

**As Presented to the 2001 Fault and Disturbance Analysis Conference at Georgia Tech
As Presented at the Spring 2001 Meeting of the IEEE Power System Relay Committee**

File Naming Convention for Time Sequence Data

Final Report of IEEE Power System Relaying Committee Working Group H8

Working Group H8: J. W. Ingleson, Chairman, M. G. Adamiak, C. H. Castro, J. W. Chadwick, A. T. Giulianti, D. K. Holstein, L. Johnson, C. P. Kinne, A. Z. Makki, K. E. Martin, H. I. Mehta, R. J. Murphy, A. P. Napikoski, C. Rogers, M. R. Rothweiler, R. P. Schulz, C. C. Shank, J. Sperr, D. M. Taylor, B. J. Tremerie, M. A. Xavier Jr.

Summary

As microprocessor based measurement and protection technologies continue to advance and replace older designs, massive volumes of electronic information about the power system are being generated. The electronic information is in the form of Time Sequence Data (TSD) saved in Personal Computer (PC) files. A typical utility can generate thousands of TSD files in a single month (transient records, relay test files, trend files, and so on). Due to the significance of useful filenames for handling very large volumes of files, the H8 working group was formed to research, develop and propose a common naming convention for TSD files. A number of conventions for assigning unique TSD filenames are currently available and are being used by various companies, but none are standard. The information used for assigning TSD filenames originates from the affected line, substation and recording device names and from the date and time of fault occurrence. Under the old DOS filenames we were limited to 11 characters but now filenames of 253 characters are generally permitted. However with compact disks (CDs) we are limited to 64 characters. Some other limitations are also discussed below. This report provides a survey of popular file naming conventions and describes the convention proposed by the H8 working group. Members of the working group have used this file naming convention under a variety of operating systems.

Significance

Filenames are essential for both operating system and user. The filename is the system's key for unlocking file contents and without friendly names the user will have trouble handling large numbers of files. In Addition, programs for analysis, trending and management applications also have to sift through and process large numbers of files directly from disk. Reading file contents requires considerable disk access especially for large files. Filenames, on the other hand, are stored separately from their contents in system files called allocation tables and can be loaded with minimal disk access. Meaningful filenames provide software developers with the ability to write programs that can manage and process large numbers of files quickly because the key information is in the filenames. The PC hardware benefits too from the reduced number of disk access operations. The alternative to meaningful filenames is to build and maintain specialized databases. However, databases require substantial programming and produce extremely large file sizes.

Filenames

A filename is composed of two parts. The first part is the name and the second part is the extension. The extension is normally used to specify the type of the file. In general, there are two different types of files. The first type is binary files such as program files (EXE) and dynamic link libraries (DLL), and the second type is ASCII files such as batch files (BAT), text files (TXT) and delimited files (CSV). Filenames are listed in system files called allocation tables (file allocation table "FAT" for DOS, and NT file system "NTFS" for Windows NT). Each entry in these tables corresponds to a single file and has a number of fields including filename, file attributes ("A" for archive and "H" for hidden) and the address of the file contents.

Filenames are limited in size. Compression, especially in DOS, may be required in order to place required key information in the filename. Compression is the art of representing a long sequence of information with a brief

sequence of codes. Popular data compression schemes include but are not limited to run-length coding, half-byte packing, Huffman coding, dictionary coding and adaptive dictionary coding. These methods are designed to compress information using the standard ASCII character set. However, certain ASCII characters (such as " ? / \ < > * | :) can not be used in naming files. Applying compression methods could produce non-valid filenames. To eliminate this problem an alternate code set is assigned to the ASCII characters that can be used in naming files. The alternate code set is normally called the filename character set.

The DOS style filenames have been the most popular and the most restrictive. The 95 and NT generation of Windows significantly relaxed the DOS restrictions. A Windows filename can be up to 253 characters long (11 for DOS) and a filename character can be any one of 245 ASCII codes (52 for DOS). DOS filenames are compatible with NT but NT filenames are not always compatible with DOS. Compatibility is lost either when the filename is longer than 11 characters or when any character in the filename is assigned an ASCII value outside the DOS range. For example: when saving the file "HELLO_MY_FRIEND.TXT" to floppy it is renamed to "HELLO_^1.TXT" which follows the DOS naming convention. If a sequence of other filenames also begins with the characters "HELLO_" then the 10th file in the sequence is named "HELLO^10.TXT" and so on.

Naming Conventions

A number of TSD naming conventions are in use today. These formats are organized in three classes. The classes are called associated, coded, and sequenced. Associated means that the filename extension defines the type of data storage format. For example, the extensions "HDR", "CFG", "DAT", and "INF" are used to indicate that the file contents are compatible with the IEEE COMTRADE standard. The non-extension part of an associated filename or the name is left at the discretion of the user and could be assigned in a coded or sequenced way.

Coded means that the filename contains some information about the event. In this case, the storage format is usually manufacturer specific. For example, certain files that are generated from digital fault recorders have the event date and time (up to 12/31/2079-23:59:59.99) and the recorder number (up to 255) coded in the filename. The recorder number is coded in the first 2 characters of the name (which is friendly up to recorder number 99) and the date and time are coded in the last 9 characters of the filename.

The sequenced filenames format is an incremental approach to naming files. This method is valid because the resulting filenames are always unique. The sequence may appear in the name or in the extension portion of the filename. The total number of attainable filenames is limited to the maximum value of the numerical sequence. When multiple devices are used then the device number is also coded in the filename. For example, some DFR filenames have the location name (up to 4 characters), the event number (4 characters = up to 9999 filenames before overwrite) and the channel group number (up to 3 characters) coded in the filename. A file named "MART1743.RCL" indicates that the event was recorded at the Martin substation and that the event number is 1743. The ".RCL" extension means that the file contents are from the lower group of the analog input channels.

Proposed Convention

The proposed TSD file naming convention defines a readable, delimited, long filename format. The delimiting character is a comma. In all cases where an alphabetical character is called for, the character can be either upper or lower case. Software should treat upper and lower case letters the same. The required contents of the filename will be as follows, in order: "Start Date, Start Time, Time Code, Station Identifier, Device Identifier, Company Name." Additional information may be added as required by the user. The working group suggests that the additional information be in the following order: "Duration, Type, Latitude, Longitude, Optional 1, Optional 2, and so on." The extension will always follow. Thus the minimum and extended formats are as follows:

Minimum: "yymmdd,hhmmssmmm,xyz,ssss,dddd,cccc.xxx"

Extended: "yymmdd,hhmmssmmm,xyz,ssss,dddd,cccc,dddhmmssmmm,t,llll,llll,oooo1,ooo2.xxx"

In order to conform with this recommendation, the first part of the file name (minimum format) is required. The additional fields described below may be used, and if used, they should be in the order shown.

Limitations

In addition to the limitations mentioned in the previous sections regarding the length and the allowable characters of the filename, members of the working group pointed out that an additional number of limitations exist while trying to save long filenames on CD and floppy. Specifically, floppies are limited to a maximum of 27 long filenames per disk regardless of the file sizes. Also, CDs do not work with filenames that are more than 64 characters in length, not counting the dot character. Given the current rate of advancement in technology it is highly possible that these limitations will vanish or become obsolete.

Beside each of the example filenames that follow we have placed the total number of characters, without including the dot character in the count.

Required Fields

Start Date (yymmdd): The year is in the form yy (two digits), followed by the month in the form mm (two digits), followed by the day of month in the form dd (two digits).

Start Time (hhmmssmmm): The time of day of file start is defined in military time format and can be specified to the required precision. For example, 170215183222, 170215183, 17021518, 1702152, and 170215 are all acceptable file start times.

Time Code (xyz): The time code is restricted to 3 formatted characters. The first two characters are a sign and a digit indicating the time difference between the time system used for start time and the Universal Time (UT, also called Greenwich Mean Time or “zulu” time). For the third character the letter “s” or “d” is used to indicate whether standard time or daylight time was in affect at the time of the recording. The third character may be set to the letter “a” indicating that the instrument automatically adjusts to standard or daylight at the conventional times for the area. In this case the time difference will always indicate the hours difference between UT and local standard time. If UT is used for file start time, this field will contain only the two letters “ut” or “UT.” Examples are as follows:

-5s	Standard Time is in use at this location and standard time is 5 hours behind UT.
-4d	Daylight Time is in use at this location and daylight time is 4 hours behind UT.
+5a	Standard Time is 5 hours ahead of UT and the device automatically adjusts to daylight time.
UT	Universal Time is used for file start time.

Station Identifier (ssss): Each user can formulate their own code for the station where the originating device is located (a unique name within the company). This is a variable length field and can contain letters, numbers, and some punctuation marks. Characters disallowed are: , ? “ / \ < > * | : (i.e. comma, question mark, quotation mark, forward slash, back slash, less than, greater than, asterisk, pipe and colon).

Device Identifier (dddd): Each user can formulate their own code for the originating device (a unique name within the station). This is a variable length field and can contain letters, numbers, and some punctuation marks. Characters disallowed are: , ? “ / \ < > * | : (i.e. comma, question mark, quotation mark, forward slash, back slash, less than, greater than, asterisk, pipe and colon).

Company Name (cccc): Each user can formulate their own code for the company (a unique name within the corporate realm) that operates the originating device. This is a variable length field and can contain letters, numbers, and some punctuation marks. Characters disallowed are: , ? “ / \ < > * | : (i.e. comma, question mark, quotation mark, forward slash, back slash, less than, greater than, asterisk, pipe and colon).

Up to this point all fields are required.

Optional Fields

Duration (dddhhmmssmmm): The duration can be specified to the precision necessary. The duration is equal to the difference between the first and last record in the file. The leading zeros are necessary. For example, a file with 99 seconds of data will have a duration code of 000000139 or 000000099. As with the start time, specify the duration with whatever precision you require.

Type (t): At the present time the working group has not formulated any specific codes for the file type field. The original idea was to use this field to describe the originating event, for example single line to ground fault. Some members of the working group have recommended using this field to describe the type of TSD. For example the code “log” can be used to specify periodic log files, “wave” for waveform files, “ser” for sequence of events files and “test” for files that are created during maintenance and testing operations.

Location Coordinates (lll,lll): Two fields are specified. The first field is to be an expression of latitude in degrees, a comma delineator, and the second is longitude in degrees. Leading zeros are necessary. These fields can be carried to the resolution desired (See the discussion on position information).

Optional (oooo): One (1) or more fields are allowed. This can be used for a comment or for any other purpose desired. Additional comment fields can also be included. These are variable length fields and can contain letters, numbers, and some punctuation marks. Characters disallowed are: , ? “ / \ < > * | : (i.e. comma, question mark, quotation mark, forward slash, back slash, less than, greater than, asterisk, pipe and colon).

File Extension

Extension (xxx): The extensions DAT, CFG, HDR and INF are reserved and used to indicate that the files are compatible with the COMTRADE format. There are no other reserved extensions, however, all of the previously noted restrictions on the filename extension will be observed. For example users should avoid using the EXE, BAT, SYS and COM extensions for naming TSD files.

Examples

The following is an example filename using the minimum format, the start time is given in millisecond resolution. It is a COMTRADE configuration file:

000809,175215183,-4d,sta80,ben717,nyiso.cfg (42)

The following is an example of a filename with start time given to a tenth of a second resolution and with the additional fields for file duration (95.9 seconds) and type:

000809,1752152,-4d,sta80,ben717,nyiso,0000000959,wave.cfg (58)

The following filename includes the additional field type and an optional field comment. Notice the back to back commas where the duration and location information was not used:

000809,175215183,-4d,sta80,ben717,nyiso,,wave,,,good relay test case.cfg (70)

Geographic Position (latitude & longitude)

These fields are optional. The working group points out that using the location coordinates will absolutely assure uniqueness of the filename. Latitude and Longitude is the most commonly used position coordinate system. Latitude is always expressed first. The letter “n” indicates latitude in the Northern Hemisphere. The letter “w” indicates a longitude in the Western Hemisphere. These letters can appear on either side of the expression. Latitude and longitude

will be expressed in decimal degrees. For degrees less than 100, leading zeros must be included.

Each degree of latitude covers 69 (statute) miles. Each degree of longitude covers 69 miles at the equator and less as you move toward the poles. Therefore .001 degrees covers 364 feet for latitude, 364 feet or less for longitude, which is more than adequate resolution for a station location. GPS position fixes with civilian equipment became more accurate and repeatable in May of 2000 when selective availability (SA) was turned off. SA was an intentional dithering of the GPS signal. Using a hand held civilian GPS receiver (after SA was turned off) at opposite corners of a large transmission substation gave these results: Entry Corner = 042610n,073905w & Opposite Corner = 042608n,073903w.

5 GPS position fixes were taken at a fixed location over a period of several weeks after SA was turned off, and it was observed that the repeatability was better than .001 degree. These results suggest to the working group that it is reasonable to specify station location using either 2 or 3 decimal places but not 4. However applications should accommodate whatever number of decimal places the user prefers. The following would be an acceptable expression of the above station's position: 04261n,07390w.

Additional Examples

Here is the same example filename as used above but with position information in the suggested form:

000809,175215183,-4d,sta80,ben717,nyiso,,,04305n,07767w.cfg (58)

The following is an example with all of the additional fields included:

000809,175215183,-4d,sta80,ben717,nyiso,0000000959,wave,04305n,07767w,comments.cfg (81)

Uniqueness

Uniqueness is essential for electronic filing systems as explained above in the significance section. Uniqueness depends on the choices made by the user. The filenames are absolutely guaranteed to be unique if the geographic position information is included. However, the working group believes that, depending on choices made by the user, the required fields alone can provide a sufficient guarantee that the filenames will be unique.

Applications

There are many applications that can be realized given a unique, informative file naming system. Possible suggested applications are: time line manager, TSD file manager, universal TSD viewer and analysis system, power system stability monitor, global/international data bank for TSD files and so on. Among the main beneficiaries of the proposed convention are the utilities and the emerging global and local energy regulatory commissions (NERC, FERC & SERC). These utilities and commissions can now realize the benefits of sharing a common electronic filing system. Take any TSD file generated from any device by any company anywhere on the planet and just seamlessly file it. Clearly, this vision can be realized only when device manufactures and third party software providers also support this convention. An IEEE standard convention for naming TSD files is therefore necessary in order to coordinate the collective industry efforts during and beyond the initial realization stages of the proposed convention.

References

"Survey of File Naming Schemes," A.&M. Makki, M. Taylor, L. Johnson, S. Semati and A. Rothweiler, Proceedings of the Fault and Disturbance Conference at Georgia Tech, May 1998.

"Overview of FAT, HPFS, and NTFS File Systems," Microsoft Windows NT Workstation 4.0 Resource Kit, Chapter 18, V.1997.

"Data Compression in Digital Systems," By Huffman, R., Chapman & Hall, New York, 1997.

"The Data Compression Book," Nelson, M., M&T Publishing, California, 1992.

"Working With MS-DOS files," User's Guide and Reference for the MS-DOS 5.0 Operating System, Chapter 4, V.1991.

"Text Compression," Bell, Cleary, and Witten, Prentice Hall, New Jersey, 1990.

"IEEE Standard Common Format for Transient Data Exchange (COMTRADE) for Power Systems," IEEE Standard P37.111, 1999..

"GPS Land Navigation," Ferguson, M., Glassford Publishing, Boise, Idaho, 1997

"Washington View: SA - Going the Way of the Dinosaur," Davis, D. A., GPS World, V. 11, No. 6, June 2000, pp 16-20.