Section E.3.1

Remote Virtual Disk Upgrade

by J. H. Saltzer

Technical Plan Section D describes the requirements for and planned uses of remote virtual disk and remote file systems. This section extends that discussion by describing in detail a plan for upgrade of the Remote Virtual Disk System originally developed by the M.I.T. Laboratory for Computer Science. A separate System Manual describes operation and use of the Remote Virtual Disk System.

This upgrade plan is organized in terms of a series of releases, with release-by-release feature lists. All those upgrade ideas that are not actually scheduled for implementation are collected together in a single release description dubbed the outplan release.

For purposes of discussing release, this document describes the Remote Virtual Disk system in four major parts:

I. the server
II. the client driver
III. the client commands
IV. the documentation

Versions 1.x of the various parts are uncoordinated; versions starting with 2.0 consist of coordinated releases of all four parts. Current target release dates:

2.0	Initial fall release	7/29/86
2.1	Cleanups, ioctl interface	8/29/86
2.2	UID's, nonces, and physical ops	10/3/86
3.0	initial Kerberos integration	11/14/86
3.1	scheduled wrapups	11/30/86
3.2	everything else	outplan

I. RVD server

Server version 1.0/1.1.

Server obtained from L.C.S. in June, 1985.

Server version 1.2 (1/1/86)

1. Bug fix—avoid crashing on delete_virtual operations.

Server version 1.3 (5/22/86)

1. Comes up with no spinups allowed; requires an allow_spinups operation to start it.
2. Logs all spindowns, including those from spindown_host and spindown_virtual.
3. All logging to the 4.3 syslogd.
4. Ignores log_truncate requests; crontab-triggered manipulation of syslog moves old logs aside.
5. Code to use RVDMASTER (old data base format) removed.
6. Display_virtual returns total number of connections and only one packet full of connection descriptions; can be told which description to start with. Can return either connections of a client or connections of a pack.
7. Allow_spinups, a new operation, controls the mode of allowable spinups on server as a whole or on specific packs.
8. Set_message and get_message permit a message-of-the-day to be posted by operations and to be read by clients.
9. Require_authorization, a new operation, tells server to read its authorization file for a password and begin accepting requests from other network nodes. If invoked while running, server refreshes its stored password from the authorization file.

Server version 1.4 (6/10/86)

1. Display_virtual has a new option to return a list of all packs that have one or more spinups, with time since most recent use.
2. Logging operations now require operations password.
3. Accepts the operations password to allow spinup of packs in otherwise disallowed modes.
4. Logs shutdowns.
5. Bughalt on receiving "network unreachable" changed to increment a statistics counter, and carry on.

Server version 1.5 (6/15/86)

1. Ported to IBM RT PC.

Server version 1.6 (6/30/86)

Never deployed—changes integrated in 2.0.

Server version 2.0—Initial Fall release)

1. Has a version number that goes into log at startup time.
2. Client errors are distinguished from server errors so that server errors can be directed to a separate log.
3. Logs attempts to do control operations with wrong password.
4. General code review.

Server version 2.1—Minor cleanups and test suite

1. Gathers and logs server queue length statistics.
2. Bughalts now log all statistics before exiting.
3. Rvdexchanges are logged correctly.
4. A server test suite provides a client package that pounds the server in any of several different ways and can be run on several clients at once. It also tests every RVD control function and every RVD and RVDCTL error condition that a client can generate. A checkout mode in the server causes it to think it is getting all possible disk error conditions from the kernel and from the network.

Server version 2.2—uid, nonces, and physical ops

1. Most calls to bughalt are replaced with more sensible responses.
2. The server accepts a -r option to mean that if a bughalt is encountered, the server should automatically restart itself.
3. Pack unique id's are provided to make rvdexch atomic. The pack UID is returned in the spinup-ack packet. The operations add_virtual, and exchange_virtual require pack UID's, and modify_ virtual allows them, as a way of renaming packs. A new respinup packet type accepts spinup with pack uid rather than name.
4. Server responds to nonces in control packets by including them in response packets.
5. Server provides delete_physical, use_physical, and disuse_physical control operations.
6. Server defines and returns a new error code ("requested mode temporarily unavailable") when allow_spinups has restricted spinup modes more tightly than that specified in add_virtual.

Server version 3.0—Initial Kerberos integration

1. Server recognizes authenticated-spinup packet type, containing a Kerberos ticket, and interprets capability of pack as the name of an access control list, and owner field of pack as name of a user with full access.
2. Server accepts Kerberos tickets in control operations, and maintains separate access control lists for operations, maintenance, and shutdown.
3. Server no longer logs "display_virtual: no such connection" incidents.

Server version 3.1—wrapups

1. Server scheduling allows control operations during heavy read/write activity.
2. Server provides get_load function.

Server Release 3.2 (outplan)

1. Spinup should report pack in use in incompatible mode even if password isn't supplied.
2. A way is needed to extract rvddb information from a running server.
3. Should have general set/get control operation pairings.
4. Server should catch and handle the signal that is associated with shutdown, by doing a graceful shutdown.
5. Measure performance: maximum service rates and fanout.
6. Server test suite: write the man page, finish writing tests, and put rvdtest into the release tree.

II. Client driver

Client version 1.0.

Driver obtained from L.C.S. in June, 1985

Client version 1.1/1.2. (1/1/86)

1. New state (misnamed "server crashed") added to client driver. Stops a workstation from continually retrying to access a formerly spunup pack after a server is restarted.

Client version 1.3 (6/15/86)

1. Ported to IBM RT PC.

Client version 1.4 (6/30/86)

Never deployed!—changes integrated in 2.0.

Client version 2.0—Initial Fall Release

1. Fix spinup to try only 5 times, then give up.
2. Adjust burst size from 32 to 16.
3. Spinup: if user tries to spinup an already spunup pack, resend the request. (Allows recovery in many cases if server crashed. In future should check returned pack uid to verify that it didn't change.)
4. Recognize hard error return code, retry a few times, then return hard error status, rather than retrying forever. [N.B.: latest version in castor:~philipp/rvd

tries to return all readable data in this burst; that version should be shaken down and installed, so that rvdcopy can take advantage of it.]

5. Redesign retry timeout on read/write. On reads and writes, retry once quickly (e.g. 500 ms.) then if no response, repeatedly double the timeout, up to a maximum of 10 seconds. Statistics should show separately the number of short and long timeouts.

6. Fix spinup code so that it passes along (and stores) all 32 allowed bytes of passwords.

7. Find and fix [vdcopy: 0 pte] bug. Occurs under heavy load with multiple processes using the client driver; generates kernel panic & crash.

Client version 2.1—new ioctl interface

1. Redesign the client interface to use ioctl calls on /dev/rvdctl rather than additional supervisor entry points.

2. Add an ioctl that returns the version number of the structure declarations of the client call interface.

3. Change client to store pack name as part of drive state, and return it on vdstats ioctl.

4. Add an ioctl that returns the number of rvd devices implemented in the driver.

5. Driver should store and vdstats should return the (not yet returned) pack uid.

6. Change spindown call to check for busy file system and reject call, but allow user to force a spindown anyway.

7. Don't send spindown packets for drives that aren't spun up.

Client version 2.2—uid, nonces, physical ops

1. spinup: stores returned pack uids (and also the originally supplied password) as part of the client driver state. If user tries to spinup an already spunup pack, sends a respinup request based on uid.

2. read/write: if "no such connection" error comes back from server, sends a respinup packet and prepares to retry the read/write if the respinup is successful.

3. Hard spinup option; if the spinup succeeds, then on reads and writes, the client retries forever if the host stops responding. On soft spinup, retries give up after 2 minutes.

4. Total timeout time on spinup/down is shortened to five retries of one second each.

Client version 3.0—Kerberos integration

1. Sends authenticated-spinup packet type containing a Kerberos ticket.

Client version 3.1—scheduled wrapups

1. Review all panics; eliminate where possible.

Client version 3.2 (outplan)

1. Create pseudo server that returns all possible error conditions, to test client driver responses. Develop standard test suite for client driver.
2. Logging via syslog to a central log service.

III. Client commands

Commands Version 1.0

Client commands obtained from L.C.S. in June, 1985.

Commands version 1.1 (6/1/86)

1. New commands rvdsetm (invokes set_message) rvdgetm (invokes get_message) and rvdallow (invokes allow_spinups). rvdshow has an option to get list of spunup packs and can cope with more than 12 spinups.
2. Up command reprogrammed in C, changed to invoke rvdgetm once for each server used and rfsck for each writeable file system mounted. Also allows multiple entries on different servers for a single drive/directory combination in /etc/rvdtab, and cycles through them till one responds.
3. Vddb now has a list operation and an exchange operation.
4. Vdstats output is a little more readable.
5. Most client commands have a -d option for debugging (displays the request and response packets.)
6. Vddb can manage a server running on a different host. (Permits central management of several servers.)
7. New rvdcopy command for fast transfer of pack contents.

Commands version 2.0—initial Fall release

1. Redesign Up/Down commands for better user interface.
2. Magic number checking subroutine that can detect attempts to mount IBM file systems on DEC processors and vice-versa.
3. Up command: change to use magic number checker, and give user-friendly comment about what is probably wrong if the bytes are out of order.
4. Up command: change to allow one-time spinups if running on a workstation or running as root. (requires new system type variable.)
5. Up command: change to look for .rvdtab in user's directory if running on a workstation.
6. Up command: when interrupted, report what state it is leaving the client in.
7. Up command: shouldn't suggest server is down if user gives wrong pack

password.

8. Down command: occasionally spins down wrong disk!

9. Make sure up and spinup allow 32-byte passwords.

10. Add option -f to rvdsetm that suppresses prompting for a password, so that it can be used in /etc/rc before require_authorization has been issued.

Commands version 2.1—new ioctl interface

1. Redesign to use driver ioctl interface; check structure version number.

2. Merge the commands into two programs that look to see what name they were invoked under, to avoid needing multiple binary copies of all the libraries.

3. Review error response and user interface design of all client commands; take advantage of pack name stored by client driver

4. Spinup no longer tries to change owner of virtual device.

Commands version 2.2—uid, nonces, and physical ops

1. Rvdexch command uses pack uid's.

2. Vddb: allows pack rename based on uid.

3. New spinup options: allows choice of hard/soft spinups, and respinup.

4. newvd: sets mode of newly created root to 555, not 500.

Commands version 3.0—Kerberos integration

1. Integrate with Kerberos.

2. vddb checks for write access to named rvddb before starting work. Doesn't require authorization password when talking to a remote system.

Commands version 3.1—scheduled wrapups

1. vddb exchange request uses nonces.

2. New control command allows use/disuse of disk partitions.

3. New control command to allow_spinups.

4. Can now install a null password with vddb.

5. vddb: add delete_physical feature.

6. up/down commands replaced with bind-based attach/detach commands. rvdflush flushes servers mentioned in attach table.

Commands version 3.2 (outplan)

1. rvdcopy: add option to suppress progress reports.

2. vddb: should do rcp of rvddb if working for a remote site.

3. vddb: figure out way to verify operations password of remote system at outset rather than at first modification.

4. Fix /etc/shutdown to shutdown rvd server, too.

5. Design method for user to change passwords and exchange packs, giving only user passwords and changing the data base.
6. Add options on savervd/restorervd.
7. newvd: verify that file system is not mounted.
8. Add nonce support to rest of RVD library.
9. Find and fix segmentation faults that occur in vddb.

IV. Documentation

Documentation release 2.0.

1. RVDCTL protocol document expanded and updated.
2. Complete review of all man pages.
3. Available documentation brought together into a single notebook; overview, installation notes, operations guide, and example standard server configuration added.

Documentation release 2.1.

1. RVDCTL protocol document updated.
2. RVD protocol document reviewed and updated.
3. Expanded operations guide.
4. Server and client statistics documentation added.

Documentation release 2.2.

1. Document limits: maximum password length, maximum pack name length, maximum number of connections, maximum number of packs, maximum burstsize allowed by server, etc.
2. Cookbook for replacing a disk used by an RVD server.

Documentation release 3.0.

1. Document Kerberos integration.