

OpenAFS Status

Jeffrey Altman
Derrick Brashear
The OpenAFS Project
27 May 2010

What we'll tell you of

- Status reports
 - Unix and OSX clients
 - Windows clients
 - Servers
- Google Summer of Code

What we'll tell you of

- New Work
- Release schedules
- Governance

In the previous year

- New stable release. (1.4.12)
- New development releases continue.
(1.5.75 soon)

1.4 releases

- 1.4.8: large number of fixes to Rx packet tracking.
- 1.4.10: more Rx fixes, including freeing and idle tracking; ubik database recovery, simple multiple realm support.
- 1.4.11: Linux cache manager error handling, dynamic vnode pool sizing, fixes to mmap() issues (deadlock/writeback failure), client PAG reporting.
- 1.4.12: more Rx fixes, store file size issue fixes, orphaned cache chunk fixes

1.5 in the past year

- Focus has been on stabilisation and code cleanup for maintainability.
- Cache manager refactored to allow extensions in the future (e.g. Rx/OSD)
- Demand Attach Fileserver bug fixes.
- Outstanding 1.5 series bugfixes.
- Compilation warning elimination.

Unix platform summary

- AIX 5 and 6 (though 6.3)
- FreeBSD 7, 8 and current
- HP-UX 11.0, 11i v1 and v2
- Irix 6.5
- Linux 2.2, 2.4, 2.6 (ia32, ia64, x86_64, ppc, ppc64, **arm**, sparc, sparc64, s390, s390x)
- MacOS 10.3, 10.4, 10.5, **10.6 (including 64 bit)**.
- **OpenBSD 4.4, 4.5, 4.6, 4.7.**
- Solaris 2.6, 7, 8, 9, 10, 11 (and OpenSolaris)

Clients

- Numerous interaction issues with GUI environments have been addressed.
 - Including Finder (more later)
- Far fewer resources are leaked during client operation. We're not perfect yet.

Linux

- Linux kernel symbols continue to be removed from our view.
- Aside from the NFS translator this has not yet been an issue for basic functionality.
- Dynamic sizing for AFS client vnode pool.

Linux

- Keyrings now authoritative for PAGs
- Possible to write a larger-than-cache mmap
()ed file.
- 1.5 series features tuning to better utilize
the Linux kernel VFS interface.
 - Performance.
 - Correctness.

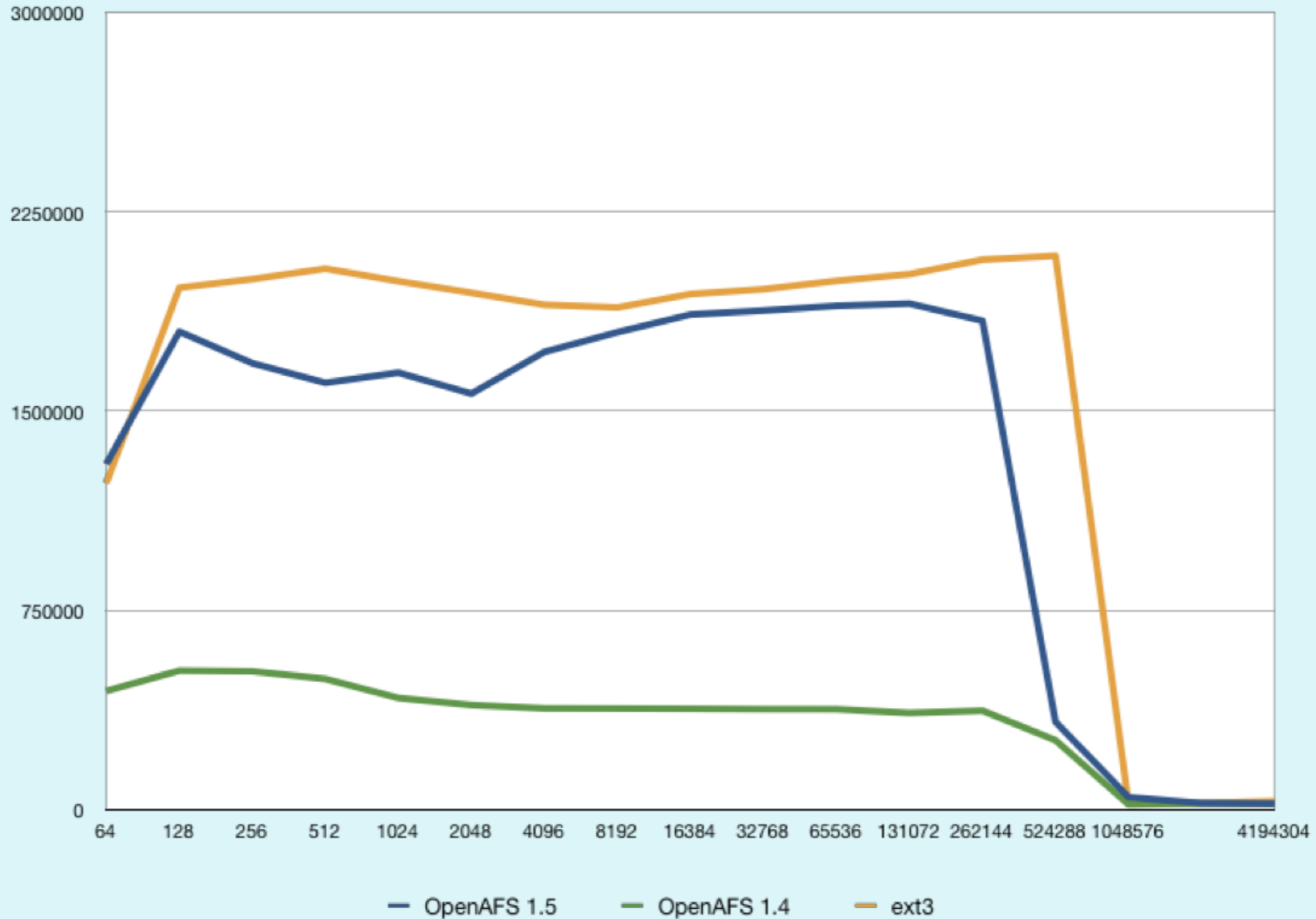
Page Cache Improvements

- Reduce the number of redundant reads by correctly using the page dirty flag
- Enable readahead when filling the page cache from disk
- Remove duplicate writes of pages to disk by telling the kernel what we're doing
- Populate the page cache with a background thread, rather than doing it during requests

Minimize Data Copies

- Copying data is expensive
- Minimise the number of copies between the network, the various caches and user space
- Significant improvements made to write-on-close
- Other cases an ongoing project

Cache read performance: AFS should match ext3 below 1GB



MacOS X

- AFSCommander integrated (Preferences Pane).
- Offers GUI configuration of many aspects of user experience with OpenAFS on Mac.
- MacOS 10.6 Snowleopard includes support for 64 bit.
 - 64 bit kext
 - 64 bit userspace (including on 32 bit kernel)

1.5 new features

- Cache bypass (Linux-only, new since 1.5.53)
 - Support for additional platforms is needed.
- Disconnected AFS
 - Supports read-write operation
 - Cached writes do not currently persist across client restart/reboot.
- Tunable cache readahead (“*fs precache*”)

1.5 new features

- Universal mountpoint-less volume addressing (`/afs/.:mount/cell:volumeid/`) is available.
 - Originally done for the Linux NFS translator.
- An extension allows any vnode to be used. (`/afs/.:mountcell:volumeid:vnodeid:uniquifier/`)
 - Needed for Finder workarounds.
- Rx NAT server ping

MacOS issues

- A fix for the MacOS “Finder cross-volume drag” issue.
- A userspace helper and the ad-hoc “reference any vnode” semantics are used.
- Still no PAG support.
- Bulkstatus now enabled.
- Finder dropbox (insert file) support.

MacOS issues

- Finder uses fsevents to change your view.
 - On authentication changes.
 - Not on callbacks yet.
- coreservicesd and leaked stat info.
- StartupItems->LaunchDaemons

Fileservers

- Not new, but you all keep asking:
- Partitions larger than 2 tb are supported.
- Old “vos” clients may report odd numbers for empty partitions.
- Large volumes are not yet properly supported (quota/disk used)

Improved network behavior

- Quotas enforced on TellMeAboutYourself / WhoAreYou calls to clients will preclude resource hogging.
- Indefinite waits are no longer possible due to idle (no data) timeout enforcement.

Demand Attach Fileserver

- Known issues:
 - Host table state can be discarded on restart erroneously.
- Testing ongoing.
- Please share other issues if you have them!

Google Summer of Code

- Last year:
 - Windows MMC management snapin, Brant Gurganus.
 - Improved OpenAFS server selection, Jake Thebault-Spieker.
 - OpenAFS features in the Linux kAFS client, Wang Lei.

Google Summer of Code

- This year:
 - Encrypted Storage, Sanket Agarwal.
 - A port of OpenAFS to NetBSD, Matt Smith.
 - Userspace interface for the Linux kAFS client, Wang Lei.
 - Extended attributes via AppleDouble files, Kelli Ireland.
 - Implementing Microsoft's Safe String (StrSafe.h) Library for UNIX/Linux, Jonas Sundberg.

Edinburgh Hackathon

- RxK5
- Rxgk
- Rx/OSD
- Rx UDP enhancements
- AFS3 protocol-wide RPC updates
- “libosi”
- Extended callbacks

Rxgk

- Draft protocol available.
- Authenticator will be revised to use new-style Rxkad authenticator for security.
- Security header from RxK5 will be added.
- Protocol draft will be updated.

Rx/OSD

- Mirroring mode found to have potential issues due to lack of transactions.
- Standardisation of new RPCs will be proposed.

Rx UDP enhancements

- Window size negotiation.
- Path MTU discovery.
 - Requires ping payload manipulation.
- Rx option negotiation including max calls.
 - Requires additional ping-ack payload field.
- Revisit jumbograms after Path MTU works.

AFS3 RPC refresh

- Intent is to future-proof any new RPCs.
 - 64 bit (100ns granular) time support.
 - 64 bit FIDs (AFS vnode identifiers).
 - Per-cell UUID.
 - Server UUID in RPCs.
 - Larger status objects to support large volumes (including quotas).

libosi

- Intended to provide OS-agnostic interfaces to common tasks.
- Previously done to support aborted trace framework.
- Will be refactored and integrated beginning with its own support framework.
- Compiler/environment detection.

Extended callbacks

- Draft published with the IETF.
- Asynchronous callback coalescing removed pending definition of AFS3 semantics.

New Work

- RxTCP.
- Byte range locking.
- PTS alternate auth name support.

Release schedules

- The Edinburgh hackathon included discussion of priorities.
- Branch for 1.6 candidate with stable DAFS was to begin near-term.
- Changes for next-stable branch will be revisited after 1.6 release.
- 1.4.12 released 2009.

Release schedules (Reality)

- Driving feature for 1.6 is DAFS
- Branch for 1.6 candidate when DAFS bug/bugs are fixed and tested.
- 1.8 series development will follow the branching.

Governance

- OpenAFS Foundation deferred.
- Likely contributors in financial pinch.
- Software Freedom Conservancy membership being pursued.
- Offers non-profit vehicle for holding and accepting resources.
- Trademark issues still being worked on.

On version control

- git was rolled out last year.
- It should be much easier to track upstream as we work with you on integrating your changes.
- Visit gerrit.openafs.org for more visibility into the contribution process and to help review incoming submissions.
- Nearly 2000 submissions pushed through gerrit.

OpenAFS Status