> 550 No file named "E"
```

From that we can conclude that both A and B were last modified at the same time (to the nearest millisecond), and that C was modified 21 days and several hours later.

The times are in GMT, so file A was modified on the 15th of June, 1998, at approximately 11am in London (summer time was then in effect), or perhaps at 8pm in Melbourne, Australia, or at 6am in New York. All of those represent the same absolute time of course. The location where the file was modified, and consequently the local wall clock time at that location, is not available.

4. File SIZE

The FTP command, SIZE OF FILE (SIZE), is used to obtain the transfer size of a file from the server-FTP process. That is, the exact number of octets (8 bit bytes) which would be transmitted over the data connection should that file be transmitted. This value will change depending on the current STRUcture, MODE and TYPE of the data connection, or a data connection which would be created were one created now. Thus, the result of the SIZE command is dependent on the currently established STRU, MODE and TYPE parameters.

The SIZE command returns how many octets would be transferred if the file were to be transferred using the current transfer structure, mode and type. This command is normally used in conjunction with the RESTART (REST) command. The server-PI might need to read the partially transferred file, do any appropriate conversion, and count the number of octets that would be generated when sending the file in order to correctly respond to this command. Estimates of the file transfer size MUST NOT be returned, only precise information is acceptable.

4.1. Syntax

The syntax of the SIZE command is:

```
size          = "Size" SP pathname CRLF
```

The server-PI will respond to the SIZE command with a 213 reply giving the transfer size of the file whose pathname was supplied, or an error response if the file does not exist, the size is unavailable, or some other error has occurred. The value returned is in a format suitable for use with the RESTART (REST) command for mode STREAM, provided the transfer mode and type are not altered.

```
size-response = "213" SP 1*DIGIT CRLF /
                error-response
```

4.2. Error responses

Where the command is correctly parsed, but the size is not available, either because the pathname identifies no existing entity, or because the entity named cannot be transferred in the current MODE and TYPE (or at all), then a 550 reply should be sent. Where the command cannot be correctly parsed, a 500 or 501 reply should be sent, as specified in [3].

4.3. FEAT response for SIZE

When replying to the FEAT command [6], a FTP server process that supports the SIZE command MUST include a line containing the single word "SIZE". This word is case insensitive, and MAY be sent in any mixture of upper or lower case, however it SHOULD be sent in upper case. That is, the response SHOULD be

```
C> FEAT
S> 211- <any descriptive text>
S>  ...
S>  SIZE
S>  ...
S> 211 END
```

The ellipses indicate placeholders where other features may be included, and are not required. The one space indentation of the feature lines is mandatory [6].

4.4. Size Examples

Consider a text file "A" stored on a Unix(TM) server where each end of line is represented by a single octet. Assume the file contains 112 lines, and 1830 octets total. Then the SIZE command would produce:

```
C> TYPE I
S> 200 Type set to I.
C> size A
S> 213 1830
C> TYPE A
S> 200 Type set to A.
C> Size A
S> 213 1942
```

Notice that with TYPE=A the SIZE command reports an extra 112 octets. Those are the extra octets that need to be inserted, one at the end of each line, to provide correct end of line semantics for a transfer using TYPE=A. Other systems might need to make other changes to the transfer format of files when converting between TYPEs and MODEs. The SIZE command takes all of that into account.

Since calculating the size of a file with this degree of precision may take considerable effort on the part of the server-PI, user-PIs should not used this command unless this precision is essential (such as when about to restart an interrupted transfer). For other uses, the "Size" fact of the MLST command (see section 8.5.7) ought be requested.

5. Restart of Interrupted Transfer (REST)

To avoid having to resend the entire file if the file is only partially transferred, both sides need some way to be able to agree on where in the data stream to restart the data transfer.

The FTP specification [3] includes three modes of data transfer, Stream, Block and Compressed. In Block and Compressed modes, the data stream that is transferred over the data connection is formatted, allowing the embedding of restart markers into the stream. The sending DTP can include a restart marker with whatever information it needs to be able to restart a file transfer at that point. The receiving DTP can keep a list of these restart markers, and correlate them with how the file is being saved. To restart the file transfer, the receiver just sends back that last restart marker, and both sides know how to resume the data transfer. Note that there are some flaws in the description of the restart mechanism in RFC 959 [3]. See section 4.1.3.4 of RFC 1123 [9] for the corrections.

5.1. Restarting in STREAM Mode

In Stream mode, the data connection contains just a stream of unformatted octets of data. Explicit restart markers thus cannot be inserted into the data stream, they would be indistinguishable from data. For this reason, the FTP specification [3] did not provide the ability to do restarts in stream mode. However, there is not really a need to have explicit restart markers in this case, as restart markers can be implied by the octet offset into the data stream.

Because the data stream defines the file in STREAM mode, a different data stream would represent a different file. Thus, an offset will always represent the same position within a file. On the other hand, in other modes than STREAM, the same file can be transferred using quite different octet sequences, and yet be reconstructed into the one identical file. Thus an offset into the data stream in transfer modes other than STREAM would not give an unambiguous restart point.

If the data representation TYPE is IMAGE, and the STRUcture is File, for many systems the file will be stored exactly in the same format as it is sent across the data connection. It is then usually very easy for the receiver to determine how much data was previously received, and notify the sender the offset where the transfer should be restarted. In other representation types and structures more effort will be required, but it remains always possible to determine the offset with finite, but perhaps non-negligible, effort. In the worst case an FTP process may need to open a data connection to itself, set the appropriate transfer type and structure, and actually transmit the file, counting the transmitted octets.

If the user-FTP process is intending to restart a retrieve, it will directly calculate the restart marker, and send that information in the RESTart command. However, if the user-FTP process is intending to restart sending the file, it needs to be able to determine how much data was previously sent, and correctly received and saved. A new FTP command is needed to get this information. This is the purpose of the SIZE command, as documented in section 4.

5.2. Error Recovery and Restart

STREAM MODE transfers with FILE STRUcture may be restarted even though no restart marker has been transferred in addition to the data itself. This is done by using the SIZE command, if needed, in combination with the RESTART (REST) command, and one of the standard file transfer commands.

When using TYPE ASCII or IMAGE, the SIZE command will return the number of octets that would actually be transferred if the file were to be sent between the two systems. I.e. with type IMAGE, the SIZE normally would be the number of octets in the file. With type ASCII, the SIZE would be the number of octets in the file including any modifications required to satisfy the TYPE ASCII CR-LF end of line convention.

5.3. Syntax

The syntax for the REST command when the current transfer mode is STREAM is:

```
rest          = "Rest" SP 1*DIGIT CRLF
```

The numeric value gives the number of octets of the immediately following transfer to not actually send, effectively causing the transmission to be restarted at a later point. A value of zero effectively disables restart, causing the entire file to be transmitted. The server-PI will respond to the REST command with a 350 reply, indicating that the REST parameter has been saved, and that another command, which should be either RETR or STOR, should then follow to complete the restart.

```
rest-response = "350" SP *TCHAR CRLF /
                error-response
```

Server-FTP processes may permit transfer commands other than RETR and STOR, such as APPE and STOU, to complete a restart, however, this is not recommended. STOU (store unique) is undefined in this usage, as storing the remainder of a file into a unique filename is rarely going to be useful. If APPE (append) is permitted, it MUST act

identically to STOR when a restart marker has been set. That is, in both cases, octets from the data connection are placed into the file at the location indicated by the restart marker value.

The REST command is intended to complete a failed transfer. Use with RETR is comparatively well defined in all cases, as the client bears the responsibility of merging the retrieved data with the partially retrieved file. If it chooses to use the data obtained other than to complete an earlier transfer, or if it chooses to re-retrieve data that had been retrieved before, that is its choice. With STOR, however, the server must insert the data into the file named. The results are undefined if a client uses REST to do other than restart to complete a transfer of a file which had previously failed to completely transfer. In particular, if the restart marker set with a REST command is not at the end of the data currently stored at the server, as reported by the server, or if insufficient data are provided in a STOR that follows a REST to extend the destination file to at least its previous size, then the effects are undefined.

The REST command must be the last command issued before the data transfer command which is to cause a restarted rather than complete file transfer. The effect of issuing a REST command at any other time is undefined. The server-PI may react to a badly positioned REST command by issuing an error response to the following command, not being a restartable data transfer command, or it may save the restart value and apply it to the next data transfer command, or it may silently ignore the inappropriate restart attempt. Because of this, a user-PI that has issued a REST command, but which has not successfully transmitted the following data transfer command for any reason, should send another REST command before the next data transfer command. If that transfer is not to be restarted, then "REST 0" should be issued.

An error-response will follow a REST command only when the server does not implement the command, or the restart marker value is syntactically invalid for the current transfer mode. That is, in STREAM mode, if something other than one or more digits appears in the parameter to the REST command. Any other errors, including such problems as restart marker out of range, should be reported when the following transfer command is issued. Such errors will cause that transfer request to be rejected with an error indicating the invalid restart attempt.

5.4. FEAT response for REST

Where a server-FTP process supports RESTart in STREAM mode, as specified here, it MUST include in the response to the FEAT command [6], a line containing exactly the string "REST STREAM". This string is not case sensitive, but SHOULD be transmitted in upper case. Where REST is not supported at all, or supported only in block or compressed modes, the REST line MUST NOT be included in the FEAT response. Where required, the response SHOULD be

```
C> feat
S> 211- <any descriptive text>
S>  ...
S>  REST STREAM
S>  ...
S> 211 end
```

The ellipses indicate placeholders where other features may be included, and are not required. The one space indentation of the feature lines is mandatory [6].

5.5. REST Example

Assume that the transfer of a largish file has previously been interrupted after 802816 octets had been received, that the previous transfer was with TYPE=I, and that it has been verified that the file on the server has not since changed.

```
C> TYPE I
S> 200 Type set to I.
C> PORT 127,0,0,1,15,107
S> 200 PORT command successful.
C> REST 802816
S> 350 Restarting at 802816. Send STORE or RETRIEVE
C> RETR cap60.pl198.tar
S> 150 Opening BINARY mode data connection
[...]
S> 226 Transfer complete.
```

6. Virtual FTP servers

It has become common in the Internet for many domain names to be allocated to a single IP address. This has introduced the concept of a "virtual host", where a host appears to exist as an independent entity, but in reality shares all of its resources with one, or more, other such hosts.

Such an arrangement presents some problems for FTP Servers, as all the FTP Server can detect is an incoming FTP connection to a particular IP address. That is, all domain names which share the IP address also share the FTP server, and more importantly, its NVFS. This means that the various virtual hosts cannot offer different virtual file systems to clients, nor can they offer different authentication systems.

No scheme can overcome this without modifications of some kind to the user-PI and the user-FTP process. That process is the only entity that knows which virtual host is required. It has performed the domain name to IP address translation, and thus has the original domain name available.

One method which could be used to allow a style of virtual host would be for the client to simply send a "CWD" command after connecting, using the virtual host name as the argument to the CWD command. This would allow the server-FTP process to implement the filestores of the virtual hosts as sub-directories in its NVFS. This is simple, and supported by essentially all server-FTP implementations without requiring any code changes.

While that method is simple to describe, and to implement, it suffers from several drawbacks. First, the "CWD" command is available only after the user-PI has authenticated itself to the server-FTP process. Thus, all virtual hosts would be required to share a common authentication scheme. Second, either the server-FTP process needs to be modified to understand the special nature of this first CWD command, negating most of the advantage of this scheme, or all users must see the same identical NVFS view upon connecting (they must connect in the same initial directory) or the NVFS must implement the full set of virtual host directories at each possible initial directory for any possible user, or the virtual host will not be truly transparent. Third, and again unless the server is specially modified, a user connecting this way to a virtual host would be able to trivially move to any other virtual host supported at the same server-FTP process, exposing the nature of the virtual host.

Other schemes overloading other existing FTP commands have also been proposed. None of those have sufficient merit to be worth discussion.

The conclusion from the examination of the possibilities seems to be that to obtain an adequate emulation of "real" FTP servers, server modifications to support virtual hosts are required. A new command seems most likely to provide the support required.

6.1. The HOST command

A new command "HOST" is added to the FTP command set to allow server-FTP process to determine to which of possibly many virtual hosts the client wishes to connect. This command is intended to be issued before the user is authenticated, allowing the authentication scheme, and set of legal users, to be dependent upon the virtual host chosen. Server-FTP processes may, if they desire, permit the HOST command to be issued after the user has been authenticated, or may treat that as an erroneous sequence of commands. The behavior of the server-FTP process which does allow late HOST commands is undefined. One reasonable interpretation would be for the user-PI to be returned to the state it existed after the TCP connection was first established, before user authentication.

Servers should note that the response to the HOST command is the ideal time to send their "welcome" message. This allows the message to be personalized for any virtual hosts that are supported, and also allows the client to have determined supported languages, or representations, for the message, and other messages, via the FEAT response, and selected an appropriate one via the OPTS command. See [7] for more information.

6.2. Syntax of the HOST command

The HOST command is defined as follows.

```
host-command     = "Host" SP hostname CRLF
hostname         = 1*DNCHAR 1*( "." 1*DNCHAR ) [ "." ]
DNCHAR           = ALPHA / DIGIT / "-" / "_" / "$" /
                   "!" / "%" / "[" / "]" / ":"
host-response    = host-ok / error-response
host-ok          = "234" [ SP *TCHAR ] CRLF
```

As with all FTP commands, the "host" command word is case independent, and may be specified in any character case desired.

The "hostname" given as a parameter specifies the virtual host to which access is desired. It should normally be the same name that was used to obtain the IP address to which the FTP control connection was made, after any client conversions to convert an abbreviated or local alias to a complete (fully qualified) domain name, but before resolving a DNS alias (owner of a CNAME resource record) to its canonical name.

If the client was given a network literal address, and consequently was not required to derive it from a hostname, it should send the HOST command with the network address, as specified to it, enclosed

in brackets (after eliminating any syntax, which might also be brackets, but is not required to be, from which the server deduced that a literal address had been specified.) That is, for example

HOST [10.1.2.3]

should be sent if the client had been instructed to connect to "10.1.2.3", or "[10.1.2.3]", or perhaps even IPv4:10.1.2.3. The method of indicating to a client that a literal address is to be used is beyond the scope of this specification.

The parameter is otherwise to be treated as a "complete domain name", as that term is defined in section 3.1 of RFC 1034 [10]. That implies that the name is to be treated as a case independent string, in that upper case ASCII characters are to be treated as equivalent to the corresponding lower case ASCII characters, but otherwise preserved as given. It also implies some limits on the length of the parameter and of the components that create its internal structure. Those limits are not altered in any way here.

RFC 1034 imposes no other restrictions upon what kinds of names can be stored in the DNS. Nor does RFC 1035. This specification, however, allows only a restricted set of names for the purposes of the HOST command. Those restrictions can be inferred from the ABNF grammar given for the "hostname".

6.3. HOST command semantics

Upon receiving the HOST command, before authenticating the user-PI, a server-FTP process should validate that the hostname given represents a valid virtual host for that server, and if so, establish the appropriate environment for that virtual host. The meaning of that is not specified here, and may range from doing nothing at all, or performing a simple change of working directory, to much more elaborate state changes, as required.

If the hostname specified is unknown at the server, or if the server is otherwise unwilling to treat the particular connection as a connection to the hostname specified, the server will respond with a 504 reply.

Note: servers may require that the name specified is in some sense equivalent to the particular network address that was used to reach the server.

6.3.1. The REIN command

As specified in [3], the REIN command returns the state of the connection to that it was immediately after the transport connection was opened. That is not changed here. The effect of a HOST command will be lost if a REIN command is performed, a new HOST command must be issued.

6.3.2. User-PI usage of HOST

A user-PI that conforms to this specification, MUST send the HOST command after opening the transport connection, or after any REIN command, before attempting to authenticate the user with the USER command.

The following state diagram shows a typical sequence of flow of control, where the "B" (begin) state is assumed to occur after the transport connection has opened, or a REIN command has succeeded. Other commands (such as FEAT [6]) which require no authentication may have intervened. This diagram is modeled upon (and largely borrowed from) the similar diagram in section 6 of [3].

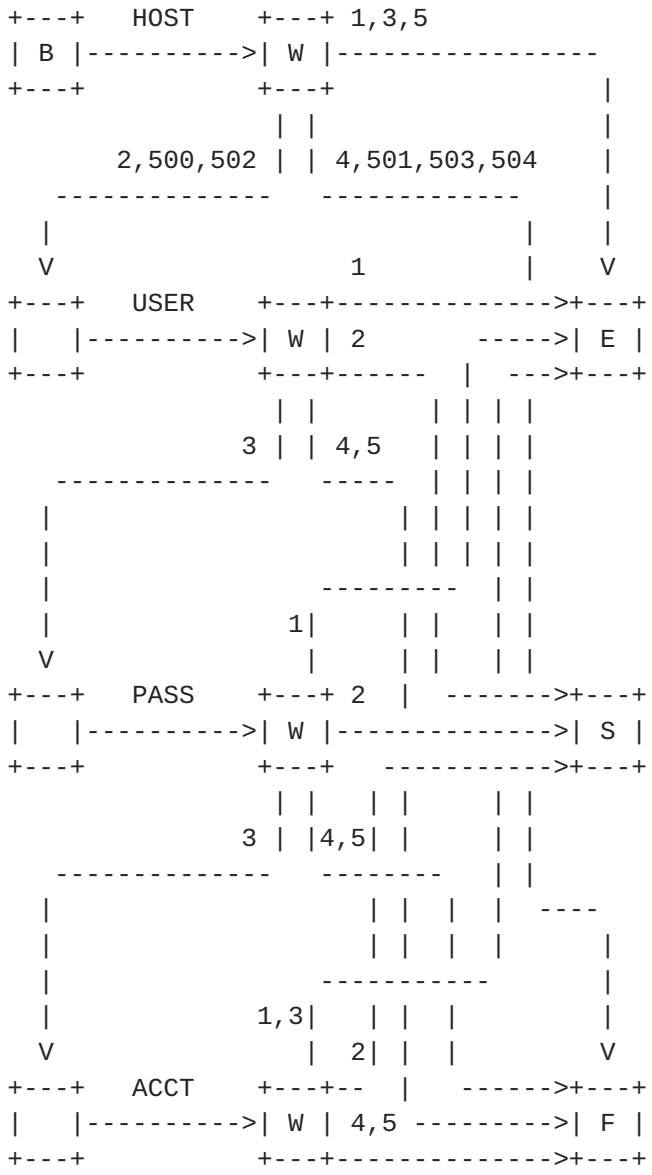
In this diagram, a three digit reply indicates that precise server reply code, a single digit on a reply path indicates any server reply beginning with that digit, other than any three digit replies that might take another path.

6.4. HOST command errors

The server-PI shall reply with a 500 or 502 reply if the HOST command is unrecognized or unimplemented. A 503 reply may be sent if the HOST command is given after a previous HOST command, or after a user has been authenticated. Alternately, the server may accept the command at such a time, with server defined behavior. A 501 reply should be sent if the hostname given is syntactically invalid, and a 504 reply if a syntactically valid hostname is not a valid virtual host name for the server.

In all such cases the server-FTP process should act as if no HOST command had been given.

A user-PI receiving a 500 or 502 reply should assume that the server-PI does not implement the HOST command style virtual server. It may then proceed to login as if the HOST command had succeeded, and perhaps, attempt a CWD command to the hostname after authenticating the user.



A user-PI receiving some other error reply should assume that the virtual HOST is unavailable, and terminate communications.

A server-PI that receives a USER command, beginning the authentication sequence, without having received a HOST command SHOULD NOT reject the USER command. Clients conforming to earlier FTP specifications do not send HOST commands. In this case the server may act as if some default virtual host had been explicitly selected, or may enter an environment different from that of all supported virtual hosts, perhaps one in which a union of all available accounts exists, and which presents a NVFS which appears to contain sub-directories containing the NVFS for all virtual hosts supported.

6.5. FEAT response for HOST command

A server-FTP process that supports the host command, and virtual FTP servers, MUST include in the response to the FEAT command [6], a feature line indicating that the HOST command is supported. This line should contain the single word "HOST". This MAY be sent in upper or lower case, or a mixture of both (it is case insensitive) but SHOULD be transmitted in upper case only. That is, the response SHOULD be

```
C> Feat
S> 211- <any descriptive text>
S>  ...
S>  HOST
S>  ...
S> 211 End
```

The ellipses indicate placeholders where other features may be included, and are not required. The one space indentation of the feature lines is mandatory [6].

7. A Trivial Virtual File Store (TVFS)

Traditionally, FTP has placed almost no constraints upon the file store (NVFS) provided by a server. This specification does not alter that. However, it has become common for servers to attempt to provide at least file system naming conventions modeled loosely upon those of the UNIX(TM) filesystem. That is, a tree structured filesystem, built of directories, each of which can contain other directories, or other kinds of files, or both. Each file and directory has a file name relative to the directory that contains it, except for the directory at the root of the tree, which is contained in no other directory, and hence has no name of its own.

That which has so far been described is perfectly consistent with the standard FTP NVFS and access mechanisms. The "CWD" command is used to move from one directory to an embedded directory. "CDUP" may be provided to return to the parent directory, and the various file manipulation commands ("RETR", "STOR", the rename commands, etc) are used to manipulate files within the current directory.

However, it is often useful to be able to reference files other than by changing directories, especially as FTP provides no guaranteed mechanism to return to a previous directory. The Trivial Virtual File Store (TVFS), if implemented, provides that mechanism.

7.1. TVFS File Names

Where a server implements the TVFS, no elementary filename shall contain the character "/". Where the underlying natural file store permits files, or directories, to contain the "/" character in their names, a server-PI implementing TVFS must encode that character in some manner whenever file or directory names are being returned to the user-PI, and reverse that encoding whenever such names are being accepted from the user-PI.

The encoding method to be used is not specified here. Where some other character is illegal in file and directory names in the underlying filestore, a simple transliteration may be sufficient. Where there is no suitable substitute character a more complex encoding scheme, possibly using an escape character, is likely to be required.

With the one exception of the unnamed root directory, a TVFS file name may not be empty. That is, all other file names contain at least one character.

With the sole exception of the "/" character, any valid IS10646 character [11] may be used in a TVFS filename. When transmitted, file name characters are encoded using the UTF-8 encoding [2].

7.2. TVFS Path Names

A TVFS "Path Name" combines the file or directory name of a target file or directory, with the directory names of zero or more enclosing directories, so as to allow the target file or directory to be referenced other than when the server's "current working directory" is the directory directly containing the target file or directory.

By definition, every TVFS file or directory name is also a TVFS path name. Such a path name is valid to reference the file from the directory containing the name, that is, when that directory is the server-FTP's current working directory.

Other TVFS path names are constructed by prefixing a path name by a name of a directory from which the path is valid, and separating the two with the "/" character. Such a path name is valid to reference the file or directory from the directory containing the newly added directory name.

Where a path name has been extended to the point where the directory added is the unnamed root directory, the path name will begin with the "/" character. Such a path is known as a fully qualified path name. Fully qualified paths may, obviously, not be further extended,

as, by definition, no directory contains the root directory. Being unnamed, it cannot be represented in any other directory. A fully qualified path name is valid to reference the named file or directory from any location (that is, regardless of what the current working directory may be) in the virtual file store.

Any path name which is not a fully qualified path name may be referred to as a "relative path name" and will only correctly reference the intended file when the current working directory of the server-FTP is a directory from which the relative path name is valid.

As a special case, the path name "/" is defined to be a fully qualified path name referring to the root directory. That is, the root directory does not have a directory (or file) name, but does have a path name. This special path name may be used only as is as a reference to the root directory. It may not be combined with other path names using the rules above, as doing so would lead to a path name containing two consecutive "/" characters, which is an undefined sequence.

7.2.1. Notes

+ It is not required, or expected, that there be only one fully qualified path name that will reference any particular file or directory.
+ As a caveat, though the TVFS file store is basically tree structured, there is no requirement that any file or directory have only one parent directory.
+ As defined, no TVFS path name will ever contain two consecutive "/" characters. Such a name is not illegal however, and may be defined by the server for any purpose that suits it. Clients implementing this specification should not assume any semantics at all for such names.
+ Similarly, other than the special case path that refers to the root directory, no TVFS path name constructed as defined here will ever end with the "/" character. Such names are also not illegal, but are undefined.
+ While any legal IS10646 character is permitted to occur in a TVFS file or directory name, other than "/", server FTP implementations are not required to support all possible IS10646 characters. The subset supported is entirely at the discretion of the server. The case (where it exists) of the characters that make up file, directory, and path names may be significant. Unless determined otherwise by means unspecified here, clients should assume that all such names are comprised of characters whose case is significant. Servers are free to treat case (or any other attribute) of a name as irrelevant, and hence map two names which appear to be distinct onto the same underlying file.

+ There are no defined "magic" names, like ".", ".." or "C:". Servers may implement such names, with any semantics they choose, but are not required to do so.
+ TVFS imposes no particular semantics or properties upon files, guarantees no access control schemes, or any of the other common properties of a file store. Only the naming scheme is defined.

7.3. FEAT Response for TVFS

In response to the the FEAT command [6] a server that wishes to indicate support for the TVFS as defined here will include a line that begins with the four characters "TVFS" (in any case, or mixture of cases, upper case is not required). Servers SHOULD send upper case.

Such a response to the FEAT command MUST NOT be returned unless the server implements TVFS as defined here.

Later specifications may add to the TVFS definition. Such additions should be notified by means of additional text appended to the TVFS feature line. Such specifications, if any, will define the extra text.

Until such a specification is defined, servers should not include anything after "TVFS" in the TVFS feature line. Clients, however, should be prepared to deal with arbitrary text following the four defined characters, and simply ignore it if unrecognized.

A typical response to the FEAT command issued by a server implementing only this specification would be:

```
C> feat
S> 211- <any descriptive text>
S>  ...
S>  TVFS
S>  ...
S> 211 end
```

The ellipses indicate placeholders where other features may be included, and are not required. The one space indentation of the feature lines is mandatory [6], and is not counted as one of the first four characters for the purposes of this feature listing.

The TVFS feature adds no new commands to the FTP command repertoire.

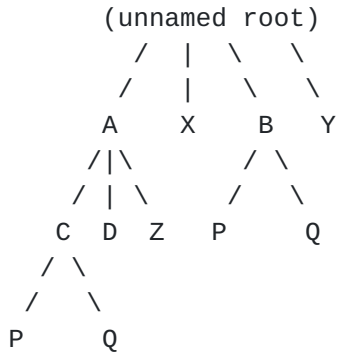
7.4. OPTS for TVFS

There are no options in this TVFS specification, and hence there is no OPTS command defined.

7.5. TVFS Examples

Assume a TVFS file store is comprised of a root directory, which contains two directories (A and B) and two non-directory files (X and Y). The A directory contains two directories (C and D) and one other file (Z). The B directory contains just two non-directory files (P and Q) and the C directory also two non-directory files (also named P and Q, by chance). The D directory is empty, that is, contains no files or directories.
11.ft 1

This structure may depicted graphically as...



Given this structure, the following fully qualified path names exist.

```
/
/A
/B
/X
/Y
/A/C
/A/D
/A/Z
/A/C/P
/A/C/Q
/B/P
/B/Q
```

It is clear that none of the paths / /A /B or /A/D refer to the same directory, as the contents of each is different. Nor do any of / /A /A/C or /A/D. However /A/C and /B might be the same directory, there is insufficient information given to tell. Any of the other path names (/X /Y /A/Z /A/C/P /A/C/Q /B/P and /B/Q) may refer to the same underlying files, in almost any combination.

If the current working directory of the server-FTP is /A then the following path names, in addition to all the fully qualified path names, are valid

C
D
Z
C/P
C/Q

These all refer to the same files or directories as the corresponding fully qualified path with "/A/" prepended.

That those path names all exist does not imply that the TVFS sever will necessarily grant any kind of access rights to the named paths, or that access to the same file via different path names will necessarily be granted equal rights.

None of the following relative paths are valid when the current directory is /A

A
B
X
Y
B/P
B/Q
P
Q

Any of those could be made valid by changing the server-FTP's current working directory to the appropriate directory. Note that the paths "P" and "Q" might refer to different files depending upon which directory is selected to cause those to become valid TVFS relative paths.

8. Listings for Machine Processing (MLST and MLSD)

The MLST and MLSD commands are intended to standardize the file and directory information returned by the Server-FTP process. These commands differ from the LIST command in that the format of the replies is strictly defined although extensible.

Two commands are defined, MLST which provides data about exactly the object named on its command line, and no others. MLSD on the other hand will list the contents of a directory if a directory is named, otherwise a 501 reply will be returned. In either case, if no object is named, the current directory is assumed. That will cause MLST to send a one line response, describing the current directory itself, and MLSD to list the contents of the current directory.

In the following, the term MLSx will be used wherever either MLST or MLSD may be inserted.

The MLST and MLSD commands also extend the FTP protocol as presented in RFC 959 [3] and RFC 1123 [9] to allow that transmission of 8-bit data over the control connection. Note this is not specifying character sets which are 8-bit, but specifying that FTP implementations are to specifically allow the transmission and reception of 8-bit bytes, with all bits significant, over the control connection. That is, all 256 possible octet values are permitted. The MLSx command allows both UTF-8/Unicode and "raw" forms as arguments, and in responses both to the MLST and MLSD commands, and all other FTP commands which take pathnames as arguments.

8.1. Format of MLSx Requests

The MLST and MLSD commands each allow a single optional argument. This argument may be either a directory name or, for MLST only, a filename. For these purposes, a "filename" is the name of any entity in the server NVFS which is not a directory. Where TVFS is supported, any TVFS relative path name valid in the current working directory, or any TVFS fully qualified path name, may be given. If a directory name is given then MLSD must return a listing of the contents of the named directory, otherwise it issues a 501 reply, and does not open a data connection. In all cases for MLST, a single fact line containing the information about the named file or directory shall be returned over the control connection, without opening a data connection.

If no argument is given then MLSD must return a listing of the contents of the current working directory, and MLST must return a listing giving information about the current working directory itself. For these purposes, the contents of a directory are whatever

filenames (not pathnames) the server-PI will allow to be referenced when the current working directory is the directory named, and which the server-PI desires to reveal to the user-PI.

No title, header, or summary, lines, or any other formatting, other than as is specified below, is ever returned in the output of an MLST or MLSD command.

If the Client-FTP sends an invalid argument, the Server-FTP MUST reply with an error code of 501.

The syntax for the MLSx command is:

```
mlst             = "MLst" [ SP pathname ] CRLF
mlsd             = "MLsD" [ SP pathname ] CRLF
```

8.2. Format of MLSx Response

The format of a response to an MLSx command is as follows:

```
mlst-response    = control-response / error-response
mlsd-response    = ( initial-response final-response ) /
                   error-response

control-response = "250-" [ response-message ] CRLF
                   1*( SP entry CRLF )
                   "250" [ SP response-message ] CRLF

initial-response = "150" [ SP response-message ] CRLF
final-response   = "226" SP response-message CRLF

response-message = *TCHAR

data-response    = *( entry CRLF )

entry            = [ facts ] SP pathname
facts            = fact *( ";" fact )
fact             = factname "=" value
factname         = "Size" / "Modify" / "Create" /
                   "Type" / "Unique" / "Perm" /
                   "Lang" / "Media-Type" / "CharSet" /
                   os-depend-fact / local-fact
os-depend-fact   = <IANA assigned OS name> "." token
local-fact       = "X." token
value            = *RCHAR
```

Upon receipt of a MLSx command, the server will verify the parameter, and if invalid return an error-response.

If valid, then for an MLST command, the server-PI will send the first (leading) line of the control response, the entry for the pathname given, or the current directory if no pathname was provided, and the terminating line. Normally exactly one entry would be returned, more entries are permitted only when required to represent a file that is to have multiple "Type" facts returned.

Note that for MLST the fact set is preceded by a space. That is provided to guarantee that the fact set cannot be accidentally interpreted as the terminating line of the control response, but is required even when that would not be possible. Exactly one space exists between the set of facts and the pathname. Where no facts are present, there will be exactly two leading spaces before the pathname. No spaces are permitted in the facts, any other spaces in the response are to be treated as being a part of the pathname.

If the command was an MLSD command, the server will open a data connection as indicated in section 3.2 of RFC959 [3]. If that fails, the server will return an error-response. If all is OK, the server will return the initial-response, send the appropriate data-response over the new data connection, close that connection, and then send the final-response over the control connection. The grammar above defines the format for the data-response, which defines the format of the data returned over the data connection established.

The data connection opened for a MLSD response shall be a connection as if the "TYPE L 8", "MODE S", and "STRU F" commands had been given, whatever FTP transfer type, mode and structure had actually been set, and without causing those settings to be altered for future commands. That is, this transfer type shall be set for the duration of the data connection established for this command only. While the content of the data sent can be viewed as a series of lines, implementations should note that there is no maximum line length defined. Implementations should be prepared to deal with arbitrarily long lines.

The facts part of the specification would contain a series of "file facts" about the file or directory named on the same line. Typical information to be presented would include file size, last modification time, creation time, a unique identifier, and a file/directory flag.

The complete format for a successful reply to the MLSD command would be:

```
facts SP pathname CRLF
facts SP pathname CRLF
facts SP pathname CRLF
...
```

Note that the format is intended for machine processing, not human viewing, and as such the format is very rigid. Implementations MUST NOT vary the format by, for example, inserting extra spaces for readability, replacing spaces by tabs, including header or title lines, or inserting blank lines, or in any other way alter this format. Exactly one space is always required after the set of facts (which may be empty). More spaces may be present on a line if, and only if, the file name presented contains significant spaces. The set of facts must not contain any spaces anywhere inside it. Facts should be provided in each output line only if they both provide relevant information about the file named on the same line, and they are in the set requested by the user-PI. There is no requirement that the same set of facts be provided for each file, or that the facts presented occur in the same order for each file.

8.3. Filename encoding

A FTP implementation supporting the MLSx commands must be 8-bit clean. This is necessary in order to transmit UTF-8 encoded filenames. This specification recommends the use of UTF-8 encoded filenames. FTP implementations SHOULD use UTF-8 whenever possible to encourage the maximum interoperability.

Filenames are not restricted to UTF-8, however treatment of arbitrary character encodings is not specified by this standard. Applications are encouraged to treat non-UTF-8 encodings of filenames as octet sequences.

Note that this encoding is unrelated to that of the contents of the file, even if the file contains character data.

Further information about filename encoding for FTP may be found in "Internationalization of the File Transfer Protocol" [7].

8.3.1. Notes about the Filename

The filename returned in the MLST response should be the same name as was specified in the MLST command, or, where TVFS is supported, a fully qualified TVFS path naming the same file. Where no argument was given to the MLST command, the server-PI may either include an empty filename in the response, or it may supply a name that refers to the current directory, if such a name is available. Where TVFS is supported, a fully qualified path name of the current directory

SHOULD be returned.

Filenames returned in the output from an MLSD command should be unqualified names within the directory named, or the current directory if no argument was given. That is, the directory named in the MLSD command SHOULD NOT appear as a component of the filenames returned.

If the server-FTP process is able, and the "type" fact is being returned, it MAY return in the MLSD response, an entry whose type is "cdir", which names the directory from which the contents of the listing were obtained. Where TVFS is supported, the name may be the fully qualified path name of the directory, or may be any other path name which is valid to refer to that directory from the current working directory of the server-FTP. Where more than one name exists, multiple of these entries may be returned. In a sense, the "cdir" entry can be viewed as a heading for the MLSD output. However, it is not required to be the first entry returned, and may occur anywhere withing the listing.

When TVFS is supported, a user-PI can refer to any file or directory in the listing by combining a type "cdir" name, with the appropriate name from the directory listing using the procedure defined in section 7.2.

Alternatively, whether TVFS is supported or not, the user-PI can issue a CWD command ([3]) giving a name of type "cdir" from the listing returned, and from that point reference the files returned in the MLSD response from which the cdir was obtained by using the filename components of the listing.

8.4. Format of Facts

The "facts" for a file in a reply to a MLSx command consist of information about that file. The facts are a series of keyword=value pairs separated by semi-colon (";") characters. An individual fact may not contain a semi-colon in its name or value. The complete series of facts may not contain the space character, nor are leading nor trailing semi-colons included. See the definition or "RCHAR" in section 2.1 for a list of the characters that can occur in a fact value. Not all are applicable to all facts.

A sample of a typical series of facts would be: (spread over two lines for presentation here only)

```
size=4161;lang=en-US;modify=19970214165800;create=19961001124534;
type=file;x.myfact=foo,bar
```

8.5. Standard Facts

This document defines a standard set of facts as follows:

```
size       -- Size in octets
modify     -- Last modification time
create     -- Creation time
type       -- Entry type
unique     -- Unique id of file/directory
perm       -- File permissions, whether read, write, execute is
              allowed for the login id.
lang       -- Language of the filename per IANA[12] registry.
media-type -- MIME media-type of file contents per IANA registry.
charset    -- Character set per IANA registry (if not UTF-8)
```

Fact names are case-insensitive. Size, size, SIZE, and SiZe are the same fact.

Further operating system specific keywords could be specified by using the IANA operating system name as a prefix (examples only):

```
OS/2.ea   -- OS/2 extended attributes
MACOS.rf  -- MacIntosh resource forks
UNIX.mode -- Unix file modes (permissions)
```

Implementations may define keywords for experimental, or private use. All such keywords MUST begin with the two character sequence "x.". As type names are case independent, "x." and "X." are equivalent. For example:

```
x.ver  -- Version information
x.desc -- File description
x.type -- File type
```

8.5.1. The type Fact

The type fact needs a special description. Part of the problem with current practices is deciding when a file is a directory. If it is a directory, is it the current directory, a regular directory, or a parent directory? The MLST specification makes this unambiguous using the type fact. The type fact given specifies information about the object listed on the same line of the MLST response.

Five values are possible for the type fact:

```
file         -- a file entry
cdir         -- the listed directory
pdir         -- a parent directory
dir          -- a directory or sub-directory
OS.name=type -- an OS or file system dependent file type
```

The syntax is defined to be:

```
type-fact       = type-label "=" type-val
type-label      = "Type"
type-val        = "File" / "cdir" / "pdir" / "dir" /
                  os-type
```

8.5.1.1. type=file

The presence of the type=file fact indicates the listed entry is a file containing non-system data. That is, it may be transferred from one system to another of quite different characteristics, and perhaps still be meaningful.

8.5.1.2. type=cdir

The type=cdir fact indicates the listed entry contains a pathname of the directory whose contents are listed. This type will only be returned as a part of the result of an MLSD command, and provides a name for the listed directory, and facts about that directory. In a sense, it can be viewed as representing the title of the listing, in a machine friendly format. It may appear at any point of the listing, it is not restricted to appearing at the start, though frequently may do so, and may occur multiple times. It MUST NOT be included if the type fact is not included, or there would be no way for the user-PI to distinguish the name of the directory from an entry in the directory.

Where TVFS is supported by the server-FTP, this name may be used to construct path names with which to refer to the files and directories returned in the same MLSD output (see section 7.2). These path names are only expected to work when the server-PI's position in the NFVS file tree is the same as its position when the MLSD command was issued, unless a fully qualified path name results.

Where TVFS is not supported, the only defined semantics associated with a "type=cdir" entry are that, provided the current working directory of the server-PI has not been changed, a pathname of type "cdir" may be used as an argument to a CWD command, which will cause the current directory of the server-PI to change so that the directory which was listed in its current working directory.

8.5.1.3. type=dir

If present, the type=dir entry gives the name of a directory. Such an entry typically cannot be transferred from one system to another using RETR, etc, but should (permissions permitting) be able to be the object of an MLSD command.

8.5.1.4. type=pdir

If present, which will occur only in the response to a MLSD command, the type=pdir entry represents a pathname of the parent directory of the listed directory. As well as having the properties of a type=dir, a CWD command that uses the pathname from this entry should change the user to a parent directory of the listed directory. If the listed directory is the current directory, a CDUP command may also have the effect of changing to the named directory. User-FTP processes should note not all responses will include this information, and that some systems may provide multiple type=pdir responses.

Where TVFS is supported, a "type=pdir" name may be a relative path name, or a fully qualified path name.

For the purposes of this type value, a "parent directory" is any directory in which there is an entry of type=dir which refers to the directory in which the type=pdir entity was found. Thus it is not required that all entities with type=pdir refer to the same directory. The "unique" fact (if supported) can be used to determine whether there is a relationship between the type=pdir entries or not.

8.5.1.5. System defined types

Files types that are specific to a specific operating system, or file system, can be encoded using the "OS." type names. The format is:

```
os-type   = "OS." os-name "=" os-type
os-name   = <an IANA registered operating system name>
os-type   = token
```

The "os-name" indicates the specific system type which supports the particular localtype. OS specific types are registered by the IANA using the procedures specified in section 11. The "os-type" provides the system dependent information as to the type of the file listed. The os-name and os-type strings in an os-type are case independent. "OS.unix=block" and "OS.Unix=BLOCK" represent the same type (or would, if such a type were registered.)

Note: Where the underlying system supports a file type which is essentially an indirect pointer to another file, the NVFS representation of that type should normally be to represent the file which the reference indicates. That is, the underlying basic file will appear more than once in the NVFS, each time with the "unique" fact (see immediately following section) containing the same value, indicating that the same file is represented by all such names. User-PIs transferring the file need then transfer it only once, and then insert their own form of indirect reference to construct alternate names where desired, or perhaps even copy the local file if that is the only way to provide two names with the same content. A file which would be a reference to another file, if only the other file actually existed, may be represented in any OS dependent manner appropriate, or not represented at all.

8.5.1.6. Multiple types

Where a file is such that it may validly, and sensibly, treated by the server-PI as being of more than one of the above types, then multiple entries should be returned, each with its own "Type" fact of the appropriate type, and each containing the same pathname. This may occur, for example, with a structured file, which may contain sub-files, and where the server-PI permits the structured file to be treated as a unit, or treated as a directory allowing the sub-files within it to be referenced.

8.5.2. The unique Fact

The unique fact is used to present a unique identifier for a file or directory in the NVFS accessed via a server-FTP process. The value of this fact should be the same for any number of pathnames that refer to the same underlying file. The fact should have different values for names which reference distinct files. The mapping between files, and unique fact tokens should be maintained, and remain consistent, for at least the lifetime of the control connection from user-PI to server-PI.

```
unique-fact  = "Unique" "=" token
```

This fact would be expected to be used by Server-FTPs whose host system allows things such as symbolic links so that the same file may be represented in more than one directory on the server. The only conclusion that should be drawn is that if two different names each have the same value for the unique fact, they refer to the same underlying object. The value of the unique fact (the token) should be considered an opaque string for comparison purposes, and is a case dependent value. The tokens "A" and "a" do not represent the same underlying object.

8.5.3. The modify Fact

The modify fact is used to determine the last time the content of the file (or directory) indicated was modified. Any change of substance to the file should cause this value to alter. That is, if a change is made to a file such that the results of a RETR command would differ, then the value of the modify fact should alter. User-PIs should not assume that a different modify fact value indicates that the file contents are necessarily different than when last retrieved. Some systems may alter the value of the modify fact for other reasons, though this is discouraged wherever possible. Also a file may alter, and then be returned to its previous content, which would often be indicated as two incremental alterations to the value of the modify fact.

For directories, this value should alter whenever a change occurs to the directory such that different filenames would (or might) be included in MLSD output of that directory.

```
modify-fact  = "Modify" "=" time-val
```

8.5.4. The create Fact

The create fact indicates when a file, or directory, was first created. Exactly what "creation" is for this purpose is not specified here, and may vary from server to server. About all that can be said about the value returned is that it can never indicate a later time than the mtime fact.

```
create-fact  = "Create" "=" time-val
```

Implementation Note: Implementors of this fact on UNIX(TM) systems should note that the unix "stat" "st_ctime" field does not give creation time, and that unix filesystems do not record creation time at all. Unix (and POSIX) implementations will normally not include this fact.

8.5.5. The perm Fact

The perm fact is used to indicate access rights the current FTP user has over the object listed. Its value is always an unordered sequence of alphabetic characters.

```
perm-fact    = "Perm" "=" pvals
pvals        = "a" / "c" / "d" / "e" / "f" /
               "l" / "m" / "p" / "r" / "w"
```

There are ten permission indicators currently defined. Many are meaningful only when used with a particular type of object. The indicators are case independent, "d" and "D" are the same indicator.

The "a" permission applies to objects of type=file, and indicates that the APPE (append) command may be applied to the file named.

The "c" permission applies to objects of type=dir (and type=pdir, type=cdir). It indicates that files may be created in the directory named. That is, that a STOU command is likely to succeed, and that STOR and APPE commands might succeed if the file named did not previously exist, but is to be created in the directory object that has the "c" permission. It also indicates that the RNTO command is likely to succeed for names in the directory.

The "d" permission applies to all types. It indicates that the object named may be deleted, that is, that the RMD command may be applied to it if it is a directory, and otherwise that the DELE command may be applied to it.

The "e" permission applies to the directory types. When set on an object of type=dir, type=cdir, or type=pdir it indicates that a CWD command naming the object should succeed, and the user should be able to enter the directory named. For type=pdir it also indicates that the CDUP command may succeed (if this particular pathname is the one to which a CDUP would apply.)

The "f" permission for objects indicates that the object named may be renamed - that is, may be the object of an RNFR command.

The "l" permission applies to the directory file types, and indicates that the listing commands, LIST, NLST, and MLSD may be applied to the directory in question, and that MLST, LIST, NLST, and STAT may be applied to objects in the directory.

The "m" permission applies to directory types, and indicates that the MKD command may be used to create a new directory within the directory under consideration.

The "p" permission applies to directory types, and indicates that objects in the directory may be deleted, or (stretching naming a little) that the directory may be purged. Note: it does not indicate that the RMD command may be used to remove the directory named itself, the "d" permission indicator indicates that.

The "r" permission applies to type=file objects, and for some systems, perhaps to other types of objects, and indicates that the RETR command may be applied to that object.

The "w" permission applies to type=file objects, and for some systems, perhaps to other types of objects, and indicates that the STOR command may be applied to the object named.

Note: That a permission indicator is set can never imply that the appropriate command is guaranteed to work - just that it might. Other system specific limitations, such as limitations on available space for storing files, may cause an operation to fail, where the permission flags may have indicated that it was likely to succeed. The permissions are a guide only.

Implementation note: The permissions are described here as they apply to FTP commands. They may not map easily into particular permissions available on the server's operating system. Servers are expected to synthesize these permission bits from the permission information available from operating system. For example, to correctly determine whether the "D" permission bit should be set on a directory for a server running on the UNIX(TM) operating system, the server should check that the directory named is empty, and that the user has write permission on both the directory under consideration, and its parent directory.

Some systems may have more specific permissions than those listed here, such systems should map those to the flags defined as best they are able. Other systems may have only more broad access controls. They will generally have just a few possible permutations of permission flags, however they should attempt to correctly represent what is permitted.

8.5.6. The lang Fact

The lang fact describes the natural language of the filename for use in display purposes. Values used here should be taken from the language registry of the IANA. See [13] for the syntax, and procedures, related to language tags.

```
lang-fact  = "Lang" "=" token
```

Server-FTP implementations MUST NOT guess language values. Language values must be determined in an unambiguous way such as filesystem tagging of language or by user configuration. Note that the lang fact provides no information at all about the content of a file, only about the encoding of its name.

8.5.7. The size Fact

The size fact applies to non-directory file types and should always reflect the approximate size of the file. This should be as accurate as the server can make it, without going to extraordinary lengths, such as reading the entire file. The size is expressed in units of octets of data in the file.

Given limitations in some systems, Client-FTP implementations must understand this size may not be precise and may change between the time of a MLST and RETR operation.

Clients that need highly accurate size information for some particular reason should use the SIZE command as defined in section 4. The most common need for this accuracy is likely to be in conjunction with the REST command described in section 5. The size fact, on the other hand, should be used for purposes such as indicating to a human user the approximate size of the file to be transferred, and perhaps to give an idea of expected transfer completion time.

```
size-fact  = "Size" "=" 1*DIGIT
```

8.5.8. The media-type Fact

The media-type fact represents the IANA media type of the file named, and applies only to non-directory types. The list of values used must follow the guidelines set by the IANA registry.

```
media-type  = "Media-Type" "=" <per IANA guidelines>
```

Server-FTP implementations MUST NOT guess media type values. Media type values must be determined in an unambiguous way such as filesystem tagging of media-type or by user configuration. This fact gives information about the content of the file named. Both the primary media type, and any appropriate subtype should be given, separated by a slash "/" as is traditional.

8.5.9. The charset Fact

The charset fact provides the IANA character set name, or alias, for the encoded pathnames in a MLSx response. The default character set is UTF-8 unless specified otherwise. FTP implementations SHOULD use UTF-8 if possible to encourage maximum interoperability. The value of this fact applies to the pathname only, and provides no information about the contents of the file.

```
charset-type  = "Charset" "=" token
```

8.5.10. Required facts

Servers are not required to support any particular set of the available facts. However, servers SHOULD, if conceivably possible, support at least the type, perm, size, unique, and modify facts.

8.6. System Dependent and Local Facts

By using an system dependent fact, or a local fact, a server-PI may communicate to the user-PI information about the file named which is peculiar to the underlying filesystem.

8.6.1. System Dependent Facts

System dependent fact names are labeled by prefixing a label identifying the specific information returned by the name of the appropriate operating system from the IANA maintained list of operating system names.

The value of an OS dependent fact may be whatever is appropriate to convey the information available. It must be encoded as a "token" as defined in section 2.1 however.

In order to allow reliable interoperation between users of system dependent facts, the IANA will maintain a registry of system dependent fact names, their syntax, and the interpretation to be given to their values. Registrations of system dependent facts are to be accomplished according to the procedures of section 11.

8.6.2. Local Facts

Implementations may also make available other facts of their own choosing. As the method of interpretation of such information will generally not be widely understood, server-PIs should be aware that clients will typically ignore any local facts provided. As there is no registration of locally defined facts, it is entirely possible that different servers will use the same local fact name to provide vastly different information. Hence user-PIs should be hesitant about making any use of any information in a locally defined fact without some other specific assurance that the particular fact is one that they do comprehend.

Local fact names all begin with the sequence "X.". The rest of the name is a "token" (see section 2.1). The value of a local fact can be anything at all, provided it can be encoded as a "token".

8.7. MLSx Examples

The following examples are all taken from dialogues between existing FTP clients and servers. Because of this, not all possible variations of possible response formats are shown in the examples. This should not be taken as limiting the options of other server implementors. Where the examples show OS dependant information, that is to be treated as being purely for the purposes of demonstration of some possible OS specific information that could be defined. As at the time of the writing of this document, no OS specific facts or file types have been defined, the examples shown here should not be trreated as in any way to be preferred over other possible similar definitions. Consult the IANA registries to determine what types and facts have been defined.

In the examples shown, only relevant commands and responses have been included. This is not to imply that other commands (including authentication, directory modification, PORT or PASV commands, or similar) would not be present in an actual connection, or were not, in fact, actually used in the examples before editing. Note also that the formats shown are those that are transmitted between client and server, not formats which would normally ever be reported to the user of the client.

In the examples, lines that begin "C> " were sent over the control connection from the client to the server, lines that begin "S> " were sent over the control connection from the server to the client, and lines that begin "D> " were sent from the server to the client over a data connection created just to send those lines and closed immediately after. No examples here show data transferred over a data connection from the client to the server. In all cases, the prefixes shown above, including the one space, have been added for the purposes of this document, and are not a part of the data exchanged between client and server.

8.7.1. Simple MLST

```
 C> PWD
 S> 257 "/tmp" is current directory.
 C> MLst cap60.pl198.tar.gz
 S> 250- Listing cap60.pl198.tar.gz
 S>  Type=file;Size=1024990;Modify=19980130010322;Perm=r /tmp/
cap60.pl198.tar.gz
 S> 250 End
```

The client first asked to be told the current directory of the server. This was purely for the purposes of clarity of this example. The client the requested facts about a specific file. The server

returned the "250-" first control-response line, followed by a single

line of facts about the file, followed by the terminating "250 " line. The text on the control-response line and the terminating line can be anything the server decides to send. Notice that the fact line is indented by a single space. Notice also that there are no spaces in the set of facts returned, until the single space before the filename. The filename returned on the fact line is a fully qualified pathname of the file listed. The facts returned show that the line refers to a file, that file contains approximately 1024990 bytes, though more or less than that may be transferred if the file is retrieved, and a different number may be required to store the file at the client's filestore, the file was last modified at 01:03:22 on January 30, 1998 (UTC), and the connected user has permission to retrieve the file but not to do anything else particularly interesting.

```
C> PWD
S> 257 "/" is current directory.
C> MLst tmp
S> 250- Listing tmp
S>  Type=dir;Modify=19981107085215;Perm=el /tmp
S> 250 End
```

Again the PWD is just for the purposes of demonstration for the example. The MLST fact line this time shows that the file listed is a directory, that it was last modified at 08:52:15 on the 7th of November, 1998 UTC, and that the user has permission to enter the directory, and to list its contents, but not to modify it in any way. Again, the fully qualified path name of the directory listed is given.

8.7.2. MLSD of a directory

```
C> MLSD tmp
S> 150 BINARY connection open for MLSD tmp
D> Type=cdir;Modify=19981107085215;Perm=el tmp
D> Type=cdir;Modify=19981107085215;Perm=el /tmp
D> Type=pdir;Modify=19990112030508;Perm=el ..
D> Type=file;Size=25730;Modify=19940728095854;Perm= capmux.tar.z
D> Type=file;Size=1830;Modify=19940916055648;Perm=r hatch.c
D> Type=file;Size=25624;Modify=19951003165342;Perm=r MacIP-02.txt
D> Type=file;Size=2154;Modify=19950501105033;Perm=r uar.netbsd.patch
D> Type=file;Size=54757;Modify=19951105101754;Perm=r iptnnladev.1.0.sit.hqx
D> Type=file;Size=226546;Modify=19970515023901;Perm=r melbcs.tif
D> Type=file;Size=12927;Modify=19961025135602;Perm=r tardis.1.6.sit.hqx
D> Type=file;Size=17867;Modify=19961025135602;Perm=r timelord.1.4.sit.hqx
D> Type=file;Size=224907;Modify=19980615100045;Perm=r uar.1.2.3.sit.hqx
D> Type=file;Size=1024990;Modify=19980130010322;Perm=r cap60.pl198.tar.gz
S> 226 MLSD completed
```

In this example notice that there is no leading space on the fact lines returned over the data connection. Also notice that two lines of "type=cdir" have been given. These show two alternate names for the directory listed, one a fully qualified pathname, and the other a local name relative to the servers current directory when the MLSD was performed. Note that all other filenames in the output are relative to the directory listed, though the server could, if it chose, give a fully qualified path name for the "type=pdir" line. This server has chose not to. The other files listed present a fairly boring set of files that are present in the listed directory. Note that there is no particular order in which they are listed. They are not sorted by filename, by size, or by modify time. Note also that the "perm" fact has an empty value for the file "capmux.tar.z" indicating that the connected user has no permissions at all for that file. This server has chosen to present the "cdir" and ir" lines before the lines showing the content of the directory, it is not required to do so. The "size" fact does not provide any meaningful information for a directory, so is not included in the fact lines for the directory types shown.

8.7.3. A more complex example

```
C> MLst test
S> 250- Listing test
S>  Type=dir;Perm=el;Unique=keVO1+ZF4 test
S> 250 End
C> MLSD test
S> 150 BINARY connection open for MLSD test
D> Type=cdir;Perm=el;Unique=keVO1+ZF4 test
D> Type=pdir;Perm=e;Unique=keVO1+d?3 ..
D> Type=OS.unix=slink:/foobar;Perm=;Unique=keVO1+4G4 foobar
D> Type=OS.unix=chr-13/29;Perm=;Unique=keVO1+5G4 device
D> Type=OS.unix=blk-11/108;Perm=;Unique=keVO1+6G4 block
D> Type=file;Perm=awr;Unique=keVO1+8G4 writeable
D> Type=dir;Perm=cpmel;Unique=keVO1+7G4 promiscuous
D> Type=dir;Perm=;Unique=keVO1+1t2 no-exec
D> Type=file;Perm=r;Unique=keVO1+EG4 two words
D> Type=file;Perm=r;Unique=keVO1+IH4  leading space
D> Type=file;Perm=r;Unique=keVO1+1G4 file1
D> Type=dir;Perm=cpmel;Unique=keVO1+7G4 incoming
D> Type=file;Perm=r;Unique=keVO1+1G4 file2
D> Type=file;Perm=r;Unique=keVO1+1G4 file3
D> Type=file;Perm=r;Unique=keVO1+1G4 file4
S> 226 MLSD completed
C> MLSD test/incoming
S> 150 BINARY connection open for MLSD test/incoming
D> Type=cdir;Perm=cpmel;Unique=keVO1+7G4 test/incoming
D> Type=pdir;Perm=el;Unique=keVO1+ZF4 ..
```

```
D> Type=file;Perm=awdrf;Unique=keVO1+EH4 bar
D> Type=file;Perm=awdrf;Unique=keVO1+LH4
D> Type=file;Perm=rf;Unique=keVO1+1G4 file5
D> Type=file;Perm=rf;Unique=keVO1+1G4 file6
D> Type=dir;Perm=cpmdelf;Unique=keVO1+!s2 empty
S> 226 MLSD completed
```

For the purposes of this example the fact set requested has been modified to delete the "size" and "modify" facts, and add the "unique" fact. First, facts about a filename have been obtained via MLST. Note that no fully qualified path name was given this time. That was because the server was unable to determine that information. Then having determined that the filename represents a directory, that directory has been listed. That listing also shows no fully qualified path name, for the same reason, thus has but a single "type=cdir" line. This directory (which was created especially for the purpose) contains several interesting files. There are some with OS dependent file types, several sub-directories, and several ordinary files.

Not much can be said here about the OS dependant file types, as none of the information shown there should be treated as any more than possibilities. It can be seen that the OS type of the server is "unix" though, which is one of the OS types in the IANA registry of Operating System names.

Of the three directories listed, "no-exec" has no permission granted to this user to access at all. From the "Unique" fact values, it can be determined that "promiscuous" and "incoming" in fact represent the same directory. Its permissions show that the connected user has permission to do essentially anything other than to delete the directory. That directory was later listed. It happens that the directory can not be deleted because it is not empty.

Of the normal files listed,
two contain spaces in their names. The file called " leading space" actually contains two spaces in its name,
one before the "l" and one between the "g" and the "s". The two spaces that separate the facts from the visible part of the path name make that clear. The file "writeable" has the "a" and "w" permission bits set, and consequently the connected user should be able to STOR or APPE to that file.

The other four file names, "file1", "file2", "file3", and "file4" all represent the same underlying file, as can be seen from the values of the "unique" facts of each. It happens that "file1" and "file2" are Unix "hard" links, and that "file3" and "file4" are "soft" or "symbolic" links to the first two. None of that information is

available via standard MLST facts, it is sufficient for the purposes of FTP to note that all represent the same file, and that the same data would be fetched no matter which of them was retrieved.

Finally, the sub-directory "incoming" is listed. Since "promiscuous" is the same directory there would be no point listing it as well. In that directory, the files "file5" and "file6" represent still more names for the "file1" file we have seen before. Notice the entry between that for "bar" and "file5". Though it is not possible to easily represent it in this document, that shows a file with a name comprising exactly three spaces (" "). A client will have no difficulty determining that name from the output presented to it however. The directroy "empty" is, as its name implies, empty, though that is not shown here. It can, however, be deleted, as can file "bar" and the file whose name is three spaces. All the files that reside in this directory can be renamed. This is a consequence of the UNIX semantics of the directory that contains them being modifiable.

8.7.4. A different server

```
C> MLST
S> 250-Begin
S>  type=dir;unique=AQkAAAAAAABCAAA /
S> 250 End.
C> MLSD .
S> 150 Opening ASCII mode data connection for MLS.
D> type=cdir;unique=AQkAAAAAAABCAAA /
D> type=dir;unique=AQkAAAAAAABEAAA bin
D> type=dir;unique=AQkAAAAAAABGAAA etc
D> type=dir;unique=AQkAAAAAAAB8AwA halflife
D> type=dir;unique=AQkAAAAAAABoAAA incoming
D> type=dir;unique=AQkAAAAAAABIAAA lib
D> type=dir;unique=AQkAAAAAAABWAEA linux
D> type=dir;unique=AQkAAAAAAABKAEA ncftpd
D> type=dir;unique=AQkAAAAAAABGAEA outbox
D> type=dir;unique=AQkAAAAAAABuAAA quake2
D> type=dir;unique=AQkAAAAAAABQAEA winstuff
S> 226 Listing completed.
C> MLSD linux
S> 150 Opening ASCII mode data connection for MLS.
D> type=cdir;unique=AQkAAAAAAABWAEA /linux
D> type=pdir;unique=AQkAAAAAAABCAAA /
D> type=dir;unique=AQkAAAAAAABeAEA firewall
D> type=file;size=12;unique=AQkAAAAAAACWAEA helo_world
D> type=dir;unique=AQkAAAAAAABYAEA kernel
D> type=dir;unique=AQkAAAAAAABmAEA scripts
D> type=dir;unique=AQkAAAAAAABkAEA security
```

```
S> 226 Listing completed.
C> MLSD linux/kernel
S> 150 Opening ASCII mode data connection for MLS.
D> type=cdir;unique=AQkAAAAAAAABYAEA /linux/kernel
D> type=pdir;unique=AQkAAAAAAAABWAEA /linux
D> type=file;size=6704;unique=AQkAAAAAAAADYAEA k.config
D> type=file;size=7269221;unique=AQkAAAAAAAACYAEA linux-2.0.36.tar.gz
D> type=file;size=12514594;unique=AQkAAAAAAAAEYAEA linux-2.1.130.tar.gz
S> 226 Listing completed.
```

Note that this server returns its "unique" fact value in quite a different format. It also returns fully qualified path names for the "pdir" entry.

8.7.5. Some IANA files

```
C> MLSD .
S> 150 BINARY connection open for MLSD .
D> Type=cdir;Modify=19990219183438 /iana/assignments
D> Type=pdir;Modify=19990112030453 ..
D> Type=dir;Modify=19990219073522 media-types
D> Type=dir;Modify=19990112033515 character-set-info
D> Type=dir;Modify=19990112033529 languages
D> Type=file;Size=44242;Modify=19990217230400 character-sets
D> Type=file;Size=1947;Modify=19990209215600 operating-system-names
S> 226 MLSD completed
C> MLSD media-types
S> 150 BINARY connection open for MLSD media-types
D> Type=cdir;Modify=19990219073522 media-types
D> Type=cdir;Modify=19990219073522 /iana/assignments/media-types
D> Type=pdir;Modify=19990219183438 ..
D> Type=dir;Modify=19990112033045 text
D> Type=dir;Modify=19990219183442 image
D> Type=dir;Modify=19990112033216 multipart
D> Type=dir;Modify=19990112033254 video
D> Type=file;Size=30249;Modify=19990218032700 media-types
S> 226 MLSD completed
C> MLSD character-set-info
S> 150 BINARY connection open for MLSD character-set-info
D> Type=cdir;Modify=19990112033515 character-set-info
D> Type=cdir;Modify=19990112033515 /iana/assignments/character-set-info
D> Type=pdir;Modify=19990219183438 ..
D> Type=file;Size=1234;Modify=19980903020400 windows-1251
D> Type=file;Size=4557;Modify=19980922001400 tis-620
D> Type=file;Size=801;Modify=19970324130000 ibm775
D> Type=file;Size=552;Modify=19970320130000 ibm866
D> Type=file;Size=922;Modify=19960505140000 windows-1258
S> 226 MLSD completed
```

```
C> MLSD languages
S> 150 BINARY connection open for MLSD languages
D> Type=cdir;Modify=19990112033529 languages
D> Type=cdir;Modify=19990112033529 /iana/assignments/languages
D> Type=pdir;Modify=19990219183438 ..
D> Type=file;Size=2391;Modify=19980309130000 default
D> Type=file;Size=943;Modify=19980309130000 tags
D> Type=file;Size=870;Modify=19971026130000 navajo
D> Type=file;Size=699;Modify=19950911140000 no-bok
S> 226 MLSD completed
C> PWD
S> 257 "/iana/assignments" is current directory.
```

This example shows some of the IANA maintained files that are relevant for this specification in MLSD format. Note that these listings have been edited by deleting many entries, the actual listings are much longer.

8.8. FEAT response for MLSx

When responding to the FEAT command, a server-FTP process that supports MLST, and MLSD, plus internationalization of pathnames, MUST indicate that this support exists. It does this by including a MLST feature line. As well as indicating the basic support, the MLST feature line indicates which MLST facts are available from the server, and which of those will be returned if no subsequent "OPTS MLST" command is sent.

```
mlst-feat     = SP "MLST" [SP factlist] CRLF
factlist      = factname ["*"] *( ";" factname ["*"] )
```

The initial space shown in the mlst-feat response is that required by the FEAT command, two spaces are not permitted. If no factlist is given, then the server-FTP process is indicating that it supports MLST, but implements no facts. Only pathnames can be returned. This would be a minimal MLST implementation, and useless for most practical purposes. Where the factlist is present, the factnames included indicate the facts supported by the server. Where the optional asterisk appears after a factname, that fact will be included in MLST format responses, until an "OPTS MLST" is given to alter the list of facts returned. After that, subsequent FEAT commands will return the asterisk to show the facts selected by the most recent "OPTS MLST".

Note that there is no distinct FEAT output for MLSD. The presence of the MLST feature indicates that both MLST and MLSD are supported.

8.8.1. Examples

```
C> Feat
S> 201- Features supported
S>  REST STREAM
S>  MDTM
S>  SIZE
S>  TVFS
S>  UTF8
S>  MLST Type*;Size*;Modify*;Perm*;Unique*;UNIX.mode;UNIX.chgd;X.hidden
S> 201 End
```

Aside from some features irrelevant here, this server indicates that it supports MLST including several, but not all, standard facts, all of which it will send by default. It also supports two OS dependant facts, and one locally defined fact. The latter three must be requested expressly by the client for this server to supply them.

```
C> Feat
S> 211-Extensions supported:
S>  CLNT
S>  MDTM
S>  MLST type*;size*;modify*;UNIX.mode*;UNIX.owner;UNIX.group;unique
S>  PASV
S>  REST STREAM
S>  SIZE
S>  TVFS
S>  Compliance Level: 19981201 (IETF mlst-05)
S> 211 End.
```

Again, in addition to some irrelevant features here, this server indicates that it supports MLST, four of the standard facts, one of which ("unique") is not enabled by default, and several OS dependant facts, one of which is provided by the server by default. This server actually supported more OS dependant facts. Others were deleted for the purposes of this document to comply with document formatting restrictions.

8.9. OPTS parameters for MLST

For the MLSx commands, the Client-FTP may specify a list of facts it wishes to be returned in all subsequent MLSx commands until another OPTS MLST command is sent. The format is specified by:

```
mlst-opts     = "OPTS" SP "MLST"
                  [ SP factname *(";" factname) ]
```

By sending the "OPTS MLST" command, the client requests the server to include only the facts listed as arguments to the command in subsequent output from MLSx commands. Facts not included in the "OPTS MLST" command MUST NOT be returned by the server. Facts that are included should be returned for each entry returned from the MLSx command where they apply. Facts requested that are not supported, or which are inappropriate to the file or directory being listed should simply be omitted from the MLSx output. This is not an error. Note that where no factname arguments are present, the client is requesting that only the file names be returned. In this case, and in any other case where no facts are included in the result, the space that separates the fact names and their values from the file name is still required. That is, the first character of the output line will be a space, (or two characters when the line is returned over the control connection,) and the file name will start immediately thereafter.

Clients should note that generating values for some facts can be possible, but very expensive, for some servers. It is generally acceptable to retrieve any of the facts that the server offers as its default set before any "OPTS MLST" command has been given, however clients should use particular caution before requesting any facts not in that set. That is, while other facts may be available from the server, clients should refrain from requesting such facts unless there is a particular operational requirement for that particular information, which ought be more significant than perhaps simply improving the information displayed to an end user.

Note, there is no "OPTS MLSD" command, the fact names set with the "OPTS MLST" command apply to both MLST and MLSD commands.

Servers are not required to accept "OPTS MLST" commands before authentication of the user-PI, but may choose to permit them.

8.9.1. Examples

```
C> Feat
S> 201- Features supported
S>  MLST Type*;Size*;Modify*;Perm*;Unique*;UNIX.mode;UNIX.chgd;X.hidden
S> 201 End
C> Opts Mlst Type;UNIX.mode;Perm
S> 200 Done
C> Feat
S> 201- Features supported
S>  MLST Type*;Size;Modify;Perm*;Unique;UNIX.mode*;UNIX.chgd;X.hidden
```

```
S> 201 End
C> Opts Mlst lang;type;charset;create
S> 200 Done
C> Feat
S> 201- Features supported
S>  MLST Type*;Size;Modify;Perm;Unique;UNIX.mode;UNIX.chgd;X.hidden
S> 201 End
C> Opts Mlst size;frogs
S> 501 Invalid MLST options
C> Feat
S> 201- Features supported
S>  MLST Type*;Size;Modify;Perm;Unique;UNIX.mode;UNIX.chgd;X.hidden
S> 201 End
```

For the purposes of this example, features other than MLST have been deleted from the output to avoid clutter. The example shows the initial default feature output for MLST. The facts requested are then changed by the client. The first change shows facts that are available from the server being selected. Subsequently FEAT output shows the altered features as being returned. The client then attempts to select some standard features which the server does not support. This is not an error, however the server simply ignores the requests for unsupported features, as the FEAT output that follows shows. Lastly, the client attempts to request a non-standard, and unsupported, feature. The server rejects that, and makes no changes at all to the fact set selected.

9. Impact On Other FTP Commands

Along with the introduction of MLST, traditional FTP commands must be extended to allow for the use of more than US-ASCII or EBCDIC character sets. In general, the support of MLST requires support for arbitrary character sets wherever filenames and directory names are allowed. This applies equally to both arguments given to the following commands and to the replies from them, as appropriate.

- CWD
- RETR
- STOR
- STOU
- APPE
- RNFR
- RNTO
- DELE
- RMD
- MKD
- PWD
- STAT

The arguments to all of these commands should be processed the same way that MLST commands and responses are processed with respect to handling embedded CRs and NULs. See section 2.2.

10. Character sets and Internationalization

FTP commands are protocol elements, and are always expressed in ASCII. FTP responses are composed of the numeric code, which is a protocol element, and a message, which is often expected to convey information to the user. It is not expected that users normally interact directly with the protocol elements, rather the user FTP-process constructs the commands, and interprets the results, in the manner best suited for the particular user. Explanatory text in responses generally has no particular meaning to the protocol. The numeric codes provide all necessary information. Server-PIs are free to provide the text in any language that can be adequately represented in ASCII, or where an alternative language and representation has been negotiated (see [7]) in that language and representation.

Pathnames are expected to be encoded in UTF-8 allowing essentially any character to be represented in a pathname. Meaningful pathnames are defined by the server NFVS.

No restrictions at all are placed upon the contents of files transferred using the FTP protocols. Unless the "media-type" fact is provided in a MLSx response nor is any advice given here which would allow determining the content type. That information is assumed to be obtained via other means.

11. IANA Considerations

This specification makes use of some lists of values currently maintained by the IANA, and creates two new lists for the IANA to maintain. It does not add any values to any existing registries.

The existing IANA registries used by this specification are modified using mechanisms specified elsewhere.

11.1. The OS specific fact registry

A registry of OS specific fact names shall be maintained by the IANA. The OS names for the OS portion of the fact name must be taken from the IANA's list of registered OS names. To add a fact name to this OS specific registry of OS specific facts, an applicant must send to the IANA a request, in which is specified the OS name, the OS specific fact name, a definition of the syntax of the fact value, which must conform to the syntax of a token as given in this

document, and a specification of the semantics to be associated with the particular fact and its values. Upon receipt of such an application, and if the combination of OS name and OS specific fact name has not been previously defined, the IANA will add the specification to the registry.

Any examples of OS specific facts found in this document are to be treated as examples of possible OS specific facts, and do not form a part of the IANA's registry merely because of being included in this document.

11.2. The OS specific filetype registry

A registry of OS specific file types shall be maintained by the IANA. The OS names for the OS portion of the fact name must be taken from the IANA's list of registered OS names. To add a file type to this OS specific registry of OS specific file types, an applicant must send to the IANA a request, in which is specified the OS name, the OS specific file type, a definition of the syntax of the fact value, which must conform to the syntax of a token as given in this document, and a specification of the semantics to be associated with the particular fact and its values. Upon receipt of such an application, and if the combination of OS name and OS specific file type has not been previously defined, the IANA will add the specification to the registry.

Any examples of OS specific file types found in this document are to be treated as potential OS specific file types only, and do not form a part of the IANA's registry merely because of being included in this document.

12. Security Considerations

This memo does not directly concern security. It is not believed that any of the mechanisms documented here impact in any particular way upon the security of FTP.

Implementing the SIZE command, and perhaps some of the facts of the MDLx commands, may impose a considerable load on the server, which could lead to denial of service attacks. Servers have, however, implemented this for many years, without significant reported difficulties.

With the introduction of virtual hosts to FTP, and the possible accompanying multiple authentication environments, server implementors will need to take some care to ensure that integrity is maintained.

The FEAT and OPTS commands may be issued before the FTP authentication has occurred [6]. This allows unauthenticated clients to determine which of the features defined here are supported, and to negotiate the fact list for MLSx output. No actual MLSx commands may be issued however, and no problems with permitting the selection of the format prior to authentication are forseen.

A general discussion of issues related to the security of FTP can be found in [14].

13. References

[1] Coded Character Set--7-bit American Standard Code for Information Interchange, ANSI X3.4-1986.

[2] Yergeau, F., "UTF-8, a transformation format of Unicode and ISO 10646", RFC 2044, October 1996.

[3] Postel, J., Reynolds, J., "File Transfer Protocol (FTP)", STD 9, RFC 959, October 1985

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

[5] Crocker, D., Overell, P., "Augmented BNF for Syntax Specifications: ABNF", RFC 2234, November 1997

[6] Hethmon, P., Elz, R., "Feature negotiation mechanism for the File Transfer Protocol", RFC 2389, August 1998

[7] Curtin, W., "Internationalization of the File Transfer Protocol", Work In Progress <draft-ietf-ftpext-intl-ftp-05.txt>, August 1998

[8] Postel, J., Reynolds, J., "Telnet protocol Specification" STD 8, RFC 854, May 1983

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

[10] Mockapetris, P., "Domain Names - Concepts and Facilities" STD 13, RFC 1034, November 1987

[11] ISO/IEC 10646-1:1993 "Universal multiple-octet coded character set (UCS) -- Part 1: Architecture and basic multilingual plane", International Standard -- Information Technology, 1993

[12] Internet Assigned Numbers Authority. http://www.iana.org Email: iana@iana.org.

[13] Alvestrand, H., "Tags for the Identification of Languages"
RFC 1766, March 1995

[14] Allman, M., Ostermann, S., "FTP Security Considerations"
Work in progress, <draft-ietf-ftpext-sec-consider-02.txt>, October 1998

Acknowledgements

This document is a product of the FTPEXT working group of the IETF.

The following people are among those who have contributed to this document:

Alex Belits
D. J. Berstein
Dave Cridland
Martin J. Duerst
Mike Gleason
Mark Harris
Alun Jones
James Matthews
Luke Mewburn
Keith Moore
Buz Owen
Stephen Tihor
and the entire FTPEXT working group of the IETF.

Apologies are offered to any inadvertently omitted.

Bernhard Rosenkraenzer suggested the HOST command, and initially described it.

The description of the modifications to the REST command and the MDTM and SIZE commands comes from a set of modifications suggested for RFC959 by Rick Adams in 1989. A draft containing just those commands, edited by David Borman, has been merged with this document.

Mike Gleason provided access to the FTP server used in some of the examples.

Copyright

Editors' Addresses

Robert Elz
University of Melbourne
Department of Computer Science
Parkville, Vic 3052
Australia

Email: kre@munnari.OZ.AU

Paul Hethmon
Hethmon Brothers
2305 Chukar Road
Knoxville, TN 37923 USA

Phone: +1 423 690 8990
Email: phethmon@hethmon.com