

[PATCH 0/7] v9fs: Plan 9 resource sharing protocol (2.0)

Source: <http://linux.derkeiler.com/Mailing-Lists/Kernel/2005-06/1644.html>

ericvh_at_gmail.com

Date: 06/07/05

Date: Tue, 7 Jun 2005 09:49:17 -0500

To: linux-kernel@vger.kernel.org

This is part [0/7] of the v9fs-2.0 patch against Linux 2.6.

Hi everyone, we're proud to announce v2.0 of our 9P2000 file system driver support for Linux.

OVERVIEW

V9FS is a distributed file system for Linux which provides an implementation of the Plan 9 resource sharing protocol 9P. It can be used to share all sorts of resources: static files, synthetic file servers (such as /proc or /sys), devices, and application file servers (such as FUSE).

BACKGROUND

Plan 9 (<http://plan9.bell-labs.com/plan9>) is a research operating system and associated applications suite developed by the Computing Science Research Center of AT&T Bell Laboratories (now a part of Lucent Technologies), the same group that developed UNIX, C, and C++. Plan 9 was initially released in 1993 to universities, and then made generally available in 1995. Its core operating systems code laid the foundation for the Inferno Operating System released as a product by Lucent Bell-Labs in 1997. The Inferno venture was the only commercial embodiment of Plan 9 and is currently maintained as a product by Vita Nuova (<http://www.vitanuova.com>). After updated releases in 2000 and 2002, Plan 9 was open-sourced under the OSI approved Lucent Public License in 2003.

The Plan 9 project was started by Ken Thompson and Rob Pike in 1985. Their intent was to explore potential solutions to some of the shortcomings of UNIX in the face of the widespread use of high-speed networks to connect machines. In UNIX, networking was an afterthought and UNIX clusters became little more than a network of stand-alone systems. Plan 9 was designed from first principles as a seamless distributed system with integrated secure network resource sharing.

Applications and services were architected in such a way as to allow for implicit distribution across a cluster of systems. Configuring an environment to use remote application components or services in place of their local equivalent could be achieved with a few simple command line instructions. For the most part, application implementations operated independent of the location of their actual resources.

Commercial operating systems haven't changed much in the 20 years since Plan 9 was conceived. Network and distributed systems support is provided by a patchwork of middle–ware, with an endless number of packages supplying pieces of the puzzle. Matters are complicated by the use of different complicated protocols for individual services, and separate implementations for kernel and application resources. The V9FS project (<http://v9fs.sourceforge.net>) is an attempt to bring Plan 9's unified approach to resource sharing to Linux and other operating systems via support for the 9P2000 resource sharing protocol.

V9FS HISTORY

V9FS was originally developed by Ron Minnich and Maya Gokhale at Los Alamos National Labs (LANL) in 1997. In November of 2001, Greg Watson setup a SourceForge project as a public repository for the code which supported the Linux 2.4 kernel.

About a year ago, I picked up the initial attempt Ron Minnich had made to provide 2.6 support and got the code integrated into a 2.6.5 kernel. I then went through a line–for–line re–write attempting to clean–up the code while more closely following the Linux Kernel style guidelines. I co–authored a paper with Ron Minnich on the V9FS Linux support including performance comparisons to NFSv3 using Bonnie and PostMark – this paper appeared at the USENIX/FREENIX 2005 conference in April 2005:

(<http://www.usenix.org/events/usenix05/tech/freenix/hensbergen.html>).

STATUS

The code is reasonably stable, although there are no doubt corner cases our regression tests haven't discovered yet. It is in regular use by several of the developers and has been tested on x86 and PowerPC (32–bit and 64–bit) in both small and large (LANL cluster) deployments. Our current regression tests include fsx, bonnie, and postmark.

It was our intention to keep things as simple as possible for this release — trying to focus on correctness within the core of the protocol support versus a rich set of features. For example: a more complete security model and cache layer are in the road map, but excluded from this release. Additionally, we have removed support for mmap operations at Al Viro's request.

PERFORMANCE

Linux-Kernel: [PATCH 0/7] v9fs: Plan 9 resource sharing protocol (2.0)

Detailed performance numbers and analysis are included in the FREENIX paper, but we show comparable performance to NFSv3 for large file operations based on the Bonnie benchmark, and superior performance for many small file operations based on the PostMark benchmark. Somewhat preliminary graphs (from the FREENIX paper) are available (<http://v9fs.sourceforge.net/perf/index.html>).

RESOURCES

The source code is available in a few different forms:

tarballs: <http://v9fs.sf.net>

CVSweb: <http://cvs.sourceforge.net/viewcvs.py/v9fs/linux-9p/>

CVS: :pserver:anonymous@cvs.sourceforge.net:/cvsroot/v9fs/linux-9p

Git: rsync://v9fs.graverobber.org/v9fs (webgit: <http://v9fs.graverobber.org>)

9P: tcp!v9fs.graverobber.org!6564

The user-level server is available from either the Plan 9 distribution or from <http://v9fs.sf.net>

Other support applications are still being developed, but preliminary version can be downloaded from sourceforge.

Documentation on the protocol has historically been the Plan 9 Man pages (<http://plan9.bell-labs.com/sys/man/5/INDEX.html>), but there is an effort under way to write a more complete Internet-Draft style specification (<http://v9fs.sf.net/rfc>).

There are a couple of mailing lists supporting v9fs, but the most used is v9fs-developer@lists.sourceforge.net -- please direct/cc your comments there so the other v9fs contributors can participate in the conversation. There is also an IRC channel: <irc://freenode.net/#v9fs>

-Eric Van Hensbergen
V9FS Project

=====
===== 06/07/2005 - ericvh ===

* Change hardcoded dirread size to negotiated packet size

=====
===== 06/06/2005 - ericvh ===

* Fix umount NULL references (from Pekka)
* Code cleanups suggested by Alexy Dobriyan

=====
===== 06/03/2005 - ericvh ===

* Fix bogus v9ses->transport assignment & free
* Add request timeout code to help with hung servers
* fix umount of hung/disconnected servers

=====
===== 06/02/2005 - ericvh ===

* convert to u64 types (from Pekka)

Linux-Kernel: [PATCH 0/7] v9fs: Plan 9 resource sharing protocol (2.0)

- * fix buffer overflow in conv (from Pekka)
- * fix dev initialization (from Ron)
- * fix rookie errors in mux and remove sed injected whitespace

=====
06/01/2005 – ericvh ===

- * split dentry routines out to vfs_dentry to reduce vfs_inode size
- * some new error codes from ron
- * release v9fs-2.0-rc7

=====
05/31/2005 – ericvh ===

- * get rid of dynamic rpc struct allocation and associated slab
- * get rid of vestigial address_space_operations (from Pekka)
- * fix type name convention to Linux kernel standard (from Pekka)

=====
05/27/2005 – ericvh ===

- * remove potential slab leak in init_v9fs
- * modify QID to prevent inodes of 0 or 1

=====
05/25/2005 – ericvh ===

- * more LKML suggested cleanups
- * change packet and mistruct slab alloc to kmalloc

=====
05/24/2005 – ericvh ===

- * remove BUG() statements in OOM cases
- * clean up more bogus kfree checks
- * clean up error cases and redundant assignments in vfs_inode
- * change the way aname (remotename) is specified and allocated

=====
05/23/2005 – ericvh ===

- * release 2.0-rc6
- * Fix some style screwups

=====
05/20/2005 – ericvh ===

- * Fix multi-user bug reported by Ron & Ollie
- * change /usr/src/linux to KDIR in CVS makefile
- * fix bogus copy_to_user error messages

=====
05/16/2005 – ericvh ===

- * release 2.0-rc5
- * Adrian Bunk style corrections
 - add statics where appropriate
 - remove unused procedures
 - pay attention to copy_from_user and copy_to_user return codes

=====
05/05/2005 – ericvh ===

- * additional cleanup of free null checks throughout code
- * fix slab leaks

=====
05/04/2005 – ericvh ===

- * release 2.0-rc3
- * use hash table in error mapping routines instead of linear search

Linux-Kernel: [PATCH 0/7] v9fs: Plan 9 resource sharing protocol (2.0)

* added all errors from fossil, vacfs, and u9fs – now we need to trim down

=====
05/03/2005 – ericvh
=====

- * switch to using lib/parser.c for command line parsing
- * change default user to nobody
- * re-enable slabs and remove slab debug
- * enable kernel slab debug options and fix corrupts slab names\
- * rework slab from session based to msize based

=====
05/02/2005 – ericvh
=====

- * clean up trans_sock a bit more
- * eliminate normal allocation/free leak detection code from mainline
- * fix double-free in t_flush code

=====
04/30/2005 – ericvh
=====

- * remove mmap support – support available in mmap branch (unmaintained)
- * clean up error paths in get_sb

=====
04/29/2005 – ericvh
=====

- * protocol field shifting for lucho – make sure you update u9fs too
- * remove transient clunks from lookup (broke linux semantics)
- * whole mess of sparse cleanup changes – multithreaded broken for now

=====
04/28/2005 – ericvh
=====

- * fix memory leaks in transport and session_init

=====
04/27/2005 – ericvh
=====

- * fix t_version tid (to -1) so we work with Plan 9 servers
- * change to C99 initializers
- * CHANGE KCONFIG option to 9P_FS from 9P (will require BK change outside CVS)

=====
04/25/2005 – ericvh
=====

- * release 2.0-rc2

=====
04/23/2005 – ericvh
=====

- * fix mount segfault problem

=====
04/22/2005 – ericvh
=====

Linux-Kernel Mentor Comments responded to:

- * clean-up Documentation (and remove/move in patch/BK)
 - * move README to Documentation/v9fs/v9fs.txt
 - * integrate AUTHORS into MAINTAINERS
 - * remove COPYING, TODO, INSTALL and CHANGELOG
 - * get rid of pointless comments in Makefile
- * clean up debug statements and legacy version hacks
 - * put debug defines in do {} while(0) loops
 - * make sure all printks have a KERN_
 - * get rid of redundant/pointless macros
- * style cleanup

Linux-Kernel: [PATCH 0/7] v9fs: Plan 9 resource sharing protocol (2.0)

- * fix function comments to match kernel_doc style
- * remove { } on single line if statements
- * get rid of big inlines
- * get rid of global rpc slab
- * return negative on error, 0 on success uniformly
- * fit entire prototype on line when possible
- * prefix all globals with v9fs_
- *

=====
04/18/2005 – ericvh
=====

- * fix atomic while sleeping warning (thanks Lucho)
- * move inline functions to top of vfs_inode to fix build warning (thanks Lucho)

=====
04/13/2005 – ericvh
=====

- * fix 9p2000.u Rerror marshalling to match RFC (thanks Russ)
- * release 2.0-rc1

=====
04/12/2005 – ericvh
=====

- * fix clunk behavior
- * fix oops during mount problems

=====
03/25/2005 – ericvh
=====

- * add extended flag
- * add nodevmap flag
- * add then remove nosuid/nosgid flag (done implicitly)

=====
03/24/2005 – ericvh
=====

- * split debug prints into debugs and errors
- * add checks for extended mode with link operations
- * allow override of extended mode with a mount option

=====
03/23/2005 – ericvh
=====

- * Code cleanup
- * Make debug level settable at mount time

=====
03/22/2005 – ericvh
=====

- * fix various problems with mmap support
- * We now pass fsx!

=====
03/20/2005 – ericvh
=====

- * add mmap addr operation support

=====
03/19/2005 – ericvh
=====

Linux-Kernel: [PATCH 0/7] v9fs: Plan 9 resource sharing protocol (2.0)

- * add support for hard links
- * mknod file (PIPES, SOCKETS, DEVICES) support

===== 03/17/2005 – ericvh =====

- * fix open modes so we actually do truncate and append correctly
- * code cleanup, do a better job of conditionalizing extensions

===== 03/16/2005 – ericvh =====

- * add support for readlink and followlink
- * add support for creation of symlinks

===== 03/14/2005 – ericvh =====

- * revert setattr to using (no-change) values in stat structure
- * revert rename to using (no-change) values in stat structure
- * fix rename bug (was trying to rename parent directory, not sub-dir)
- * fix bug with chown that wouldn't set uid right
(wasn't setting string name, only n_uid and u9fs was favoring string name)

===== 03/11/2005 – ericvh =====

- * fix setuid and setgid

===== 03/03/2005 – ericvh =====

- * rearrange dir structure to make parsing a bit easier
 - * debug numeric id and unix[n] extension
- NOTICE: make sure you use a new u9fs, or thing will break

===== 03/02/2005 – ericvh =====

- * Add 9P2000.u error semantics
- * Add 9P2000.u numeric ids to directory structures
- * Provision for variable length extension string (unix[n])

===== 03/01/2005 – ericvh =====

- * add version check for definition of vfs_permission versus generic_permission
- * add new 9P2000.u changes to 9p.h

===== 02/25/2005 – ericvh =====

- * fix t_wstat in set_attr so chmod & chown works
- * fix t_wstat implementation in rename (but u9fs still broken for rename)

===== 02/23/2005 – ericvh =====

Fixed some permissions bugs associated with mounting actual Plan 9 file servers. Also fixed a bug which crept into attach where we were sending server address as the aname.

We need to start doing a better job about keeping this file up-to-date.

=====
12/15/2004 – ericvh
=====

Some updates (mostly from lucho) which fix up the re-write. Seems to pass bonnie and postmark versus u9fs server with same performance as before the rewrite.

=====
12/03/2004 – ericvh
=====

WARNING: CVS TIP IS UNSTABLE – use test3 if you want some level of stability

- * line for line rewrite
- * major source reorganization
- * move to using slab allocators for several structures (fcall, pkts, mistat)
- * removal of all cacheing operations including addr_ops (required for mmap)

Sorry for making a somewhat unstable commit. Didn't want to hold up others from making progress with the code while I fixed up the rewrite. Code should be much more approachable now and closer to LKML guidelines.

=====
11/21/2004 – lucho (via patch to ericvh)
=====

- * add info to 9P debug prints
- * move socket receive processing to a kernel thread
- * send t_flush in response to user-interrupted transactions

=====
11/09/2004 – ericvh
=====

- * restructure debug prints

removed a bunch of noisy debug prints and restructured with a more simple set of sections. Disabled logging as it appears to have some form of corruption, may remove immediately soon.

- * Big Kernel Locks

Intended to remove these, but they seem to still be necessary to guarantee serialization. Ended up adding more – will work hard to get rid of these in the next round of changes.

- * Sleep/Wakeup in p9client was causing multithreaded problems. Sprinkled liberally with wakeup calls which seems to unweave threads (assuming BKLs in place). At the very least it passes Bonnie now

Linux-Kernel: [PATCH 0/7] v9fs: Plan 9 resource sharing protocol (2.0)

v9fs-2.0-final

```
---
commit ac45b5118f193193aba4c328fa7afebcaal9bd9c
tree 6f6b5048e0ff1d1d2e04c9af2790dc0ea93f9956
parent 7cef5677ef3a8084f2588ce0a129dc95d65161f6
author <ericvh@ericvhG5.(none)> Tue, 07 Jun 2005 09:08:03 -0500
committer <ericvh@ericvhG5.(none)> Tue, 07 Jun 2005 09:08:03 -0500
Documentation/filesystems/v9fs.txt |      88 ++
MAINTAINERS                        |       11
fs/9p/9p.c                          |     357 ++++++++
fs/9p/9p.h                          |     339 ++++++++
fs/9p/Makefile                      |       17
fs/9p/conv.c                        |     725 ++++++++
fs/9p/conv.h                        |       34 +
fs/9p/debug.h                       |       68 ++
fs/9p/error.c                       |       92 ++
fs/9p/error.h                       |      176 ++++
fs/9p/fid.c                         |     233 ++++++
fs/9p/fid.h                         |       55 +
fs/9p/idpool.c                      |     150 ++++
fs/9p/idpool.h                      |       40 +
fs/9p/mux.c                         |     452 ++++++++
fs/9p/mux.h                         |       37 +
fs/9p/trans_sock.c                 |     272 ++++++
fs/9p/transport.h                  |       43 +
fs/9p/v9fs.c                        |     408 ++++++++
fs/9p/v9fs.h                        |       86 ++
fs/9p/v9fs_vfs.h                   |       51 +
fs/9p/vfs_dentry.c                 |     139 ++++
fs/9p/vfs_dir.c                    |     240 ++++++
fs/9p/vfs_file.c                   |     420 ++++++++
fs/9p/vfs_inode.c                  |    1418 ++++++++
fs/9p/vfs_super.c                  |     259 ++++++
fs/Kconfig                         |       11
fs/Makefile                         |        1
28 files changed, 6222 insertions(+), 0 deletions(-)
```

-
To unsubscribe from this list: send the line "unsubscribe linux-kernel" in
the body of a message to majordomo@vger.kernel.org
More majordomo info at <http://vger.kernel.org/majordomo-info.html>
Please read the FAQ at <http://www.tux.org/lkml/>