CVS Client/Server

1 Goals

- Do not assume any access to the repository other than via this protocol. It does not depend on NFS, rdist, etc.
- Providing a reliable transport is outside this protocol. It is expected that it runs over TCP, UUCP, etc.
- Security and authentication are handled outside this protocol (but see below about '`cvs kserver`').
- This might be a first step towards adding transactions to CVS (i.e. a set of operations is either executed atomically or none of them is executed), improving the locking, or other features. The current server implementation is a long way from being able to do any of these things. The protocol, however, is not known to contain any defects which would preclude them.
- The server never has to have any CVS locks in place while it is waiting for communication with the client. This makes things robust in the face of flaky networks.
- Data is transferred in large chunks, which is necessary for good performance. In fact, currently the client uploads all the data (without waiting for server responses), and then waits for one server response (which consists of a massive download of all the data). There may be cases in which it is better to have a richer interraction, but the need for the server to release all locks whenever it waits for the client makes it complicated.

2 Notes on the Current Implementation

The client is built in to the normal `cvs` program, triggered by a `CVSROOT` variable containing a colon, for example `cygnus.com:/rel/cvsfiles`.

The client stores what is stored in checked-out directories (including '`CVS`'). The way these are stored is totally compatible with standard CVS. The server requires no storage other than the repository, which also is totally compatible with standard CVS.

The server is started by `cvs server`. There is no particularly compelling reason for this rather than making it a separate program which shares a lot of sources with cvs.

The server can also be started by `cvs kserver`, in which case it does an initial Kerberos authentication on stdin. If the authentication succeeds, it subsequently runs identically to `cvs server`.

The current server implementation can use up huge amounts of memory when transmitting a lot of data. Avoiding this would be a bit tricky because it is not acceptable to have the server block on the network (which may be very slow) when it has locks open. The buffer code has been rewritten so that this does not appear to be a serious problem in practice. However, if it is seen to be a problem several solutions are possible. The two-pass design would involve first noting what versions of everything we need (with locks in place) and then sending the data, blocking on the network, with no locks needed. The lather-rinse-repeat design would involve doing things as it does now until a certain amount of server memory is being used (10M?), then releasing locks, and trying the whole update again (some of it is presumably already done). One problem with this is getting merges to work right.

3 How to add more remote commands

It's the usual simple twelve step process. Let's say you're making the existing `cvs fix` command work remotely.

- Add a declaration for the `fix` function, which already implements the `cvs fix` command, to '`server.c`'.
- Now, the client side. Add a function `client_fix` to '`client.c`', which calls `parse_cvsroot` and then calls the usual `fix` function.
- Add a declaration for `client_fix` to '`client.h`'.
- Add `client_fix` to the "fix" entry in the table of commands in '`main.c`'.
- Now for the server side. Add the `serve_fix` routine to '`server.c`'; make it do:

```
static void
serve_fix (arg)
    char *arg;
{
    do_cvs_command (fix);
}
```

- Add the server command `"fix"` to the table of requests in '`server.c`'.
- The `fix` function can now be entered in three different situations: local (the old situation), client, and server. On the server side it probably will not need any changes to cope. Modify the `fix` function so that if it is run when the variable `client_active` is set, it starts the server, sends over parsed arguments and possibly files, sends a "fix" command to the server, and handles responses from the server. Sample code:

```
    if (!client_active) {
        /* Do whatever you used to do */
    } else {
        /* We're the local client.  Fire up the remote server.  */
        start_server ();

        if (local)
            if (fprintf (to_server, "Argument -l\n") == EOF)
                error (1, errno, "writing to server");
        send_option_string (options);

        send_files (argc, argv, local);

        if (fprintf (to_server, "fix\n") == EOF)
```

```
            error (1, errno, "writing to server");
        err = get_responses_and_close ();
    }
```

- Build it locally. Copy the new version into somewhere on the remote system, in your path so that `rsh host cvs` finds it. Now you can test it.
- You may want to set the environment variable `CVS_CLIENT_PORT` to -1 to prevent the client from contacting the server via a direct TCP link. That will force the client to fall back to using `rsh`, which will run your new binary.
- Set the environment variable `CVS_CLIENT_LOG` to a filename prefix such as '`/tmp/cvslog`'. Whenever you run a remote CVS command, the commands and responses sent across the client/server connection will be logged in '`/tmp/cvslog.in`' and '`/tmp/cvslog.out`'. Examine them for problems while you're testing.

This should produce a good first cut at a working remote `cvs fix` command. You may have to change exactly how arguments are passed, whether files or just their names are sent, and how some of the deeper infrastructure of your command copes with remoteness.

4 Notes on the Protocol

A number of enhancements are possible:

- The `Modified` request could be speeded up by sending diffs rather than entire files. The client would need some way to keep the version of the file which was originally checked out, which would double client disk space requirements or require coordination with editors (e.g. maybe it could use emacs numbered backups). This would also allow local operation of `cvs diff` without arguments.
- Have the client keep a copy of some part of the repository. This allows all of `cvs diff` and large parts of `cvs update` and `cvs ci` to be local. The local copy could be made consistent with the master copy at night (but if the master copy has been updated since the latest nightly re-sync, then it would read what it needs to from the master).
- Provide encryption using kerberos.
- The current procedure for `cvs update` is highly sub-optimal if there are many modified files. One possible alternative would be to have the client send a first request without the contents of every modified file, then have the server tell it what files it needs. Note the server needs to do the what-needs-to-be-updated check twice (or more, if changes in the repository mean it has to ask the client for more files), because it can't keep locks open while waiting for the network. Perhaps this whole thing is irrelevant if client-side repositories are implemented, and the rcsmerge is done by the client.

5 The CVS client/server protocol

5.1 Entries Lines

Entries lines are transmitted as:

`/` *name* `/` *version* `/` *conflict* `/` *options* `/` *tag_or_date*

tag_or_date is either '`T`' *tag* or '`D`' *date* or empty. If it is followed by a slash, anything after the slash shall be silently ignored.

version can be empty, or start with '`0`' or '`-`', for no user file, new user file, or user file to be removed, respectively.

conflict, if it starts with '`+`', indicates that the file had conflicts in it. The rest of *conflict* is '`=`' if the timestamp matches the file, or anything else if it doesn't. If *conflict* does not start with a '`+`', it is silently ignored.

5.2 Modes

A mode is any number of repetitions of

mode-type `=` *data*

separated by '`,`'.

mode-type is an identifier composed of alphanumeric characters. Currently specified: '`u`' for user, '`g`' for group, '`o`' for other, as specified in POSIX. If at all possible, give these their POSIX meaning and use other mode-types for other behaviors. For example, on VMS it shouldn't be hard to make the groups behave like POSIX, but you would need to use ACLs for some cases.

data consists of any data not containing '`,`', '`\0`' or '`\n`'. For '`u`', '`g`', and '`o`' mode types, data consists of alphanumeric characters, where '`r`' means read, '`w`' means write, '`x`' means execute, and unrecognized letters are silently ignored.

5.3 Requests

File contents (noted below as *file transmission*) can be sent in one of two forms. The simpler form is a number of bytes, followed by a newline, followed by the specified number of bytes of file contents. These are the entire contents of the specified file. Second, if both client and server support '`gzip-file-contents`', a '`z`' may precede the length, and the 'file contents' sent are actually compressed with '`gzip`'. The length specified is that of the compressed version of the file.

In neither case are the file content followed by any additional data. The transmission of a file will end with a newline iff that file (or its compressed form) ends with a newline.

`Root` *pathname* `\n`
: Response expected: no. Tell the server which `CVSROOT` to use.

`Valid-responses` *request-list* `\n`
: Response expected: no. Tell the server what responses the client will accept. request-list is a space separated list of tokens.

`valid-requests \n`
: Response expected: yes. Ask the server to send back a `Valid-requests` response.

`Repository` *repository* `\n`
: Response expected: no. Tell the server what repository to use. This should be a directory name from a previous server response. Note that this both gives a default for `Entry` and `Modified` and also for `ci` and the other commands; normal usage is to send a `Repository` for each directory in which there will be an `Entry` or `Modified` , and then a final `Repository` for the original directory, then the command.

`Directory` *local-directory* `\n`
: Additional data: *repository* `\n`. This is like `Repository`, but the local name of the directory may differ from the repository name. If the client uses this request, it affects the way the server returns pathnames; see Section 5.4 [Responses], page 11. *local-directory* is relative to the top level at which the command is occurring (i.e. the last `Directory` or `Repository` which is sent before the command).

`Max-dotdot` *level* `\n`
: Tell the server that *level* levels of directories above the directory which `Directory` requests are relative to will be needed. For example, if the client is planning to use a `Directory` request for '`../../foo`', it must send a `Max-dotdot` request with a *level* of at least 2. `Max-dotdot` must be sent before the first `Directory` request.

`Static-directory \n`
: Response expected: no. Tell the server that the directory most recently specified with `Repository` or `Directory` should not have additional files checked out unless explicitly

requested. The client sends this if the `Entries.Static` flag is set, which is controlled by the `Set-static-directory` and `Clear-static-directory` responses.

`Sticky` *tagspec* `\n`
: Response expected: no. Tell the server that the directory most recently specified with `Repository` has a sticky tag or date *tagspec*. The first character of *tagspec* is '`T`' for a tag, or '`D`' for a date. The remainder of *tagspec* contains the actual tag or date.

`Checkin-prog` *program* `\n`
: Response expected: no. Tell the server that the directory most recently specified with `Directory` has a checkin program *program*. Such a program would have been previously set with the `Set-checkin-prog` response.

`Update-prog` *program* `\n`
: Response expected: no. Tell the server that the directory most recently specified with `Directory` has an update program *program*. Such a program would have been previously set with the `Set-update-prog` response.

`Entry` *entry-line* `\n`
: Response expected: no. Tell the server what version of a file is on the local machine. The name in *entry-line* is a name relative to the directory most recently specified with `Repository`. If the user is operating on only some files in a directory, `Entry` requests for only those files need be included. If an `Entry` request is sent without `Modified`, `Unchanged`, or `Lost` for that file the meaning depends on whether `UseUnchanged` has been sent; if it has been it means the file is lost, if not it means the file is unchanged.

`Modified` *filename* `\n`
: Response expected: no. Additional data: mode, `\n`, file transmission. Send the server a copy of one locally modified file. *filename* is relative to the most recent repository sent with `Repository`. If the user is operating on only some files in a directory, only those files need to be included. This can also be sent without `Entry`, if there is no entry for the file.

`Lost` *filename* `\n`
: Response expected: no. Tell the server that *filename* no longer exists. The name is relative to the most recent repository sent with `Repository`. This is used for any case in which `Entry` is being sent but the file no longer exists. If the client has issued the `UseUnchanged` request, then this request is not used.

`Unchanged` *filename* `\n`
: Response expected: no. Tell the server that *filename* has not been modified in the checked out directory. The name is relative to the most recent repository sent with `Repository`. This request can only be issued if `UseUnchanged` has been sent.

`UseUnchanged \n`

Response expected: no. Tell the server that the client will be indicating unmodified files with `Unchanged`, and that files for which no information is sent are nonexistent on the client side, not unchanged. This is necessary for correct behavior since only the server knows what possible files may exist, and thus what files are nonexistent.

`Argument` *text* `\n`

Response expected: no. Save argument for use in a subsequent command. Arguments accumulate until an argument-using command is given, at which point they are forgotten.

`Argumentx` *text* `\n`

Response expected: no. Append `\n` followed by text to the current argument being saved.

`Global_option` *option* `\n`

Transmit one of the global options '`-q`', '`-Q`', '`-l`', '`-t`', '`-r`', or '`-n`'. *option* must be one of those strings, no variations (such as combining of options) are allowed. For graceful handling of `valid-requests`, it is probably better to make new global options separate requests, rather than trying to add them to this request.

`expand-modules \n`

Response expected: yes. Expand the modules which are specified in the arguments. Returns the data in `Module-expansion` responses. Note that the server can assume that this is checkout or export, not rtag or rdiff; the latter do not access the working directory and thus have no need to expand modules on the client side.

`co \n`
`update \n`
`ci \n`
`diff \n`
`tag \n`
`status \n`
`log \n`
`add \n`
`remove \n`
`rdiff \n`
`rtag \n`
`import \n`
`admin \n`
`export \n`

`history \n`

`release \n`
: Response expected: yes. Actually do a cvs command. This uses any previous `Argument`, `Repository`, `Entry`, `Modified`, or `Lost` requests, if they have been sent. The last `Repository` sent specifies the working directory at the time of the operation. No provision is made for any input from the user. This means that `ci` must use a `-m` argument if it wants to specify a log message.

`update-patches \n`
: This request does not actually do anything. It is used as a signal that the server is able to generate patches when given an `update` request. The client must issue the `-u` argument to `update` in order to receive patches.

`gzip-file-contents` *level* `\n`
: This request asks the server to filter files it sends to the client through the '`gzip`' program, using the specified level of compression. If this request is not made, the server must not do any compression.

  This is only a hint to the server. It may still decide (for example, in the case of very small files, or files that already appear to be compressed) not to do the compression. Compression is indicated by a '`z`' preceding the file length.

  Availability of this request in the server indicates to the client that it may compress files sent to the server, regardless of whether the client actually uses this request.

other-request text `\n`
: Response expected: yes. Any unrecognized request expects a response, and does not contain any additional data. The response will normally be something like '`error unrecognized request`', but it could be a different error if a previous command which doesn't expect a response produced an error.

When the client is done, it drops the connection.

5.4 Responses

After a command which expects a response, the server sends however many of the following responses are appropriate. Pathnames are of the actual files operated on (i.e. they do not contain '`,v`' endings), and are suitable for use in a subsequent `Repository` request. However, if the client has used the `Directory` request, then it is instead a local directory name relative to the directory in which the command was given (i.e. the last `Directory` before the command). Then a newline and a repository name (the pathname which is sent if `Directory` is not used). Then the slash and the filename. For example, for a file '`i386.mh`' which is in the local directory '`gas.clean/config`' and for which the repository is '`/rel/cvsfiles/devo/gas/config`':

```
gas.clean/config/
/rel/cvsfiles/devo/gas/config/i386.mh
```

Any response always ends with '`error`' or '`ok`'. This indicates that the response is over.

`Valid-requests` *request-list* `\n`
: Indicate what requests the server will accept. *request-list* is a space separated list of tokens. If the server supports sending patches, it will include '`update-patches`' in this list. The '`update-patches`' request does not actually do anything.

`Checked-in` *pathname* `\n`
: Additional data: New Entries line, `\n`. This means a file *pathname* has been successfully operated on (checked in, added, etc.). name in the Entries line is the same as the last component of *pathname*.

`New-entry` *pathname* `\n`
: Additional data: New Entries line, `\n`. Like `Checked-in`, but the file is not up to date.

`Updated` *pathname* `\n`
: Additional data: New Entries line, `\n`, mode, `\n`, file transmission. A new copy of the file is enclosed. This is used for a new revision of an existing file, or for a new file, or for any other case in which the local (client-side) copy of the file needs to be updated, and after being updated it will be up to date. If any directory in pathname does not exist, create it.

`Merged` *pathname* `\n`
: This is just like `Updated` and takes the same additional data, with the one difference that after the new copy of the file is enclosed, it will still not be up to date. Used for the results of a merge, with or without conflicts.

`Patched` *pathname* `\n`
: This is just like `Updated` and takes the same additional data, with the one difference that instead of sending a new copy of the file, the server sends a patch produced by '`diff -u`'. This client must apply this patch, using the '`patch`' program, to the existing file. This will only be used when the client has an exact copy of an earlier revision of a file. This response is only used if the `update` command is given the '`-u`' argument.

`Checksum` *checksum*`\n`
: The *checksum* applies to the next file sent over via `Updated`, `Merged`, or `Patched`. In the case of `Patched`, the checksum applies to the file after being patched, not to the patch itself. The client should compute the checksum itself, after receiving the file or patch, and signal an error if the checksums do not match. The checksum is the 128 bit MD5 checksum represented as 32 hex digits. This response is optional, and is only used if the client supports it (as judged by the `Valid-responses` request).

`Copy-file` *pathname* `\n`
: Additional data: *newname* `\n`. Copy file *pathname* to *newname* in the same directory where it already is. This does not affect `CVS/Entries`.

`Removed` *pathname* `\n`
: The file has been removed from the repository (this is the case where cvs prints '`file foobar.c is no longer pertinent`').

`Remove-entry` *pathname* `\n`
: The file needs its entry removed from `CVS/Entries`, but the file itself is already gone (this happens in response to a `ci` request which involves committing the removal of a file).

`Set-static-directory` *pathname* `\n`
: This instructs the client to set the `Entries.Static` flag, which it should then send back to the server in a `Static-directory` request whenever the directory is operated on. *pathname* ends in a slash; its purpose is to specify a directory, not a file within a directory.

`Clear-static-directory` *pathname* `\n`
: Like `Set-static-directory`, but clear, not set, the flag.

`Set-sticky` *pathname* `\n`
: Additional data: *tagspec* `\n`. Tell the client to set a sticky tag or date, which should be supplied with the `Sticky` request for future operations. *pathname* ends in a slash; its purpose is to specify a directory, not a file within a directory. The first character of *tagspec* is '`T`' for a tag, or '`D`' for a date. The remainder of *tagspec* contains the actual tag or date.

`Clear-sticky` *pathname* `\n`
: Clear any sticky tag or date set by `Set-sticky`.

`Set-checkin-prog` *dir* `\n`
: Additional data: *prog* `\n`. Tell the client to set a checkin program, which should be supplied with the `Checkin-prog` request for future operations.

`Set-update-prog` *dir* `\n`
: Additional data: *prog* `\n`. Tell the client to set an update program, which should be supplied with the `Update-prog` request for future operations.

`Module-expansion` *pathname* `\n`
: Return a file or directory which is included in a particular module. *pathname* is relative to cvsroot, unlike most pathnames in responses.

`M` *text* `\n`
: A one-line message for the user.

`E` *text* `\n`
: Same as `M` but send to stderr not stdout.

`error` *errno-code* ' ' *text* `\n`
: The command completed with an error. *errno-code* is a symbolic error code (e.g. `ENOENT`); if the server doesn't support this feature, or if it's not appropriate for this particular message, it just omits the errno-code (in that case there are two spaces after '`error`'). Text is an error message such as that provided by strerror(), or any other message the server wants to use.

`ok \n`
: The command completed successfully.

5.5 Example

Lines beginning with '`c>`' are sent by the client; lines beginning with '`s>`' are sent by the server; lines beginning with '`#`' are not part of the actual exchange.

```
c> Root /rel/cvsfiles
# In actual practice the lists of valid responses and requests would
# be longer
c> Valid-responses Updated Checked-in M ok error
c> valid-requests
s> Valid-requests Root co Modified Entry Repository ci Argument Argumentx
s> ok
# cvs co devo/foo
c> Argument devo/foo
c> co
s> Updated /rel/cvsfiles/devo/foo/foo.c
s> /foo.c/1.4/Mon Apr 19 15:36:47 1993 Mon Apr 19 15:36:47 1993//
s> 26
s> int mein () { abort (); }
s> Updated /rel/cvsfiles/devo/foo/Makefile
s> /Makefile/1.2/Mon Apr 19 15:36:47 1993 Mon Apr 19 15:36:47 1993//
s> 28
s> foo: foo.c
s>         $(CC) -o foo $<
s> ok
# In actual practice the next part would be a separate connection.
# Here it is shown as part of the same one.
c> Repository /rel/cvsfiles/devo/foo
# foo.c relative to devo/foo just set as Repository.
c> Entry /foo.c/1.4/Mon Apr 19 15:36:47 1993 Mon Apr 19 15:36:47 1993//
c> Entry /Makefile/1.2/Mon Apr 19 15:36:47 1993 Mon Apr 19 15:36:47 1993//
c> Modified foo.c
c> 26
c> int main () { abort (); }
# cvs ci -m <log message> foo.c
```

```
c> Argument -m
c> Argument Well, you see, it took me hours and hours to find this typo and I
c> Argumentx searched and searched and eventually had to ask John for help.
c> Argument foo.c
c> ci
s> Checked-in /rel/cvsfiles/devo/foo/foo.c
s> /foo.c/1.5/ Mon Apr 19 15:54:22 CDT 1993//
s> M Checking in foo.c;
s> M /cygint/rel/cvsfiles/devo/foo/foo.c,v  <--  foo.c
s> M new revision: 1.5; previous revision: 1.4
s> M done
s> ok
```