

Parchive

Parchive (a portmanteau of **parity archive**, and formally known as **Parity Volume Set Specification**^[1]) is an erasure code system that produces **par** files for checksum verification of data integrity, with the capability to perform data recovery operations that can repair or regenerate corrupted or missing data.

Parchive was originally written to solve the problem of reliable file sharing on Usenet,^[2] but it is now commonly used for protecting any kind of data from data corruption, disc rot, bit rot, and accidental or malicious damage. Despite the name, Parchive uses more advanced techniques that do not use simplistic parity methods of error detection and correction.

As of 2014, **PAR1** is obsolete, **PAR2** is mature for widespread use, and **PAR3** is an experimental version being developed by MultiPar author Yutaka Sawada.^{[3][4][5][6]} The original SourceForge Parchive project has been inactive since November 9, 2010.^[7]

Contents

History

1

Versions

2

- Parity Volume Set Specification 1.0
 - 2.1
- Parity Volume Set Specification 2.0
 - 2.2
- Parity Volume Set Specification 3.0
 - 2.3

Software

3

- Windows
 - 3.1
- Mac OS X
 - 3.2
- Linux
 - 3.3
- FreeBSD
 - 3.4
- POSIX
 - 3.5

See also

4

References

5

External links

6

History

Parchive was intended to increase the reliability of transferring files via Usenet newsgroups. Usenet was originally designed for informal conversations, and the underlying protocol, NNTP was not designed to transmit arbitrary binary data. Another limitation, which was acceptable for conversations but not for files, was that messages were normally fairly short in length and limited to 7-bit ASCII text.^[8]

Various techniques were devised to send files over Usenet, such as uuencoding and Base64. Later Usenet software allowed 8 bit Extended ASCII, which permitted new techniques like yEnc. Large files were broken up to reduce the effect of a corrupted download, but the unreliable nature of Usenet remained.

With the introduction of Parchive, parity files could be created that were then uploaded along with the original data files. If any of the data files were damaged or lost while being propagated between Usenet servers, users could download parity files and use them to reconstruct the damaged or missing files. Parchive included the construction of small index files (*.par in version 1 and *.par2 in version 2) that do not contain any recovery data. These indexes contain file hashes that can be used to quickly identify the target files and verify their integrity.

Because the index files were so small, they minimized the amount of extra data that had to be downloaded from Usenet to verify that the data files were all present and undamaged, or to determine how many parity volumes were required to repair any damage or reconstruct any missing files. They were most useful in version 1 where the parity volumes were much larger than the short index files. These larger parity volumes contain the actual recovery data along with a duplicate copy of the information in the index files (which allows them to be used on their own to verify the integrity of the data files if there is no small index file available).

In July 2001, Tobias Rieper and Stefan Wehlius proposed the Parity Volume Set specification, and with the assistance of other project members, version 1.0 of the specification was published in October 2001.^[9] Par1 used Reed–Solomon error correction to create new recovery files. Any of the recovery files can be used to rebuild a missing file from an incomplete download.

Version 1 became widely used on Usenet, but it did suffer some limitations:

- It was restricted to handle at most 255 files.
- The recovery files had to be the size of the largest input file, so it did not work well when the input files were of various sizes. (This limited its usefulness when not paired with the proprietary RAR compression tool.)
- The recovery algorithm had a bug, due to a flaw^[10] in the academic paper^[11] on which it was based.
- It was strongly tied to Usenet and it was felt that a more general tool might have a wider audience.

In January 2002, Howard Fukada proposed that a new Par2 specification should be devised with the significant changes that data verification and repair should work on blocks of data rather than whole files, and that the algorithm should switch to using 16 bit numbers rather than the 8 bit numbers that PAR 1 used. Michael Nahas and Peter Clements took up these ideas in July 2002, with additional input from Paul Nettle and Ryan Gallagher (who both wrote Pari clients). Version 2.0 of the Parchive specification was published by Michael Nahas in September 2002.^[12]

Peter Clements then went on to write the first two Par2 implementations, QuickPar and par2cmdline. Abandoned since 2004, Paul Houle created phpar2 to supersede par2cmdline. Yutaka Sawada created MultiPar to supersede QuickPar. Multipar uses par2j.exe (which is partially based on par2cmdline's optimization techniques) to use as MultiPar's backend engine.

Versions

Versions 1 and 2 of the file format are incompatible. (However, many clients support both.)

Parity Volume Set Specification 1.0

For Par1, the files *f1*, *f2*, ..., *fn*, the Parchive consists of an index file (*f.par*), which is CRC type file with no recovery blocks, and a number of "parity volumes" (*f.p01*, *f.p02*, etc.). Given all of the original files except for one (for example, *f2*), it is possible to create the missing *f2* given all of the other original files and any one of the parity volumes. Alternatively, it is possible to recreate two missing files from any two of the parity volumes and so forth.^[13]

Par1 supports up to a total of 256 source and recovery files.

Parity Volume Set Specification 2.0

Par2 files generally use this naming/extension system: *filename.vol000+01.PAR2*, *filename.vol001+02.PAR2*, *filename.vol003+04.PAR2*, *filename.vol007+06.PAR2*, etc. The +01, +02, etc. in the filename indicates how many blocks it contains, and the vol000, vol001, vol003 etc. indicates the number of the first recovery block within the PAR2 file. If an index file of a download states that 4 blocks are missing, the easiest way to repair the files would be by downloading *filename.vol003+04.PAR2*. However, due to the redundancy, *filename.vol007+06.PAR2* is also acceptable. There is also an index file *filename.PAR2*, it is identical in function to the small index file used in PAR1.

Par2 specification supports up to 32768 source blocks and up to 65535 recovery blocks. Input files are split into multiple equal-sized blocks so that recovery files do not need to be the size of the largest input file.

Although Unicode is mentioned in the PAR2 specification as an option, most PAR2 implementations do not support Unicode.

Directory support is included in the PAR2 specification, but most or all implementations do not support it.

Parity Volume Set Specification 3.0

The Par3 specification was originally planned to be published as an enhancement over the Par2 specification. However, till date, it has remained closed source by specification owner Yutaka Sawada.

Software

Windows

- MultiPar (https://www.livebusinesschat.com/smf/index.php?board=396.0) (freeware) — Builds upon QuickPar's features and GUI, and Yutaka Sawada's par2j.exe as the PAR2 backend. MultiPar supports multiple languages by Unicode. The name of MultiPar was derived from "multi-lingual PAR client". MultiPar is able to add recovery data to ZIP and 7-Zip^[14] files, with a few minor caveats.^[15] MultiPar is also verified to work with Wine under TrueOS, and may work with other operating systems too.^[16] Although the Par2 components are (or will be) open source, the MultiPar GUI on top of them is currently not open source.^[17]
- QuickPar (freeware) — unmaintained since 2004, superseded by MultiPar.
- par2+ttbb (https://web.archive.org/web/20110311041855/http://chuchusoft.com/par2_tbb/) (GPLv2) — a concurrent (multithreaded) version of par2cmdline 0.4 using TBB.
- Par-N-Rar (GPL)
- phpar2 (http://paulhoule.com/phpar2/index.php) — advanced par2cmdline with multithreading and highly optimized assemblercode (about 66% faster than QuickPar 0.9.1)
- Rarslave (GPLv2)
- SmartPAR (freeware) — Unmaintained since 2002 and obsolete as this application written for Microsoft Windows only works with the original Par1 (PAR) Parchive format parity files. Superseded by QuickPar. It uses Reed-Solomon error correction to create new recovery files. SmartPAR is able to correct errors and recover missing parts of distributed files from PAR files.^[18] Last stable release 0.13d1 dated January 22, 2002^[19]
- Mirror (http://www.wehlius.de/mirror/index.html) — First PAR implementation, unmaintained since 2001.
- Original par2cmdline (http://parchive.sourceforge.net/) — (obsolete).
- par2cmdline (https://github.com/Parchive/par2cmdline) by BlacklkeEagle.

Mac OS X

- MacPAR deLuxe 4.2 (https://gp.home.xs4all.nl/Site/MacPAR_deLuxe.html)
- UnRarX (http://www.unrarx.com/)
- par2+ttbb (https://web.archive.org/web/20110311041855/http://chuchusoft.com/par2_tbb/) is a concurrent (multithreaded) version of par2cmdline 0.4 using TBB, GPLv2, or later.

Linux

- The par2 (https://github.com/Parchive/par2cmdline) utility, which is a maintained fork of par2cmdline.
- PyPar2 1.4 (http://pypar2.silent-blade.org/index.php?n=Main.HomePage), a frontend for par2.
- GPar2 2.03 (http://sourceforge.net/projects/parchive/)
- par2+ttbb (https://web.archive.org/web/20110311041855/http://chuchusoft.com/par2_tbb/) is a concurrent (multithreaded) version of par2cmdline 0.4 using TBB, GPLv2, or later.
- par2cmdline-mt (https://github.com/jkansanen/par2cmdline-mt) is another multithreaded version of par2cmdline using OpenMP, GPLv2, or later.

FreeBSD

- par2+ttbb (https://web.archive.org/web/20110311041855/http://chuchusoft.com/par2_tbb/) is a concurrent (multithreaded) version of par2cmdline 0.4 using TBB, GPLv2, or later. It is available in the FreeBSD Ports system as par2cmdline-tbb (https://www.freshports.org/archivers/par2cmdline-tbb/).
- par2cmdline (http://parchive.sourceforge.net/) is available in the FreeBSD Ports system as par2cmdline (https://www.freshports.org/archivers/par2cmdline/).

POSIX

Software for POSIX conforming operating systems:

- Par2 for KDE 4 (http://sourceforge.net/projects/ekpar2/)

See also

- Bit rot
- Disc rot
- Data corruption
- Checksum
- Comparison of file archivers - Some file archivers are capable of integrating parity data into their formats for error detection and correction:
- RAID - RAID levels at and above RAID 5 make use of parity data to detect and repair errors.

References

- ↑ **Re: Correction to Parchive on Wikipedia** (https://www.livebusinesschat.com/smf/index.php?topic=5736.msg38234#msg38234), reply #3, by Yutaka Sawada: "Their formal title are "Parity Volume Set Specification 1.0" and "Parity Volume Set Specification 2.0."
- ↑ "**Parchive: Parity Archive Volume Set**" (http://parchive.sourceforge.net/#desc). Retrieved 2009-10-29. "The original idea behind this project was to provide a tool to apply the data-recovery capability concepts of RAID-like systems to the posting and recovery of multi-part archives on Usenet."
- ↑ possibility of new PAR3 file (http://www.livebusinesschat.com/smf/index.php?topic=5098.0)
- ↑ Question about your usage of PAR3 (http://www.livebusinesschat.com/smf/index.php?topic=3339.0)
- ↑ Risk of undetectable intended modification (http://www.livebusinesschat.com/smf/index.php?topic=5025.msg29912;topicseen#msg29912)
- ↑ PAR3 specification proposal not finished as of April 2011 (http://www.livebusinesschat.com/smf/index.php?topic=3527.msg8850;topicseen#msg8850)
- ↑ "Parchive: Parity Archive Tool" (http://sourceforge.net/projects/parchive/). Retrieved 2012-09-02.
- ↑ Kantor, Brian; Lapsley, Phil (February 1986). "Character Codes" (https://tools.ietf.org/html/rfc977#section-2.2). *Network News Transfer Protocol* (https://tools.ietf.org/html/rfc977). IETF. p. 5. sec. 2.2. RFC 977. https://tools.ietf.org/html/rfc977#section-2.2. Retrieved 2009-10-29.
- ↑ Nahas, Michael (2001-10-14). "Parity Volume Set Specification v1.0" (http://parchive.sourceforge.net/docs/specifications/parity-volume-spec-1.0/article-spec.html). Retrieved 2017-06-19.
- ↑ Plank, James S.; Ding, Ying (April 2003). "**Note: Correction to the 1997 Tutorial on Reed-Solomon Coding**" (http://web.eecs.utk.edu/~plank/plank/papers/CS-03-504.html). Retrieved 2009-10-29.
- ↑ Plank, James S. (September 1997). "A Tutorial on Reed-Solomon Coding for Fault-Tolerance in RAID-like Systems" (http://web.eecs.utk.edu/~plank/plank/papers/SPE-9-97.html). Retrieved 2009-10-29.
- ↑ Nahas, Michael; Clements, Peter; Nettle, Paul; Gallagher, Ryan (2003-05-11). "Parity Volume Set Specification 2.0" (http://parchive.sourceforge.net/docs/specifications/parity-volume-spec/article-spec.htm ml). Retrieved 2009-10-29.
- ↑ Wang, Wallace (2004-10-25). "Finding movies (or TV shows): Recovering missing RAR files with PAR and PAR2 files". *Steal this File Sharing Book* (https://books.google.com/books?id=FGfMS5kymmCC&pg=PT183) (1st ed.). San Francisco, California: No Starch Press. pp. 164 – 167. ISBN 1-59327-050-X. Retrieved 2009-09-24.
- ↑ "7-Zip" (https://sourceforge.net/p/sevenzip/feature-requests/1006/).
- ↑ How to add recovery record to ZIP or 7-Zip archive (http://www.livebusinesschat.com/smf/index.php?topic=4922.0)
- ↑ MultiPar works with PCBSD 9.0 (http://www.livebusinesschat.com/smf/index.php?topic=4902.0)
- ↑ contacted you, asking about sourcecode (https://www.livebusinesschat.com/smf/index.php?topic=5402.0)
- ↑ Wang, Wallace (2004-10-25). "Finding movies (or TV shows): Recovering missing RAR files with PAR and PAR2 files". *Steal this File Sharing Book* (https://books.google.com/books?id=FGfMS5kymmCC&pg=PT183) (1st ed.). San Francisco, California: No Starch Press. pp. 164 – 167. ISBN 1-59327-050-X. Retrieved 2009-09-24.
- ↑ "Parchive: Parity archive tool" (http://parchive.sourceforge.net/). Retrieved 2009-09-26.

External links

- Parchive project - full specifications and math behind it (http://parchive.sourceforge.net/)
- Introduction to PAR and PAR2 (http://www.ydecode.com/page_articles003.htm)
- Slyck's Guide To The Usenet Newsgroups: PAR & PAR2 Files (http://www.slyck.com/Newsgroups_Guide_PAR_PAR2_Files)
- Guide to repair files using PAR2 (http://www.binaries4all.com/quickpar/repair.php)
- par2+ttbb (https://web.archive.org/web/20100912073937/http://chuchusoft.com/par2_tbb/)
- Par-N-Rar (http://www.milow.net/public/projects/parnrar/parnrar.html)
- Rarslave (http://www.irasnyder.com/devel/#rarslave)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Parchive&oldid=803415025"

