Aegis
A Project Change Supervisor

User Guide

Peter Miller
pmiller@opensource.org.au

.

DEDICATIONS

This user guide is dedicated to my wife
Mary Therese Miller
for all her love and support
despite the computers.

And to my grandmother
Jean Florence Pelham
1905 — 1992
Always in our hearts.

This document describes Aegis version 4.25
and was prepared 26 November 2012.

Table of Contents

1. Introduction

Aegis is a CASE tool with a difference. In the spirit of the UNIX® Operating System, Aegis is a small component designed to work with other programs.

Many CASE systems attempt to provide everything, from bubble charts to source control to compilers. Users are trapped with the components supplied by the CASE system, and if you don't like one of the components (it may be too limited, for instance), then that is just tough.

In contrast, UNIX provides many components of a CASE system − compilers, editors, dependency tools (such as make), source control (such as SCCS). You may substitute the tool of your choice − gcc, jove, cake, rcs (to name a few) if you don't like the ones supplied with the system.

Aegis adds to this list with software configuration management (SCM), and consistent with UNIX philosophy, Aegis does not dictate the choice of any of the other tools (although it may stretch them to their limits).

1.1. Year 2000 Status

Aegis does not suffer from Year 2000 problems.

• Aegis stores dates internally in Unix style (*i.e.* seconds offset), so internal storage of times and dates does not suffer from any Y2K problems.

• Aegis always uses the ANSI C standard `strftime` function to display times and dates. (This assumes that your vendor has supplied a compliant `strftime`.) This means that displaying dates does not assume fixed field widths, nor will it display the year 2000 as "100".

• There is no user-input of years at any time, so there is no issue surrounding "guessing" the century.

1.2. What does aegis do?

Just what is software configuration management? This question is sufficiently broad as to require a book in answer. In essence, the aegis program is a project change supervisor. It provides a framework within which a team of developers may work on many changes to a program independently, and the aegis program coordinates integrating these changes back into the master source of the program, with as little disruption as possible. Resolution of contention for source files, a major headache for any project with more than one developer, is one of the aegis program's major functions.

It should be noted that the aegis program is a developer's tool, in the same sense as make or RCS are developer's tools. It is not a manager's tool − it does not provide progress tracking or help with work allocation.

1.3. Why use aegis?

So why should you use the aegis program? The aegis program uses a particular model of the development of software projects. This model has a master source (or baseline) of a project, consisting of several (possibly several hundred) files, and a team of developers creating changes to be made to this baseline. When a change is complete, it is integrated with the baseline, to become the new baseline. Each change must be atomic and self-contained, no change is allowed to cause the baseline to cease to work. "Working" is defined as passing its own tests. The tests are considered part of the baseline. Aegis provides support for the developer so that an entire copy of the baseline need not be taken to change a few files, only those files which are to be changed need to be copied.

The win in using the aegis program is that there are *O(n)* interactions between developers and the baseline. Contrast this with a master source which is being edited directly by the developers − there is *O(n!)* interactions between developers − this makes adding "just one" more developer a potential disaster.

Another win is that the project baseline always works. Always having a working baseline means that a version is always available for demonstrations, or those "pre-release snapshots" we are always forced to provide.

The above advantages are all very well − for management types. Why should Joe Average Programmer use the aegis program? Recall that RCS provides file locking, but only for one file at a time. The aegis program provides the file locking, atomically, for the set of files in the change. Recall also that correct RCS usage locks the file the instant you start editing it. This makes popular files a project bottleneck. The aegis program allows concurrent editing, and a resolution mechanism just before the change must be integrated, meaning fewer delays for J.A.Programmer.

1.4. How to use this manual

This manual assumes the reader is already familiar with the UNIX operating system, and with developing software using the UNIX operating system and the tools available; terms such as *RCS* and *SCCS* and *make*(1) are not explained.

There is also the assumption that the reader is familiar with the issues surrounding team development of software; coordination and multiple version issues, for example, are not explained.

This manual is broken into a number of sections.

Chapter 2
: describes how aegis works and some of the reasoning behind the design and implementation of the aegis program. Look here for answers to "Why does it..." questions.

Chapter 3
: is a worked example of how particular users interact with the aegis program. Look here for answers to "How do I..." questions.

Chapter 4
: is a discussion of how aegis interacts with the History Tool, and provides templates and suggestions for history tools known to work with aegis.

Chapter 5
: is a discussion of how aegis interacts with the Dependency Maintenance Tool (DMT), and provides templates and suggestions for DMTs known to work with aegis.

Chapter 6
: is a discussion of how aegis interacts with the Difference Tools, and provides templates and suggestions for difference tools known to work with aegis.

Chapter 7
: describes the project attributes and how the various parameters may be used for particular projects.

Chapter 8
: describes managing tests and testing with Aegis.

Chapter 9
: describes the branching mechanism used in Aegis.

Chapter 10
: is a collection of helpful hints on how to use aegis effectively, based on real-world experience. This is of most use when initially placing projects under the supervision of the aegis program.

Chapter 11
: describes how to manage geographically distributed development using Aegis.

Appendix A
: is a quick reference for placing an existing project under aegis.

Appendix B
: is a glossary of terms.

Appendix D
: is a description of why Aegis must be set-uid-root, for system administrators who are concerned about the security issues.

Appendix I
: is a brief look at internationalization and localization if Aegis.

1.5. GNU GPL

Aegis is distributed under the terms and conditions of the GNU General Public License. Programs which are developed using Aegis are not automatically subject to the GNU GPL. Only programs which are derivative works based on GNU GPL code are automatically subject to the GNU GPL. We still encourage software authors to distribute their work under terms like those of the GNU GPL, but doing so is not required to use Aegis.

2. How Aegis Works

Before you will be able to exploit Aegis fully, you will need to know what Aegis does and why.

The Aegis program provides a change control mechanism and a repository, a subset of the functionality which CASE vendors call Software Configuration Management (SCM). In order to fit into a software engineering environment, or any place software is written, Aegis needs a clearly defined place in the scheme of things.

This chapter describes the model of the software development process embodied in the Aegis program, some of the deliberate design decisions made for Aegis, some of the things Aegis will and wont do for you, and the situations where Aegis is most and least useful.

2.1. The Model

The model of the software development process used by Aegis evolved and grew with time in a commercial software development environment, and it has continued to be used and developed.

The model described here is generic, and can be adapted in a variety of ways. These will be described at the relevant points in the text.

2.1.1. The Baseline

Most CASE systems revolve around a repository: a place where *stuff* is kept. This *stuff* is the raw material that is processed in some way to produce the final product, whatever that may be. This *stuff* is the preferred form for editing or composing or whatever.

In the Aegis program, the repository is known as the *baseline* and the units of *stuff* are UNIX files. The Aegis program makes no distinction between text and binary files, so both are supported.

The history mechanism which must be included in any repository function is not provided by the Aegis program. It is instead provided by some other per-project configurable software, such as RCS. This means that the user may select the history tool most suited to any given project. It also means that Aegis is that much smaller to test and maintain.

The structure of the baseline is dictated by the nature of each project. The Aegis program attempts to make as few arbitrary rules as possible. There is one mandatory file in the your project baseline. The mandatory file is called *aegis.conf* by default, and contains the per-project configuration information. The name of this file may be changed if you want to call it something different. It ios also common (though not mandatory, and the name may be changed) to have a directory called *test* which contains all of the test scripts. The contents and structure of the *test* directory (or whatever you call it) are controlled by a test filename pattern you supply to Aegis. Tests are treated just like any other source file, and are subject to the same process.

The baseline in Aegis has one particular attribute: it always works. It is always there to show off to visiting big-wigs, it is always there to grab a copy of and ship a "pre-release snapshot" to some overly anxious customer, it is always there to let upper management "touch and feel" the progress being made towards the next release.

You may claim that "works" is comfortably fuzzy, but it is not. The baseline contains not only the source of a project, but also the tests for a project. Tests are treated just like any other source file, and are subject to the same process. A baseline is defined to "work" if and only if it passes all of its own tests. The Aegis program has mandatory testing, to ensure that all changes to the baseline are accompanied by tests, and that those tests have been run and are known to pass. This means that no change to the baseline may result in the baseline ceasing to work[1].

The model may be summarized briefly: it consists of a *baseline* (master source), updated through the agency of an *integrator*, who is in turn fed *changes* by a team of *developers*. These terms will be explained in the following sections. See figure 1 for a picture of how files flow around the system.

The baseline is a set of files including the source files for a projects, and also all derived files (such as generated code, binary files from the compiler, etc), and all of the tests. Tests are treated just like any other source file, and are subject to the same process. All files in the baseline are consistent with each other.

Thus the baseline may be considered to be the *closure* of the source files, in mathematical terms. That is, it is the source files and all implications

[1] Well, mostly. It is possible for this restriction to be relaxed if you feel there are special circumstances for a particular change. The danger is that a change will be integrated with the baseline when that change is not actually of acceptable quality.

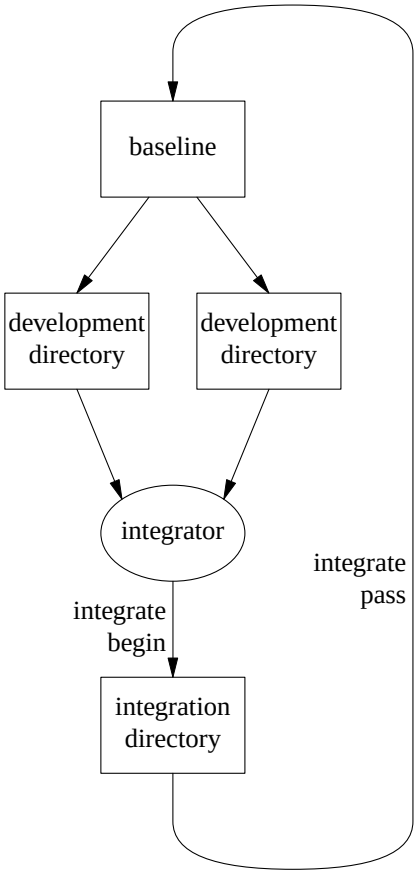


Figure 1: Flow of Files through the Model

flowing from those source files, such as object files and executables. All files in the baseline are consistent with each other; this means that development builds can take object files from the baseline rather than rebuild them within the development directory.

The baseline is readable by all staff, and usually writable by no-one. When it is necessary to write to the baseline, this is done by Aegis, as will be shown below.

In many ways, the baseline may be thought of as a database, and all derived files are projections (views) of the source files. Passing its own tests may be thought of as input validation of fields. This is a powerful concept, and indeed the implementation of the Aegis program performs many of the locking and synchronization tasks demanded of a database engine.

All of the files forming this database are text files. This means that they may be repaired with an ordinary text editor, should remedial action be necessary. The format is documented in section 5 of the reference manual. Should you wish to perform some query not yet available in Aegis, the files are readily accessible to members of the appropriate UNIX group.

Tests are treated just like any other source file, and are subject to the same process.

2.1.2. The Change Mechanism

Any changes to the baseline are made by atomic increments, known (unoriginally) as "changes". A change is a collection of files to be added to, modified in, or deleted from, the baseline. These files must all be so altered simultaneously for the baseline to continue to "work".[2]

For example, if the calling interface to a function were changed in one file, all calls to that function in any other file must also change for the baseline to continue to work. All of the files must be changed simultaneously, and thus must all be included in the one change. Other files which would logically be included in such an change include the reference manual entry for the function, the design document relating to that area of functionality, the relevant user documentation, tests would have to be included for the functionality, and existing tests may need to be revised.

Changes must be accompanied by tests. These tests will either establish that a bug has been fixed (in the case of a bug fix) or will establish that new functionality works (in the case of an enhancement).

Tests are shell scripts, and as such are capable of testing anything which has functionality accessible from the command line. The ability to run background processes allows even client-server models to be tested. Tests are thus text files, and are treated as source files; they may be modified by the same process as any other source file. Tests usually need to be revised as a project grows and adapts to changing requirements, or to be extended as functionality is extended. Tests can even be deleted if the functionality they test has been deleted; tests are deleted by the same process as any other source file.

[2] Whether to allow several logically independent changes to be included in the one change is a policy decision for individual projects to make, and is not dictated by the Aegis program. It is a responsibility of reviewers to ensure that all new and changed functionality is tested and documented.

2.1.3. Change States

As a change is developed using Aegis, it passes through six states. Many Aegis commands relate to transitions between these states, and Aegis performs any validation at these times.

The six states of a change are described as follows, although the various state transitions, and their conditions, will be described later.

2.1.3.1. Awaiting Development

A change is in this state after it has been created, but before it has been assigned to a developer. This state can't be skipped: a change can't be immediately assigned to a developer by an administrator, because this disempowers the staff.

The Aegis program is not a progress tracking tool, nor is it a work scheduling tool; plenty of both already exist.

2.1.3.2. Being Developed

A change is in this state after it has been assigned to a developer, by the developer. This is the coal face: all development is carried out in this state. Files can be edited in no other state, this particularly means that only developers can develop, reviewers and integrators only have the power to veto a change. Staff roles will be described more fully in a later section.

To advance to the next state, the change must build successfully, it must have tests, and it must pass those tests.[3]

The new tests must also *fail* against the baseline; this is to establish that tests for bug-fixes actually reproduce the bug and then demonstrate that it is gone. New functionality added by a change will naturally fail when tested in the old baseline, because it is not there.

When these conditions are met, the Aegis program marks all of the changes files as locked, simultaneously. If any one of them is already locked, you can't leave the *being developed* state, because the file is part of a change which is somewhere between *being reviewed* and *being integrated*.

If any one of them is out-of-date with respect to the baseline, the lock is not taken, either. Locking the files at this state transition means that popular files may be modified simultaneously in many changes, but that only differences to the latest version are ever submitted for integration. The Aegis program provides a mechanism, described later, for bringing out-of-date files in changes up-to-date without losing the edits made by the developer.

[3] It is possible for these testing requirements to be waived on either a per-project or per-change basis. How is described in a later section. The power to waive this requirement is not automatically granted to developers, as experience has shown that it is usually abused.

2.1.3.3. Awaiting Review

The default configuration for an Aegis project does not use this state, because for small-ish projects it can be tedious. For larger projects, however, it assists in coordinating reviewers when you use email notification that a review is required to several potential reviewers.

To enable this state, you need to change the *develop_end_action* field of the project attributes. See *aepa*(1) for more information, or *tkaepa*(1) for a GUI interface.

It is also possible, by a different setting of the same project attribute, to skip the code review step altogether. This can be of benefit to one-person projects where independent code review would be impossible.

The rest of this description will assume the *awaiting review* state is not being used, but code reviews *are* being used, to simplify matters. Once you are more familiar with Aegis, enabling the use of the *awaiting review* state will be simple and will behave intuitively.

2.1.3.4. Being Reviewed

A change is in this state after a developer has indicated that development is complete. The change is inspected, usually by a second party (or parties), to ensure that it matches the change description as to what it is meant to be doing, and meets other project or company standards you may have.

The style of review, and who may review, is not dictated by the Aegis program. A number of alternative have been observed:

• You may have a single person who coordinates review panels of, say, 4 peers, with this coordinator the only person allowed to sign-off review passes or fails.

• You may allow any of the developers to review any other developer's changes.

• You may require that only senior staff, familiar with large portions of the code, be allowed to review.

The Aegis program enforces that a developer may not review their own code. This ensures that at least one person other than the developer has scrutinized the code, and eliminates a rather obvious conflict of interest. It is possible to turn this requirement off on a per-project basis, but this is only desirable for projects with a one person team (or maybe two). The Aegis program has no way of knowing that the user passing a review has actually looked at, and understood, the code.

The reviewer knows certain things about a change for it to reach this state: it has passed all of the conditions required to reach this state. The change compiles, it has tests and it passes those tests, and the changes are to the current version of the baseline. The reviewer may thus concentrate on issues of completeness, implementation, and standards − to name only a few.

2.1.3.4.1. Customizing Code Review Policy

It is possible to require more than one reviewer for a change. By setting the *review_policy_command* of the project configuration file, you can pass a shell script (or other command) the relevant change details, and the exit status will be used to determine of the change advances to the *awaiting integration* state, or requires additional code reviewers first.

Because it is a program, it is possible to implement almost any policy you can think of, including particular reviewers for particular areas of code, or that there must be 3 different reviewers, *etc*.

2.1.3.5. Awaiting Integration

A change is in this state after a reviewer has indicated that a change is acceptable to the reviewer(s). This is essentially a queue, as there may be many developers, but only one integration may proceed at any one time.

The issue of one integration at a time is a philosophical one: all of the changes in the queue are physically independent; because of the *Develop End* locking rules they do not have intersecting sets of files. The problem comes when one change would break another, in these cases the integrator needs to know which to bounce and which to accept. Integrating one change at a time greatly simplifies this, and enforces the "only change one thing at a time" maxim, occasionally at the expense of integrator throughput.

2.1.3.6. Being Integrated

A change is in this state when the integration of the change back into the baseline is commenced. A (logical) copy of the baseline is taken, and the change is applied to that copy. In this state, the change is compiled and tested once again.

The additional compilation has two purposes: it ensures that the successful compile performed by the developer was not a fluke of the developer's environment, and it also allows the baseline to be the closure of the sources files. That is, all of the implications flowing from the source files, such as object files and linked programs or libraries. It is not possible for Aegis to know which files these are in the development directory, because Aegis is decoupled from the build mechanism (this will discussed later).

To advance to the next state, the integration copy must have been compiled, and the tests included in the change must have been run and passed.

The integrator also has the power of veto. A change may fail an integration because it fails to build or fails tests, and also just because the integrator says so. This allows the *being integrated* state to be another review state, if desired. The *being integrated* state is also the place to monitor the quality of reviews and reviewers.

Should a faulty change manage to reach this point, it is to be hoped that the integration process, and the integrator's sharp eyes, will detect it.

While most of this task is automated, this step is necessary to ensure that some strange quirk of the developer's environment was not responsible for the change reaching this stage. The change is built once more, and tested once more. If a change fails to build or test, it is returned to the developer for further work; the integrator may also choose to fail it for other reasons. If the integrator passes that change, the integrated version becomes the new baseline.

2.1.3.7. Completed

A change reaches this state when integration is complete. The (logical) copy of the baseline used during integration has replaced the previous copy of the baseline, and the file histories have been updated. Once in this state, a change may never leave it, unlike all other states.

If you wish to remove a change which is in this state from the baseline, you will have to submit another change.

2.1.4. The Software Engineers

The model of software development used by Aegis has four different roles for software engineers to fill. These four roles may be overlapping sets of people, or be distinct, as appropriate for your project.

2.1.4.1. Developer

This is the coal-face. This role is where almost everything is done. This is the only role allowed to edit a source file of a project.

Most staff will be developers. There is nothing stopping a developer from also being an administrator, except for the possible conflict of interests with respect to testing exemptions.

A developer may edit many of the attributes of a change while it is being developed. This is mostly useful to update the description of the change to say why it was done and what was actually done. A developer may not grant testing exemptions (but they may be relinquished).

2.1.4.2. Reviewer

The role of the reviewer is to check a developer's work. This review may consist of a peer examining the code, or it may be handled by a single member of staff setting up and scheduling multi-person review panels. The Aegis program does not mandate what style of review, it only requires that a reviewer pass or fail each change. If it passes, an integrator will handle it next, otherwise it is returned to the developer for further work.

In a large team, the reviewers are usually selected from the more senior members of the team, because of their depth of experience at spotting problems, but also because this is an opportunity for more senior members of staff to coach juniors on the finer points of the art.

The Aegis programs makes some of the reviewer's task easier, because the reviewer knows several specific things about a change before it comes up for review: it builds, it has tests, and they have run successfully. There is also optional (per project) additional conditions imposed at the end of development, such as line length limits, or anything else which is automatically testable. The Aegis program also provides a difference listing to the reviewer, so that each and every edit, to each and every file, can be pointed out to the reviewer.

There is nothing stopping a reviewer from being either an administrator or a developer. The Aegis program specifically prevents a developer from reviewing his own work, to avoid conflicts of interest. (It is possible for this restriction to be waived, but that only makes sense for one person projects.)

It will occasionally be necessary to arbitrate between a developer and a reviewer. The appropriate person to do this would have line responsibility above both staff involved. Thus it is desirable that supervisors/managers not be reviewers, except in very small teams.

2.1.4.3. Integrator

The role of the integrator is to take a change which has already been reviewed and integrate it with the baseline, to form a new baseline. The integrator is thus the last line of defense for the baseline.

There is nothing preventing an integrator from being an administrator, a developer or a reviewer. The Aegis program specifically prevents a developer or reviewer from integrating his own work, eliminating any conflict of interests. (It is possible for this restriction to be waived, but that only makes sense for one and two person projects.)

It will occasionally be necessary to arbitrate between an integrator and a reviewer and/or a developer. The appropriate person to do this would have line responsibility above all of the staff involved. Thus it is desirable that supervisors/mangers not be integrators, except in very small teams.

The baseline is readable by all developers, but not writable. All updates of the baseline to reflect changes produced by developers are performed through the agency of the integrator.

2.1.4.4. Administrator

The project administrator has the following duties:

• Create new changes. These may be the result of some customer bug reporting mechanism, it may be the result of new functionality being requested.

• Grant testing exemptions. By default, Aegis insists that all changes be accompanied by tests. The project administrator may grant case-by-case exemptions, or a project-wide exemption.

• Add or remove staff. The four roles described in this section may be assigned to, or removed from, specific UNIX logins by the project administrator.

• Edit project attributes. There are many attributes attached to a project, only a project administrator may alter them.

• Edit change attributes. There are many attributes attached to a change, only a project administrator may alter all of them.

A project usually has only one or two administrators at any one time.

2.1.5. The Change Process

This section will examine the progression of a change through the six change states. Most of the attention will be given to the conditions which must be met in order to progress from one state to the next, as this is where the software development model employed by Aegis is most often expressed.

See figure 2 for a picture of how all of the states and transitions fit together.

2.1.5.1. New Change

A project administrator creates a change. This change will consist mostly of a description at this time. The project administrator is not able (through Aegis) to assign it to a specific developer.

The change is awaiting development; it is in the awaiting development state.

2.1.5.2. New Change Undo

It is possible to abandon a change if it is in the *awaiting development* state. All record of the change, including its description, will be deleted.

It is called *new change undo* to emphasize the state it must be in to delete it.

2.1.5.3. Develop Begin

A developer, for whatever reason, scans the list of changes awaiting development. Having selected a change, the developer then assigns that change to herself.

The change is now being developed; it is in the being developed state.

A number of Aegis commands only work in this state, including commands to include files and tests in the change (be they new files to be added to the baseline, files in the baseline to be modified, or files to be deleted from the baseline), commands to build the change, commands to test the change, and commands to difference the change.

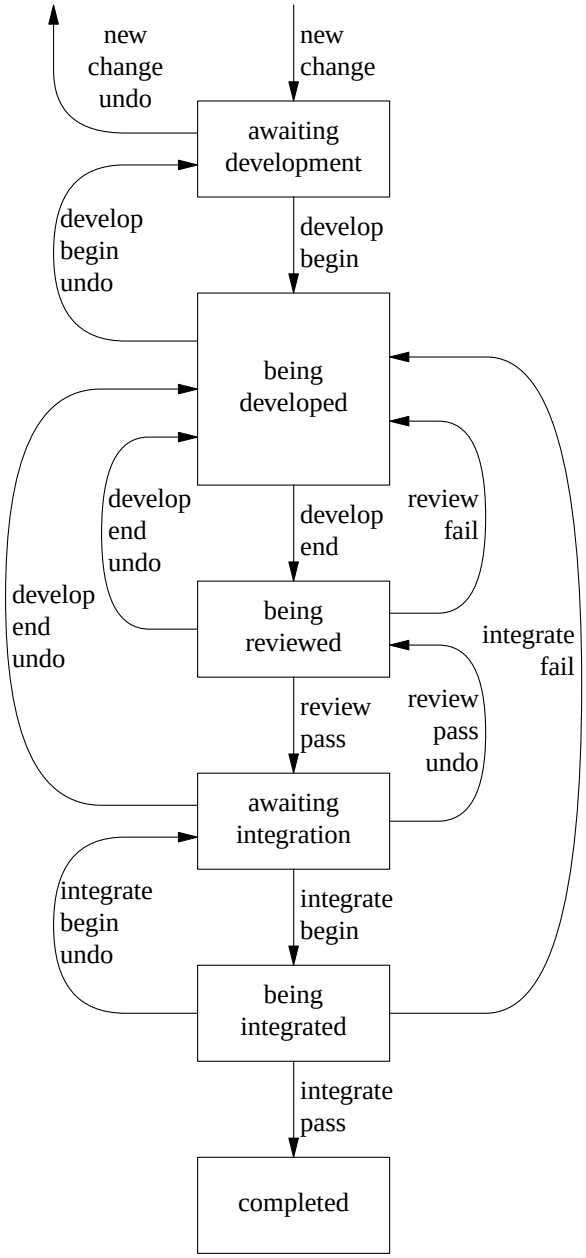


Figure 2: Change States and Transitions

The process of taking sources files, the preferred form for editing of a project, and transforming them, through various manipulations and translations, into a "finished" product is known as building. In the UNIX world this usually means things like compiling and linking a program, but as fancy graphical programs become more widespread, the source files could be the binary output from a graphical Entity-Relationship-Diagram editor, which would then be run through a

database schema generator.

The process of testing a change has three aspects. The most intuitive is that a test must be run to determine of the functionality works. The second requirement is that the test be run against the baseline and fail; this is to ensure that bugs are not just fixed, but reproduced as well. The third requirement is optional: all or some of the tests already in the baseline may also be run. Tests consist of UNIX shell scripts − anything that can be done in a shell script can be tested.

In preparation for review, a change is differenced. This usually consists of automatically comparing the present contents of the baseline with what the change proposes to do to the baseline, on a file-by-file basis. The results of the difference, such as UNIX *diff −c* output, is kept in a difference file, for examination by the reviewer(s). The benefit of this procedure is that reviewers may examine these files to see every change the developer made, rather than only the obvious ones. The differencing commands are per-project configurable, and other validations, such as line length restrictions, may also be imposed at this time.

To leave this state, the change must have source files, it must have tests, it must have built successfully, it must have passed all its own tests, and it must have been differenced.

2.1.5.4. Develop Begin Undo

It is possible to return a change from the being developed state to the awaiting development state. This is usually desired if a developer selected the wrong change by mistake. It also provides a method to start over on a change for some other reason.

2.1.5.5. Develop End

When the conditions for the end of development have been met (the change must have source files, it must have tests, it must have built successfully, it must have passed all its own tests, and it must have been differenced) the developer may cause the change to leave the being developed state and enter the being reviewed state. The Aegis program will check to see that all the conditions are met at this time. There is no history kept of unsuccessful develop end attempts.

Most of these preconditions are determined by the use of time stamps which are recorded for various operations, in addition to file system timestamps on the files themselves. Logical sequencing (*e.g.* tests being run after building after editing) is also verified.

Note that there are 3 kinds of tests

1. If a change contains a new test or a test which is being modified, this test must pass against the code compiled and linked in the change. This is simply referred to as a “test”. Changes may be granted an exemption from such tests.
2. If a change contains a new test and the change is a bug fix, this test must *fail* against the old code in the baseline. This is to confirm that the bug has been fixed. This is referred to as a “baseline test”. Changes may be granted an exemption from such tests.
3. Tests which already exist in the baseline may be run against the code compiled and linked in the change. These tests must pass. This is to confirm that the project has not regressed, which is why these tests are referred to as “regression tests”. Changes may be granted an exemption from such tests.

A successful develop end command results in the change advancing from the *being developed* state to the *being reviewed* state. (It is also possible to advance to the *awaiting review* state or the *awaiting integration* state. See *aede*(1) or *aepattr*(5) for more information.)

2.1.5.6. Develop End Undo

There are many times when a developer thinks that a change is completed, and goes hunting for a reviewer. Half way down the hall, she thinks of something that should have been included.

It is possible for a developer to rescind a *Develop End* to allow further work on a change. No reason need be given. This request may be issued to a change in either the *being reviewed* or *awaiting integration* states.

2.1.5.7. Review Pass

This event is used to notify Aegis that the change has been examined, by a method unspecified as discussed above, and has been found to be acceptable.

2.1.5.8. Review Pass Undo

The reviewer of a change may rescind a *Review Pass* while the change remains in the *awaiting integration* state. No reason need be supplied. The change will be returned to the *being reviewed* state.

2.1.5.9. Review Fail

This event is used to notify Aegis that the change has been examined, by a method unspecified as discussed above, and has been found to be unacceptable.

A file containing a brief summary of the problems must be given, and will be included in the change's history.

The change will be returned to the *being developed* state for further work.

It is not the responsibility of any reviewer to fix a defective change.

2.1.5.10. Integrate Begin

This command is used to commence integration of a change into the project baseline.

Whether a physical copy of the baseline is used, or a logical copy using links, is controlled by the project configuration file. The change is then applied to this copy.

The integrator must then issue build and test commands as appropriate. This is not automated as some integrator tasks may be required in and around these commands.

2.1.5.11. Integrate Begin Undo

This command is used to return a change to the integration queue, without prejudice. No reason need be given.

This is usually done when a particularly important change is in the queue, and the current integration is expected to take a long time.

2.1.5.12. Integrate Pass

This command is used to notify Aegis that the change being integrated is acceptable.

The current baseline is replaced with the integration copy, and the history is updated.

2.1.5.13. Integrate Fail

This command is used to notify Aegis that an integration is unacceptable, usually because it failed to build or test in some way, or sometimes because the integrator found a deficiency.

A file containing a *brief* summary of the problems must be given, and the summary will be included in the change's history.

The change will be returned to the *being developed* state for further work. The integration copy of the baseline is deleted, leaving the original baseline unchanged.

It is not the responsibility of any integrator to fix a defective change, or even diagnose what the defect may be.

2.2. Philosophy

The philosophy is simple, and that makes some of the implementation complex.

• When a change is in the *being developed* state, the aegis program is a developer's tool. Its purpose is to make it as easy for a developer to develop changes as possible.

• When a change leaves (or attempts to leave) the *being developed* state, the aegis program is protecting the project baseline, and does not exist to make the developer happy.

• The aegis program attempts to adhere to the UNIX minimalist philosophy. Least unnecessary output, least command line length, least dependence on *specific* 3rd party tools.

• No overlap in functionality of cooperating tools. (I.e. no internal build mechanism, no internal history mechanism, etc.)

2.2.1. Development

During the development of a change, the aegis program exists to help the developer. It helps him navigate around his change and the project, it copies file for him, and keeps track of the versions. It can even tell him what changes he has made.

2.2.2. Post Development

When a change has left the "being developed" state, or when it is attempting to leave that state, the aegis program ceases to attempt to help the developer and proceeds to defend the project baseline. The model used by aegis states that "the baseline always works", and aegis attempts to guarantee this.

2.2.3. Minimalism

The idea of minimalism is to help the user out. It is the intention that the aegis program can work out unstated command line options for itself, in cases where it is "safe" to do so. This means a number of defaulting mechanisms, all designed to help the user.

2.2.4. Overlap

It was very tempting while writing the aegis program to have it grow and cover source control and dependency maintenance roles. Unfortunately, this would have meant that the user would have been trapped with whatever the aegis program provided, and the aegis program is already plenty big. To add this functionality would have diverted effort, resulting in an inferior result. It would also have violated the underlying UNIX philosophy.

2.2.5. Design Goals

A number of specific ideas molded the shape of the aegis program. These include:

The UNIX philosophy of writing small tools for specific tasks with little or no overlap. Tools should be written with the expectation of use in pipes or scripts, or other combinations.

• Stay out of the way. If it is possible to let a project do whatever it likes, write the code to let it. It is not possible to anticipate even a fraction of the applications of a software tool.

• People. The staff using aegis should be in charge of the development process. They should not feel that some machine is giving them orders.

• Users aren't psychic. Feedback must be clear, accurate and appropriate.

2.3. Security

Access to project data is controlled by the UNIX group mechanism. The group may be selected as suitable for your project, as may the umask.

All work done by developers (build, difference, etc) is all with a default group of the project's group, irrespective of the user's default group. Directories (when BSD semantics are available) are all created so that their contents default to the correct group. This ensures that reviewers and integrators are able to examine the change.

Other UNIX users not in the project's group may be excluded, or not, by the appropriate setting of the project umask. This umask is used by all Aegis actions, assuring appropriate default behaviour.

A second aspect of security is to ensure that developers are unable to deliberately deceive Aegis. Should the files be tampered with at any later date, Aegis will notice.

2.4. Scalability

How big can a project get before Aegis chokes? There are a huge number of variables in this question.

The most obvious bottleneck is the integrator. An artificial "big project" example: Assume that the average integration takes an hour to build and test. A full-time integrator could be expected to get 7 or 8 of these done per day (this was the observed average on one project the author was involved in). Assume that the average time for a developer to develop a change is two weeks; a figure recommended by many text books as "*the most you can afford to throw away*". These two assumptions mean that for this "big project" Aegis can cope with 70 to 80 developers, before integrations become a bottleneck.

The more serious bottle neck is the dependency maintenance tool. Seventy developers can churn out a staggering volume of code. It takes a very long time to wade through the file times and the rules, just to find the one or two files which changed. This can easily push the integrate build time past the one hour mark. Developers also become very vocal when build times are this long.

2.5. When (not) to use Aegis

The aegis program is not a silver bullet; it will not solve all of your problems. Aegis is suitable for some kinds of projects, useful for others, and useless for a few.

The software development process embodied by Aegis has the following attributes:

- Each change set is applied atomically.
- Each change set must build successfully before it will be accepted. (This can be trivial, if desired.)
- Each change set must test successfully before it will be accepted. (This can be disabled, if desired.)
- Each change set must pass a peer review before it will be accepted. (This can be a rubber stamp, if desired.)

The most difficult thing about Aegis program is that it takes management buy-in. It takes effort to convince many people that the model used by aegis has benefits, and you need management backing you up when some person comes along with a way of developing software "without the extra work" imposed by the model used by Aegis.

2.5.1. Building

If the source code to your software product doesn’t build, it isn’t a product. However, many software shops commit changes to their repository without preconditions, and then do a daily build (or worse, weekly). The problem here is that "pollution" by defective changes is already *in your product* before it is detected. Aegis will not let it be committed in the first place.

If your product is entirely composed of scripts or HTML, you can make the build step completely trivial: "exit 0" is usually used for this purpose. Thus, this requirement, while usually highly desirable, may be avoided if desired.

2.5.2. Testing

There is extra up-front work: writing tests. The win is that the tests hang around forever, catching minor and major slips before they become embarrassing "features" in a released product. Prevention is cheaper than cure in this case, the tests save work down the track. See the *testing* chapter for more information.

2.5.3. Reviewing

Code reviews of some sort are normal in most software houses. Often, unfortunately, time pressures or other political pressures mean that code reviews manage not to happen. Yet the literature repeatedly cites reviews as one of the most important factors in removing defects before they reach your code repository. Aegis requires a code review before it will commit code into your product; again, the idea is to remove defects *before* they pollute the product.

2.6. Further Work

The Aegis program is far from finished. A number of features are known to be lacking.

At the date of this writing, Aegis is being actively supported and improved.

2.6.1. Code Coverage Tool

It would be very helpful if a code coverage tool could be used to analyze tests included with changes to ensure that the tests actually exercised the lines of code changed in the change.

Another use of the code coverage tool would be to select regression tests based on the object files recompiled by a change, and those regression tests which exercise those files.

While there is freeware C code coverage tool available, based on GNU C, the interfacing and semantics still need more thought.

Note: A fairly good approximation is already available using the −*suggest* option of the *aet*(1) command. It works on the correlation of sources file versus tests in the various change sets. See *aet*(1) for more information.

2.6.2. Virtual File System

There is almost sufficient information in the Aegis data base to create a virtual file system, overlaying the development directory atop the baseline[4]. This could be implemented similarly to automounters, intercepting file system operations by pretending to be an NFS server. Many commercial CASE products provide such a facility.

Such a virtual file system has a number of advantages: you don’t need such a capable DMT, for starters; it only needs the dynamic include dependencies, and does not need a search path[5]. Second, many horrible and dumb compilers, notably FORTRAN and "fourth" GLs, don’t have adequate include semantics; overlaying the two directories make this much easier to deal with[6]. Many graphical tools, such as bubble chart drawers, etc, when they do actually have include files, have no command line specifiable search path.

The disadvantage is that this adds significant complexity to an already large program. Also, implementation is limited to NFS capable systems, or would have to be rewritten for a variety of other systems. The semantics of interactions between the daemon and other Aegis commands, while clearly specifiable, are challenging to implement. Performance could also be a significant factor.

The question is "is it really necessary?" If the job can be done without it, does the effort of writing such a beast result in significant productivity gains?

The addition of the *create_symlinks_before_build* field to the project configuration file has greatly reduced the need for this functionality. However, it does not provide copy-on-write semantics, nor automatic *aecp* functionality; which a virtual file system could do.

[4] Reminiscent of Sun’s TFS, but not the same. Similar to AT&T’s 3D-FS. Similar to TeamNet. Similar to ClearCase, but I wasn’t thinking of the time-travel aspects which they implement.

[5] Discussed in the *Dependency Maintenance Tool* chapter.

[6] There are other ways, discussed in the *Tips and Traps* chapter.

3. The Change Development Cycle

As a change to a project is developed using Aegis, it passes through several states. Each state is characterized by different quality requirements, different sets of applicable Aegis commands, and different responsibilities for the people involved.

These people may be divided into four categories: developers, reviewers, integrators and administrators. Each has different responsibilities, duties and permissions; although one person may belong to more than one category, depending on how a project is administered.

This chapter looks at each of these categories, by way of an example project undergoing its first four changes. This example will be examined from the perspective of each category of people in the following sections.

There are six hypothetical users in the example: the developers are Pat, Jan and Sam; the reviewers are Robyn and Jan; the integrator is Isa; and the administrator is Alex[7]. There need not have been this many people involved, but it keeps things slightly cleaner for this example.

The project is called "example". It implements a very simple calculator. Many features important to a quality product are missing, checking for divide-by-zero for example. These have been omitted for brevity.

[7] The names are deliberately gender-neutral. Finding such a name starting with "I" is not easy!

3.1. The Developer

The developer role is the coal face[8]. This is where new software is written, and bugs are fixed. This example shows only the addition of new functionality, but usually a change will include modifications of existing code, similar to bug-fixing activity.

3.1.1. Before You Start

Have you configured your account to use Aegis? See the *User Setup* section of the *Tips and Traps* chapter for how to do this.

3.1.2. The First Change

While the units of change, unoriginally, are called "changes", this also applies to the start of a project − a change to nothing, if you like. The developer of this first change will be Pat.

First, Pat has been told by the project administrator that the change has been created. How Alex created this change will be detailed in the "Administrator" section, later in this chapter. Pat then acquires the change and starts work.

```
pat% aedb -l -p example.1.0
Project "example.1.0"
List of Changes

Change  State           Description
------- -------         -------------
  10    awaiting_       Create initial skeleton.
        development
pat% aedb example.1.0 10
aegis: project "example.1.0": change 10: development directory "/u/pat/
       example.1.0.C010"
aegis: project "example.1.0": change 10: user "pat" has begun development
pat% aecd
aegis: project "example.1.0": change 10: /u/pat/example.1.0.C010
pat%
```

At this point Aegis has created a development directory for the change and Pat has changed directory to the development directory[9].

Five files will be created by this change.

```
pat% aenf aegis.conf Howto.cook gram.y lex.l main.c
aegis: project "example.1.0": change 10: file "Howto.cook" added
aegis: project "example.1.0": change 10: file "aegis.conf" added
aegis: project "example.1.0": change 10: file "gram.y" added
aegis: project "example.1.0": change 10: file "lex.l" added
aegis: project "example.1.0": change 10: file "main.c" added
pat%
```

The contents of the *aegis.conf* file will not be described in this section, mostly because it is a rather complex subject; so complex it requires four chapters to describe: the *History Tool* chapter, the *Dependency Maintenance Tool* chapter, the *Difference Tools* chapter and the *Project Attributes* chapter. The contents of the *Howto.cook* file will not be described in this section, as it is covered in the *Dependency Maintenance Tool* chapter.

[8] I thought this expression was fairly common English usage, until I had a query. "The Coal Face" is an expression meaning "where the *real* work is done" in reference to old-style coal mining which was hard, tiring, hot, very dangerous, and bad for your health even if you were lucky enough not to be killed. It was a 14-hour per day job, and you walked to and from work in the dark, even in summer. Unlike the mine owners, who rode expensive horses and saw sunlight most days of the week.

[9] The default directory in which to place new development directories is configurable for each user.

The file *main.c* will have been created by Aegis as an empty file. Pat edits it to look like this

```
#include <stdio.h>

static void
usage()
{
        fprintf(stderr, "usage: example\n");
        exit(1);
}

void
main(argc, argv)
        int     argc;
        char    **argv;
{
        if (argc != 1)
                usage();
        yyparse();
        exit(0);
}
```

The file *gram.y* describes the grammar accepted by the calculator. This file was also created empty by Aegis, and Pat edits it to look like this:

```
%token  DOUBLE
%token  NAME

%union
{
        int     lv_int;
        double  lv_double;
}

%type <lv_double> DOUBLE expr
%type <lv_int> NAME

%left '+' '-'
%left '*' '/'
%right UNARY

%%
```

```
example
        : /* empty */
        | example command '\n'
                { yyerrflag = 0; fflush(stderr); fflush(stdout); }
command
        : expr
                { printf("%g\n", $1); }
        | error
expr
        : DOUBLE
                { $$ = $1; }
        | '(' expr ')'
                { $$ = $2; }
        | '-' expr
                %prec UNARY
                { $$ = -$2; }
        | expr '*' expr
                { $$ = $1 * $3; }
        | expr '/' expr
                { $$ = $1 / $3; }
        | expr '+' expr
                { $$ = $1 + $3; }
        | expr '-' expr
                { $$ = $1 - $3; }
```

The file *lex.l* describes a simple lexical analyzer. It will be processed by *lex*(1) to produce C code implementing the lexical analyzer. This kind of simple lexer is usually hand crafted, but using lex allows the example to be far smaller. Pat edits the file to look like this:

```
%{
#include <math.h>
#include <libaegis/gram.h>
%}
%%
[ \t]+      ;
[0-9]+(\.[0-9]*)?([eE][+-]?[0-9]+)?    {
                yylval.lv_double = atof(yytext);
                return DOUBLE;
        }
[a-z]   {
                yylval.lv_int = yytext[0] - 'a';
                return NAME;
        }
\n      |
.       return yytext[0];
```

Note how the *gram.h* file is included using the `#include <filename>` form. This is very important for builds in later changes, and is discussed more fully in the *Using Cook* section of the *Dependency Maintenance Tool* chapter.

The files are processed, compiled and linked together using the *aeb* command; this is known as *building* a change. This is done through Aegis so that Aegis can know the success or failure of the build. (Build success is a precondition for a change to leave the *being developed* state.) The build command is in the *aegis.conf* file so vaguely described earlier. In this example it will use the *cook*(1) command which in turn will use the *Howto.cook* file, also alluded to earlier. This file describes the commands and dependencies for the various processing, compiling and linking.

```
pat% aeb
aegis: project "example.1.0": change 10: development build started
aegis: cook −b Howto.cook project=example.1.0 change=10
        version=1.0.C010 −nl
cook: yacc −d gram.y
cook: mv y.tab.c gram.c
cook: mv y.tab.h gram.h
cook: cc −I. −I/projects/example/branch.1./branch0/baseline −O −c gram.c
cook: lex lex.l
cook: mv lex.yy.c lex.c
cook: cc −I. −I/projects/example/branch.1/branch.0/baseline −O −c lex.c
cook: cc −I. −I/projects/example/baseline −O −c main.c
cook: cc −o example gram.o lex.o main.o −ll −ly
aegis: project "example.1.0": change 10: development build complete
pat%
```

The example program is built, and Pat could even try it out:

```
pat% example
1 + 2
3
^D
pat%
```

At this point the change is apparently finished. The command to tell Aegis this is the *develop end* command:

```
pat% aede
aegis: project "example.1.0": change 10: no current 'aegis −DIFFerence'
        registration
pat%
```

It didn't work, because Aegis thinks you have missed the difference step.

The difference step is used to produce files useful for reviewing changes, mostly in the form of context difference files between the project baseline and the development directory. Context differences allow reviewers to see exactly what has changed, and not have to try to track them down and inevitably miss obscure but important edits to large or complex files.

```
pat% aed
aegis: set +e; diff −c /dev/null /u/pat/example.1.0.C010/Howto.cook >
        /u/pat/example.1.0.C010/Howto.cook,D; test $? −eq 0 −o $? −eq 1
aegis: set +e; diff −c /dev/null /u/pat/example.1.0.C010/aegis.conf >
        /u/pat/example.1.0.C010/aegis.conf,D; test $? −eq 0 −o $? −eq 1
aegis: set +e; diff −c /dev/null /u/pat/example.1.0.C010/gram.y >
        /u/pat/example.1.0.C010/gram.y,D; test $? −eq 0 −o $? −eq 1
aegis: set +e; diff −c /dev/null /u/pat/example.1.0.C010/lex.l >
        /u/pat/example.1.0.C010/lex.l,D; test $? −eq 0 −o $? −eq 1
aegis: set +e; diff −c /dev/null /u/pat/example.1.0.C010/main.c >
        /u/pat/example.1.0.C010/main.c,D; test $? −eq 0 −o $? −eq 1
aegis: project "example.1.0": change 10: difference complete
pat%
```

Doing a difference for a new file may appear a little pedantic, but when a change consists of tens of files, so modifications of existing files and some new, there is a temptation for reviewers to use "more *,D" and thus completely miss the new files if it were not for this pedanticism[10].

So that reviewers, and conscientious developers, may locate and view all of these difference files, the command

[10] This is especially true when you use a tool such as *fcomp*(1) which gives a complete file listing with the inserts and deletes marked in the margin. This tool is also available from the author of Aegis.

```
pat% more ‘find . −name "*,D" −print | sort‘
...examines each file...
pat%
```

could be used, however this is a little too long winded for most users, and so the *aedmore* alias does exactly this. There is a similar *aedless* alias for those who prefer the *less*(1) command.

So now Pat is done, let’s try to sign off again:

```
pat% aede
aegis: project "example.1.0": change 10: no current ’aegis −Test’
        registration
pat%
```

It didn’t work, again. This time Aegis is reminding Pat that every change must be accompanied by at least one test. This is so that the project team can be confident at all times that a project works[11]. Making this a precondition to leave the *being developed* state means that a reviewer can be sure that a change builds and passes its tests before it can ever be reviewed. Pat adds the truant test:

```
pat% aent
aegis: project "example.1.0": change 10: file "test/00/t0001a.sh" new
        test
pat%
```

The test file is in a weird place, eh? This is because many flavors of UNIX are slow at searching directories, and so Aegis limits itself to 100 tests per directory. Whatever the name, Pat edits the test file to look like this:

```
#!/bin/sh
#
# test simple arithmetic
#
tmp=/tmp/$$
here=‘pwd‘
if [ $? −ne 0 ]; then exit 1; fi

fail()
{
        echo FAILED 1>&2
        cd $here
        rm −rf $tmp
        exit 1
}

pass()
{
        cd $here
        rm −rf $tmp
        exit 0
}
trap "fail" 1 2 3 15

mkdir $tmp
if [ $? −ne 0 ]; then exit 1; fi
cd $tmp
if [ $? −ne 0 ]; then fail; fi
```

[11] As discussed in the *How Aegis Works* chapter, aegis has the objective of ensuring that projects always work, where "works" is defined as passing all tests in the project’s baseline. A change "works" if it passes all of its accompanying tests.

```
#
# with input like this
#
cat > test.in << 'foobar'
1
(24 - 22)
-(4 - 7)
2 * 2
10 / 2
4 + 2
10 - 3
foobar
if [ $? -ne 0 ]; then fail; fi

#
# the output should look like this
#
cat > test.ok << 'foobar'
1
2
3
4
5
6
7
foobar
if [ $? -ne 0 ]; then fail; fi

#
# run the calculator
# and see if the results match
#
$here/example < test.in > test.out
if [ $? -ne 0 ]; then fail; fi
diff test.ok test.out
if [ $? -ne 0 ]; then fail; fi

#
# this much worked
#
pass
```

There are several things to notice about this test file:

- It is a Bourne shell script. All test files are Bourne shell scripts because they are the most portable.[12] (Actually, Aegis likes test files not to be executable, it passes them to the Bourne shell explicitly when running them.)
- It makes the assumption that the current directory is either the development directory or the baseline. This is valid, aegis always runs tests this way; if you run one manually, you must take care of this yourself.
- It checks the exit status of each and every command. It is essential that even unexpected and impossible failures are handled.
- A temporary directory is created for temporary files. It cannot be assumed that a test will be run from a directory which is writable; it is also easier to clean up after strange errors, since you need only throw the directory away, rather than track down individual temporary files. It mostly protects against rogue programs scrambling files in the current directory, too.

[12] Portable for Aegis' point of view: Bourne shell is the most widely available shell. Of course, if you are writing code to publish on USENET or for FTP, portability of the tests will be important from the developer's point of view also.

- Every test is self-contained. The test uses auxiliary files, but they are not separate source files (figuring where they are when some are in a change and some are in the baseline can be a nightmare). If a test wants an auxiliary file, it must construct the file itself, in a temporary directory.
- Two functions have been defined, one for success and one for failure. Both forms remove the temporary directory. A test is defined as passing if it returns a 0 exit status, and failing if it returns anything else.
- Tests are treated just like any other source file, and are subject to the same process. They may be altered in another change, or even deleted later if they are no longer useful.

The most important feature to note about this test, after ignoring all of the trappings, is that it doesn't do much you wouldn't do manually! To test this program manually you would fire it up, just as the test does, you would give it some input, just as the test does, and you would compare the output against your expectations of what it will do, just as the test does.

The difference with using this test script and doing it manually is that most development contains many iterations of the "build, test, *think*, edit, build, test..." cycle. After a couple of iterations, the manual testing, the constant re-typing, becomes obviously unergonomic. Using a shell script is more efficient, doesn't forget to test things later, and is preserved for posterity (i.e. adds to the regression test suite).

This efficiency is especially evident when using commands[13] such as

```
pat% aeb && aet ; vi aegis.log
...
pat% !aeb
...
pat%
```

It is possible to talk to the shell extremely rarely, and then only to re-issue the same command, using a work pattern such as this.

As you have already guessed, Pat now runs the test like this:

```
pat% aet
aegis: sh /u/pat/example.1.0.C010/test/00/t0001a.sh
aegis: project "example.1.0": change 10: test "test/00/t0001a.sh"
        passed
aegis: project "example.1.0": change 10: passed 1 test
pat%
```

Finally, Pat has built the change, prepared it for review and tested it. It is now ready for sign off.

```
pat% aede
aegis: project "example.1.0": change 10: no current 'aegis -Build'
        registration
pat%
```

Say what? The problem is that the use of *aent* canceled the previous build registration. This was because Aegis is decoupled from the dependency maintenance tool (*cook* in this case), and thus has no way of knowing whether or not the new file in the change would affect the success or failure of a build[14]. All that is required is to re-build, re-test, re-difference (yes, the test gets differenced, too) and sign off.

[13] This is a *csh* specific example, unlike most others.

[14] Example: in addition to the executable file "example" shown here, the build may also produce an archive file of the project's source for export. The addition of one more file may push the size of this archive beyond a size limit; the build would thus fail because of the addition of a test.

```
pat% aeb
aegis: logging to "/u/pat/example.1.0.C010/aegis.log"
aegis: project "example.1.0": change 10: development build started
aegis: cook -b Howto.cook project=example.1.0 change=10
        version=1.0.C001 -nl
cook: "all" is up-to-date
aegis: project "example.1.0": change 10: development build complete
pat% aet
aegis: logging to "/u/pat/example.1.0.C010/aegis.log"
aegis: sh /u/pat/example..1.0.C010/test/00/t0001a.sh
aegis: project "example.1.0": change 10: test "test/00/t0001a.sh"
        passed
aegis: project "example.1.0": change 10: passed 1 test
pat% aed
aegis: logging to "/u/pat/example.1.0.C010/aegis.log"
aegis: set +e; diff -c /dev/null /u/pat/example.1.0.C010/test/00/
        t0001a.sh > /u/pat/example.1.0.C010/test/00/t0001a.sh,D; test
        $? -eq 0 -o $? -eq 1
aegis: project "example.1.0": change 10: difference complete
pat% aede
aegis: sh /usr/local/lib/aegis/de.sh example.1.0 10 pat
aegis: project "example.1.0": change 10: development completed
pat%
```

The change is now ready to be reviewed. This section is about developers, so we will have to leave this change at this point in its history. Some time in the next day or so Pat receives electronic mail that this change has passed review, and another later to say that it passed integration. Pat is now free to develop another change, possibly for a different project.

3.1.3. The Second Change

The second change was created because someone wanted to name input and output files on the command line, and called the absence of this feature a bug. When Jan arrived for work, and lists the changes awaiting development, the following list appeared:

```
jan% aedb -l -p example.1.0
Project "example.1.0"
List of Changes

Change  State           Description
------  ------          ------------
  11    awaiting_       Add input and output file names to the
        development     command line.
  12    awaiting_       add variables
        development
  13    awaiting_       add powers
        development
jan%
```

The first on the list is chosen.

```
jan% aedb -c 11 -p example.1.0
aegis: project "example.1.0": change 11: development directory "/u/
        jan/example.1.0.C011"
aegis: project "example.1.0": change 11: user "jan" has begun
        development
jan% aecd
aegis: project "example.1.0": change 11: /u/jan/example.002
jan%
```

The best way to get details about a change is to used the "change details" listing.

```
jan% ael cd
Project "example.1.0", Change 11
Change Details

NAME
        Project "example.1.0", Change 11.

SUMMARY
        file names on command line

DESCRIPTION
        Optional input and output files may be specified on the
        command line.

CAUSE
        This change was caused by internal_bug.

STATE
        This change is in 'being_developed' state.

FILES
        Change has no files.

HISTORY
        What            When            Who     Comment
        ------          ------          -----   -------
        new_change      Fri Dec 11      alex
                        14:55:06 1992
        develop_begin   Mon Dec 14      jan
                        09:07:08 1992
jan%
```

Through one process or another, Jan determines that the *main.c* file is the one to be modified. This file is copied into the change:

```
jan% aecp main.c
aegis: project "example.1.0": change 11: file "main.c" copied
jan%
```

This file is now extended to look like this:

```
#include <stdio.h>

static void
usage()
{
        fprintf(stderr, "usage: example [ <infile> [ <outfile> ]]\n");
        exit(1);
}

void
main(argc, argv)
        int     argc;
        char    **argv;
{
        char    *in = 0;
        char    *out = 0;
        int     j;
```

```
        for (j = 1; j < argc; ++j)
        {
                char *arg = argv[j];
                if (arg[0] == '-')
                        usage();
                if (!in)
                        in = arg;
                else if (!out)
                        out = arg;
                else
                        usage();
        }

        if (in && !freopen(in, "r", stdin))
        {
                perror(in);
                exit(1);
        }
        if (out && !freopen(out, "w", stdout))
        {
                perror(out);
                exit(1);
        }

        yyparse();
        exit(0);
}
```

A new test is also required,

```
jan% aent
aegis: project "example.1.0": change 11: file "test/00/t0002a.sh" new
        test
jan%
```

which is edited to look like this:

```
#!/bin/sh
#
# test command line arguments
#
tmp=/tmp/$$
here='pwd'
if [ $? -ne 0 ]; then exit 1; fi

fail()
{
        echo FAILED 1>&2
        cd $here
        rm -rf $tmp
        exit 1
}

pass()
{
        cd $here
        rm -rf $tmp
        exit 0
}
trap "fail" 1 2 3 15
```

```
mkdir $tmp
if [ $? −ne 0 ]; then exit 1; fi
cd $tmp
if [ $? −ne 0 ]; then fail; fi

#
# with input like this
#
cat > test.in << ’foobar’
1
(24 − 22)
−(4 − 7)
2 * 2
10 / 2
4 + 2
10 − 3
foobar
if [ $? −ne 0 ]; then fail; fi

#
# the output should look like this
#
cat > test.ok << ’foobar’
1
2
3
4
5
6
7
foobar
if [ $? −ne 0 ]; then fail; fi

#
# run the calculator
# and see if the results match
#
# (Use /dev/null for input in case input redirect fails;
# don’t want the test to hang!)
#
$here/example test.in test.out < /dev/null
if [ $? −ne 0 ]; then fail; fi
diff test.ok test.out
if [ $? −ne 0 ]; then fail; fi
$here/example test.in < /dev/null > test.out.2
if [ $? −ne 0 ]; then fail; fi
diff test.ok test.out.2
if [ $? −ne 0 ]; then fail; fi

#
# make sure complains about rubbish
# on the command line
#
$here/example −trash < test.in > test.out
if [ $? −ne 1 ]; then fail; fi

#
# this much worked
#
pass
```

Now it is time for Jan to build and test the change. Through the magic of static documentation, this works first time, and here is how it goes:

```
jan% aeb
aegis: logging to "/u/pat/example.1.0.C011/aegis.log"
aegis: project "example.1.0": change 11: development build started
aegis: cook −b /projects/example/baseline/Howto.cook
        project=example.1.0 change=11 version=1.0.C011 −nl
cook: cc −I. −I/projects/example/baseline −O −c main.c
cook: cc −o example main.o /projects/example/baseline/gram.o
        /projects/example/baseline/lex.o −ll −ly
aegis: project "example.1.0": change 11: development build complete
jan% aet
aegis: logging to "/u/pat/example.1.0.C011/aegis.log"
aegis: sh /u/jan/example.1.0.C011/test/00/t0002a.sh
aegis: project "example.1.0e": change 11: test "test/00/t0002a.sh"
        passed
aegis: project "example.1.0": change 11: passed 1 test
jan%
```

All that remains if to difference the change and sign off.

```
jan% aed
aegis: logging to "/u/pat/example.1.0.C011/aegis.log"
aegis: set +e; diff −c /projects/example/main.c /u/jan/
        example.1.0.C011/main.c > /u/jan/example.1.0.C011/main.c,D; test $?
        −eq 0 −o $? −eq 1
aegis: project "example.1.0": change 11: difference complete
jan% aedmore
...examines the file...
jan%
```

Note how the context difference shows exactly what has changed. And now the sign-off:

```
jan% aede
aegis: project "example.1.0": change 11: no current 'aegis −Test
        −BaseLine' registration
jan%
```

No, it wasn't enough. Tests must not only pass against a new change, but must fail against the project baseline. This is to establish, in the case of bug fixes, that the bug has been isolated *and* fixed. New functionality will usually fail against the baseline, because the baseline can't do it (if it could, you wouldn't be adding it!). So, Jan needs to use a variant of the *aet* command.

```
jan% aet −bl
aegis: sh /u/jan/example.1.0.C011/test/00/t0002a.sh
usage: example
FAILED
aegis: project "example.1.0": change 11: test "test/00/t0002a.sh" on
        baseline failed (as it should)
aegis: project "example.1.0": change 11: passed 1 test
jan%
```

Running the regression tests is also a good idea

```
jan% aet −reg
aegis: logging to "/u/pat/example.1.0.C011/aegis.log"
aegis: sh /projects/example/baseline/test/00/t0001a.sh
aegis: project "example.1.0": change 11: test "test/00/t0001a.sh"
        passed
aegis: project "example.1.0": change 11: passed 1 test
jan%
```

Now Aegis will be satisfied

```
jan% aede
aegis: sh /usr/local/lib/aegis/aegis/de.sh example.1.0 11 jan
aegis: project "example.1.0": change 11: development completed
jan%
```

Like Pat in the change before, Jan will receive email that this change passed review, and later that it passed integration.

3.1.4. The Third and Fourth Changes

This section will show two people performing two changes, one each. The twist is that they have a file in common.

First Sam looks for a change to work on and starts, like this:

```
sam% aedb -l
Project "example.1.0"
List of Changes

Change  State           Description
------- -------         -------------
  12    awaiting_       add powers
        development
  13    awaiting_       add variables
        development
sam% aedb 12
aegis: project "example.1.0": change 12: development directory "/u/
        sam/example.1.0.C012"
aegis: project "example.1.0": change 12: user "sam" has begun
        development
sam% aecd
aegis: project "example.1.0": change 12: /u/sam/example.1.0.C012
sam%
```

A little sniffing around reveals that only the *gram.y* grammar file needs to be altered, so it is copied into the change.

```
sam% aecp gram.y
aegis: project "example.1.0": change 12: file "gram.y" copied
sam%
```

The grammar file is changed to look like this:

```
%token DOUBLE
%token NAME
%union
{
        double  lv_double;
        int     lv_int;
};

%type <lv_double> DOUBLE expr
%type <lv_int> NAME
%left '+' '-'
%left '*' '/'
%right '^'
%right UNARY

%%
example
        : /* empty */
        | example command '\n'
                { yyerrflag = 0; fflush(stderr); fflush(stdout); }
        ;
```

```
command
        : expr
                { printf("%g\n", $1); }
        | error
        ;

expr
        : DOUBLE
        | '(' expr ')'
                { $$ = $2; }
        | '-' expr
                %prec UNARY
                { $$ = -$2; }
        | expr '^' expr
                { $$ = pow($1, $3); }
        | expr '*' expr
                { $$ = $1 * $3; }
        | expr '/' expr
                { $$ = $1 / $3; }
        | expr '+' expr
                { $$ = $1 + $3; }
        | expr '-' expr
                { $$ = $1 - $3; }
        ;
```

The changes are very small. Sam checks to make sure using the difference command:

```
sam% aed
aegis: logging to "/u/sam/example.1.0.C012/aegis.log"
aegis: set +e; diff -c /projects/example/baseline/gram.y /u/sam/
        example.1.0.C012/gram.y > /u/sam/example.1.0.C012/gram.y,D; test $?
        -eq 0 -o $? -eq 1
aegis: project "example.1.0": change 12: difference complete
sam% aedmore
...examines the file...
sam%
```

The difference file looks like this

```
*** /projects/example/baseline/gram.y
--- /u/sam/example.1.0.C012/gram.y
***************
*** 1,5 ****
--- 1,6 ----
  %{
  #include <stdio.h>
+ #include <math.h>
  %}
  %token DOUBLE
  %token NAME

***************
*** 13,18 ****
--- 14,20 ----
  %type <lv_int> NAME
  %left '+' '-'
  %left '*' '/'
+ %right '^'
  %right UNARY
  %%
  example
```

```
***************
*** 32,37 ****
--- 34,41 ----
        | '-' expr
                %prec UNARY
                { $$ = -$2; }
+       | expr '^' expr
+               { $$ = pow($1, $3); }
        | expr '*' expr
                { $$ = $1 * $3; }
        | expr '/' expr
```

These are the differences Sam expected to see.

At this point Sam creates a test. All good software developers create the tests first, don't they?

```
sam% aent
aegis: project "example.1.0": change 12: file "test/00/t0003a.sh" new
        test
sam%
```

The test is created empty, and Sam edit it to look like this:

```
:
here=`pwd`
if test $? -ne 0 ; then exit 1; fi
tmp=/tmp/$$
mkdir $tmp
if test $? -ne 0 ; then exit 1; fi
cd $tmp
if test $? -ne 0 ; then exit 1; fi

fail()
{
        echo FAILED 1>&2
        cd $here
        chmod u+w `find $tmp -type d -print`
        rm -rf $tmp
        exit 1
}

pass()
{
        cd $here
        chmod u+w `find $tmp -type d -print`
        rm -rf $tmp
        exit 0
}
trap "fail" 1 2 3 15

cat > test.in << 'end'
5.3 ^ 0
4 ^ 0.5
27 ^ (1/3)
end
if test $? -ne 0 ; then fail; fi

cat > test.ok << 'end'
1
2
3
end
if test $? -ne 0 ; then fail; fi
```

```
$here/example test.in < /dev/null > test.out 2>&1
if test $? -ne 0 ; then fail; fi

diff test.ok test.out
if test $? -ne 0 ; then fail; fi

$here/example test.in test.out.2 < /dev/null
if test $? -ne 0 ; then fail; fi

diff test.ok test.out.2
if test $? -ne 0 ; then fail; fi

# it probably worked
pass
```

Everything is ready. Now the change can be built and tested, just like the earlier changes.

```
sam% aeb
aegis: logging to "/u/sam/example.1.0.C012/aegis.log"
aegis: project "example1.0": change 12: development build started
aegis: cook -b /projects/example/baseline/Howto.cook
        project=example.1.0 change=12 version=1.0.C012 -nl
cook: yacc -d gram.y
cook: mv y.tab.c gram.c
cook: mv y.tab.h gram.h
cook: cc -I. -I/projects/example/baseline -O -c gram.c
cook: cc -I. -I/projects/example/baseline -O -c /projects/
        example/baseline/lex.c
cook: cc -o example gram.o lex.o /projects/example/baseline/
        main.o -ll -ly -lm
aegis: project "example": change 3: development build complete
sam%
```

Notice how the yacc run produces a *gram.h* which logically invalidates the *lex.o* in the baseline, and so the *lex.c* file in the baseline is recompiled, using the *gram.h* include file from the development directory, leaving a new *lex.o* in the development directory. This is the reason for the use of

```
#include <filename>
```

directives, rather then the double quote form.

Now the change is tested.

```
sam% aet
aegis: logging to "/u/sam/example.1.0.C012/aegis.log"
aegis: sh /u/sam/example.1.0.C012/test/00/t0003a.sh
aegis: project "example.1.0": change 12: test "test/00/t0003a.sh"
        passed
aegis: project "example.1.0": change 12: passed 1 test
sam%
```

The change must also be tested against the baseline, and fail. Sam knows this, and does it here.

```
sam% aet -bl
aegis: logging to "/u/sam/example.1.0.C012/aegis.log"
aegis: sh /u/sam/example.1.0.C012/test/00/t0003a.sh
1,3c1,6
< 1
< 2
< 3
---
> syntax error
> 5.3
> syntax error
> 4
> syntax error
> 27
FAILED
aegis: project "example.1.0": change 12: test "test/00/t0003a.sh" on
        baseline failed (as it should)
aegis: project "example.1.0": change 12: passed 1 test
sam%
```

Running the regression tests is also a good idea.

```
sam% aet -reg
aegis: logging to "/u/sam/example.1.0.C012/aegis.log"
aegis: sh /projects/example/baseline/test/00/t0001a.sh
aegis: project "example.1.0": change 12: test "test/00/t0001a.sh"
        passed
aegis: sh /projects/example/baseline/test/00/t0002a.sh
aegis: project "example.1.0": change 12: test "test/00/t0002a.sh"
        passed
aegis: project "example.1.0": change 12: passed 2 tests
sam%
```

A this point Sam has just enough time to get to the lunchtime aerobics class in the staff common room.

Earlier the same day, Pat arrived for work a little after Sam, and also looked for a change to work on.

```
pat% aedb -l
Project "example.1.0"
List of Changes

Change  State           Description
------- -------         -------------
  13    awaiting_       add variables
        development
pat%
```

With such a wide choice, Pat selected change 13.

```
pat% aedb 13
aegis: project "example.1.0": change 13: development directory "/u/
        pat/example.1.0.C013"
aegis: project "example.1.0": change 13: user "pat" has begun
        development
pat% aecd
aegis: project "example.1.0": change 13: /u/pat/example.1.0.C013
pat%
```

To get more information about the change, Pat then uses the "change details" listing:

```
pat% ael cd
Project "example.1.0", Change 13
Change Details
```

```
NAME
        Project "example.1.0", Change 13.

SUMMARY
        add variables

DESCRIPTION
        Enhance the grammar to allow variables. Only single
        letter variable names are required.

CAUSE
        This change was caused by internal_enhancement.

STATE
        This change is in 'being_developed' state.

FILES
        This change has no files.

HISTORY
        What            When            Who     Comment
        ------          ------          -----   -------
        new_change      Mon Dec 14      alex
                        13:08:52 1992
        develop_begin   Tue Dec 15      pat
                        13:38:26 1992
pat%
```

To add the variables the grammar needs to be extended to understand them, and a new file for remembering and recalling the values of the variables needs to be added.

```
pat% aecp gram.y
aegis: project "example.1.0": change 13: file "gram.y" copied
pat% aenf var.c
aegis: project "example.1.0": change 13: file "var.c" added
pat%
```

Notice how Aegis raises no objection to both Sam and Pat having a copy of the *gram.y* file. Resolving this contention is the subject of this section.

Pat now edits the grammar file.

```
pat% vi gram.y
...edit the file...
pat% aed
aegis: logging to "/u/pat/example.1.0.C013/aegis.log"
aegis: set +e; diff -c /projects/example/baseline/gram.y /u/pat/
        example.1.0.C013/gram.y > /u/pat/example.1.0.C013/gram.y,D; test $?
        -eq 0 -o $? -eq 1
aegis: project "example.1.0": change 13: difference complete
pat%
```

The difference file looks like this

. . . hey, someone fill me in!...

The new *var.c* file was created empty by Aegis, and Pat edits it to look like this:

```
static double memory[26];
```

```
void
assign(name, value)
        int     name;
        double  value;
{
        memory[name] = value;
}

double
recall(name)
        int     name;
{
        return memory[name];
}
```

Little remains except to build the change.

```
pat% aeb
aegis: logging to "/u/pat/example.1.0.C013/aegis.log"
aegis: cook −b /example.proj/baseline/Howto.cook
        project=example.1.0 change=13 version=1.0.C013 −nl
cook: yacc −d gram.y
cook: mv y.tab.c gram.c
cook: mv y.tab.h gram.h
cook: cc −I. −I/projects/example/baseline −O −c gram.c
cook: cc −I. −I/projects/example/baseline −O −c /projects/
        example/baseline/lex.c
cook: cc −I. −I/projects/example/baseline −O −c var.c
cook: cc −o example gram.o lex.o /projects/example/baseline/
        main.o var.o −ll −ly −lm
aegis: project "example.1.0": change 13: development build complete
pat%
```

A new test for the new functionality is required and Pat creates one like this.

```
:
here=‘pwd‘
if test $? −ne 0 ; then exit 1; fi
tmp=/tmp/$$
mkdir $tmp
if test $? −ne 0 ; then exit 1; fi
cd $tmp
if test $? −ne 0 ; then exit 1; fi

fail()
{
        echo FAILED 1>&2
        cd $here
        chmod u+w ‘find $tmp −type d −print‘
        rm −rf $tmp
        exit 1
}
pass()
{
        cd $here
        chmod u+w ‘find $tmp −type d −print‘
        rm −rf $tmp
        exit 0
}
trap "fail" 1 2 3 15
```

```
cat > test.in << 'end'
a = 1
a + 1
c = a * 40 + 5
c / (a + 4)
end
if test $? -ne 0 ; then fail; fi

cat > test.ok << 'end'
2
9
end
if test $? -ne 0 ; then fail; fi

$here/example test.in < /dev/null > test.out 2>&1
if test $? -ne 0 ; then fail; fi

diff test.ok test.out
if test $? -ne 0 ; then fail; fi

$here/example test.in test.out.2 < /dev/null
if test $? -ne 0 ; then fail; fi

diff test.ok test.out.2
if test $? -ne 0 ; then fail; fi

# it probably worked
pass
```

The new files are then differenced:

```
pat% aed
aegis: logging to "/u/pat/example.1.0.C013/aegis.log"
aegis: set +e; diff -c /projects/example/baseline/gram.y /u/pat/
        example.1.0.C013/gram.y > /u/pat/example.1.0.C013/gram.y,D; test $?
        -eq 0 -o $? -eq 1
aegis: set +e; diff -c /dev/null /u/pat/example.1.0.C013/test/00/
        t0004a.sh > /u/pat/example.1.0.C013/test/00/t0004a.sh,D; test
        $? -eq 0 -o $? -eq 1
aegis: set +e; diff -c /dev/null /u/pat/example.1.0.C013/var.c > /u/
        pat/example.1.0.C013/var.c,D; test $? -eq 0 -o $? -eq 1
aegis: project "example.1.0": change 13: difference complete
pat%
```

Notice how the difference for the *gram.y* file is still current, and so is not run again.

The change is tested.

```
pat% aet
aegis: logging to "/u/pat/example.1.0.C013/aegis.log"
aegis: sh /u/pat/example.1.0.C013/test/00/t0001a.sh
aegis: project "example.1.0": change 13: test "test/00/t0004a.sh"
        passed
aegis: project "example.1.0": change 13: passed 2 tests
pat%
```

The change is tested against the baseline.

```
pat% aet −bl
aegis: logging to "/u/pat/example.1.0.C013/aegis.log"
aegis: sh /u/pat/example.1.0.C013/test/00/t0001a.sh
1,2c1,4
< 2
< 9
−−−
> syntax error
> syntax error
> syntax error
> syntax error
FAILED
aegis: project "example.1.0": change 13: test "test/00/t0004a.sh" on
        baseline failed (as it should)
pat%
```

And the regression tests

```
pat% aet −reg
aegis: logging to "/u/pat/example.1.0.C013/aegis.log"
aegis: sh /projects/example/baseline/test/00/t0001a.sh
aegis: project "example.1.0": change 13: test "test/00/t0001a.sh"
        passed
aegis: sh /projects/example/baseline/test/00/t0002a.sh
aegis: project "example.1.0": change 13: test "test/00/t0002a.sh"
        passed
aegis: project "example.1.0": change 13: passed 2 tests
pat%
```

Note how test 3 has not been run, in any form of testing. This is because test 3 is part of another change, and is not yet integrated with the baseline.

All is finished for this change,

```
pat% aede
aegis: sh /usr/local/lib/aegis/de.sh example.1.0 13 pat
aegis: project "example.1.0": change 13: development completed
pat%
```

Anxious to get this change into the baseline, Pat now wanders down the hall in search of a reviewer, but more of that in the next section.

Some time later, San returns from aerobics feeling much improved. All that is required for change 12 is to do develop end, or is it?

```
sam% aede
aegis: project "example.1.0": change 12: file "gram.y" in baseline
        has changed since last 'aegis −DIFFerence' command
sam%
```

A little sleuthing on Sam's part with the Aegis list command will reveal how this came about. The way to resolve this problem is with the difference command, but the merge variant − this will merge the new baseline version, and Sam's edit together.

```
sam% aem
aegis: logging to "/u/pat/example.1.0.C012/aegis.log"
aegis: co −u’1.1’ −p /projects/example/history/gram.y,v > /tmp/
        aegis.14594
/projects/example/history/gram.y,v  −−>  stdout revision 1.1 (unlocked)
aegis: (diff3 −e /projects/example/baseline/gram.y /tmp/
        aegis.14594 /u/sam/example.003/gram.y | sed −e ’/^w$/d’
        −e ’/^q$/d’;  echo ’1,$p’ ) | ed − /projects/example/
        baseline/gram.y,B > /u/sam/example.003/gram.y
aegis: project "example.1.0": change 12: merge complete
aegis: project "example.1.0": change 12: file "gram.y" was out of
        date and has been merged, see "gram.y,B" for original source
aegis: new ’aegis −Build’ required
sam%
```

This was caused by the conflict between change 13, which is now integrated, and change 12; both of which are editing the *gram.y* file. Sam examines the *gram.y* file, and is satisfied that it contains an accurate merge of the edit done by change 13 and the edits for this change. The merged source file looks like this:

```
%{
#include <stdio.h>
#include <math.h>
%}
%token DOUBLE
%token NAME
%union
{
        double  lv_double;
        int     lv_int;
};

%type <lv_double> DOUBLE expr
%type <lv_int> NAME
%left ’+’ ’−’
%left ’*’ ’/’
%right ’^’
%right UNARY

%%
example
        : /* empty */
        | example command ’\n’
                { yyerrflag = 0; fflush(stderr); fflush(stdout); }
        ;

command
        : expr
                { printf("%g\n", $1); }
        | NAME ’=’ expr
                { assign($1, $3); }
        | error
        ;
```

```
expr
        : DOUBLE
        | NAME
                { extern double recall(); $$ = recall($1); }
        | '(' expr ')'
                { $$ = $2; }
        | '-' expr
                %prec UNARY
                { $$ = -$2; }
        | expr '^' expr
                { $$ = pow($1, $3); }
        | expr '*' expr
                { $$ = $1 * $3; }
        | expr '/' expr
                { $$ = $1 / $3; }
        | expr '+' expr
                { $$ = $1 + $3; }
        | expr '-' expr
                { $$ = $1 - $3; }
        ;
```

The automatic merge worked because most such conflicts are actually working on logically separate portions of the file. Two different areas of the grammar in this case. In practice, there is rarely a real conflict, and it is usually small enough to detect fairly quickly.

Sam now rebuilds:

```
sam% aeb
aegis: logging to "/u/sam/example.1.0.C012/aegis.log"
aegis: project "example.1.0": change 12: development build started
aegis: cook -b /projects/example/baseline/Howto.cook
        project=example.1.0 change=12 version=1.0.C012 -nl
cook: rm gram.c
cook: rm gram.h
cook: yacc -d gram.y
cook: mv y.tab.c gram.c
cook: mv y.tab.h gram.h
cook: rm gram.o
cook: cc -I. -I/projects/example/baseline -O -c gram.c
cook: rm lex.o
cook: cc -I. -I/projects/example/baseline -O -c /projects/
        example/baseline/lex.c
cook: rm example
cook: cc -o example gram.o lex.o /projects/example/baseline/
        main.o /projects/example/baseline/var.o -ll -ly -lm
aegis: project "example.1.0": change 12: development build complete
sam%
```

Notice how the list of object files linked has also adapted to the addition of another file in the baseline, without any extra work by Sam.

All that remains is to test the change again.

```
sam% aet
aegis: /bin/sh /u/sam/example.1.0.C012/test/00/t0003a.sh
aegis: project "example.1.0": change 12: test "test/00/t0003a.sh"
        passed
aegis: project "example.1.0": change 12: passed 1 test
sam%
```

And test against the baseline,

```
sam% aet −bl
aegis: /bin/sh /u/sam/example.1.0.C012/test/00/t0003a.sh
1,3c1,6
< 1
< 2
< 3
−−−
> syntax error
> 5.3
> syntax error
> 4
> syntax error
> 27
FAILED
aegis: project "example.1.0": change 12: test "test/00/t0003a.sh" on
        baseline failed (as it should)
aegis: project "example.1.0": change 12: passed 1 test
sam%
```

Perform the regression tests, too. This is important for a merged change, to make sure you didn't break the functionality of the code you merged with.

```
sam% aet −reg
aegis: logging to "/u/sam/example.1.0.C012/aegis.log"
aegis: /bin/sh /projects/example/baseline/test/00/
        t0001a.sh
aegis: project "example.1.0": change 12: test "test/00/t0001a.sh"
        passed
aegis: /bin/sh /projects/example/baseline/test/00/
        t0002a.sh
aegis: project "example.1.0": change 12: test "test/00/t0002a.sh"
        passed
aegis: /bin/sh /projects/example/baseline/test/00/
        t0004a.sh
aegis: project "example.1.0": change 12: test "test/00/t0004a.sh"
        passed
aegis: project "example.1.0": change 12: passed 3 tests
sam%
```

All done, or are we?

```
sam% aede
aegis: project "example.1.0": change 12: no current 'aegis −Diff'
        registration
sam%
```

The difference we did earlier, which revealed that we were out of date, does not show the differences since the two changes were merged, and possibly further edited.

```
sam% aed
aegis: logging to "/u/sam/example.1.0.C012/aegis.log"
aegis: set +e; diff /projects/example/baseline/gram.y /u/pat/
        example.1.0.C012/gram.y > /u/pat/example.1.0.C012/gram.y,D;
        test $? −le 1
aegis: project "example.1.0": change 12: difference complete
sam%
```

This time everything will run smoothly,

```
sam% aede
aegis: project "example.1.0": change 12: development completed
sam%
```

Some time soon Sam will receive email that this change passed review, and later that it passed integration.

Within the scope of a limited example, you have seen most of what Aegis can do. To get a true feeling for the program you need to try it in a similarly simple case. You could even try doing this example manually.

3.1.5. Developer Command Summary

Only a few of the Aegis commands available to developers have been used in the example. The following table (very tersely) describes the Aegis commands most useful to developers.

Command	Description
aeb	Build
aeca	edit Change Attributes
aecd	Change Directory
aeclean	Clean a development directory
aeclone	copy a whole change
aecp	Copy File
aecpu	Copy File Undo
aed	Difference
aedb	Develop Begin
aedbu	Develop Begin Undo
aede	Develop End
aedeu	Develop End Undo
ael	List Stuff
aenf	New File
aenfu	New File Undo
aent	New Test
aentu	New Test Undo
aerm	Remove File
aermu	Remove File Undo
aet	Test

You will want to read the manual entries for all of these commands. Note that all Aegis commands have a *−Help* option, which will give a result very similar to the corresponding *man*(1) output. Most Aegis commands also have a *−List* option, which usually lists interesting context sensitive information.

/* vim: set ts=8 sw=4 et : */

3.2. The Reviewer

The role of a reviewer is to check another user's work. You are helped in this by Aegis, because changes can never reach the *being reviewed* state without several preconditions:

• The change is known to build. You know that it compiled successfully, so there is no need to search for syntax errors.

• The change has tests, and those tests have been run, and have passed.

This information allows you to concentrate on implementation issues, completeness issues, and local standards issues.

To help the reviewer, a set of "comma D" files is available in the change development directory. Every file which is to be added to the baseline, removed from the baseline, or changed in some way, has a corresponding "comma D" file.

3.2.1. Before You Start

Have you configured your account to use Aegis? See the *User Setup* section of the *Tips and Traps* chapter for how to do this.

3.2.2. The First Change

Robyn finds out what changes are available for review by asking Aegis:

```
robyn% aerpass −l −p example.1.0

Project "example.1.0"
List of Changes

Change  State           Description
−−−−−−− −−−−−−−         −−−−−−−−−−−−
  10    being_reviewed  Place under Aegis
robyn%
```

Any of the above changes could be reviewed, Robyn chooses the first.

```
robyn% aecd −p example.1.0 −c 10
aegis: project "example": change 1: /u/pat/example.1.0.C010
robyn% aedmore
...examines each file...
robyn%
```

The *aedmore* command walks the development directory tree to find all of the "comma D" files, and displays them using *more*(1). There is a corresponding *aedless* for those who prefer the *less*(1) command.

Once the change has been reviewed and found acceptable, it is passed:

```
robyn% aerpass −p example.1.0 10
aegis: sh /usr/local/lib/aegis/rp.sh example.1.0 10 pat robyn
aegis: project "example.1.0": change 10: passed review
robyn%
```

Some time soon Isa will notice the email notification and commence integration of the change.

3.2.3. The Second Change

Most reviews have the same pattern as the first.

```
robyn% aerpass -l -p example.1.0

Project "example.1.0"
List of Changes

Change  State           Description
------- -------         -------------
  11    being_reviewed  file names on command line
robyn%
```

Always change directory to the change's development directory, otherwise you will not be able to review the files.

```
robyn% aecd -p example.1.0 -c 11
aegis: project "example.1.0": change 11: /u/jan/example.1.0.C011
robyn%
```

Another useful way of finding out about a change is the "list change details" command, viz:

```
robyn% ael cd -p example.1.0 -c 11

Project "example.1.0", Change 11
Change Details

NAME
        Project "example.1.0", Change 11.

SUMMARY
        file names on command line

DESCRIPTION
        Optional input and output files may be specified on
        the command line.

CAUSE
        This change was caused by internal_bug.

STATE
        This change is in 'being_reviewed' state.

FILES
        Type    Action  Edit    File Name
        ------- ------- ------- -----------
        source  modify  1.1     main.c
        test    create          test/00/t0002a.sh

HISTORY
        What            When            Who     Comment
        ------          ------          -----   -------
        new_change      Fri Dec 11      alex
                        14:55:06 1992
        develop_begin   Mon Dec 14      jan
                        09:07:08 1992
        develop_end     Mon Dec 14      jan
                        11:43:23 1992
robyn%
```

Once Robyn knows what the change is meant to be doing, the files are then examined:

```
robyn% aedmore
...examines each file...
robyn%
```

Once the change is found to be acceptable, it is passed:

```
robyn% aerpass -p example.1.0 11
aegis: sh /usr/local/lib/aegis/rp.sh example.1.0 11 jan robyn
aegis: project "example.1.0": change 11: passed review
robyn%
```

Some time soon Isa will notice the email notification and commence integration of the change.

The reviews of the third and fourth changes will not be given here, because they are almost identical to the other changes. If you want to know how to fail a review, see the *aerfail*(1) manual entry.

3.2.4. Reviewer Command Summary

Only a few of the Aegis commands available to reviewers have been used in this example. The following table (very tersely) describes the Aegis commands most useful to reviewers.

Command	Description
aecd	Change Directory
aerpass	Review Pass
aerpu	Review Pass Undo
aerfail	Review Fail
ael	List Stuff

You will want to read the manual entries for all of these commands. Note that all Aegis commands have a *−Help* option, which will give a result very similar to the corresponding *man*(1) output. Most Aegis commands also have a *−List* option, which usually lists interesting context sensitive information.

3.3. The Integrator

This section shows what the integrator must do for each of the changes shown to date. The integrator does not have the ability to alter anything in the change; if a change being integrated is defective, it is simply failed back to the developer. This documented example has no such failures, in order to keep it manageably small.

3.3.1. Before You Start

Have you configured your account to use Aegis? See the *User Setup* section of the *Tips and Traps* chapter for how to do this.

3.3.2. The First Change

The first change of a project is often the trickiest, and the integrator is the last to know. This example goes smoothly, and you may want to consider using the example project as a template.

The integrator for this example project is Isa. Isa knows there is a change ready for integration from the notification which arrived by email.

```
isa% aeib −l −p example.1.0

Project "example.1.0"
List of Changes

Change  State           Description
−−−−−−− −−−−−−−         −−−−−−−−−−−−
  10    awaiting_       Place under Aegis
        integration
isa% aeib example.1.0 10
aegis: project "example.1.0": change 10: link baseline to integration
        directory
aegis: project "example.1.0": change 10: apply change to integration
        directory
aegis: project "example.1.0": change 10: integration has begun
isa%
```

The integrator must rebuild and retest each change. This ensures that it was no quirk of the developer's environment which resulted in the success at the development stage.

```
isa% aeb
aegis: logging to "/projects/example/delta.001/aegis.log"
aegis: project "example.1.0": change 10: integration build started
aegis: cook −b Howto.cook project=example.1.0 change=10
        version=1.0.D001 −nl
cook: yacc −d gram.y
cook: mv y.tab.c gram.c
cook: mv y.tab.h gram.h
cook: cc −I. −O −c gram.c
cook: lex lex.l
cook: mv lex.yy.c lex.c
cook: cc −I. −O −c lex.c
cook: cc −I. −O −c main.c
cook: cc −o example gram.o lex.o main.o −ll −ly
aegis: project "example.1.0": change 10: integration build complete
isa%
```

Notice how the above build differed from the builds that were done while in the *being developed* state; the extra baseline include is gone. This is because the integration directory will shortly be the new baseline, and must be entirely internally consistent and self-sufficient.

You are probably wondering why this isn't all rolled into the one Aegis command. It is not because there may be some manual process to be performed after the build and before the test. This may be making a command set-uid-root (as in the case of Aegis itself) or it may require some tinkering with the local oracle or ingress database. Instructions for the integrator may be placed in the description field of the change

attributes.

The change is now re-tested:

```
isa% aet
aegis: logging to "/projects/example/delta.001/aegis.log"
aegis: sh /project/example/delta.001/test/00/t0001a.sh
aegis: project "example": change 1: test "test/00/t0001a.sh"
        passed
aegis: project "example": change 1: passed 1 test
isa%
```

The change builds and tests. Once Isa is happy with the change, perhaps after browsing the files, Isa then passes the integration, causing the history files to be updated and the integration directory becomes the baseline.

```
isa% aeipass
aegis: logging to "/projects/example/delta.001/aegis.log"
aegis: ci -u -m/dev/null -t/dev/null /projects/example/delta.001/
        Howto.cook /projects/example/history/Howto.cook,v;
        rcs -U /projects/example/history/Howto.cook,v
/projects/example/history/Howto.cook,v  <--
        /projects/example/delta.001/Howto.cook
initial revision: 1.1
done
RCS file: /projects/example/history/Howto.cook,v
done
aegis: rlog -r /projects/example/history/Howto.cook,v | awk
        '/^revision/ {print $2}' > /tmp/aegis.15309
...lots of similar RCS output...
aegis: project "example.1.0": change 10: remove development directory
aegis: sh /usr/local/lib/aegis/ip.sh example.1.0 10 pat robyn isa
aegis: project "example.1.0": change 10: integrate pass
isa%
```

All of the staff involved, will receive email to say that the change has been integrated. This notification is a shell script, so USENET could be usefully used instead.

You should note that the development directory has been deleted. It is expected that each development directory will only contain files necessary to develop the change. You should keep "precious" files somewhere else.

3.3.3. The Other Changes

There is no difference to integrating any of the later changes. The integration process is very simple, as it is a cut-down version of what the developer does, without all the complexity.

Your project may elect to have the integrator also monitor the quality of the reviews. An answer to "who will watch the watchers" if you like.

It is also a good idea to rotate people out of the integrator position after a few weeks in a busy project, this is a very stressful position. The position of integrator gives a unique perspective to software quality, but the person also needs to be able to say "NO!" when a cruddy change comes along.

3.3.4. Integrator Command Summary

Only a few of the Aegis commands available to integrators have been used in this example. The following table (very tersely) describes the Aegis commands most useful to integrators.

Command	Description
aeb	Build
aecd	Change Directory
aed	Difference
aeib	Integrate Begin
aeibu	Integrate Begin Undo
aeifail	Integrate Fail
ael	List Stuff
aet	Test
aeipass	Integrate Pass

You will want to read the manual entries for all of these commands. Note that all Aegis commands have a *−Help* option, which will give a result very similar to the corresponding *man*(1) output. Most Aegis commands also have a *−List* option, which usually lists interesting context sensitive information.

3.3.5. Minimum Integrations

The **aegis −integrate-begin** command provides a **−minimum** option which may be used for various reasons. The term **minimum** may be a bit counter intuitive. One might think it means to the **minimum** amount of work, however it actually means use a **minimum** of files from the baseline in populating the *delta* directory. This normally leads to actually building everything in the project from sources and, as such, might be considered the most robust of builds.

Note that any change which removes a file, whether by *aerm* or *aemv*, results in an implicit **minimum** integration. This is intended to ensure nothing in the project references the removed file.

A project may adopt a policy that a product release should be based on a minimum integration. Such a policy may be a reflection of local confidence, or lack therof, in the projects DMT (Dependency Maintenance Tool) or build system. Or it may be based on a validation process wishing to make a simple statement on how the released package was produced.

Another, more transient, reason a to require a minimum integration might be when upgrading a third party library, compiler or maybe even OS level. Any of these events would signal the need for a minimum integration to ensure everything is rebuilt using the new resources.

The cost of a **minimum** integration varies according to type and size of the project. For very large projects, especially those building large numbers of binaries, the cost can be large. However large projects also require significant time to fully populate the delta directory. A minimum integration only copies those files under aegis control, skipping all “produced” files. In the case where a file upon which everything depends is changed, everything will be built anyway so the copy of the already built files is a waste of time. This means that sometimes a minimum can be as cheap as a normal integration.

3.4. The Administrator

The previous discussion of developers, reviewers and integrators has covered many aspects of the production of software using Aegis. The administrator has responsibility for everything they don't, but there is very little left.

These responsibilities include:

• access control: The administrator adds and removes all categories of user, including administrators. This is on a per-project basis, and has nothing to do with UNIX user administration. This simply nominates which users may do what.

• change creation: The administrator adds (and sometimes removes) changes to the system. At later stages, developers may alter some attributes of the change, such as the description, to say what they fixed.

• project creation: Aegis does not limit who may create projects, but when a project is created the user who created the project is set to be the administrator of that project.

All of these things will be examined

3.4.1. Before You Start

Have you configured your account to use Aegis? See the *User Setup* section of the *Tips and Traps* chapter for how to do this.

3.4.2. The First Change

Many things need to happen before development can begin on the first change; the project must be created, the staff but be given access permissions, the branches created, and the change must be created.

```
alex% aenpr example -dir /projects/example -version -
aegis: project "example": project directory "/projects/example"
aegis: project "example": created
alex%
```

Once the project has been created, the project attributes are set. Alex will set the desired project attributes using the **-Edit** option of the **aepa** command. This will invoke an editor (*vi*(1) by default) to edit the project attributes. Alex edits them to look like this:

```
description = "Aegis Documentation Example Project";
developer_may_review = false;
developer_may_integrate = false;
reviewer_may_integrate = false;
```

The project attributes are set as follows:

```
alex% aepa -edit -p example
...edit as above...
aegis: project "example.1.0": attributes changed
alex% ael p
List of Projects

Project     Directory           Description
-------     -----------         -------------
example     /projects/example   Aegis Documentation Example
                                Project
alex%
```

The various staff must be added to the project. Developers are the only staff who may actually edit files.

```
alex% aend pat jan sam -p example
aegis: project "example": user "pat" is now a developer
aegis: project "example": user "jan" is now a developer
aegis: project "example": user "sam" is now a developer
alex%
```

Reviewers may veto a change. There may be overlap between the various categories, as show here for Jan:

```
alex% aenrv robyn jan -p example
aegis: project "example": user "robyn" is now a reviewer
aegis: project "example": user "jan" is now a reviewer
alex%
```

The next role we need to fill is an integrator.

```
alex% aeni isa -p example
aegis: project "example": user "isa" is now an integrator
alex%
```

Once the staff have been given access, it is time to create the working branch. Branches inherit their attributes and staff lists from their parent branches when they are first created, which is why we set all that stuff first.

```
alex% aegis -nbr -p example 1
aegis: project "example.1": created
alex% aegis -nbr -p example.1 0
aegis: project "example.1.0": created
alex%
```

This is for versioning; see the *Branching* chapter for more information. For the moment, we will simply work on branch 1.0. Notice how the branches appear as projects in the project listing; in general branches can be used interchangeably with projects.

```
alex% ael p
List of Projects

Project     Directory           Description
-------     -----------         -------------
example     /projects/example   Aegis Documentation Example
                                Project
example.1   /projects/example/  Aegis Documentation Example
            branch.1            Project, branch.1.
example.1.0 /projects/example/  Aegis Documentation Example
            branch.1/branch.0   Project, branch.1.0.
alex%
```

Once the working branch has been created, Alex creates the first change. The **-Edit** option of the **aenc** command is used, to create the attributes of the change. They are edited to look like this:

```
brief_description = "Create initial skeleton.";
description = "A simple calculator using native \
floating point precision.  \
The four basic arithmetic operators to be provided, \
using conventional infix notation.  \
Parentheses and negation also required.";
cause = internal_enhancement;
```

The change is created as follows:

```
alex% aenc -edit -p example.1.0
...edit as above...
aegis: project "example.1.0": change 10: created
alex%
```

Notice that the first change number is “10”. This is done so that changes 1 to 9 could be used as bug-fix branches at some future time. See the *Branching* chapter for more information. You can over-ride this is you need to.

The above was written almost a decade ago. There is a newer command, *tkaenc*, which uses a GUI and is much easier to use, with a much less fiddly interface. You may want to try that command, instead, for most routine change creation.

At this point, Alex walks down the hall to Pat’s office, to ask Pat to develop the first change. Pat has had some practice using Aegis, and can be relied on to do the rest of the project configuration speedily.

3.4.3. The Second Change

Some time later, Alex patiently sits through the whining and grumbling of an especially pedantic user. The following change description is duly entered:

```
brief_description = "file names on command line";
description = "Optional input and output files may be \
specified on the command line.";
cause = internal_bug;
```

The pedantic user wanted to be able to name files on the command line, rather than use I/O redirection. Also, having a bug in this example is useful. The change is created as follows:

```
alex% aenc -edit -p example.1.0
...edit as above...
aegis: project "example.1.0": change 11: created
alex%
```

At some point a developer will notice this change and start work on it.

3.4.4. The Third Change

Other features are required for the calculator, and also for this example. The third change adds exponentiation to the calculator, and is described as follows:

```
brief_description = "add powers";
description = "Enhance the grammar to allow exponentiation.  \
No error checking required.";
cause = internal_enhancement;
```

The change is created as follows:

```
alex% aenc -edit -p example.1.0
...edit as above...
aegis: project "example.1.0": change 12: created
alex%
```

At some point a developer will notice, and this change will be worked on.

3.4.5. The Fourth Change

A fourth change, this time adding variables to the calculator is added.

```
brief_description = "add variables";
description = "Enhance the grammar to allow variables.  \
Only single letter variable names are required.";
cause = internal_enhancement;
```

The change is created as follows:

```
alex% aenc -edit -p example.1.0
...edit as above...
aegis: project "example.1.0": change 13: created
alex%
```

At some point a developer will notice, and this change will be worked on.

3.4.6. Administrator Command Summary

Only a few of the Aegis commands available to administrators have been used in this example. The following table lists the Aegis commands most useful to administrators.

Command	Description
aeca	edit Change Attributes
ael	List Stuff

aena	New Administrator
aenc	New Change
aencu	New Change Undo
aend	New Developer
aeni	New Integrator
aenpr	New Project
aenrv	New Reviewer
aepa	edit Project Attributes
aera	Remove Administrator
aerd	Remove Developer
aeri	Remove Integrator
aermpr	Remove Project
aerrv	Remove Reviewer

You will want to read the manual entries for all of these commands. Note that all Aegis commands have a *−Help* option, which will give a result very similar to the corresponding *man*(1) output. Most Aegis commands also have a *−List* option, which usually lists interesting context sensitive information.

3.5. What to do Next

This chapter has given an overview of what using Aegis feels like. As a next step in getting to know Aegis, it would be a good idea if you created a project and went through this same exercise. You could use this exact example, or you could use a similar small project. The idea is simply to run through many of the same steps as in the example. Typos and other natural events will ensure that you come across a number of situations not directly covered by this chapter.

If you have not already done do, a printed copy of the section 1 and 5 manual entries will be invaluable. If you don't want to use that many trees, they will be available on-line, by using the "−Help" option of the appropriate command variant. Try:

```
% aedb -help
...manual entry...
%
```

Note that this example has not demonstrated all of the available functionality. One item of particular interest is that tests, like any other source file, may be copied into a change and modified, or even deleted, just like any other source file.

3.6. Common Questions

There are a number of questions which are frequently asked by people evaluating Aegis. This section attempts to address some of them.

3.6.1. Insulation

The repository model used by Aegis is of the "push" type − that is, changes to the baseline are "pushed" onto the developer as soon as they are integrated. Many configuration management systems have a "pull" model, where the developer elects when to cope with changes in the repository. At first glance, Aegis does not appear to have a "pull" equivalent.

It is possible to insulate your change from the baseline as much or as little as required. The *aecp*(1) command, used to copy files into a change, has a `-read-only` option. The files copied in this way are marked as insulation (*i.e.* you don't intend to change them). If you have not un-copied them at develop end time, the *aede*(1) command will produce a suitable error message, reminding you to un-copy the insulation and verify that your change still builds and tests successfully with the (probably) now-different baseline.

3.6.1.1. Copy Read-Only

It is possible to select the degree of insulation. By using "`aecp .`" at the top of a development directory, the complete project source tree will be copied, thus completely insulating you. Mind you, it comes at the cost of a complete build.

If you are working on a library, and only want the rest of the library to remain fixed, simply copy the whole library (`aecp library/fred`), and allow the rest to track the baseline. This comes at a smaller cost, because more of the baseline's object files can be taken advantage of.

3.6.1.2. Branches

It is also possible to create a sub-branch (see the *Branching* chapter). This does not itself automatically insulate, however the first change of a branch intended to insulate would copy and integrate *but not modify* the files to be insulated. You need to remember to perform a *cross-branch merge* with the parent branch before integrating the branch back into the parent branch.

3.6.1.3. Builds

You can also insulate yourself from baseline change by being selective about what you choose to build. You can do this by giving specific build targets on the *aeb*(1) command line, or you could copy the build tool's configuration file and butcher it. Remember to change it back before you *aede*(1) your change!

3.6.1.4. Mix-and-Match

Some or all of the above techniques may be combined to provide an insulation technique best suited to your project and development policy. *E.g.* changing the build configuration file for a branch dedicated to working on a small portion of a large project; towards the ed of the development, change the build configuration file back and perform integration testing.

3.6.1.5. Disadvantages

There is actually a down-side to insulating your changes from the evolution of the baseline. By noticing and adapting to the baseline, you have much less merging to do at the end of your change set. Each integration will typically be be modest, but the cumulative effect could be

substantial, and add a huge unexpected (and unbudgeted for) time penalty.

It also means that if there are integration problems between your work and the changes which were integrated before yours, or if your work shows up a bug in their work, the project find this out late, rather than early. The literature, based on industrial experience, indicates that the earlier problems are found the cheaper they are to fix.

Insulated development directories also use more disk space. While disk space is relatively cheap these days, it can still add up to a substantial hit for a large development team. Un-insulated development directories can take advantage of the pre-compiled objects and libraries in the baseline.

3.6.2. Partial Check-In

In the course of developing new functionality, it is very common to come across a pre-existing bug which the new functionality exposes. It is common for such bugs to be fixed by the developer in the same development directory, in order to get the new functionality to a testable state.

There are two common courses of action at this point: simply include the bug fix with the rest of the change, or integrate the bug fix earlier than the rest of the change. Combining the bug fix with the rest of the change can have nasty effects on statistics: it can hide the true bug level from your metrics program, and it also denies Aegis the opportunity of having accurate test correlations (see *aet*(1) for more information.) It also denies the rest of the development team the use of the bug fix, or worse, it allows the possibility that more than one team member will fix the bug, wasting development effort and time.

Many configuration management systems allow you to perform a partial check-in of a work area. This means that you can check-in just the bug fix, but continue to work on the unfinished portions of the functionality you are implementing.

Because Aegis insists on atomic change sets which are known to build and test successfully, such a partial check-in is not allowed − because Aegis can't know for certain that it works.

Instead, you are able to *clone* a change (see *aeclone*(1) for more information). This gives you a new change, and a second development directory, with exactly the same files. You then remove from this second change all of the files not related to the bug fix (using *aecpu*(1), *aenfu*(11), *etc*). You then create a test, build, difference, run the test, develop end, all as usual.

The original change will then need to be merged with the baseline, because the bug fix change will have been integrated before it. Usually this is straight-forward, as you already have the changes (some merge tools make this harder than others). Often, all that is required is to merge, and then say "`aecpu -unch`" to un-copy all files which are (now) unchanged when compared to the current baseline.

3.6.3. Multiple Active Branches

Some companies have multiple branches active at the same time, for different customers or distributions, *etc*.

They often need to make the same change to more than one branch. Some configuration management systems allow you to check-in the same file multiple times, once to each active branch. Aegis does not let you do this, because you need to convince Aegis that the change set will build and test cleanly on each branch. It is quite common for the change to require non-trivial edits to work on each branch.

3.6.3.1. Cloning

Aegis instead provides two mechanisms to handle this. The first, and simplest to understand, is to clone the change onto each relevant branch (rather than onto the same branch, as mentioned above for bug fixes). Then build and test as normal.

3.6.3.2. Ancestral

The second technique is more subtle. Perform the fix as a change to the common ancestor of both branches. This assumes that neither branch is insulated against the relevant area of code, and that earlier changes to the branch do not mask it in some way (otherwise a cross-branch merge with the common ancestor will be needed to propagate the fix).

3.6.4. Collaboration

It is often the case that difficult problems are tackled by small groups of 2 or 3 staff working together. In order to do this, they often work in a shared work area with group-writable or global-write permissions. However, this tends to give security auditor heart attacks.

Aegis has several different ways to achieve the same ends, and still not give the auditors indigestion.

3.6.4.1. Change Owner

The simplest method available is to change the ownership of a change from one developer to the next. A new development directory is created for the new developer, and the source files are copied across[15]. This allows a kind of serial collaboration between developers.

3.6.4.2. Branch

The other possibility is to create a branch to perform the work in, rather than a simple change. (A branch in Aegis is literally just a big change, which has lots of sub-changes.) This allows parallel collaboration between developers.

3.6.4.3. View Path Hacking

Aegis usually provides a "view path" to the build tool. This specifies where to look for source files and derived files, in order to union together the development directory, and the baseline, and the branch's ancestors' baselines. If you run the build by hand, rather than through Aegis, you can add another developer's development directory to the view path, after your own development directory, but before the baseline.

This has many of the advantages of the branch method, but none of the safeguards. In particular, if the other developer edits a file, and it no longer compiles, your development directory will not, either.

[15] For the technically minded: the *chown*(2) system call has semantics which vary too widely between UNIX variants and file-systems to be useful.

4. The History Tool

Aegis is decoupled from the history mechanism. This allows you to use the history mechanism of your choice, SCCS or RCS, for example. You may even wish to write your own.

The intention of this is that you may use a history mechanism which suits your special needs, or the one that comes free with your flavour of UNIX operating system.

Aegis uses the history mechanism for file *history* and so does not require many of the features of SCCS or RCS. This simplistic approach can sometimes make the interface to these utilities look a little strange.

4.1. History File Names

In order to track project source file renames and yet preserve a continuous history, the name of each source file and the name of each corresponding history file have nothing in common. The history file will have the same name (both on the local repository and any remote repository it is in) no matter how many times the source file is renamed.

Each source file is assigned universally unique identifier (UUID) when it is first created. This attribute, unlike the source file's name, is immutable and thus is suitable for use when forming the name of the history file.

4.2. Interfacing

The history mechanism interface is found in the project configuration file called *aegis.conf*, relative to the root of the baseline. It is a source file and subject to the same controls as any other source file. The history fields of the file are described as follows

4.2.1. history_create_command

This field is used to create a new history. The command is always executed as the project owner. Substitutions available for the command string are:

${Input}
: absolute path of source file

${History}
: absolute path of history file

In addition, all substitutions described in *aesub*(5) are available.

This command should be identical to the *history_put_command* otherwise mysterious things can happen when branches are ended.

4.2.2. history_get_command

This field is used to get a file from history. The command may be executed by developers. Substitutions available for the command string are:

${History}
: absolute path of history file

${Edit}
: edit number, as given by the history_query_command.

${Output}
: absolute path of destination file

In addition, all substitutions described in *aesub*(5) are available.

4.2.3. history_put_command

This field is used to add a new change to the history. The command is always executed as the project owner. Substitutions available for the command string are:

${Input}
: absolute path of source file

${History}
: absolute path of history file

In addition, all substitutions described in *aesub*(5) are available.

This command should be identical to the *history_create_command* otherwise mysterious things can happen when branches are ended.

4.2.4. history_query_command

This field is used to query the topmost edit of a history file. Result to be printed on the standard output. This command may be executed by developers. Substitutions available for the command string are:

${History}
: absolute path of history file

In addition, all substitutions described in *aesub*(5) are available.

4.2.5. history_content_limitation

This field describes the content style which the history tool is capable of working with.

ascii_text
: The history tool can only cope with files

which contain printable ASCII characters, plus space, tab and newline. The file must end with a newline. This is the default.

international_text
: The history tool can only cope with files which do not contain the NUL character. The file must end with a newline.

binary_capable
: The history tool can cope with files without any limitation on the form of the contents.

When a file is added to the history (by either the *history_create_command* or the *history_put_command* field) it is examined for conformance to this limitation. If there is a problem, the file is encoded in either the MIME quoted printable or the MIME Base 64 encoding (see RFC 1521), whichever is smaller, before being given to the history tool. The file in the baseline is unchanged.

On extract (the *history_get_command* field) the encoding is reversed, using information attached to the change file information. This is because each put could use a different encoding (although in practice, file contents rarely change that dramatically, and the same encoding is likely to be deduced every time).

4.2.6. history_tool_trashes_file

Many history tools (e.g. RCS) can modify the contents of the file when it is committed. While there are usually options to turn this off, they are seldom used. The problem is: if the commit changes the file, the source in the repository now no longer matches the object file in the repository − *i.e.* the history tool has compromised the referential integrity of the repository.

By default, when this happens Aegis issues a fatal error (at *intergate pass* time). You can turn this into a warning if you are convinced this is irrelevant. This would only make sense if the substition only ever occurs in comments. See *aepconf*(5) for more information on the values for this field.

4.2.7. Quoting Filenames

The default setting is for Aegis to reject filenames which contain shell special characters. This ensures that filenames may be substituted into the commands without worrying about whether this is safe. If you set the *shell_safe_filenames* field of the project *aegis.conf* file to `false`, you will need to surround filenames with the `${quote` *filename*`}` substitution. This will only quote filenames which actually need to be quoted, so users usually will not notice. This applies to all of the various filenames in the commands in the sections which follow.

4.2.8. Templates

The source distribution contains numerous configuration examples in a directory called *lib/config.example/* which is installed into */usr/local/share/aegis/config.example/* by default. In the interests of accuracy, it may be best to copy configurations from there, rather than copy-type the ones below.

4.3. Using aesvt

The *aesvt*(1) command is distributed with Aegis. It supports binary files, has versy small history files, and has good end-to-end behaviour. The entries for the commands are listed below.

4.3.1. history_create_command

This command is used to create a new file history. This command is always executed as the project owner.

The following substitutions are available:

${Input}
: absolute path of the source file

${History}
: absolute path of the history file

The entry in the *aegis.conf* file looks like this:

```
history_create_command =
  "aesvt -checkin "
  "-history $history "
  "-f $input"
  ;
```

4.3.2. history_put_command

It is essential that the *history_create_command* and the *history_put_command* are identical. It is a historical accident that there are two separate commands: before Aegis supported branches, this was not a requirement.

4.3.3. history_get_command

This command is used to get a specific edit back from history. This command is always executed as the project owner.

The following substitutions are available:

${History}
: absolute path of the history file

${Edit}
: edit number, as given by history_query_-command

${Output}
: absolute path of the destination file

The entry in the *aegis.conf* file looks like this:

```
history_get_command =
  "aesvt -checkout "
  "-history $history "
  "-edit $edit "
  "-o $output"
  ;
```

4.3.4. history_query_command

This command is used to query what the history mechanism calls the top-most edit of a history file. The result may be any arbitrary string, it need not be anything like a number, just so long as it uniquely identifies the edit for use by the *history_get_command* at a later date. The edit number is to be printed on the standard output. This command is always executed as the project owner.

The following substitutions are available:

${History}
: absolute path of the history file

The entry in the *aegis.conf* file looks like this:

```
history_query_command =
  "aesvt -query "
  "-history $history"
  ;
```

4.3.5. Templates

The *lib/config.example/aesvt* file in the Aegis distribution (installed as */usr/local/share/aegis/config.example/aesvt* by default) contains all of the above commands, so that you may readily insert them into your project configuration file.

Also, there are some subtleties to writing the commands, which are not present in the above examples. In particular, being able to support file names which contain characters which are special to the shell requires the use of the ${quote} substitution around all of the files names in the commands.

In addition, it is possible to store meta-data with each version. For example: "`Description=${quote ($version) ${change description}}`" inserts the version number and the brief description into the file's log. This means that using the *aesvt -list* option will provide quite useful summaries.

4.3.6. Binary Files

The *aesvt*(1) command is able to cope with binary files. Set

```
history_content_limitation =
    binary_capable;
```

so that Aegis knows that no encoding is required.

4.4. Using SCCS

The entries for the commands are listed below. SCCS uses a slightly different model than Aegis wants, so some maneuvering is required. The command strings in this section assume that the SCCS command *sccs* is in the command search PATH, but you may like to hard-wire the path, or set PATH at the start of each command. (It is also possible that you need to say “delta” instead of “sccs delta”. if this is the case, this command needs to be in the path.) You should also note that the strings are always handed to the Bourne shell to be executed, and are set to exit with an error immediately a sub-command fails.

One further assumption is that the *ae-sccs-put*(1) command, which is distributed with Aegis, is in the command search path. This insulates some of the weirdness that SCCS carries on with, and makes the commands below comprehensible.

4.4.1. history_create_command

This command is used to create a new project history. The command is always executed as the project owner.

The following substitutions are available:

${Input}
: absolute path of the source file

${History}
: absolute path of the history file

The entry in the *aegis.conf* file looks like this:

```
history_create_command =
  "ae-sccs-put -y$version -G$input "
  " ${d $h}/s.${b $h}";
```

It is important that the *history_create_command* and the *history_put_command* be the same. This is necessary for branching to work correctly.

4.4.2. history_get_command

This command is used to get a specific edit back from history. The command may be executed by developers.

The following substitutions are available:

${History}
: absolute path of the history file

${Edit}
: edit number, as given by history_query_command

${Output}
: absolute path of the destination file

The entry in the *aegis.conf* file looks like this:

```
history_get_command =
  "get -r'$e' -s -p -k "
  " ${d $h}/s.${b $h} > $o";
```

4.4.3. history_put_command

This command is used to add a new "top-most" entry to the history file. This command is always executed as the project owner.

The following substitutions are available:

${Input}
: absolute path of source file

${History}
: absolute path of history file

The entry in the *aegis.conf* file looks like this:

```
history_put_command =
  "ae-sccs-put -y$version -G$input "
  " ${d $h}/s.${b $h}";
```

Note that the SCCS file is left in the *not-edit* state, and that the source file is left in the baseline.

It is important that the *history_create_command* and the *history_put_command* be the same. This is necessary for branching to work correctly.

4.4.4. history_query_command

This command is used to query what the history mechanism calls the top-most edit of a history file. The result may be any arbitrary string, it need not be anything like a number, just so long as it uniquely identifies the edit for use by the *history_get_command* at a later date. The edit number is to be printed on the standard output. This command may be executed by developers.

The following substitutions are available:

${History}
: absolute path of the history file

The entry in the *aegis.conf* file looks like this:

```
history_query_command =
  "get -t -g ${d $h}/s.${b $h}";
```

Note that "get" reports the edit number on stdout.

4.4.5. Templates

The *lib/config.example/sccs* file in the Aegis distribution contains all of the above commands (installed as */usr/local/share/aegis/example.config/sccs* by default) so that you may readily insert them into your project configuration file (called *aegis.conf* by default, see *aepconf(5) for how to call it something else).*

Also, there are some subtleties to writing the commands, which are not present in the above examples. In particular, being able to support file names which contain characters which are special to the shell requires the use of the ${quote} substitution around all of the files names in the commands.

In addition, it is possible to have a much more useful description for the **-y** option. For example: "`-y${quote ($version) ${change description}}`" inserts the version number and the brief description into the file's log. This means that using the *sccs prs*(1) command will provide quite useful summaries.

4.4.6. Binary Files

SCCS is unable to cope with binary files. However, Aegis will transparently encode all such files, if you leave the *history_content_limitation* field unset.

4.5. Using RCS

The entries for the commands are listed below. RCS uses a slightly different model than aegis wants, so some maneuvering is required. The command strings in this section assume that the RCS commands *ci* and *co* and *rcs* and *rlog* are in the command search PATH, but you may like to hard-wire the paths, or set PATH at the start of each. You should also note that the strings are always handed to the Bourne shell to be executed, and are set to exit with an error immediately a sub-command fails.

In these commands, the RCS file is kept unlocked, since only the owner will be checking changes in. The RCS functionality for coordinating shared access is not required.

One advantage of using RCS version 5.6 or later is that binary files are supported, should you want to have binary files in the baseline.

4.5.1. history_create_command

This command is used to create a new file history. This command is always executed as the project owner.

The following substitutions are available:

${Input}
: absolute path of the source file

${History}
: absolute path of the history file

The entry in the *aegis.conf* file looks like this:

```
history_create_command =
  "ci -u -d -M -m$c -t/dev/null \
  $i $h,v; rcs -U $h,v";
```

The "`ci -u`" option is used to specify that an unlocked copy will remain in the baseline. The "`ci -d`" option is used to specify that the file time rather than the current time is to be used for the new revision. The "`ci -M`" option is used to specify that the mode date on the original file is not to be altered. The "`ci -t`" option is used to specify that there is to be no description text for the new RCS file. The "`ci -m`" option is used to specify that the change number is to be stored in the file log if this is actually an update (typically from *aenf* after *aerm* on the same file name). The "`rcs -U`" option is used to specify that the new RCS file is to have unstrict locking.

It is essential that the *history_create_command* and the *history_put_command* are identical. It is a historical accident that there are two separate commands: before Aegis supported branches, this was not a requirement.

4.5.2. history_get_command

This command is used to get a specific edit back from history. This command is always executed as the project owner.

The following substitutions are available:

${History}
: absolute path of the history file

${Edit}
: edit number, as given by history_query_command

${Output}
: absolute path of the destination file

The entry in the *aegis.conf* file looks like this:

```
history_get_command =
  "co -r'$e' -p $h,v > $o";
```

The "`co -r` option is used to specify the edit to be retrieved. The "`co -p` option is used to specify that the results be printed on the standard output; this is because the destination filename will *never* look anything like the history source filename.

4.5.3. history_put_command

This command is used to add a new "top-most" entry to the history file. This command is always executed as the project owner.

The following substitutions are available:

${Input}
: absolute path of source file

${History}
: absolute path of history file

The entry in the *aegis.conf* file looks like this:

```
history_put_command =
  "ci -u -d -M -m$c -t/dev/null \
  $i $h,v; rcs -U $h,v";
```

Uses ci to deposit a new revision, using -d and -M as described for history_create_command. The -m flag stores the change number in the file log, which allows *rlog*(1) to be used to find the Aegis change numbers to which each revision of the file corresponds.

The "`ci -u`" option is used to specify that an unlocked copy will remain in the baseline. The "`ci -d`" option is used to specify that the file time rather than the current time is to be used for the new revision. The "`ci -M`" option is used to specify that the mode date on the original file is

not to be altered. The "`ci -m`" option is used to specify that the change number is to be stored in the file log, which allows *rlog* to be used to find the change numbers to which each revision of the file corresponds. You might want to use `-m$p,$c` instead which stores both the project name and the change number. Or `-m$version`, which will be composed of the branch and the delta. These make it much easier to track changes across branches.

It is essential that the *history_create_command* and the *history_put_command* are identical. It is a historical accident that there are two separate commands: before Aegis supported branches, this was not a requirement.

4.5.4. history_query_command

This command is used to query what the history mechanism calls the top-most edit of a history file. The result may be any arbitrary string, it need not be anything like a number, just so long as it uniquely identifies the edit for use by the *history_get_command* at a later date. The edit number is to be printed on the standard output. This command is always executed as the project owner.

The following substitutions are available:

${History}
: absolute path of the history file

The entry in the *aegis.conf* file looks like this:

```
history_query_command =
  "rlog -r $h,v | "
  "awk '/^revision/ {print $$2}'";
```

4.5.5. merge_command

RCS also provides a *merge* program, which can be used to provide a three-way merge.

All of the command substitutions described in *aesub*(5) are available. In addition, the following substitutions are also available:

${ORiginal}
: The absolute path name of a file containing the version originally copied. Usually in a temporary file.

${Most_Recent}
: The absolute path name of a file containing the most recent version. Usually in the baseline.

${Input}
: The absolute path name of the edited version of the file. Usually in the development directory. Aegis usually moves the original source file aside, so that the output may have the source file's name.

${Output}
: The absolute path name of the file in which to write the difference listing. Usually in the development directory, usually the name of a change source file.

The entry in the *aegis.conf* file looks like this:

```
merge_command =
  "set +e; "
  "merge -p -L baseline -L C$c "
  " $mr $orig $in > $out; "
  "test $? -le 1";
```

The "`merge -L`" options are used to specify labels for the baseline and the development directory, respectively, when conflict lines are inserted into the result. The "`merge -p`" options is used to specify that the results are to be printed on the standard output.

It is important that this command does not move its input and output files around, otherwise this contradicts the warnings Aegis may issue to the user. (In previous versions of Aegis, this was necessary, however this is no longer the case.)

Warning: The version of *diff3*(1) available to RCS *merge*(1) has a huge impact on its performance and utility. You need to grab and install GNU diff to get the best results. *Unfortunately* the diff tool used by RCS *merge*(1) is determined at compile time. This means that you need to build and install GNU diff package *before* you build and install GNU RCS package.

4.5.6. Referential Integrity

Many history tools (including RCS) can modify the contents of the file when it is committed. While there are usually options to turn this off, they are seldom used. The problem is: if the commit changes the file, the source in the repository now no longer matches the object file in the repository − *i.e.* the history tool has compromised the referential integrity of the repository.

```
history_put_trashes_file = warn;
```

If you use RCS keyword substitution, you will need this line. (The default is to report a fatal error.)

Another reason for this option is that it tells Aegis it needs to recalculate the file's fingerprint after a checkin.

4.5.7. Templates

The *lib/config.example/rcs* file in the Aegis distribution (installed as */usr/local/share/aegis/config.example/rcs* by default) contains all of the above commands, so that you may readily insert them into your project configuration file.

Also, there are some subtleties to writing the commands, which are not present in the above examples. In particular, being able to support file names which contain characters which are special to the shell requires the use of the ${quote} substitution around all of the files names in the commands.

In addition, it is possible to have a much more useful description for the **-m** option. For example: "`-m${quote ($version) ${change description}}`" inserts the version number and the brief description into the file's log. This means that using the *rlog*(1) command will provide quite useful summaries.

4.5.8. Binary Files

RCS (version 5.6 and later) is able to cope with binary files. It does so by saving a whole copy of the file at each check-in.

If you want Aegis to transparently encode all such files, simply leave the *history_content_limitation* field unset.

If you want to check-in binary files, add the `-kb` option to each of the `rcs -U` commands in the fields above, and also set

```
history_content_limitation =
    binary_capable;
```

so that Aegis knows that no encoding is desired.

4.5.9. history_put_trashes_files

If you use RCS keywords, such as `$id$` or `$log$`, this will result in the file in the baseline being changed by RCS at integrate pass. This is *after* the build. The result is that the source files no longer match the object files. Oops.

While such mechanism are essential when using only a simple history tool, far more information may be obtained using the file history report (`aer file_history` *filename*), rendering such crude methods unnecessary.

In addition to expected expansions in file header comments, this can also be very destructive if, for example, such a string appeared in a uuencoded or MIME base 64 encoded file.

If you wish to prevent RCS from performing keyword expansion, used the `rcs -kb` option.

If, however, you wish to keep using keyword expansion, set

```
history_tool_trashes_file = warning;
```

to cause Aegis to warn you, rather than fail.

4.6. Using fhist

The *fhist* program was written by David I. Bell and is admirably suited to providing a history mechanism with out the "cruft" that SCCS and RCS impose.

Please note that the `[# edit #]` feature needs to be avoided, or the *−Forced_Update* (−fu) flag needs to be used in addition to the *−Conditional_Update* (−cu) flag, otherwise updates will complain that “Input file "*XXX*" contains edit *A* instead of *B* for module "*YYY*"”

The *history_create_command* and the *history_put_command* are intentionally identical. This minimizes problems when using branches.

4.6.1. history_create_command

This command is used to create a new project history. The command is always executed as the project owner.

The following substitutions are available:

${Input}
: absolute path of the source file

${History}
: absolute path of the history file

The entry in the *aegis.conf* file looks like this:

```
history_create_command =
  "fhist ${b $h} −create −cu "
  "−i $i −p ${d $h} −r";
```

Note that the source file is left in the baseline.

4.6.2. history_get_command

This command is used to get a specific edit back from history. The command may be executed by developers.

The following substitutions are available:

${History}
: absolute path of the history file

${Edit}
: edit number, as given by history_query_-command

${Output}
: absolute path of the destination file

The entry in the *aegis.conf* file looks like this:

```
history_get_command =
  "fhist ${b $h} −e '$e' −o $o "
  "−p ${d $h}";
```

Note that the destination filename will *never* look anything like the history source filename, so the −p is essential.

4.6.3. history_put_command

This command is used to add a new "top-most" entry to the history file. This command is always executed as the project owner.

The following substitutions are available:

${Input}
: absolute path of source file

${History}
: absolute path of history file

The entry in the *aegis.conf* file looks like this:

```
history_put_command =
  "fhist ${b $h} −create −cu "
  "−i $i −p ${d $h} −r";
```

Note that the source file is left in the baseline.

4.6.4. history_query_command

This command is used to query what the history mechanism calls the "top-most" edit of a history file. The result may be any arbitrary string, it need not be anything like a number, just so long as it uniquely identifies the edit for use by the *history_get_command* at a later date. The edit number is to be printed on the standard output. This command may be executed by developers.

The following substitutions are available:

${History}
: absolute path of the history file

The entry in the *aegis.conf* file looks like this:

```
history_query_command =
  "fhist ${b $h} −l 0 "
  "−p ${d $h} −q";
```

4.6.5. Templates

The *lib/config.example/fhist* file in the Aegis distribution (installed as */usr/local/share/aegis/config.example/fhist* by default) contains all of the above commands, so that you may readily insert them into your project configuration file.

4.6.6. Capabilities

By default, FHist is unable to cope will NUL characters in its input files, however this is the only limitation. By default, Aegis expects that history tools are only able to cope with printable ASCII text. To tell it ontherwise, set

```
history_content_limitation =
    international_text;
```

in the project *aegis.conf* file.

Aegis will transparently encode binary files (files which contain NUL characters) on entry and exit from the history tool. This means that you may have binary files in your project without configuring anything special.

4.6.7. Binary Files

FHist (version 1.7 and later) has support for binary files. The *fhist −binary* option may be used to specify that the file is binary, that it may contain NUL characters. It is essential that you have consistent presence or absence of the −binary option for each file when combined with the −CReate, −Update, −Conditional_Update and −Extract options. Failure to do so will produce inconsistent results.

This means that you have to *always* use the −binary option in the *history_create_command* and *history_put_command* fields. You have to decide right at the very beginning if your project *history* will ever have binary files, or will never have binary files. You can't change your mind later. If you choose to use the −binary option, set

```
history_content_limitation =
    binary_capable;
```

However, Aegis would transparently encode all such files, if you leave the *history_content_limitation* field set for international text. In some cases, Aegis' encoding will be more efficient than fhist's. And you have the advantage of being able to change your mind later.

4.7. Detecting History File Corruption

When you have files which exist for long periods of time, particularly files such as the ones typically used by history tools, which are generally appended to, without modification of the bulk of the file, there is a very real possibility that a block of the file could become corrupted over the years.[16] Unless you access the file versions contained within that block, you have no way of knowing whether or not the history file is OK. (Arguably, the operating system should check for this, but many do not, and in any case the error may not be detectable at that level.)

Using Aegis, you can add a simple checksum to your history files which will detect many cases of corruption such as this, for all of the commonly used history tools. Note: it cannot detect all corruptions (nothing can) but it will detect more than many operating systems will.

You don't need to use this technique with SCCS or *aesvt*(1), they already have checksums in their files.

4.7.1. General Method

In general, you need to do three things:

1. You need to create some kind of checksum of your history file each time you modify it. Something like *md5sum*(1) from the GNU Fileutils would be good. Store the checksum in a file next to the history file. This would be done in the *history_create_command* and *history_put_command* fields of the project *aegis.conf* file.
2. Each time the file is read, you need to verify the file's checksum. Use the same checksum utility as before, and then compare it using, say, *cmp*(1); it it fails (either an IO error, or the checksum doesn't compare equal) then don't proceed with the history file access. You may need to repair or replace the disk. You will need to restore from backup (yesterday's backup, see below). This would be done at the beginning of the *history_create_command*, *history_put_command*, *history_get_command* and *history_query_command* fields of the project *aegis.conf* file.
3. Because you may not actually interact with the file for years at a time, you need to check the file fingerprints much more often. Daily or at least weekly is suggested. You do this with a *cron*(1) job run nightly which compares all of the history files with their *md5sum*(1) checksums. Email failures to the system administrator and the project administrators. By doing this nightly, you not only avoid backing-up corrupted files, you will always know on which backup tape the good copy resides − yesterday's.

[16] See also Saltzer, J.H. *et al* (1981) *End-to-end arguments in system design*, http://web.mit.edu/-Saltzer/www/publications/endtoend/endtoend.pdf

4.7.2. Configuration Commands

In order to implement this, you need to modify some fields of your project *aegis.conf* file as follows:

history_create_command
: You need to test if the history file and its checksum file exist, and check the checksum if this is the case. Then, use whichever history tool you choose (see the previous sections of this chapter). If it succeeds, run *md5sum*(1) over the history file (*not* the source file) and store the checksum in a file next to the history tool's file. Using the same filename plus a `.md5sum` extension makes the *cron*(1) job easier to write.

history_put_command
: You need to test if the file exists (it may, for example, be an old project to which you have recently added this technique) and check the checksum if this is the case. Then, use your history tool as normal. If it succeeds, run *md5sum*(1) over the history file (*not* the source file) as in the create case.

history_get_command
: You need to test if the file exists (it may, for example, be an old project to which you have recently added this technique) and check the checksum if this is the case. Then use your history tool as normal.

history_head_command
: This command is only used at *aeipass* file, immediately after one of the *history_create_command* or *history_put_command* commands. It is up to you whether you think you need to add a guard as for the *history_get_command field.*

4.7.3. An Alternative

Rather than run *md5sum*(1) on the history files each time you modify them, you could use *gzip*(1) to obtain some minor compression, but it also provides and Adler32 checksum of the file. For files with long histories, this can be tedious to

unpack every time you need to extract an old version, but such operations are frequently I/O bound, and so there may be no perceived slowness by the user..

4.7.4. Aegis' Database

In addition to your history files, Aegis maintains a database of file meta-data. In order to add a checksum to the various file making up the database, turn on the *compressed_database* project attribute. In addition to compressing the database (a minor savings) it also adds an Adler32 checksum.

You can check this in the *cron*(1) job by using *gzcat*(1) sent to */dev/null*.

5. The Dependency Maintenance Tool

Aegis can place heavy demands on the dependency maintenance tool, so it is important that you select an appropriate one. This chapter talks about what features a dependency maintenance tool requires, and gives examples of how to use the various alternatives.

5.1. Required Features

The heart of any DMT is an *inference engine*. This inference engine accepts a *goal* of what you want it to construct and a set of *rules* for how to construct things, and attempts to construct what you asked for given the rules you specified. This is exactly a description of an expert system, and the DMT needs to be an expert system for constructing files. Something like PROLOG is probably ideal.

Aegis is capable of supporting a wide variety of development directory styles. The different development directory styles place different demands on the dependency maintenance tool. Development directory styles will be described in the next section, but here is a quick summary:

copy of all sources:
: This is what CVS does, and what many other VC tool do. Because you have a complete copy of all source files, the dependency maintenance tool only needs to be aware of one directory tree.

copy of everything:
: This is a small optimization of the previous case to cut down the time required for that first build,because the derived files from the integration build can be reused.

link all sources
: The is an optimization of the "copy all sources" case, because linking a file is significantly faster than making a copy of a file. The dependency maintenance tool only needs to be aware of one directory tree.

link everything
: This is an optimization of the previous case, again reusing derived files from the integration build, except that you need to ensure that your dependency maintenance tool is configured to remove the derived file outputs of each rule before creating them, to avoid corrupting the baseline or getting "permission denied" error.

view path
: This is the most efficient development directory style, and it scales much better than any of the above, but the dependency maintenance tool must be able to cope with a hierarchy of parallel source directory trees. These trees for a "view path", a list of directories that programs search below to find the files of interest. The `vpath` statements of GNU Make are almost, but not quite, capable of being used in this way.

5.1.1. View Paths

For the union of all files in a project and all files in a change (remembering that a change only copies those files it is modifying, plus it may add or remove files) for all files you must be able to say to the dependency maintenance tool,

> "If and only if the file is up-to-date in the baseline, use the baseline copy of the file, otherwise construct the file in the development directory".

The presence of a source file in the change makes the copy in the baseline out-of-date.

Most DMTs with this capability implement it by using some sort of search path, allowing a hierarchy of directories to be scanned with little or no modification to the rules.

If your DMT of choice does not provide this functionality, the *development_directory_style.source_file_symlink* field of the project configuration file may be set to *true*, which tells Aegis to maintain symbolic links in the development directory for all source files in the baseline which are not present in the development directory. (See *aepconf*(5) and *aeb*(1) for more information.) This incurs a certain amount of overhead when Aegis maintains these links, but a similar amount of work is done within DMTs which have search path functionality.

5.1.2. Dynamic Include File Dependencies

Include file dependencies are very important, because a change may alter an include file, and all of the sources in the baseline which use that include file must be recompiled.

Consider the example given earlier: the include file describing the interface definition of a

function is copied into a change and edited, and so is the source file defining the function. It is essential that all source files in the baseline which include that file are recompiled, which will usually result in suitable diagnostic errors if any of the clients of the altered function have yet to be included in the change.

There are two ways of handling include file dependencies:

• They can be kept in a file, and the file can be maintained by suitable programs (maintaining it manually never works, that's just human nature).

• They can be determined by the DMT when it is scanning the rules to determine what needs updating.

5.1.2.1. Static File

Keeping include dependencies in a file has a number of advantages:

• Most existing DMTs have the ability to include other rules files, so that when performing a development build from a baseline rules file, it could include a dependencies file in the development directory.

• Reading a file is much faster than scanning all of the source files.

Keeping include dependencies in a file has a number of disadvantages:

• The file is independent of the DMT, it is either generated before the DMT is invoked, in which case it may do more work than is necessary, or it may be invoked after the DMT (or after the DMT has scanned its rules), in which case it may well be out-of-date when the DMT needs it.

For example, the use of *gcc −M* produces "dot d" files, which may be merged to construct such an includable dependency file. This happens after the DMT has read and applied the rules, but possibly before the DMT has finished executing.[17]

• Many tools which can generate this information, such as the *gcc −M* option, are triggered by source files, and are unable to manage a case where it is an include file which is changing, to include a different set of other include files. In this case, the inaccurate dependencies file may contain references to the old set of nested include files, some of which may no longer exist, This causes the DMT to incorrectly generate an error stating that the old include file is missing, when it is actually no longer required.

If a DMT can only support this kind of include file dependencies, it is not suitable for use with Aegis.

5.1.2.2. Dynamic

In order for a DMT to be suitable for use with Aegis, it is essential that rules for the DMT may be specified in such a way that include file dependencies are determined "on the fly" when the DMT is determining if a given rule is applicable, and before the rule is applied.

This method suffers from the problem being rather slow; but this is amenable to some caching and the losses of performance are not as bad as could be imagined.

This method has the advantage of correctness in all cases, where a static file may at times be out-of-date.

5.2. Development Directory Style

The project configuration file, usually called *aegis.conf*, contains a field called *development_directory_style* which controls how the project sources are presented to the DMT.

See *aepconf*(5) for a complete description of this field.

There is a corresponding *integration_directory_style* field, which defaults to the same value as the *development_directory_style*. It is usually a very bad idea if these two are different.

5.2.1. View Path

By not setting *development_directory_style* at all the only source files present in the development directory are source files being create and/or modified.

By using information provided by the *$search_path* substitution, the build can access the unchanged source files in the branch baseline and deeper branch baselines. The great thing about this approach is that there are also "precompiled" object files on the viewpath, so if an object file does not need to be compiled (there are no soure files in the development directory that have anything to do with it) then the build can simply link the unchanged object files in the baseline without recompiling.

[17] See the *Using Make* section for how GNU Make may be used. It effectively combines both methods: keeping *.d* files and dynamically updating them. Because it combines both methods, it has some of the advantages and disadvantages of both.

This build method scales the best, and is the Aegis default.

The difficulties of finding a DMT which is capable of coping with a view path means that this is not the only work area style. All other methods scale less well than a view path; some scale *much* less well.

5.2.2. Link the Baseline

The first two sub-fields of interest in the *development_directory_style* are *source_file_link* and *source_file_symlink*.

source_file_link = true; This field is true if hard links are to be used for project source files (which are not part of the change) so that the work area has a complete set of source files.

source_file_symlink = true; This field is true if symbolic links are to be used for project source files (which are not part of the change) so that the work area has a complete set of source files.

By using these settings, all source files are present in the development directory. They will be read-only. As you decide to modify files in the change set, the *aecp* command will remove the link and replace it with a read-write copy of the file.

You need both these sub-fields set, because hard links are not allowed to cross file system boundaries. Aegis will use hard links in preference to soft links when it can.

Maintaining the hard links can be time consuming for large projects, and add quite a noticeable delay before builds start doing anything. But see the *−assume-symbolic-links* option of the *aeb*(1) command; use it sparingly.

The biggest penalty with this method is that the initial build for a change set for a large project can be *very* time consuming. Recall that the baseline has a complete "prebuild" already available. To take advantage of these pre-built derived files, there are a few more sub-fields:

derived_file_copy = true; This field is true if copies are to be used for non-source files which are present in the project baseline but which are not present in the work area, so that the work area has a complete set of derived files.

derived_at_start_only = true; This settign causes the above fields controling the appearance of derived files to be acted upon only when the development directory is created (at *aedb*(1) time).

Copying files can be very time consuming and also eats a lot of disk space. If you are prepared to change your build slightly, it is possible to use the following fields:

derived_file_link = true; This field is true if hard links are to be used for non-source files which are present in the project baseline but which are not present in the work area, so that the work area has a complete set of derived files.

derived_file_symlink = true; This field is true if symbolic links are to be used for non-source files which are present in the project baseline but which are not present in the work area, so that the work area has a complete set of derived files.

Just as for source files, hard links will be used in preference to symbolic links if possible.

Note that *every* rule in your Makefile (or whatever your DMT uses) **must** remove its outputs before doing enything else, to break the links to the files in the baseline, otherwise you will corrupt the baseline. Aegis tries very hard to ensure that the baseline files (and thus the links) are read-only, so that you get an error from the build if you forget to break a link.

This development directory style is called "arch style" after Tom Lord's *arch* (tla) which does something very similar.

If you are placing an existing project under Aegis, do the above three things one step at a time. First get the source files available and integrate that. In a second change set get derived file copies working. In a third change set (if you do it at all) change the build and use derived file links.

5.2.3. Copy All Sources

The sub-fields of interest in the *development_directory_style* is *source_file_copy*.

source_file_copy = true; This field says copies are to be used for project source files (which are not part of the change) so that the work area has a complete set of source files. File modification time attributes will be preserved.

By using this setting, all source files are present in the development directory. They will be read-only. As you decide to modify files in the change set, the *aecp* command will remove the file and

replace it with a read-write copy of the file.

Maintaining the copies can be time consuming for large projects, and add quite a noticeable delay before builds start doing anything. But see the *−assume-symbolic-links* option of the *aeb*(1) command; use it sparingly (yes, it applies to copies as well).

The biggest penalty with this method is that the initial build for a change set for a large project can be *very* time consuming. Recall that the baseline has a complete "prebuild" already available. To take advantage of these pre-built derived files, there are a few more sub-fields:

derived_file_copy = true; This says copies are to be used for non-source files which are present in the project baseline but which are not present in the work area, so that the work area has a complete set of derived files.

derived_at_start_only = true; This setting causes the above fields controlling the appearance of derived files to be acted upon only when the development directory is created (at *aedb*(1) time).

This development directory style is called "CVS style" after GNU CVS which does something very similar.

5.2.4. Obsolete Features

There are several fields in the *aegis.conf* file which are obsolete. Aegis will automatically transfer these to create a *development_directory_style* if you haven't specified one.

create_symlinks_before_build: This is like setting both *development_directory_style.source_file_symlink* and *development_directory_style.derived_file_symlink* at the same time.

remove_symlinks_after_build: This is like setting the *development_directory_style.during_build_only* field.

create_symlinks_before_integration_build: This is like setting both *integration_directory_style.source_file_symlink* and *integration_directory_style.derived_file_symlink* at the same time.

remove_symlinks_after_integration_build: This is like setting the *integration_directory_style.during_build_only* field.

Aegis will print a warning if you use any of these fields.

5.3. Using Cook

The *Cook* program is the only dependency maintenance tool, known to the author, which is sufficiently capable to supply Aegis' needs.[18] Tools such as *cake* and *GNU Make* are described later. They need a special tweak to make them work.

This section describes appropriate contents for the *Howto.cook* file, input to the *cook*(1) program. It also discusses the *build_command* and *integrate_build_command* and *link_baseline* and *change_file_command* and *project_file_command* and *link_integration_directory* fields of the configuration file. See *aepconf*(5) for more information about this file.

5.3.1. Invoking Cook

The *build_command* field of the configuration file is used to invoke the relevant build command. In this case, it is set as follows

```
build_command =
  "cook -b ${s Howto.cook} -nl\
   project=$p change=$c version=$v";
```

This command tells Cook where to find the recipes. The `${s Howto.cook}` expands to a path into the baseline during development if the file is not in the change. Look in *aesub*(5) for more information about command substitutions.

The recipes which follow will all remove their targets before constructing them, which qualifies them for the next entry in the configuration file:

```
link_integration_directory = true;
```

The links must be removed first, otherwise the baseline would cease to be self-consistent.

5.3.2. The Recipe File

The file containing the recipes is called *Howto.cook* and is given to Cook on the command line.

The following items are preamble to the rest of the file; they ask Aegis for the source files of the project and change so that Cook can determine what needs to be compiled and linked.

```
project_files =
  [collect_lines aelpf
    -p [project] -c [change]];
change_files =
  [collect_lines aelcf
    -p [project] -c [change]];
source_files =
  [stringset [project_files]
    [change_files]];
```

This example continues the one from chapter 3, and thus has a single executable to be linked from all the object files

```
object_files =
  [fromto %.y %.o [match_mask %.y
    [source_files]]]
  [fromto %.l %.o [match_mask %.l
    [source_files]]]
  [fromto %.c %.o [match_mask %.c
    [source_files]]]
  ;
```

It is necessary to determine if this is a development build, and thus has the baseline for additional ingredients searches, or an integration build, which does not. The version supplied by Aegis will tell us this information, because it will be *major.minor.*C*change* for development builds and *major.minor.*D*delta* for integration builds.

```
if [match_mask %1C%2 [version]] then
{
  baseline = [collect aegis -cd -bl
    -p [project]];
  search_list = . [baseline];
}
```

The *search_list* variable in Cook is the list of directories to search for dependencies; it defaults to only the current directory. The *resolve* builtin function of Cook may be used to ask Cook for the name of the file actually used to resolve dependencies, so that recipe bodies may reference the appropriate file:

```
example: [object_files]
{
  [cc] -o example
    [resolve [object_files]]
    -ly -ll;
}
```

This recipe says that to Cook the example program, you need the object files determined earlier, and them link them together. Object files which were up to date in the baseline are used wherever possible, but files which were out of date are constructed in the current directory and those will be linked.

[18] The version in use when writing this section was 1.5. All versions from 1.3 onwards are known to work with the recipes described here.

5.3.3. The Recipe for C

Next we need to tell Cook how to manage C sources. On the surface, this is a simple recipe:

```
%.o: %.c
{
  rm %.o;
  [cc] [cc_flags] -c %.c;
}
```

Unfortunately it has forgotten about finding the include file dependencies. The Cook package includes a program called *c_incl* which is used to find them. The recipe now becomes

```
%.o: %.c: [collect c_incl -eia %.c]
{
  rm %.o;
  [cc] [cc_flags] -c %.c;
}
```

The file may not always be present to be removed (causing a fatal error), and it is irritating to execute a redundant command, so the remove is mangled to look like this:

```
%.o: %.c: [collect c_incl -eia %.c]
{
  if [exists %.o] then
    rm %.o
      set clearstat;
  [cc] [cc_flags] -c %.c;
}
```

The "set clearstat" clause tells Cook that the command will invalidate parts of its *stat* cache, and to look at the command for what to invalidate.

Another thing this recipe needs is to use the baseline for include files not in a change, and so the recipe is altered again:

```
%.o: %.c: [collect c_incl -eia
  [prepost "-I" "" [search_list]]
    %.c]
{
 if [exists %.o] then
    rm %.o
      set clearstat;
  [cc] [cc_flags] [prepost "-I" ""
    [search_list]] -c %.c;
}
```

See the *Cook Reference Manual* for a description of the *prepost* builtin function, and other Cook details.

There is one last change that must be made to this recipe, it must use the resolve function to reference the appropriate file once Cook has found it on the search list:

```
%.o: %.c: [collect c_incl -eia
  [prepost "-I" "" [search_list]]
    [resolve %.c]]
{
  if [exists %.o] then
    rm %.o
      set clearstat;
  [cc] [cc_flags] [prepost "-I" ""
    [search_list]] -c [resolve %.c];
}
```

Only use this last recipe for C sources, the others are only shown so that the derivation of the recipe is clear; while it is very similar to the original, it looks daunting at first.

5.3.3.1. C Include Semantics

The semantics of C include directives make the

```
#include "filename"
```

directive dangerous in a project developed with the Aegis program and Cook.

Depending on the age of your compiler, whether it is AT&T traditional C or newer ANSI C, this form of directive will search first in the current directory and then along the search path, or in the directory of the including file and then along the search path.

The first case is fairly benign, except that compilers are rapidly becoming ANSI C compliant, and an operating system upgrade could result in a nasty surprise.

The second case is bad news. If the source file is in the baseline and the include file is in the change, you don’t want the source file to use the include file in the baseline.

Always use the

```
#include <filename>
```

form of the include directive, and set the include search path explicitly on the command line used by Cook.

Cook is able to dynamically adapt to include file dependencies, because they are not static. The presence of an include file in a change means that any file which includes this include file, whether that source file is in the baseline or in the change, must have a dependency on the change’s include file. Potentially, files in the baseline will need to be recompiled, and the object file stored in the change, not the baseline. Subsequent linking needs to pick up the object file in the change, not from the baseline.

5.3.4. The Recipe for Yacc

Having explained the complexities of the recipes in the above section about C, the recipe for yacc will be given without delay:

```
%.c %.h: %.y
{
  if [exists %.c] then
    rm %.c
      set clearstat;
  if [exists %.h] then
    rm %.h
      set clearstat;
  [yacc] [yacc_flags] -d
    [resolve %.y];
  mv y.tab.c %.c;
  mv y.tab.h %.h;
}
```

This recipe could be jazzed up to cope with the listing file, too, if that was desired, but this is sufficient to work with the example.

Cook's ability to cope with transitive dependencies will pick up the generated .c file and construct the necessary .o file.

5.3.5. The Recipe for Lex

The recipe for lex is vary similar to the recipe for yacc.

```
%.c: %.l
{
  if [exists %.c] then
    rm %.c
      set clearstat;
  [lex] [lex_flags] -d [resolve %.l];
  mv lex.yy.c %.c;
}
```

Cook's ability to cope with transitive dependencies will pick up the generated .c file and construct the necessary .o file.

5.3.6. Recipes for Documents

You can format documents, such as user guides and manual entries with Aegis and Cook, and the recipes are similar to the ones above.

```
%.ps: %.ms: [collect c_incl -r -eia
  [prepost "-I" "" [search_list]]
  [resolve %.ms]]
{
  if [exists %.ps] then
    rm %.ps
      set clearstat;
  roffpp [prepost "-I" ""
    [search_list]] [resolve %.ms]
    | groff -p -t -ms
    > [target];
}
```

This recipe says to run the document through groff, with the *pic*(1) and *tbl*(1) filters, use the *ms*(7) macro package, to produce PostScript output. The *roffpp* program comes with Cook, and is like *soelim*(1) but it accepts include search path options on the command line.

Manual entries may be handled in a similar way

```
%.cat: %.man: [collect c_incl -r -eia
  [prepost "-I" "" [search_list]]
  [resolve %.man]]
{
  if [exists %.cat] then
    rm %.cat
      set clearstat;
  roffpp [prepost "-I" ""
    [search_list]] [resolve %.man]
    | groff -Tascii -t -man
    > [target];
}
```

5.3.7. Templates

The *lib/config.example/cook* file in the Aegis distribution contains all of the above commands, so that you may readily insert them into your project configuration file.

5.4. Using Cake

This section describes how to use *cake* as the dependency maintenance tool. The *cake* package was published in the *comp.sources.unix* USENET newsgroup volume 12, around February 1988, and is thus easily accessible from the many archives around the internet.

It does not have a search path of any form, not even something like *VPATH*. It does, however, have facilities for dynamic include file dependencies.

5.4.1. Invoking Cake

The *build_command* field of the configuration file is used to invoke the relevant build command. In this case, it is set as follows

```
build_command =
  "cake -f ${s Cakefile} \
   -DPROJECT=$p -DCHANGE=$c \
   -DVERSION=$v";
```

This command tells *cake* where to find the rules. The `${s Cakefile}` expands to a path into the baseline during development if the file is not in the change. Look in *aesub*(5) for more information about command substitutions.

The rules which follow will all remove their targets before constructing them, which qualifies them for the next entry in the configuration file:

```
link_integration_directory = true;
```

The links must be removed first, otherwise the baseline would be corrupted by integration builds.

Another field to be set in this file is

```
development_directory_style =
{
    source_file_symlink = true;
};
```

which tells Aegis to maintain symbolic links between the development directory and the baseline. This also requires that rules remove their targets before constructing them, to ensure that rules do not attempt to write their results onto the read-only versions in the baseline.

5.4.2. The Rules File

The file containing the rules is called *Cakefile* and is given to cake on the command line.

The following items are preamble to the rest of the file; they ask Aegis for the source files of the project and change so that cake can determine what needs to be compiled and linked.

```
#define project_files \
  [[aelpf -p PROJECT \
    -c CHANGE]];
#define change_files \
  [[aelcf -p PROJECT \
   -c CHANGE]];
#define source_files \
  project_files change_files

#define CC     gcc
#define CFLAGS -O
```

This example parallels the one from chapter 3, and thus has a single executable to be linked from all the object files

```
#define object_files \
  [[sub -i X.c %.o source_files]] \
  [[sub -i X.y %.o source_files]] \
  [[sub -i X.l %.o source_files]]
```

Constructing the program is straightforward

```
example: object_files
  rm -f example
  CC -o example object_files
```

This rule says that to construct the example program, you need the object files determined earlier, and them link them together. Object files which were up to date in the baseline are used wherever possible, but files which were out of date are constructed in the current directory and those will be linked.

5.4.3. The Rule for C

Next we need to tell cake how to manage C sources. On the surface, this is a simple rule:

```
%.o: %.c
  CC CFLAGS -c %.c
```

paralleling that found in most makes, however it needs to delete the target first, and to avoid deleting the *.o* file whenever cake thinks it is transitive.

```
%.o!: %.c
  rm -f %.o
  CC CFLAGS -c %.c
```

The *-f* option to the *rm* command is because the file does not always exist.

Unfortunately this rule omits finding the include file dependencies. The cake package includes a program called *ccincl* which is used to find them. The rule now becomes

```
%.o!: %.c* [[ccincl %.c]]
  rm -f %.o
  CC CFLAGS -c %.c
```

This rule is a little quirky about include files which do not yet exist, but must be constructed by

some other rule. You may want to use *gcc −MM* instead, which is almost as quirky when used with cake. Another alternative, used by the author with far more success, is to use the *c_incl* program from the *cook* package, mentioned in an earlier section. The *gcc −MM* understands C include semantics perfectly, the *c_incl* command caches its results and thus goes faster, so you will need to figure which you most want.

5.4.3.1. Include Directives

Unlike *cook* described in an earlier section, using *cake* as described here allows you to continue using the

```
#include "filename"
```

form of the include directive. This is because the development directory appears, to the compiler, to be a complete copy of the baseline.

5.4.4. The Rule for Yacc

Having explained the complexities of the rules in the above section about C, the rule for yacc will be given without delay:

```
#define YACC yacc
#define YFLAGS

%.c! %.h!: %.y  if exist %.y
  rm −f %.c %.h y.tab.c y.tab.h
  YACC YFLAGS −d %.y
  mv y.tab.c %.c
  mv y.tab.h %.h
```

This rule could be jazzed up to cope with the listing file, too, if that was desired, but this is sufficient to work with the example.

Cake's ability to cope with transitive dependencies will pick up the generated *.c* file and construct the necessary *.o* file.

5.4.5. The Rule for Lex

The rule for lex is vary similar to the rule for yacc.

```
#define LEX lex
#define LFLAGS

%.c!: %.l  if exist %.l
  rm −f %.c
  LEX LFLAGS %.l
  mv lex.yy.c %.c
```

Cake's ability to cope with transitive dependencies will pick up the generated *.c* file and construct the necessary *.o* file.

5.4.6. Rules for Documents

You can format documents, such as user guides and manual entries with Aegis and cake, and the rules are similar to the ones above.

```
%.ps!: %.ms* [[soincl %.ms]]
  rm −f %.ps
  groff −s −p −t −ms %.ms > %.ps
```

This rule says to run the document through groff, with the *soelim*(1) and *pic*(1) and *tbl*(1) filters, use the *ms*(7) macro package, to produce PostScript output.

This suffers from many of the problems with include files which need to be generated, as does the C rule, above. You may want to use *c_incl −r* from the *cook* package, rather than the *soincl* supplied by the *cake* package.

Manual entries may be handled in a similar way

```
%.cat!: %.man* [[soincl %.man]]
  rm −f %.cat
  groff −Tascii −s −t −man %.man \
    > %.cat
```

5.5. Using Make

The *make*(1) program exists in many forms, usually one is available with each UNIX version. The one used in the writing of this section is *GNU Make 3.70*, available by anonymous FTP from your nearest GNU archive site. GNU Make was chosen because it was the most powerful, it is widely available (usually for little or no cost) and discussion of the alternatives (SunOS make, BSD 4.3 make, etc), would not be universally applicable. "Plain vanilla" make (with no transitive closure, no pattern rules, no functions) is not sufficiently capable to satisfy the demands placed on it by Aegis.

With the introduction of the *development_directory_style* field of the project configuration file, any project which is currently using a "plain vanilla" make may continue to use it, and still manage the project using Aegis.

As mentioned earlier in this chapter, *make* is not really sufficient, because it lacks dynamic include dependencies. However, GNU Make has a form of dynamic include dependencies, and it has a few quirks, but mostly works well.

The other feature lacking in *make* is a search path. While GNU Make has functionality called *VPATH*, the implementation leaves something to be desired, and can't be used for the search path functionality required by Aegis. Because of this, the *development_directory_style.source_file_symlink* field of the project configuration file is set to *true* so that Aegis will arrange for the development directory to be full of symbolic links, making it appear that the entire project source is in each change's development directory.

5.5.1. Invoking Make

The *build_command* field of the project configuration file is used to invoke the relevant build command. In this case, it is set as follows

```
build_command =
  "gmake -f ${s Makefile} project=$p \
   change=$c version=$v";
```

This command tells make where to find the rules. The `${s Makefile}` expands to a path into the baseline during development if the file is not in the change. Look in *aesub*(5) for more information about command substitutions.

The rules which follow will all remove their targets before constructing them, which qualifies them for the next entry in the configuration file:

```
link_integration_directory = true;
```

The files must be removed first, otherwise the baseline would be corrupted by integration builds (or even by developer builds, if your aren't using a separate user for the project owner).

Note: if you are migrating an existing project **do not set this field**; only set it *after* you have changed *all* of the Make rules. If in doubt, **don't** set this field.

Another field to be set in this file is

```
development_directory_style =
{
    source_file_symlink = true;
};
```

which tells Aegis to maintain symbolic links between the development directory and the baseline for source files (but not derived files). See *aepconf*(5) for more information.

5.5.2. The Rule File

The file containing the rules is called *Makefile* and is given to make on the command line.

The following items are preamble to the rest of the file; they ask Aegis for the source files of the project and change so that make can determine what needs to be compiled and linked.

```
project_files := \
  $(shell aelpf -p $(project) \
    -c $(change))
change_files := \
  $(shell aelcf -p $(project) \
    -c $(change))
source_files := \
  $(sort $(project_files) \
    $(change_files))
CC := gcc
CFLAGS := -O
```

This example parallels the one from chapter 3, and thus has a single executable to be linked from all the object files

```
object_files := \
  $(patsubst %.y,%.o,$(filter \
    %.y,$(source_files)))     \
  $(patsubst %.l,%.o,$(filter \
    %.l,$(source_files)))     \
  $(patsubst %.c,%.o,$(filter \
    %.c,$(source_files)))
```

Constructing the program is straightforward, remembering to remove the target first.

```
example: $(object_files)
  rm -f example
  $(CC) -o example $(object_files) \
    -ly -ll
```

This rule says that to make the example program, you need the object files determined earlier, and them link them together. Object files which were up to date in the baseline are used wherever possible, but files which were out of date are constructed in the current directory and those will be linked.

5.5.3. The Rule for C

Next we need to tell make how to manage C sources. On the surface, this is a simple rule:

```
%.o: %.c
  $(CC) $(CFLAGS) -c $*.c
```

This example matches the built-in rule for most *make*s. But it forgets to remove the target before constructing it.

```
%.o: %.c
  rm -f $*.o
  $(CC) $(CFLAGS) -c $*.c
```

The target may not yet exist, hence the *-f* option.

Something missing from this rule is finding the include file dependencies. The GNU Make User Guide describes a method for obtaining include file dependencies. A set of dependency files are constructed, one per *.c* file.

```
%.d: %.c
  rm -f %.d
  $(CC) $(CFLAGS) -MM $*.c \
  | sed 's/^\(.*\).o :/\1.o \1.d :/' \
  > $*.d
```

These dependency files are then included into the *Makefile* to inform GNU Make of the dependencies.

```
include $(patsubst \
  %.o,%.d,$(object_files))
```

GNU Make has the property of making sure all its include files are up-to-date. If any are not, they are made, and then GNU Make starts over, and re-reads the Makefile and the include files from scratch, before proceeding with the operation requested. In this case, it means that our dependency construction rule will be applied before any of the sources are constructed.

This method is occasionally quirky about absent include files which you have yet to write, or which are generated and don't yet exist, but this is usually easily corrected, though you do need to watch out for things which will stall an integration.

The *-MM* option to the $(CC) command means that this rule requires the *gcc* program in order to work correctly. It may be possible to use *c_incl*(1) from cook, or *ccincl*(1) from cake to build the dependency lists instead; but they don't understand the conditional preprocessing as well as *gcc* does.

This method also suffers when heterogeneous development is performed. If you include different files, depending on the environment being compiled within, the *.d* files may be incorrect, and GNU Make has no way of knowing this.

5.5.3.1. Include Directives

Unlike *cook* described in an earlier section, using GNU Make as described here allows you to continue using the

```
#include "filename"
```

form of the include directive. This is because the development directory appears, to the compiler, to be a complete copy of the baseline.

5.5.4. The Rule for Yacc

Having explained the complexities of the rules in the above section about C, the rule for yacc will be given without delay:

```
%.c %.h: %.y
  rm -f $*.c $*.h y.tab.c y.tab.h
  $(YACC) $(YFLAGS) -d $*.y
  mv y.tab.c $*.c
  mv y.tab.h $*.h
```

This rule could be jazzed up to cope with the listing file, too, if that was desired, but this is sufficient to work with the example.

GNU Make's ability to cope with transitive closure will pick up the generated *.c* file and construct the necessary *.o* file.

To prevent GNU Make throwing away the transitive files, and thus slowing things down in some cases, make them precious:

```
.PRECIOUS: \
  $(patsubst %.y,%.c,$(filter \
    %.y,$(source_files)))     \
  $(patsubst %.y,%.h,$(filter \
    %.y,$(source_files)))
```

5.5.5. The Rule for Lex

The rule for lex is vary similar to the rule for yacc.

```
%.c: %.l
  rm -f $*.c lex.yy.c
  $(LEX) $(LFLAGS) $*.l
  mv lex.yy.c $*.c
```

GNU Make's ability to cope with transitive closure will pick up the generated *.c* file and construct the necessary *.o* file.

To prevent GNU Make throwing away the transitive files, and thus slowing things down in some cases, make them precious:

```
.PRECIOUS: \
  $(patsubst %.l,%.c,$(filter \
    %.l,$(source_files)))
```

5.5.6. Rules for Documents

You can format documents, such as user guides and manual entries with Aegis and GNU Make, and the rules are similar to the ones above.

```
%.ps: %.ms
  rm -f $*.ps
  groff -p -t -ms $*.ms > $*.ps
```

This rule says to run the document through groff, with the *pic*(1) and *tbl*(1) filters, use the *ms*(7) macro package, to produce PostScript output.

This omits include file dependencies. If this is important to you, the *c_incl* program from *cook* can be used to find them. Filtering its output can then produce the necessary dependency files to be included, rather like the C rules, above.

Manual entries may be handled in a similar way

```
%.cat: %.man
  rm $*.cat
  groff -Tascii -s -t -man $*.man \
    > $*.cat
```

5.5.7. Other Makes

All of the above discussion assumes that GNU Make and GCC are used. If you do not want to do this, or may not do this because of internal company politics, it is possible to perform all of the automated features manually.

This may, however, rapidly become spectacularly tedious. For example: if a user needs to copy the *Makefile* into their change for any reason, they will need to constantly use *aed*(1) to "catch up" with integrations into the baseline.

Reviewers are also affected: they must check that each change to the *Makefile* accurately reflects the object list and the dependencies of each source file.

5.5.8. Templates

The *lib/config.example/make* file in the Aegis distribution contains all of the above commands, so that you may readily insert them into your project configuration file.

5.5.9. GNU Make VPATH Patch

Version 3.76 and later of GNU Make include this patch, so you don't need to read this section unless you have GNU Make 3.75 or earlier.

There is a patch available for GNU Make 3.75 and earlier which gives it improved VPATH semantics. At the time it was not maintained by the same person who maintained GNU Make. Since then, the maintaier changed, and the patch has been incorporated.

The patch is the work of Paul D. Smith <psmith@BayNetworks.com> and may be fetched By Anonymous FTP from

Host: ftp.wellfleet.com
Dir: /netman/psmith/gmake
File: vpath+.README
File: vpath+.patch.*version*

The version numbers track the GNU Make version numbers.

For a description of the VPATH problem, and how this patch addresses it, see the README file referenced.

5.5.10. GNU Make's VPATH+

In theory, using GNU Make 3.76 or later (or a suitable patched earlier version) is similar to using Cook. The project configuration file now requires

```
link_integration_directory = false;
```

which is the default. The *Makefile* now requires

```
VPATH . bl
```

Assuming that *bl* is a symbolic link to the baseline. The `.d` files continue to be used.

5.6. Building Executable Scripts

Aegis treats source files as, well, source files. This means that it forgets any executable bits (and any other mode bits) you may set on the file. Usually, this isn't a problem − except for scripts.

So, just how *do* you get Aegis to give you an executable script? Well, you add a build rule. However, since it can't depend on itself, it needs to depend on something else.

Using a Cook example, we could write

```
bin/%: script/%.sh
{
    /* copy the script */
    cp script/%.sh bin/%;
    /* make it executable */
    chmod a+rx bin/%;
    /* syntax check */
    bash −n bin/%;
}
```

There is a small amout of value-added here: we did a syntax check along the way, which catches all sorts of problems.

The same technique also works for Perl

```
bin/%: script/%.pl
{
    cp script/%.pl bin/%;
    chmod a+rx bin/%;
    perl −cw bin/%;
}
```

The same technique also works for TCL

```
bin/%: script/%.tcl
{
    cp script/%.rcl bin/%;
    chmod a+rx bin/%;
    procheck −nologo bin/%;
}
```

The *procheck*(1) command is part of the TclPro package.

Many tools have a similar options.

You can also combine this with GNU Autoconf to produce architecture specific shell scripts from architecture neutral sources.

/* vim: set ts=8 sw=4 et : */

5.7. GNU Autoconf

If your projects uses GNU Make, GNU Autoconf and GNU Automake, here is a quick and simple method to import your project into Aegis and have it running fairly quickly.

5.7.1. The Sources

Once you have create and Aegis project to, your first change set should simply contain all of the source files, without removing or adding anything. The only additonal file is the Aegis project configuration file, usually called *aegis.conf* and usually located in the top-level directory.

Follow the directions in the section, above, on *using Make* for how to fill out this file.

Note that if you are working from a tarball, they usually contain several *derived* file. That is, files which are not primary sourec files, but are instead derived from other files. This is a convenience for the end-user but a nuisance at this point. Exanple of derived files in source tarballs include `configure`, `Makefile.in`, `config.h.in`, *etc*. You will need to exclude them form the first change set.

In this first change set, you don't even try to build anything.

```
build_command = "exit 0";
```

Which will allow the Aegis process to complete.

5.7.2. Building

You actually get your project to buld in the second change set. Once you have started development, you will see all of the source files in the development directory (well, symlinks to them).

In order to get you build to work, you have to bootstrap the `Makefile`. Using the usual GNU tool chain, this file is generated from `Makefile.in` which is in turn generated from `Makefile.am`, and this is not presently in the development directory.

This is done by creating a new primary source file called `makefile` at the top level

```
$ aenf makefile
$
```

and setting its contents to be

```
include Makefile

ifndef srcdir

bogus−default−target: Makefile
        $(MAKE) $(MAKEFLAGS) $(MAKECMDGOALS)

Makefile: configure Makefile.in config.h.in
        rm −f config.cache
        ./configure
```

```
configure: configure.ac
        autoconf

config.h.in: configure.ac
        autoheader

Makefile.in: Makefile.am
        automake

endif
```

This works because *make*(1) looks for `makefile` before `Makefile`, but also because our bootstrapping `makefile` includes the real `Makefile` if it exists, and the real file's rules will take precedence. At this point, GNU Make has a very useful feature: it will rebuild include files which are out-of-date before it does anything else. In oue new development directory, this will result in the necessary files being automagically generated and then acted upon.

Things that can go wrong: many projects include files such as *install-sh* and *missing* and *mkinstalldirs* in the directibution. You will need to include rules for these files in the conditonal part of your bootstrapping `makefile` rules.

```
AUTOMAKE_DIR=/usr/share/automake-1.7

install-sh: $(AUTOMAKE_DIR)/install-sh
        cp $^ $@

missing: $(AUTOMAKE_DIR)/missing
        cp $^ $@

mkinstalldirs: $(AUTOMAKE_DIR)/mkinstalldirs
        cp $^ $@
```

You will have to tell the *configure* rule that it depends on these files as well.

Other things that can go wrong: some projects use different rules for constructing the `config.h` file. You should read the generated `Makefile.in` file for how, and duplicate into the bootstrapping `makefile` file. You may also need a rule for the `aclocal.m4` file, and tell the *configure* rule it depends on it.

There is a template `makefile` installed in the */usr/local/share/aegis/config.example* directory.

Now you can set the build command field of the project configuration file:

```
build_command =
        "make "
        "project=$project "
        "change=$change "
        "version=$version";
```

Aegis watches the eist status of the build command. Be aware that many build systems which use recursive make report false positives, because the exist status of the sub-make is often ignored by the top-level Makefile. This means that Aegis may think your project compiles when, in fact, it does not.

If, while trying to get it to build, you discover more derived files which should not be primary source files, simply use the *aerm*(1) command. The *aeclean*(1) command may come in handly, too.

Once this second change set builds, integrate it via the usual Aegis process.

5.7.3. Testing

If the project you are importing has tests, they are probably executed by saying

```
$ make check
lots of output
$
```

or something similar. Aegis expects each test to be in a separate shell script. Usually this is simple enough to arrange. See the chapter on *Testing* for some hints.

5.7.4. An Optimization

The first build in a new development directory can be quite time consuming. It is possible to short-ciruit this by using the pre-built object files in the baseline. To do this, use the following setting in the project configuration file:

```
development_directory_style =
{
    source_file_symlink = true;
    derived_file_copy = true;
    derived_at_start_only = true;
};
```

This causes Aegis to copy all of the derived file into your development directory at *aedb* time. This is usually much faster than compiling and linking all over again.

5.7.5. Signed-off-by

It is possible to get the Aegis process to automatically append `Signed-off-by` lines to the change description. Set the following field in the project configuration file:

```
signed_off_by = true;
```

Only open source projects should use this field. The OSDL definition of the Developer's Certificate of Origin 1.0 can be found at http://-www.osdl.org/newsroom/press_releases/2004/-2004_05_24_dco.html and is defined to mean:

"By making a contribution to this project, I certify that:

(a) The contribution was created in whole or in part by me and I have the right to submit it under the open source license indicated in the file; or

(b) The contribution is based upon previous work that, to the best of my knowledge, is covered under an appropriate open source license and I have the right under that license to submit that work with modifications, whether created in whole or in part by me, under the same open source license (unless I am permitted to submit under a different license), as indicated in the file; or

(c) The contribution was provided directly to me by some other person who certified (a), (b) or (c) and I have not modified it."

5.7.6. Importing the Next Upstream Tarball

If you are using Aegis to track your local changes, but the master sourecs are elsewhere, you will need to track upstream changes when they are released.

It is tempting to use the *aetar*(1) command, but it will not be able to detect derived files which have been added to the tarball. You will need to uppack the tarball and remove them manually.

Create a change set in the usual way, and *aecd*(1) into it. Copy the entire project into your change set, because you don't yet know what the tarball will want to change (and it will include unchanged files).

```
$ aecd
$ aecp .
$
```

(Yes, that dot is part of the command.) Now you can unpack the tarball. You need to strip off the leading directory somehow (most polite projects use a prefix). The author uses the *tardy*(1) command, like this:

```
$ zcat project-x.y.tar.gz | \
  tardy -rp project-x.y -now | \
  tar xf -
$
```

It pays to change that the tarball is the shape you expect *before* running this command.

At this point you have to once again remove all of the files which are in the tarball, but which are not primary source files, such as *configure* and *Makefile.in* and the like.

```
$ rm -f configure Makefile.in config.h.in etc
$ rm -f aegis.log
$
```

It is useful if you place the *rm*(1) command in a shell script, and tell Aegis it is a source file, because you will have to do this every time.

Now you can have Aegis add any new files by using the follwoing command:

```
$ aenf .
$
```

(Yes, that dot is part of the command.) Note that if there are no new files, this command will give you an error, this is expected.

You will have to work out moved and removed files for yourself, and use the *aemv*(1) and *aerm*(1) commands.

At this point you should remove all the files which were present in the tarball but which dod not actually change from the change set. The follwoing command does this quickly and simply:

```
$ aecpu -unchanged
$
```

You change set now contains the minimum set of differences. Go ahead and complete it using the usual Aegis process.

5.7.7. Importing the Next Upstream Patch

In contrast to tarballs, patches tend to be far easier to cope with. In general, all that is necessary is to use the *aepatch*(1) command, something like this:

```
$ aepatch -receive -file project-x.y.diff
$
```

which will create a change set, check-out only those file the patch alters, and copes with creates and removes automagically.

There are two problems with this method. The largest problem is patches whicg contains diff for derived fiels as well. This is unfortunately *very* common.

The simplest way of coping with this is to add the *aepatch -trojan* option, which will leave the change in the *being developed* state, where you can examine it and use the *aenfu*(1) command for any derived files it insisted on creating as primary source files.

The second problem is much simpler: if a patch only contains new files, Aegis can't work out how much of the leading path it should ignore on the filenames in the patch. You will need to use the *aepatch -remove-prefix* option in this case.

5.8. No Build Required

For some projects, particularly web sites and those written exclusively in interpreted languages, it may not be necessary to ever actually build your project.

For this kind of project you add the following line to the project configuration file:

```
build_command = "exit 0";
```

For a project configured in this way, the *aede*(1) and *aeipass*(1) commands will not check that a build has been performed.

5.8.1. Why This May Not Be Such A Good Idea

It isn't always desirable to configure a project this way, even when it may initialy appear to be a good idea.

Web sites:

> You can use the build stage to check the HTML files against the relevant standards and DTDs. You can also check that all of you (internal) links are valid, and don't point to non-existant pages or anchors. Sometimes, if you have the space, you can resolve server side includes, to make it faster for Apache, by serving static pages.

Interpreted Languages:

> A whole lot of simple errors, such as syntax errors, can be caught by a static check of the source files. For example, the `perl -c` option can syntax check your Perl files without executing them. See also the GNU `awk -lint` option, the Python built-in `compile()` function, and the `php -l` (lower case L) option. You can also check that all include files referenced actually exist.

Documentation:

> Many systems allow documentation to be extracted from the source files, and turned into HTML or PDF files (*e.g.* Doxygen). This is a sensable thing to do at build time.

6. The Difference Tools

This chapter describes the difference commands in the project configuration file. Usually these commands are used by the *aegis −DIFFerence* command when differencing files, but they may be used to accomplish some other things.

The default setting is for Aegis to reject filenames which contain shell special characters. This ensures that filenames may be substituted into the commands without worrying about whether this is safe. If you set the *shell_safe_filenames* field of the project *aegis.conf* file to `false`, you will need to surround filenames with the `${quote` *filename*`}` substitution. This will only quote filenames which actually need to be quoted, so users usually will not notice. This command applies to all of the various filenames in the sections which follow.

6.1. Binary Files

Aegis doesn't particularly care whether your files are binary or text. However, your difference and merge tools certainly will. In general, you need format-specific difference and merge tools for each of the file formats used in your project. Unfortunately, most vendors of software which make use of proprietary file formats do not supply difference and merge tools.

The simplest compromise is to treat all files as text, with manual repairs for binary files.

A more elegant solution is to use a shell script invoked by the *diff_command* in the project *aegis.conf* file. This shell script examines the file to determine the file format, and then runs the appropriate difference tool. Similar considerations apply to the *merge_command* field.

Please note that this support is not present in Aegis itself because (a) it would cause code bloat, and (b) it is entirely possible to do with a shell script launched by *diff_command*.

6.2. Interfacing

The diff command is configured by a field of the project configuration file (*aegis.conf*).

6.2.1. diff_command

This command is used by *aed*(1) to produce a difference listing when file in the development directory was originally copied from the current version in the baseline[19].

All of the command substitutions described in *aesub*(5) are available. In addition, the following substitutions are also available:

${ORiginal}
: The absolute path name of a file containing the version originally copied. Usually in the baseline.

${Input}
: The absolute path name of the edited version of the file. Usually in the development directory.

${Output}
: The absolute path name of the file in which to write the difference listing. Usually in the development directory.

An exit status of 0 means successful, even of the files differ (and they usually do). An exit status which is non-zero means something is wrong.

The non-zero exit status may be used to overload this command with extra tests, such as line length limits. The difference files must be produced in addition to these extra tests.

6.2.2. merge_command

This command is used by *aed*(1) to produce a difference listing when file in the development directory is out of date compared to the current version in the baseline.

All of the command substitutions described in *aesub*(5) are available. In addition, the following substitutions are also available:

${ORiginal}
: The absolute path name of a file containing the version originally copied. Usually in a temporary file.

${Most_Recent}
: The absolute path name of a file containing the most recent version. Usually in the baseline.

${Input}
: The absolute path name of the edited version of the file. Usually in the development directory. Aegis usually moves the source file aside, so that the output can replace the source file.

${Output}
: The absolute path name of the file in which

[19] Or this is logically the case.

to write the difference listing. Usually in the development directory. This is usually the name of a change source file.

An exit status of 0 means successful, even of the files differ (and they usually do). An exit status which is non-zero means something is wrong.

6.3. When No Diff is Required

It is possible to configure a project to omit the diff step as unnecessary, by the following setting:

```
diff_command = "exit 0";
```

This disables all generation, checking and validation of difference files for each change source file. The merge functions of the *aediff*(1) command are unaffected by this setting.

6.4. Using diff and merge

These two tools are available with most flavours of UNIX, but often in a very limited form. One severe limitation is the *diff3*(1) command, which often can only cope with 200 lines of differences. The best alternative is to use GNU diff, which has context differences available, and a far more robust *diff3*(1) implementation.

See the earlier *Interfacing* section for substitution details.

6.4.1. diff_command

The entry in the configuration file looks like this:

```
diff_command =
  "set +e; diff -c $original "
  "$input > $output; test $? -le 1";
```

This needs a little explanation:
• This command is always executed with the shell's **-e** option enabled, causing the shell to exit on the first error. The "set +e" turns this off.
• The *diff*(1) command exits with a status of 0 if the files are identical, and a status of 1 if they differ. Any other status means something horrible happened. The "test" command is used to change this to the exit status aegis expects.

The **-c** option says to produce a context diff. You may choose to use the **-u** option, to produce unidiffs, if your *diff* command supports it.

You may also wish to consider ignoring white space in comparisons, as these tend to be cosmetic changes and not very interesting to code reviewers. The **-b** option of GNU Diff will ignore changes to the amount of white space, and the **-w** option will ignore white space altogether.

Binary files will often cause modern versions of GNU Diff to exit with an exit status of 2, which is probably reasonable most of the time. If your project contains binary files, you may want to treat all files as text files. Use the GNU Diff **-a** option in this case.

A useful alternative, available with more recent versions of GNU Diff, is the **-U** option. This is a more compact form than the **-c** option, and is able to give the whole file as context.

```
diff_command =
  "set +e; diff -U999999 $original "
  "$input > $output; test $? -le 1";
```

The exit status must once again be taylored, however the output will be the whole source for context, with changes marked by '+' and '-' in the left margin. This, reviewers need only search for `/^[-+]/` in order to see all edit made to the file.

6.4.2. merge_command

Note: The *merge*(1) command is better than this use of the *diff3*(1) command. See the RCS chapter for more details.

The entry in the configuration file looks like this:

```
merge_command =
  "(diff3 -e $MostRecent $original  \
   $input | sed -e '/^w$$/d' -e     \
   '/^q$$/d'; echo '1,$$p' ) | ed - \
   $MostRecent > $output";
```

This needs a lot of explanation.
• The *diff3*(1) command is used to produce an edit script that will incorporate into $MostRecent, all the changes between $original and $input. You may want the **-a** option, to treat all files as ACSII.
• The *sed*(1) command is used to remove the "write" and "quit" commands from the generated edit script.
• The *ed*(1) command is used to apply the generated edit script to the $MostRecent file, and print the results on the standard output, which are redirected into the $output file.

6.5. Using fhist

The **fhist** program by David I. Bell also comes with two other utilities, *fcomp* and *fmerge*, which use the same minimal difference algorithm.

See the earlier *Interfacing* section for substitution details.

6.5.1. diff_command

The entry in the configuration file looks like this:

```
diff_command =
  "fcomp -w $original $input "
  "-o $output";
```

The **-w** option produces an output of the entire file, with insertions and deletions marked by "change bars" in the left margin. This is superior to context difference, as it shows the entire file as context.

For more information, see the *fcomp*(1) manual entry.

6.5.2. merge_command

The entry in the configuration file looks like this:

```
merge_command =
  "fmerge $original $MostRecent \
   $input -o $output -c /dev/null";
```

The output of this command is similar to the output of the merge_command in the last section. Conflicts are marked in the output. For more information, see the *fmerge*(1) manual entry.

7. The Project Attributes

The project attributes are manipulated using the *aepa*(1) command. This command reads a project attributes file to set the project attributes. This file can be thought of as having several sections, each of which will be covered by this chapter. You should see the *aepattr*(5) manual entry for more details.

7.1. Description and Access

The *description* field is a string which contains a description of the project. Large amounts of prose are not required; a single line is sufficient.

The *default_development_directory* field is a string which contains the pathname of where to place new development directories. The pathname must be absolute. This field is only consulted if the uconf(5) field of the same name is not set. Defaults to $HOME.

The *umask* field is an integer which is set to the file permission mode mask. See *umask*(2) for more information. This value will always be OR'ed with 022, because aegis is paranoid.

7.2. Notification Commands

The *develop_end_notify_command* field is a string which contains a command to be used to notify that a change requires reviewing. All of the substitutions described in *aesub*(5) are available. This field is optional, if it is not specified no notification will be issued. This command could also be used to notify other management systems, such as progress and defect tracking, in addition to notifying users.

The *develop_end_undo_notify_command* field is a string containing a command used to notify that a change has been withdrawn from review for further development. All of the substitutions described in *aesub*(5) are available. This field is optional, if it is not specified no notification will be issued. This command could also be used to notify other management systems, such as progress and defect tracking, in addition to notifying users.

The *review_pass_notify_command* field is a string containing the command to notify that the review has passed. All of the substitutions described in *aesub*(5) are available. This field is optional, if it is not specified no notification will be issued. This command could also be used to notify other management systems, such as progress and defect tracking, in addition to notifying users.

The *review_pass_undo_notify_command* field is a string containing the command to notify that a review pass has has been rescinded. All of the substitutions described in *aesub*(5) are available. This field is optional, and defaults to the *develop_end_notify_command* field if not specified. If neither is specified, no notification will be issued. This command could also be used to notify other management systems, such as progress and defect tracking, in addition to notifying users.

The *review_fail_notify_command* field is a string containing the command to notify that the review has failed. All of the substitutions described in *aesub*(5) are available. This field is optional, if it is not specified no notification will be issued. This command could also be used to notify other management systems, such as progress and defect tracking, in addition to notifying users.

The *integrate_pass_notify_command* field is a string containing the command to notify that the integration has passed. All of the substitutions described in *aesub*(5) are available. This field is optional, if it is not specified no notification will be issued. This command could also be used to notify other management systems, such as progress and defect tracking, in addition to notifying users.

The *integrate_fail_notify_command* field is a string containing the command to notify that the integration has failed. All of the substitutions described in *aesub*(5) are available. This field is optional, if it is not specified no notification will be issued. This command could also be used to notify other management systems, such as progress and defect tracking, in addition to notifying users.

7.2.1. Notification by email

The aegis command is distributed with a set of shell scripts to perform these notifications by email. They are installed into the */usr/local/lib/aegis* directory, by default; the actual installed directory at your site is available as the *${DATa_DIRectory}* substitution. The entries in the project attribute file look like this:

```
develop_end_notify_command =
  "$sh $datadir/de.sh $project $change";
develop_end_undo_notify_command =
  "$sh $datadir/deu.sh $project $change";
review_pass_notify_command =
  "$sh $datadir/rp.sh $project $change \
  $developer $reviewer";
review_pass_undo_notify_command =
  "$sh $datadir/rpu.sh $project $change \
  $developer";
review_fail_notify_command =
  "$sh $datadir/rf.sh $project $change \
  $developer $reviewer";
integrate_pass_notify_command =
  "$sh $datadir/ip.sh $project $change \
  $developer $reviewer $integrator";
integrate_fail_notify_command =
  "$sh $datadir/if.sh $project $change \
  $developer $reviewer $integrator";
```

Please note: the exit status of all these commands will be ignored.

7.2.2. Notification by USENET

The aegis command is distributed with a set of shell scripts to perform these notifications by USENET. They are installed into the */usr/local/lib/aegis* directory, by default; the actual installed directory at your site is available as the *${DATa_DIRectory}* substitution. The entries in the project attribute file look like this:

```
develop_end_notify_command =
  "$sh $datadir/de.inews.sh $p $c alt.$p";
develop_end_undo_notify_command =
  "$sh $datadir/deu.inews.sh $p $c alt.$p";
review_pass_notify_command =
  "$sh $datadir/rp.inews.sh $p $c alt.$p";
review_pass_undo_notify_command =
  "$sh $datadir/rpu.inews.sh $p $c alt.$p";
review_fail_notify_command =
  "$sh $datadir/rf.inews.sh $p $c alt.$p";
integrate_pass_notify_command =
  "$sh $datadir/ip.inews.sh $p $c alt.$p";
integrate_fail_notify_command =
  "$sh $datadir/if.inews.sh $p $c alt.$p";
```

The last argument to each command is the newsgroup to post the article in, you may want to use some other group. Note that "$p" is an abbreviation for "$project" and "$c" is an abbreviation for "$change".

7.3. Exemption Controls

The *developer_may_review* field is a boolean. If this field is true, then a developer may review her own change. This is probably only a good idea for projects of less than 3 people. The idea is for as many people as possible to critically examine a change.

The *developer_may_integrate* field is a boolean. If this field is true, then a developer may integrate her own change. This is probably only a good idea for projects of less than 3 people. The idea is for as many people as possible to critically examine a change.

The *reviewer_may_integrate* field is a boolean. If this field is true, then a reviewer may integrate a change she reviewed. This is probably only a good idea for projects of less than 3 people. The idea is for as many people as possible to critically examine a change.

The *developers_may_create_changes* field is a boolean. If this field is true then developers may create changes, in addition to administrators. This tends to be a very useful thing, since developers find most of the bugs.

The *default_test_exemption* field is a boolean. This field contains what to do when a change is created with no test exemption specified. The default is "false", i.e. no testing exemption, tests must be provided.

This kind of blanket exemption should only be set when a project has absolutely no functionality available from the command line; examples include X11 programs. The program could possibly be improved by providing access to the functionality from the command line, thus allowing tests to be written.

7.3.1. One Person Projects

The entries in the project attributes file for a one person project look like this:

```
developer_may_review = true;
developer_may_integrate = true;
reviewer_may_integrate = true;
developers_may_create_changes = true;
```

All of the staff roles (administrator, developer, reviewer and integrator) are all set to be the same user.

7.3.2. Two Person Projects

A two person project has the opportunity for each to review the other's work.

The entries in the project attributes file for a two person project look like this:

```
developer_may_review = false;
developer_may_integrate = true;
reviewer_may_integrate = true;
developers_may_create_changes = true;
```

All of the staff roles (developer, reviewer and integrator) are all set to allow both users.

7.3.3. Larger Projects

Once you have 3 or more staff on a project, you can assign all of the roles to separate people. The idea is for the greatest number of eyes to see each change and detect flaws before they reach the baseline.

The entries in the project attributes file for a three person project look like this:

```
developer_may_review = false;
developer_may_integrate = false;
reviewer_may_integrate = false;
developers_may_create_changes = true;
```

For smaller teams, everyone may be a reviewer. As the teams get larger, the more experienced staff are often the reviewers, rather than everyone.

7.3.4. RSS Feeds

Aegis has the ability to publish RSS 2.0 items to an RSS channel when changesets transition to a new state. This is an optional feature that must be enabled and configured via the project-specific attributes.

Project administrators can configure each change of state to cause an RSS item to be added to a specified RSS channel. Each transition is individually controlled, allowing each transition to be recorded in separate channels, or all transitions in the same channel, or some combination thereof.

Generating RSS items for particular state transitions is enabled by the *rss:feedfilename* project-specific attibute. The format of this attribute is:

```
name = "rss:feedfilename-<filename>";
value = "<state> [<state> <state>]";
```

The *name* part of this attribute includes a *filename*, which is the name of the RSS feed file (channel) to which the item is to be added. The *value* part of the attribute is a space-separated list of states that will cause an RSS item to be added to the specified file. For example,

```
name = "rss:feedfilename-foo.xml";
value = "awaiting_review
         awaiting_integration";
```

will cause items to be added to the channel stored in the file "foo.xml" when a changeset transitions *into* the awaiting_review and awaiting_integration states.

The channel description can be specified using the *rss:feeddescription* attribute. The format of this attribute is:

```
name = "rss:feeddescription-<filenane>";
value = "<Some description>";
```

For example,

```
name = "rss:feeddescription-foo.xml";
value = "This is a description";
```

will cause the <description> sub-element of the <channel> element stored in the file foo.xml to have the value "*This is a description*". If this attribute is not used, the default description is: "*Feed of changes in state...*"

The channel title can be specified using the *rss:feedtitle* attribute. The format of this attribute is:

```
name = "rss:feedtitle-<filename>";
value = <Some title>;
```

For example,

```
name = "rss:feedtitle-foo.xml";
value = "This is a title";
```

will cause the <title> sub-element of the <channel> element stored in the file foo.xml to have the value "Project ...: This is a title" The title will always start with the word "Project" and the project name. If this attribute is supplied, this default title is appended with the test provided.

The channel language can be specified using the *rss:feedlanguage* attribute. The format of this attribute is:

```
name = "rss:feedlanguage-filename";
value = "language";
```

For example,

```
name = "rss:feedlanguage-foo.xml";
value = "en-AU";
```

will cause the <language> sub-element of the <channel> element stored in the file foo.xml to have the value *en-AU* If not specified, the default value of the language sub-element is "en-US".

7.3.4.1. Serving RSS Channels

aeget is able to serve up RSS channels, with an appropriate URL. An example URL is

```
http://somehost/cgi-bin/aeget/proj.1.0/?rss+foo.xml
```

The key aspect of the URL shown is the "*?rss+foo.xml*" modifier. "*foo.xml*" should obviously be replaced with the name of your RSS channel feed (that is, the filename specified with the "*rss:feedfilename*" project-specified attribute(s).

In order to read the RSS channels, it is recommended to point your RSS aggregator of choice to the appropriate URL. In order to make determining the URL easy, aeget will also place "RSS" icons next to the individual state links on the main project web page (*"proj.1.0/?menu"*) if there is an RSS channel configured to include that changeset state.

7.3.4.2. Links in RSS Channels

Links within RSS feed files are stored using a placeholder ("@@SCRIPTNAME@@") instead of the serving script in URLs. This is done because the code that knows about hte URL of a particular installation is encapsulated within aeget.

The placeholder is replaced with the real scriptname when the file is served by aeget.

8. Testing

This chapter discusses testing, and using Aegis to manage your tests and testing.

8.1. Why Bother?

Writing tests is extra work, compared to the way many small (and some not-so-small) software shops operate. For this reason, the testing requirement may be turned off.

The win is that the tests hang around forever, catching minor and major slips before they become embarrassing "features" in a released product. Prevention is cheaper than cure in this case, the tests save work down the track.

All of the "extra work" of writing tests is a long-term win, where old problems never again reappear. All of the "extra work" of reviewing changes means that another pair of eyes sees the code and finds potential problems before they manifest themselves in shipped product. All of the "extra work" of integration ensures that the baseline always works, and is always self-consistent. All of the "extra work" of having a baseline and separate development directories allows multiple parallel development, with no inter-developer interference; and the baseline always works, it is never in an "in-between" state. In each case, not doing this "extra work" is a false economy.

The existence of these tests, though, is what determines which projects are most suited to Aegis and which are not. It should be noted that suitability is a continuous scale, not black-and-white. With effort and resources, almost anything fits.

8.1.1. Projects for which Aegis' Testing is Most Suitable

Projects most suited to supervision by Aegis are straight programs. What the non-systems-programmers out there call "tools" and sometimes "applications". These are programs which take a pile of input, chew on it, and emit a pile of output. The tests can then compare actual outputs with expected outputs.

As an example, you could be writing a *sed*(1) look-alike, a public domain clone of the UNIX sed utility. You could write tests which exercise every feature (insertion, deletion, etc.) and generate the expected output with the real UNIX sed. You write the code, and run the tests; you can immediately see if the output matches expectations.

This is a simple example. More complex examples exist, such as Aegis itself. The Aegis program is used to supervise its own development. Tests consist of sequences of commands and expected results are tested for.

Other types of software have been developed using Aegis: compilers and interpreters, client-server model software, magnetic tape utilities, graphics software such as a ray-tracer. The range is vast, but it is not all types of software.

8.1.2. Projects for which Aegis' Testing is Useful

For many years there have been full-screen applications on text terminals. In more recent times there is increasing use of graphical interfaces.

In developing these types of programs it is still possible to use Aegis, but several options need to be explored.

8.1.2.1. Testing Via Emulators

There are screen emulators for both full-screen text and X11 available. Using these emulators, it is possible to test the user interface, and test via the user interface. As yet, the author knows of no freely available emulators suitable for testing via Aegis. If you find one, please let me know.

8.1.2.2. Limited Testing

You may choose to use Aegis simply for its ability to provide controlled access to a large source. You still get the history and change mechanisms, the baseline model, the enforced review. You simply don't test all changes, because figuring out what is on the screen, and testing it against expectations, is too hard.

If the program has a command line interface, in addition to the full-screen or GUI interface, the functionality accessible from the command line may be tested using Aegis.

It is possible that "limited testing" actually means "no testing", if you have no functionality accessible from the command line.

8.1.2.3. Testing Mode

Another alternative is to provide hooks into your program allowing you to substitute a file for user input, and to be able to trigger the dump of a "screen image". The simulated user input can then be fed to the program, and the screen dump (in some terminal-independent form) can be

compared against expectations.

This is easier for full-screen applications, than for X11 applications. You need to judge the cost-benefit trade-off. Cost of development, cost of storage space for X11 images, cost of *not* testing.

8.1.2.4. Manual Tests

The Aegis program provides a manual test facility. It was originally intended for programs which required some physical action from a user, such as "unplug Ethernet cable now" or "mount tape XG356B now". It can also be used to have a user confirm that some on-screen activity has happened.

The problem with manual tests is that they simply don’t happen. It is far more pleasant to say "run the automatic tests" and go for a cup of coffee, than to wait while the computer thinks of mindless things to ask you to do. This is human nature: if it can be automated, it is more likely to happen.

8.1.2.5. Unit Tests

Many folks think of testing as taking the final product and testing it. It is also possible to build specialized unit tests, which exercise specific portions of the code. These tests can then be administrated by Aegis, even if the full-blown GUI cannot be.

8.1.3. Projects for which Aegis’ Testing is Least Useful

Another class of software is things like operating system kernels and firmware; things which are "stand alone". This isolated nature makes it the most difficult to test: to test it you want to provide physical input and watch the physical output. By its very nature, it is hard to put into a shell script, and thus hard to write an Aegis test for.

The above chapter was written in 1991. At this writing (1999) there are projects like ×Linux and operating systems like VxWorks. These are all embedded, and all have excellent network and download support. It is entirely possible (with design support!) to write automatically testable embedded systems.

8.1.3.1. Operating Systems

It is not impossible, just that few of us have the resources to do it. You need to have a test system and a testing system: the test system has all of its input and outputs connected to the outputs and inputs of the testing system. That is, the testing system controls and drives the test system, and watches what happens.

For example, in the olden days before everyone had PC and graphics terminals, there were only serial interfaces available. Many operating system vendors tested their products by using computers connected to each serial line to simulate "user input". The system can be rebooted this way, and using dual-ported disks allows different versions of a kernel to be tried, or other test conditions created.

For software houses which write kernels, or device drivers for kernels, or some other kernel work, this is bad news: the Aegis program is probably not for you. It is possible, but there may be more cost-effective development strategies. Of course, you could always use the rest of Aegis, and ignore the testing part.

However, Aegis has been used quite successfully to develop Linux kernel modules. With suitable *sudo*(1) configuration to permit access to *insmod*(1) &co, developers can write test scripts which load device drivers, try them out, and unload them again, all without universal *root* access.

Also, the advent of modern tools, such as VMware, which allow one operating system to "host" another, may also permit straightforward testing of kernels and operating systems.

8.1.3.2. Firmware

Firmware is a similar deal: you need some way to download the code to be tested into the test system, and write-protect it to simulate ROM, and have the necessary hardware to drive the inputs and watch the outputs.

As you can see, this is generally not available to run-of-the-mill software houses, but then they rarely write firmware, either. Those that do write firmware usually have the download capabilities, and some kind of remote operation facility.

However, this omits the possibility of not only cross compiling your code for the target system, but also compiling your code to run natively on the build system. The firmware (in the host incarnation) then falls into one of the categories above, and may be readily tested. This does not relieve you of also testing the firmware, but it increases the probability that the firmware isn’t completely useless before you download it.

By using an object oriented language, such as C++, the polymorphism necessary to cope with multiple environments can be elegantly hidden behind a pure abstract base class. Alternatively, by using a consistent API, you can accomplish the necessary sleight-of-hand at link time.

The unit test method mentioned earlier is also very useful for firmware, even if the device "as a whole" cannot be tested.

8.2. Writing Tests

This section describes a number of general guidelines for writing better tests, and some pitfalls to be avoided.

There are also a number of suggestions for portability of tests in specific scripting languages; this will definitely be important if you are writing software to publish on WWW or for FTP. Portability is often required *within* an organization, also. Examples include a change in company policy from one 386 UNIX to another (e.g. company doesn't like Linux, now you must use AT&T's SVR4 offering), or the development team use *gcc* until the company finds out and forces you to use the prototype-less compiler supplied with the operating system, or even that the software being developed must run under both UNIX and Windows NT.

Note, also, that when using Aegis' heterogeneous build support, portability will again feature prominently.

8.2.1. Contributors

I'd like to thank Steven Knight <knight@baldmt.com> for writing portions of this information.

If other readers have additional testing techniques, or use other scripting languages, contributions are welcome.

8.2.2. General Guidelines

This section lists a number of general guidelines for all aegis tests, regardless of implementation language. Use this section to guide how you write tests if the scripting language you choose is not specifically covered in greater detail below.

8.2.2.1. Choice of Scripting Language

The aegis program uses the *test_command* field of the project *aegis.conf* file to specify how tests are executed. The default value of the *test_command* field:

```
test_command = "$shell $file_name";
```

specifies that tests be Bourne shell scripts. You may, however, change the value of *test_command* to specify some other scripting language interpreter, which allows you to write your test scripts in whatever scripting language is appropriate for your project. The Perl or Python scripting languages, for example, could be used to create test scripts that are portable to systems other than UNIX systems.

This means that if you can write it in your scripting language of choice, you can test it. This includes such things as client-server model interfaces, and multi-user synchronization testing.

8.2.2.2. No Execute Permission

Under aegis, script files do not have execute permission set, so they should always be invoked by passing the script file to the interpreter, not executing the test directly:

```
sh filename
perl filename
```

8.2.2.3. No Command-Line Arguments

Tests should not expect command line arguments. Tests are not passed the name of the project nor the number of the change.

8.2.2.4. Identifying the Scripting Language

Even though aegis does not execute the test script directly, it is a good idea to put some indication of its scripting language into the test script. See the sections below for suggested "magic number" identification of scripts in various languages.

8.2.2.5. Current Directory

Tests are always run with the current directory set to either the development directory of the change under test when testing a change, or the integration directory when integrating a change, or the baseline when performing independent tests.

A test must not make assumptions about where it is being executed from, except to the extent that it is somewhere a build has been performed. A test must not assume that the current directory is writable, and must not try to write to it, as this could damage the source code of a change under development, potentially destroying weeks of work.

8.2.2.6. Check Exit Status and Return Values

A test script should check the exit status or return value of every single command or function call, even those which cannot fail. Checking the exit status or return value of every statement in the script ensures that strange permission settings, or disk space problems, will cause the test to fail, rather than plow on and produce spurious results. See the sections below for specific suggestions on checking exit status or return values in various scripting languages.

8.2.2.7. Temporary Directory

Tests should create a temporary subdirectory in the operating system's temporary directory (typically */tmp* on UNIX systems) and then change its working directory (*cd*) to this directory. This isolates any vandalism that the program under test may indulge in, and serves as a place to write temporary files.

At the end of the test, it is sufficient to change directory out of the temporary subdirectory and then remove the entire temporary subdirectory hierarchy, rather than track and remove all test files which may or may not be created.

Some UNIX systems provide other temporary directories, such as */var/tmp*, which may provide a better location for a temporary subdirectory for testing (more file system space available, administrator preference, etc.). Test scripts wishing to accomodate alternate temporary directories should use the TMPDIR environment variable (or some other environment variable appropriate to the operating system hosting the tests) as the location for creating their temporary subdirectory, with */tmp* as a reasonable default if TMPDIR is not set.

8.2.2.8. Trap Interrupts

Test scripts should catch appropriate interrupts (1 2 3 and 15 on UNIX systems) and cause the test to fail. The interrupt handler should perform any cleanup the test requires, such as removing the temporary subdirectory.

8.2.2.9. PAGER

If the program under test invokes pagers on its output, a la *more*(1) et al, it should be coded to use the PAGER environment variable. Tests of such programs should always set PAGER to *cat* so that tests always behave the same, irrespective of invocation method (either by aegis or from the command line).

8.2.2.10. Auxiliary Files

If a test requires extra files as input or output to a command, it must construct them itself from in-line data. (See the sections below for more specific information about how to use in-line data in various scripting languages to create files.)

It is almost impossible to determine the location of an auxiliary file, if that auxiliary file is part of the project source. It could be in either the change under test or the baseline.

8.2.2.11. New Test Templates

Regardless of your choice of scripting language, it is possible to specify most of the repetitious items above in a *file template* used every time a user creates a new test. See the *aent*(1) command for more information.

Having the machine do it for you means that you are more likely to do it.

8.2.3. Bourne Shell

The Bourne shell is available on all flavors of the UNIX operating system, which allows Bourne shell scripts to be written portably across those systems. Here are some specific guidelines for writing aegis tests using Bourne shell scripts.

8.2.3.1. Magic Number

Some indication that the test is a Bourne shell script is a good idea. While many systems accept that a first line starting with a colon is a Bourne shell "magic number", a more widely understood "magic number" is

```
#! /bin/sh
```

as the first line of the script file.

8.2.3.2. Check Exit Status

A Bourne shell test script should check the exit status of every single command, even those which cannot fail. Do not rely on, or use, the *set −e* shell option (it provides no ability to clean up on error).

Checking the exit status involves testing the contents of the **$?** shell variable. Do not use an *if* statement wrapped around an execution of the program under test as this will miss core dumps and other terminations caused by signals.

8.2.3.3. Temporary Directory

Bourne shell test scripts should create a temporary subdirectory in */tmp* (or the directory specified by the TMPDIR environment variable) and then *cd* into this directory. At the end of the test, or on interrupt, the script should *cd* out of the temporary subdirectory and then *rm −rf* it.

8.2.3.4. Trap Interrupts

Use the *trap* statement to catch interrupts 1 2 3 and 15 and cause the test to fail. This should perform any cleanup the test requires, such as removing the temporary directory.

8.2.3.5. Auxiliary Files

If a test requires extra files as input or output to a command, it must construct them itself, using *here* documents:

```
cat <<EOF >file
contents
of the
file
EOF
```

See *sh*(1) for more information.

8.2.3.6. [test]

You should always use the *test* command, rather than the square bracket form, as many systems do not have the square bracket form, if you publish to USENET or for FTP.

8.2.3.7. Other Bourne Shell Portability Issues

The above list covers the most common Bourne shell issues that are relevant to most aegis tests. The documentation for the GNU autoconf utility, however, contains a more exhaustive list of Bourne shell portability issues. If you want (or need) to make your tests as portable as possible, see the documentation for GNU autoconf.

8.2.4. Perl

Perl is a popular open-source scripting language available on a number of operating systems. Here are some specific guidelines for writing aegis tests using Perl scripts.

8.2.4.1. Magic Number

Some indication that the test is a Perl script is a good idea. Because Perl is not installed in the same location on all UNIX systems, a first-line "magic number" such as:

```
#! /usr/local/bin/perl
```

that hard-codes the Perl path name will not be portable if you publish your tests.

If the *env*(1) program is available, a more portable "magic number" for Perl is:

```
#! /usr/bin/env perl
```

8.2.4.2. Check Return Values

A Perl test script should check the return value from every subroutine, even those which cannot fail.

A Perl test script should also check the exit status of every command it executes. Checking the exit status involves testing the contents of the **$?** variable. See the Perl documentation on "Predefined Variables" for details.

8.2.4.3. Temporary Directory

Perl test scripts should create a temporary subdirectory in */tmp* (or the directory specified by the **$ENV{TMPDIR}** environment variable) and then *chdir* into this directory. At the end of the test, or on interrupt, the script should *chdir* out of the temporary subdirectory and then remove it and its hierarchy. A portable way to do this within a Perl script:

```
use File::Find;
finddepth(sub { if (-d $_) {
                rmdir($_)
        } else {
                unlink($_)
        } },
        $dir);
```

8.2.4.4. Trap Interrupts

Use Perl's *$SIG* hash to catch interrupts for HUP, INT, QUIT and TERM and cause the test to fail. This should perform any cleanup the test requires, such as removing the temporary directory. A very simple example:

```
$SIG{HUP} =
$SIG{INT} =
$SIG{QUIT} =
$SIG{TERM} =
        sub { &cleanup; exit(2) };
```

8.2.4.5. Auxiliary Files

If a test requires extra files as input or output to a command, it must construct them itself, using in-line data such as *here* documents See the Perl documentation for more information.

8.2.4.6. Exit Values

Aegis expects tests to exit with a status of 0 for success, 1 for failure, and 2 for no result. The following code fragment will map all failed (non-zero) exit values to an exit status of 1, regardless of what Perl module called exit:

```
END { $? = 1 if $? }
```

A more complete example could check conditions and set the exit status to 2 to indicate NO RESULT.

8.2.4.7. Modules

Perl supports the ability to re-use modules of common routines, and to search several directories for modules. This makes it convenient to write modules to share code among the tests in a project.

Any modules that are used by your test scripts (other than the standard modules included by Perl) should be checked in to the project as source files. Test scripts should then import the module(s) via the normal Perl mechanism:

```
use MyTest;
```

When a test is run, the module file may actually be in the baseline directory, not the development or integration directories. To make sure that the test invocation finds the module, the *test_command* field in the project *aegis.conf* file should use the Perl **−I** option to search first the local directory and then the baseline:

```
test_command =
    "perl −I. −I${BaseLine} \
    ${File_Name}"
```

or, alternatively, if you had created your Perl test modules in a subdirectory named **aux**:

```
test_command =
    "perl −I./aux −I${BaseLine}/aux \
    ${File_Name}"
```

For details on the conventions involved in writing your own modules, consult the Perl documentation or other reference work.

Actually, you need to use the ${search_path} substitution. I'll have to fix this one day.

8.2.4.8. The Test::Cmd Module

A Test::Cmd module is available on CPAN (the Comprehensive Perl Archive Network) that makes it easy to write Perl scripts that conform to aegis test requirements. The Test::Cmd module supports most of the guidelines mentioned above, including creating a temporary subdirectory, cleaning up the temporary subdirectory on exit or interrupt, writing auxiliary files from in-line contents, and provides methods for exiting on success, failure, or no result. The following example illustrates some of its capabilities:

```
#! /usr/bin/env perl
use Test::Cmd;
$test = Test::Cmd−>new(prog
        => 'program_under_test',
                workdir => '');
$ret = $test−>write('aux_file', <<EOF);
contents of file
EOF
$test−>no_result(! $ret =>
        sub { print STDERR
        "Couldn't write file: $!\\n"});
$test−>run(args => 'aux_file');
$test−>fail($? != 0);
$test−>pass;
```

The various methods supplied by the Test::Cmd module have a number of options to control their behavior.

The Test::Cmd module manipulates file and path names using the operating-system-independent File::Spec module, so the Test::Cmd module can be used to write tests that are portable to any operating system that runs Perl and the program under test.

The Test::Cmd module is available on CPAN. See the module's documentation for details.

8.2.4.9. The Test and Test::Harness Modules

Perl supplies two modules, Test and Test::Harness, to support its own testing infrastructure. Perl's tests use different conventions than aegis tests; specifically, Perl tests do not use the exit status to indicate the success or failure of the test, like aegis expects. The Test::Harness module expects that Perl tests report the success or failure of individual sub-tests on standard output, and always exit with a status of 0 to indicate the script tested everything it was supposed to.

This difference makes it awkward to use the Test and Test::Harness modules for aegis tests. In some circumstances, though, you may be forced to write tests using the Test and Test::Harness modules − for example, if you use aegis to develop a Perl module for distribution − but still wish to have the tests conform to aegis conventions during development.

This can be done by writing each test to use an environment variable to control whether its exit status should conform to aegis or Perl conventions. This is easy when using the Test module to write tests, as its *onfail* method provides an appropriate place to set the exit status to non-zero if the appropriate environment variable is set. The following code fragment at or near the beginning of each Perl test script accomplishes this:

```
use Test;
BEGIN { plan tests => 3,
    onfail => sub {
        $? = 1 if $ENV{AEGIS_TEST}
        }
    }
```

(See the documentation for the Test module for information about using it to write tests.)

There then needs to be a wrapper Perl script around the execution of the tests to set the environment variable. The following script (called *mytest.pl* for the sake of example) sets the **AEGIS_TEST** environment variable expected by the previous code fragment:

```
use Test::Harness;
$ENV{AEGIS_TEST} = 1;
open STDOUT, ">/dev/null" || exit (2);
runtests(@ARGV);
END { $? = 1 if $?;
    print STDERR $?
        ? "FAILED" : "PASSED",
        "\n"; }
```

It also makes its output more nearly conform to aegis' examples by redirecting standard output to **/dev/null** and restricting its reporting of results to a simple **FAILED** or PASSED on standard error output.

The last piece of the puzzle is to modify the *test_command* field of the project *aegis.conf* file to have the *mytest.pl* script call the test script:

```
test_command =
    "perl -I. -I${BaseLine} mytest.pl \
    ${File_Name}"
```

The Test and Test::Harness modules are part of the standard Perl distribution and do not need to be downloaded from anywhere. Because these modules are part of the standard distribution, they can be used by test scripts without being checked in to the project.

8.2.4.10. Granularity By *Steven Knight <knight@baldmt.com>*

The granularity of Perl and Aegis tests mesh very well at the individual test file (.t) level. Aegis and Test::Harness are simply different harnesses that expect slightly different conventions from the tests they execute: Aegis uses the exit code to communicate an aggregate pass/fail/no result status, Test::Harness examines the output from tests to decide if a failure occurred.

It's actually pretty easy to accomodate both conventions. You can do this as easily as setting the test_command variable in the project configuration file to something like the following:

```
test_command =
        "perl -MTest::Harness -e 'runtests(\"$fn\"); \
        END {$$? = 1 if $$? }'";
```

In reality, you'll likely need to add variable expansions to generate -I or other Perl options for the full Aegis search path. The END block takes care of mapping any non-zero Test::Harness exit code to the '1' that Aegis expects to indicate a failure.

The only thing you really lose here is the Test::Harness aggregation of results and timing at the end of a multi-test run. This is more than offset by having Aegis track which tests need to be run for a given change.

Alternatively, you can execute the .t files directly, not through Test::Harness::runtests. This is easily accomodated using the onfail method from the standard Perl Test module in each test. Here's a standard opening block for .t tests

```
use Test;
BEGIN { $| = 1; plan tests => 19,
        onfail => sub { $? = 1 if $ENV{AEGIS_TEST} }
}
END {print "not ok 1\n" unless $loaded;}
use Test::Cmd;
$loaded = 1;
ok(1);
```

That's it (modulo specifying the appropriate number of tests). My .t tests now use the proper exit status to report a failure back to Aegis. The only other piece is configuring the project's "test_command" value to set the AEGIS_TEST environment variable.

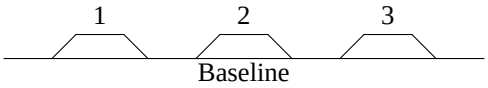
You can also use an intermediate script that also redirects the tests's STDOUT to /dev/null, if you are used to and like the coarser PASSED/FAILED status.

8.2.5. Batch Testing

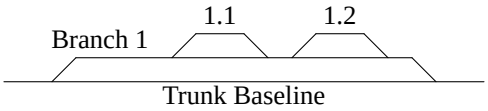
The usual "*test_command*" field of the project *aegis.conf* file runs a single test at a time. When you have a multi-CPU machine, or are able to distribute the testing load across a range of machines, it is often desirable to do so. The "*batch_test_command*" of the project configuration file is for this purpose. See *aepconf*(5) for more information.

9. Branching

This chapter describes the concept of branching implemented by Aegis. The process described in previous chapters makes changes to a project's master source.



Branching generalizes this change model, by allowing the baseline to be treated as a change, or the ability to treat a change as a baseline.



Since branchs are sometimes considered as a changes it is useful to expand on the differences. A branch, or trunk, baseline may have children which are either *changes* or deeper *branches.* From this perspective the difference is that nothing may be modified directly in a branch. A branch is an integrated baseline with all the associated protection. To modify a branch one must open a change under that branch.

Looking upward from a change under a branch, its *parent* is the branch baseline, and its *grand*parent is the *parent* of the branch. We will see this used later when we talk about ending a branch.

A significant feature of Aegis branches is that, because they are an extension of the *change* concept, they are expected to end, and be integrated back into *their* baseline, or *parent* ,in time.

The most common case of this is in project releases.

A branch in the *being developed* state may have changes made to it, and/or deeper branches. This may recurse to any level. Once a branch is complete, no further deeper branches may be created from that branch.

9.1. How To Use Branching

To access a project branch, the project name has the branch appended, separated by a dot or a hyphen. For example: branch 1 of project "aegis" is referred to as "aegis.1". To reference changes on this branch, use this compound project name wherever you would normally use a project name.

Traditional 2-level project release names are obtained by using a further level of branching. For example: by creating branch 0 of project "aegis.1", there would be a branch accessed as project "aegis.1.0".

By default, these two level of project branching are created automatically when the *aenpr*(1) command is used. You need to use the **−VERSion** option to make this deeper or shallower, or have different numbering.

Command	Branches Created
aenpr foo	foo, foo.1, foo.1.0
aenpr foo −vers 2.4.1	foo, foo.2, foo.2.4, foo.2.4.1
aenpr foo.7	foo, foo.7
aegis −npr foo −vers −	foo

The last is a special case, to enable a project to be created with no default branches (it's also hard to get the empty string past the alias).

To add branching and release level management to an existing project on uses the *aenbr*(1) command at any level. Say we already have `foo.1.0`, which represents version 1.0 of our software. One method of release level management would be to integrate *foo.1.0* into its *parent foo.1* and then do *aenbr −p foo.1* would create *foo.1.1* representing version 1.1. Eventually we might want to make a major version release and would integrate *foo.1* into its *parent foo* and then do *aenbr −p foo*, which would create *foo.2*. Then if we do *aenbr −p foo.2* we create *foo.2.0*, for development of version 2.0 of our software.

9.2. Transition Using aenrls

To convert a project from the old-style to the new branching style, use the *aenrls*(1) command.

If you give no second project name, the new name is generated by removing the numeric suffixes. If you did not give a −VERSion option, the numeric suffixes will be used to determine the next version, by adding one to the previous minor version number. The new project is then created rather like the *aenpr*(1) command.

The files of the old project are copied across as an implicit change on the newly created branch within the new project. If the new project name already exists, and is a new-style project, the *aenrls*(1) command will attempt to make the appropriate numbered branches. If the new project already exists and is an old-style project, or it exists and the branch number(s) are already in use, *aenrls*(1) will emit an error and fail. The

aenrls(1) command only works on old-style projects, and always converts them to new style projects.

Planning you branch numbers is essential. If you want to use 3-level branch numbers (e.g. "aegis.2.3.1") at some time in the future, then you must use 3-level version numbers all the way through (e.g. "aegis.2.3.0"). This is because change numbers and branch numbers come from the same common pool of numbers. Once change one has been used (e.g. "aegis.2.3.C001"), then branch one is no longer available (e.g. "aegis.2.3.1.C042" conflicts).

9.3. Cross Branch Merge

From time to time you will want to merge the changes from one branch into a change. This may be done using a cross-branch merge. This is done by specifying the **−BRanch** option to the *aegis −diff −merge-only* command.

The most common cross branch merge is when the project's files are out-of-date. Because it is not possible to use *aegis −diff −merge-only* directly on the branch, this must be in a change on the branch. As a short-cut, the branch may be specified using the *−grandparent* option.

9.4. Multiple Branch Development

It is very common for a bug fix to need to be applied to more than one branch at once. The change could be applied to the common ancestor branch, however this may not be effective in the branch immediately. An alternative is to use the *aegis −clone* command, which can be used to identically reproduce a change on another branch.

9.5. Hierarchy of Projects

It would be nice if there was some way to use one project as a sort of "super change" to a "super project", so that large teams (say 1000 people) could work as lots of small teams (say 100 people). As a small team gets their chunk ready, using the facilities provided to-date by Aegis, the small team's baseline is treated as a change to be made to the large team baseline.

This idea can be extended quite naturally to any depth of layering.

After reading *Transaction Oriented Configuration Management: A Case Study* Peter Fieler, Grace Downey, CMU/SEI-90-TR-23, this is not a new idea. It also provides some ideas for how to do branching sensibly, and was influential in the design of Aegis' branching.

9.5.1. Fundamentals

Aegis has everything you need to have a super project and a number of sub-projects. All you need to do is create an active branch for each sub-project. Each branch gets a separate baseline, *viz*

```
% aenpr gizmo.0.1
project "gizmo": created
project "gizmo.0": created
project "gizmo.0.1": created
%
```

Now, for each of your desired sub-projects, create another branch

```
aenbr −p gizmo.0.1 1 # for the foo project
aenbr −p gizmo.0.1 2 # for the bar project
aenbr −p gizmo.0.1 3 # for the baz project
```

Now, the guys on the *foo* project set their AEGIS_PROJECT environment variable to to *gizmo.0.1.1*, the *bar* guys use *gizmo.0.1.2*, and *baz* uses *gizmo.0.1.3*. From the developer's point of view they are separate projects. From one level up, though, they are just part of a bigger project.

It helps if you design and implement the build system first. You do this as a change set on the common parent branch. Once it is completed each branch can inherit it from the common parent. This makes integration easier, when it comes time to integrate the sub-projects together.

9.5.2. Incremental Integration

It is very common that not all of the sub-projects will be ready to be integrated at the same time. This is the normal situation with Aegis branching, and is handled cleanly and simply.

In Aegis each branch is literally a change, all the way down into the internal data structures. Just as each change gets its own development directory, each branch gets its own baseline. Just as a development directory inherits everything its doesn't have from the baseline, so branches inherit everything *they* don't have from their parent branch (or ultimately from the trunk). Just as you incrementally integrate changes into a branch, you incrementally integrate branches into their parent.

The branches only influence each other when they are integrated, just as changes only influence each other when they are integrated.

There are times when a branch being integrated into its parent is found to be inadequate. Aegis provides a simple mechanism to "bounce" a

branch integration. Recall that, for Aegis, branches are the same as changes. Just as you "develop end" a change (see *aede*(1) for more information) you also *aede* a branch when development on it is completed.

Once a branch has develop-end (stops being an *active* branch), it is reviewed as a normal change, and integrated as a normal change. If integration failed, it returns to "being developed" and becomes an active branch once again, until the next *aede*. As you can see, it is as easy to bounce a branch integration as it is to bounce a change integration.

An unsuccessful branch integration leaves the repository unchanged, just as an unsuccessful change integration leaves it unchanged.

9.5.3. Super-Project Branching

In many real-world situations it is very important to be able to branch at any point in the past history of the super-project to fix (integration specific) bugs or to customize more the older states of the super-project.

You can create a branch at any time, on any active branch or active branch ancestor. You can populate that branch with historical versions (from any other branch, actually, not just the ancestral line). The method is a little fussy − you can't *aecp* into a branch directly, you need to do this via a change to that branch. Files not changed by a change on a branch are inherited from the current (*i.e.* active) state of the parent branch. See the section on *Insulation*, above.

9.5.4. Super-Project Testing

Many folks see Aegis' testing features as useful for unit testing individual files or change sets. For large projects, it is common that a specific test tool will be written. However, even large scale integration testing is possible using Aegis.

You can change the test command from being a shell script to being anything to you want − see the *test_command* field in *aepconf*(5). Or run the test tool from the shell script. If the integration tests can be automated, it makes sense to preserve them in the repository − they are some of the most valuable regression tests for developers, because they describe correct behavior outside the "box" the developer usually works in.

9.5.5. The Next Cycle

Once you have a fully-integrated product, what happens on the next cycle? Well, first you may want to finish *gizmo.0.1* and integrate it into *gizmo.0*, and then *aenbr −p gizmo.0 2*

Then what? Same deal as before, but anything not changed in one of the sub-project branches gets inherited from the ancestor.

```
aenbr -p gizmo.0.2 1 # for the foo project
aenbr -p gizmo.0.2 2 # for the bar project
aenbr -p gizmo.0.2 3 # for the baz project
```

Most folks find doing the whole mega-project-build every time tiresome − so don't. Temporarily (via a change set) hack the build configuration to build only the bit you want − obviously a different hack on each sub-project's branch. Just remember to un-hack it (via another change set) before you integrate the sub-project.

9.5.6. Bug Fixing

The *aeclone*(1) command lets you clone a change set from one branch to another. So if you have a bug fix that needs to be done on each active branch you can clone it (once you have it fixed the first time). You still have to build review and integrate *n* times (branches often differ non-trivially). Providing it isn't already in use, you can even ask for the same change number on each branch − handy for syncing with an external bug tracking system.

Alternatively, fix bugs in the common ancestor, and the sub-projects will inherit the fix the next time they integrate something on their branch (assuming they aren't insulated against it).

9.6. Conflict Resolution

A development directory becomes out of date, compared to the project, when another change is integrated which has a file in common. This situation is detected automatically by *aede*(1) and you resolve it using *aed*(1), usually with something like the *−merge-only* option. Additionally, you can see if you have an out-of-date file from the *change files* listing, because it will show you the current baseline version in parentheses if you are out-of-date.

Aegis implements branches as very long changes, with sub-changes. A side effect of this is that a branch can become out-of-date in the same way that a development director becomes out of date. When it comes time to *aede*(1) the branch, you will be told if there are any out-of-date files.

Additionally, the *project files* listing will show out of date files in exactly the same way that the *change file* listing does.

9.6.1. Cross Branch Merge

However, unlike a simple change, if you attempt to use the *aed −merge-only* command in the branch baseline, you will get an error message! How, then, do you resolve the apparent impasse?

The *aed*(1) command has a number of options designed for just this purpose.

- The *−branch* option may be used to specify another branch of the same project, as a source of the file to be differenced against. This is almost what you need.
- The *−grandparent* option is a special case of the *−branch* option, and it means the parent branch of project.
- The *−trunk* option is also a special case of the *−branch* option, and it means the base branch from which the entire branch tree springs.

By creating a new change on the out-of-date branch, and copying in the out-of-date files, you have almost everything required. All that is necessary is to perform a cross-branch merge against the project grandparent, and the necessary merging will be performed. **In addition** Aegis will remember that it was a cross-branch merge, and once *aeipass* completes successfully, the branch will be up-to-date once more.

□ Create a new change on the out-of-date branch

□ Use a simple *aecp* command to copy the out-of-date files. (Do *not* use any *−branch* or *−delta* options.)

□ Use the "*aed −merge-only −grandparent*" command to perform the merge.

At this point, if you use the "*ael cf*" command, you will notice that this file is tagged in the listing with the new branch edit origin, to be used during *aeipass*. If it isn't, you have made a mistake.

□ As usual, use your favourite editor to check the merge results, and resolve any conflicts.

□ Build and test as usual.

□ Complete the change as usual.

□ Once *aeipass* is successful, the branch will be up-to-date (for the files in the change).

9.6.2. Insulation

One of the stated benefits of using a branch is the insulating effects which branches can provide. However, when you have multiple simultaneous active branches, that insulation will inevitably lead to out-of-date branch files. Now that *how* to merge them has been described, *when* should you merge?

In a simple change's development directory, there are times when an *aeipass* will result in all developers needing to recompile. Depending on what files you are working on, it may be that you need to merge some of your change files immediately, or *aecp* an earlier version of the files which changed in the project.

Branches can also suffer from exactly the same problems, and are mended by exactly the same alternatives.

9.6.2.1. Branch Insulated Against Project

If you created a branch to insulate the work being done on the branch from other activities in the project, it follows that when such build problems occurred, you would use an "*aecp −delta" command to continue insulating.*

This action defers the labour of merging until towards the end of the branch development, sometimes with a quite visible schedule impact.

9.6.2.2. Project Insulated Against Branch

If you created a branch to insulate the project from work being done in the branch, it follows that you would do a cross branch immediately.

This action amortizes the labour of merging across the life of the branch, often with a number of small delays and less schedule impact.

9.6.2.3. Mix 'n' Match

Of course, we usually have both these motives, and some more besides, so the answer is usually "it depends".

9.7. Ending A Branch

"OK, I give up. I do not understand the ending of branches."

Usually, you end development of a branch the same way you end development of a simple change. In this example, branch *example.1.42* will be ended:

```
% aede -p example 1 -c 42
aegis: project "example.1": change
42: file "fubar" in the baseline
has changed since the last 'aegis
-DIFFerence' command, you need to
do a merge
%
```

Oops. Something went wrong. Don't panic!

I'm going to assume, for the purposes of explanation, that there have been changes in one of the ancestor branches, and thus require a merge, just like file *fubar*, above.

You need to bring file *fubar* up-to-date. How? You do it in a change set, like everything else.

At his point you need to do 5 things: (1) create a new change on example.1.42, (2) copy *fubar* into it, (3) merge *fubar* with branch "example.1" (4) end development of the change and integrate it, and (5) now you can end example.1.42

The -GrandParent option is a special case of the -BRanch option. You are actually doing a cross-branch merge, just up-and-down rather than sideways.

```
% aem -gp fubar
%
```

And manually fix any conflicts... naturally.

At this point, have a look at the file listing, it will show you something you have never seen before - it will show you what it is *going to* set the branch's edit_number_origin to for each file, at *aeipass*.

```
% ael cf
Type   Action Edit        File Name
------ ------ -------     -----------
source modify 1.3         fubar
              {cross 1.2}
```

Now finish the change as usual... *aeb*, *aed*, *aede*, *aerpass*, *aeib*, ..., *aeipass* nothing special here.

One small tip: merge the files one at a time. It makes keeping track of where you are up to much easier.

Now you can end development of the branch, because all of the files are up-to-date.

In some cases, Aegis has detected a logical conflict where you, the human, know there is none. Remember that the *aem* command saves the old version of the file with a `,B` suffix ('B' for backup). If you have a file like this, just use

```
% mv fubar,B fubar
%
```

to discard everything from the ancestor branch, and use the current branch.

10. Tips and Traps

This chapter contains hints for how to use the aegis program more efficiently and documents a number of pitfalls you may encounter.

This chapter is at present very "ad hoc" with no particular ordering. Fortunately, it is, as yet, rather small. The final size of this chapter is expected to be quite large.

10.1. Renaming Include Files

Renaming include files can be a disaster, either finding all of the clients, or making sure the new copy is used rather than the old copy still in the baseline.

Aegis provides some assistance. When the *aemv*(1) command is used, a file in the development directory is created in the *old* location, filled with garbage. Compiles will fail very diagnostically, and you can change the reference in the source file, probably after *aecp*(1)ing it first.

If you are moving an include file from one directory to another, but leaving the basename unchanged, create a link[20] between the new and old names, but only in the development directory (i.e. replacing the "garbage" file aegis created for you). Create the link after *aemv*(1) has succeeded. This insulates you from a number of nasty Catch-22 situations in writing the dependency maintenance tool's rules file.

10.2. Symbolic Links

If you are on a flavor of UNIX which has symbolic links, it is often useful to create a symbolic link from the development directory to the baseline. This can make browsing the baseline very simple.

Assuming that the project and change defaults are appropriate, the following command

```
ln -s 'aegis -cd -bl' bl
```

is all that is required to create a symbolic link called *bl* pointing to the baseline. Note that the *aecd* alias is inappropriate in this case.

This can be done automatically for every change, by placing the line

```
develop_begin_command =
        "ln -s $baseline bl";
```

into the project configuration file.

10.3. User Setup

There are a number of things which users of aegis can do to make it more useful, or more user friendly. This section describes just a few of them.

10.3.1. The .cshrc or .profile files

The aliases for the various user commands used throughout this manual are obtained by appending a line of the form

```
. /usr/local/share/aegis/profile
```

to the *.profile* file in the user's home directory, if they use the *sh*(1) shell or the *bash*(1) shell.

If the user uses the *csh*(1) shell, append a line of the form

```
source /usr/local/share/aegis/cshrc
```

to the *.cshrc* file in the user's home directory.

These days, many systems also provide an */etc/profile.d* directory, which has symbolic links to the start-up scripts for various packages. These are run automatically for all users. If your system has such a thing, arrange for symbolic links

```
ln -s /usr/local/share/aegis/profile \
        /etc/profile.d/aegis.sh
ln -s /usr/local/share/aegis/cshrc \
        /etc/profile.d/aegis.csh
```

and you will not need to edit every user's *.cshrc* or *.profile* file.

10.3.2. The AEGIS_PATH environment variable

If users wish to use aegis for their own projects, in addition to the "system" projects, the *AEGIS_PATH* environment variable forms a colon separated search path of aegis "library" directories. The */usr/local/lib/aegis* directory is always implicitly added to this list.

The user should not create this library directory, but let aegis do this for itself (otherwise you will get an error message).

The *AEGIS_PATH* environment variable should be set in the *.cshrc* or *.profile* files in the user's home directory. Typical setting is

```
setenv AEGIS_PATH ~/lib/aegis
```

and this is the default used in the */usr/local/share/aegis/cshrc* file.

[20] A hard link uses fewer disk blocks. Symbolic links survive the subject file being deleted and recreated.

10.3.3. The .aegisrc file

The *.aegisrc* file in the user's home directory contains a number of useful fields. See *aeuconf*(5) for more information.

10.3.4. The defaulting mechanism

In order for you to specify the minimum possible information on the command line, aegis has been designed to work most of it out itself.

The default project is the project which you are working on changes for, if there is only one, otherwise it is gleaned from the *.aegisrc* file. The command line overrides any default.

The default change is the one you are working on within the (default or specified) project, if there is only one. The command line overrides any default.

10.4. The Project Owner

For the greatest protection from accidental change, it is best if the project is owned by a UNIX account which is none of the staff. This account is often named the same as the project, or sometimes there is a single umbrella account for all projects.

When an aegis project is created, the owner is the user creating the project, and the group is the user's default group. The creating user is installed as the project's first administrator.

A new project administrator should be created − an actual user account. The UNIX password should then be disabled on the project account − it will never be necessary to use it again.[21]

The user nominated as project administrator many then assign all of the other staff roles. Aegis takes care of ensuring that the baseline is owned by the project account, not any of the other staff, while development directories always belong to the developer (but the group will always be the project group, irrespective of the developer's default group).

All of the staff working on a project should be members of the project's group, to be able to browse the baseline, for reviewers to be able to review changes. This use of UNIX groups means that projects may be as secure or open as desired.

[21] Unless bugs in aegis corrupt the database, in which case repairs can be accomplished as the project account using a text editor.

10.5. USENET Publication Standards

If you are writing software to publish on USENET, a number of the source newsgroups have publication standards. This section describes ways of generating the following files, required by many of the newsgroups' moderators:

Makefile	How to build the distribution.
CHANGES	What happened for this distribution.
patchlevel.h	An identification of this distribution.

Each of these files may be generated from information known to aegis, with the aid of some fairly simple shell scripts.

10.5.1. CHANGES

Write this section.

Look in the *aux/CHANGES.sh* file included in the aegis distribution for an example of one way to do this.

10.5.2. Makefile

Write this section.

Look in the *aux/Makefile.sh* and *aux/Makefile.awk* files included in the aegis distribution for an example of one way to do this.

10.5.3. patchlevel.h

Write this section.

Look in the *aux/Howto.cook* file included in the aegis distribution for an example of one way to do this.

10.5.4. Building Patch Files

The *patch* program by Larry Wall is one of the enduring marvels of USENET. This section describes how to build input files for this miracle program.

Write this section.

Look in the *aux/patches.sh* file included in the aegis distribution for an example of one way to do this.

10.6. Heterogeneous Development

The aegis program has support for heterogeneous development. It will enforce that each change be built and tested on each of a list of architectures. It determines which architecture it is currently executing on by using the *uname*(2) system call.

The *uname*(2) system call can yield uneven results, depending on the operating systems vendor's interpretation of what it should return[22]. To cope with this, each required architecture for a project is specified as a name and a pattern.

The name is used by aegis internally, and is also available in the *${ARCHitecture}* substitution (see *aesub*(5) for more information).

The patterns are simple shell file name patterns (see *sh*(1) for more information) matched against the output of the *uname*(2) system call.

The result of *uname*(2) has four fields of interest: *sysname*, *release*, *version* and *machine*. These are stitched together with hyphens to form an architecture *variant* to be matched by the pattern.

For example, a system the author commonly uses is "SunOS-4.1.3-8-sun4m" which matches the "SunOS-4.1*-*-sun4*" pattern. A solaris system, a very different beast, matches the "SunOS-5.*-*-sun4*" pattern. Sun's 386 version of Solaris matches the "SunOS-5.*-*-i86pc" pattern. A convex system matches the "ConvexOS-*-10.*-convex" pattern.

10.6.1. Project *aegis.conf* File

To require a project to build and test on each of these architectures, the *architecture* field of the project *aegis.conf* file is set. See *aepconf*(5) for more details on this file. The above examples of architectures could be represented as

```
architecture =
[
  {
    name = "sun4";
    pattern = "SunOS-4.1*-*-sun4*";
  },
  {
    name = "sun5";
    pattern = "SunOS-5.*-*-sun4*";
  },
  {
    name = "sun5pc";
    pattern = "SunOS-5.*-*-i86pc";
  },
  {
    name = "convex";
    pattern = "ConvexOS-*-10.*-*";
  }
];
```

This would require that all changes build and test on each of the "sun4", "sun5", "sun5pc" and "convex" architectures.

It is also possible to have *optional* architectures. This may be used to recognise an environment, but not mandate that it be built every time.

```
  {
    name = "solaris-8-sparc";
    pattern = "SunOS-5.8*-*-sun4*";
    mode = optional;
  },
```

However, once an architecture name appears in a change's architecture list, it is mandatory for that change.

If the *architecture* field does not appear in the project *aegis.conf* file, it defaults to

```
architecture =
[
  {
    name = "unspecified";
    pattern = "*";
  }
];
```

Setting the architectures is usually done as part of the first change of a project, but it also may be done to existing projects. This information is kept in the project *aegis.conf* file, rather than as a project attribute, because it requires that the DMT configuration file and the tests have corresponding details (see below).

The *lib/config.example/architecture* file in the Aegis distribution contains many architecture variations, so that you may readily insert them into your project configuration file.

10.6.2. Change Attribute

The *architecture* attribute is inherited by each new change. A project administrator may subsequently edit the change attributes to grant exemptions for specific architectures. See *aeca*(1) for how to do this.

A build must be successfully performed on each of the target architectures. Similarly, the tests must be performed successfully on each. These

[22] For example, SCO 3.2 returns the nodename in the sysname field, when it should place "SCO" there; Convex and Pyramid scramble it even worse.

requirements are because there is often conditional code present to cope with the vagaries of each architecture, and this needs to be compiled and tested in each case.

This multiple build and test requirement includes both development and integration states of each change.

10.6.3. Network Files

This method of heterogeneous development assumes that the baseline and development directories are available as the same pathname in all target architectures. With software such as NFS, this does not present a great problem, however NFS locking must also work.

There is also an assumption that all the hosts remotely mounting NFS file systems will agree on the time, because aegis uses time stamps to record that various tasks have been performed. Software such as *timed*(8) is required[23].

10.6.4. DMT Implications

This method of heterogeneous development assumes that the baseline will have a copy of all object files for all target architectures *simultaneously*.

This means that the configuration file for the DMT will need to distinguish all the variations of the object files in some way. The easiest method is to have a separate object tree for each architecture[24]. To facilitate this, there is an *${ARCHitecture}* substitution available, which may then be passed to the DMT using the *build_command* field of the project *aegis.conf* file.

The architecture name used by aegis needs to be used by the DMT, so that both aegis and the DMT can agree on which architecture is currently targeted.

10.6.4.1. Cook Example

As and example of how to do this, the cook recipes from the DMT chapter are modified as appropriate. First, the *build_command* field of the project *aegis.conf* file is changed to include the *${ARCHitecture}* substitution:

```
build_command =
  "cook -b ${s Howto.cook} \
  project=$p change=$c \
  version=$v arch='$arch' -nl";
```

Second, the C recipe must be changed to include the architecture in the path of the result:

```
[arch]/%.o: %.c: [collect c_incl
  -eia [prepost "-I" ""
  [search_list]] [resolve %.c]]
{
  if [not [exists [arch]]] then
    mkdir [arch]
      set clearstat;
  if [exists [target]] then
    rm [target]
      set clearstat;
  [cc] [cc_flags] [prepost "-I"
    "" [search_list]] -c
    [resolve %.c];
  mv %.o [target];
}
```

Third, the link recipe must be changed to include the architecture in the name of the result:

```
[arch]/example: [object_files]
{
  if [not [exists [arch]]] then
    mkdir [arch]
      set clearstat;
  if [exists [target]] then
    rm [target]
      set clearstat;
  [cc] -o [target] [resolve
    [object_files]] -ly -ll;
}
```

The method used to determine the `object_files` variable is the same as before, but the object file names now include the architecture:

```
object_files =
  [fromto %.y [arch]/%.o
    [match_mask %.y [source_files]]]
  [fromto %.l [arch]/%.o
    [match_mask %.l [source_files]]]
  [fromto %.c [arch]/%.o
    [match_mask %.c [source_files]]]
  ;
```

Note that the form of these recipes precludes performing a build in each target architecture simultaneously, because intermediate files in the recipes may clash. However, aegis prevents simultaneous build, for this and other reasons.

10.6.5. Test Implications

Tests will need to know in which directory the relevant binary files reside. The *test_command* field of the project *aegis.conf* file may be changed from

[23] Some sites manage by running *rdate*(8) from *cron*(8) every 15 minutes.

[24] A tree the same shape as the source tree makes navigation easier, and users need not think of file names unique across all directories.

the default

```
test_command =
  "$shell $file_name";
```

to pass the architecture name to the test

```
test_command =
  "$shell $file_name $arch";
```

This will make the architecture name available as `$1` within the shell script. Tests should fail elegantly when the architecture name is not given, or should assume a sensible default.

10.6.6. Cross Compiling

If you are cross compiling to a number of different target architectures, you would not use aegis' heterogeneous development support, since it depends on the *uname*(2) system call, which would tell it nothing useful when cross compiling. In this case, simply write the DMT configuration file to cross compile to all architectures in every build.

10.6.7. File Version by Architecture

There is no intention of ever providing the facility where a project source file may have different versions depending on the architecture, but all of these versions overload the same file name[25].

The same effect may be achieved by naming files by architecture, and using the DMT to compile and link those files in the appropriate architecture.

This has the advantage of making it clear that several variations of a file exist, one for each architecture, rather than hiding several related but independent source files behind the one file name.

10.7. Reminders

This section documents some scripts available for reminding users of changes which require their attention. These scripts are installed into the */usr/local/share/aegis/remind* directory, and may be run by *cron*(8) at appropriate intervals. You will almost certainly want to customize them for your site.

10.7.1. Awaiting Development

The */usr/local/share/aegis/remind/awt_dvlp.sh* script takes a project name as argument. It is placed in the project leader's per-user crontab. It is suggested that this script be run weekly, at 8AM on Monday. This script will send all developers of the named project email if there are any changes in the *awaiting development* state in the named project. No mail is sent if there are no changes outstanding.

10.7.2. Being Developed

The */usr/local/share/aegis/remind/bng_dvlpd.sh* script takes no arguments. It is placed in each user's per-user crontab. It is suggested that this script be run weekly, at 8AM on Monday. This script takes no arguments, and sends email to the user if they have any changes in the *being developed* or *being integrated* states. No mail is sent if there are no changes outstanding.

10.7.3. Being Reviewed

The */usr/local/share/aegis/remind/bng_rvwd.sh* script takes a project name as argument. It is placed in the project leader's per-user crontab. It is suggested that this script be run daily at 8AM. This script will send all reviewers of the named project email if there are any changes in the *being reviewed* state in the named project. No mail is sent if there are no changes outstanding.

10.7.4. Awaiting Integration

The */usr/local/share/aegis/remind/awt_intgrtn.sh* script takes a project name as argument. It is placed in the project leader's per-user crontab. It is suggested that this script be run daily at 8AM. This script will send all integrators of the named project email if there are any changes in the *awaiting integration* state in the named project. No mail is sent if there are no changes outstanding.

[25] Some other SCM tools provide a repository with this facility.

11. Geographically Distributed Development

This chapter describes various methods of collaboratively developing software using Aegis, where the collaborating sites are separated by administrative domains or even large physical distances.

While many Open Source projects on the Internet typify such development, this chapter will also describe techniques suitable for commercial enterprises who do not wish to compromise their intellectual property.

11.1. Introduction

The core of the distribution method is the *aedist*(1) command. In its simplest form, the command

```
aedist −send | aedist −receive
```

will clone a change set locally. This may appear less than useful (after all, the *aeclone*(1) command already exists) until you consider situations such as

```
aedist −send | e−mail | aedist −receive
```

where *e-mail* represents the sending, transport and receiving of e-mail. In this example, the change set would be reproduced on the e-mail recipient's system, rather than locally. Similar mechanisms are also possible for web distribution.

11.1.1. Risk Reduction

Receiving change sets in the mail, however, comes with a number of risks:

- You can't just commit it to your repository, because it may not even compile.
- Even if it does compile, you want to run some tests on it first, to make sure it is working and doesn't break anything.
- Finally, you would always check it out, to make sure it was appropriate, and didn't do more subtle damage to the source.

While these are normal concerns for distributing source over the Internet, and also internally within companies, they are the heart of the process employed by Aegis. All of these checks and balances are already present. The receive side simply creates a normal Aegis change, and applies the normal Aegis process to it.

- The change set format is unpacked into a private work area, not directly into the repository. This is a normal Aegis function.
- The change set is then confirmed to build against the repository. All implications flowing from the change are exercised. Build inconsistencies will flag the change for attention by a human, and the change set will not be committed to the repository. This is a normal Aegis function.
- The change set is tested. If it came accompanied by tests, these are run. Also, relevant tests from the repository are run. Test inconsistencies will flag the change for attention by a human, and the change set will not be committed to the repository. This is a normal Aegis function.
- Once the change set satisfies these requirements, it must still be reviewed by a human before being committed, to validate the change set for suitability and completeness. This is a normal Aegis function.

11.1.2. What to Send

While there are many risks involved in receiving change sets, there also problems in figuring out what to send.

At the core of Aegis' design is a transaction. Think of the source files as rows in a database table, and each change-set as a transaction against that table. The build step represents maintaining referential integrity of the database, but also represents an input validation step, as does the review. And like databases, the transactions are all-or-nothing affairs, it is not possible to commit "half" a transaction.

As you can see, Aegis changes are already elegantly validated, recorded and tracked, and ideally suited to being packaged and sent to remote repositories.

11.1.3. Methods and Topologies

In distributed systems such as described in this chapter, there are two clear methods of distribution:

- The "push" method has the change set producer automatically send the change-set to a registered list of interested consumers. This is supported by Aegis and *aedist*.
- The "pull" method has the change set producer make the change sets available for interested consumers to come and collect. This is supported by Aegis and *aedist*.

These are two ends of a continuum, and it is possible and common for a mix-and-match approach to be taken.

There are also many ways of arranging how distribution is accomplished, and many of the distribution arrangements (commonly called topologies, when you draw the graphs) are supported by Aegis and *aedist*:

- The star topology has a central master repository, surrounded by contributing satellite repositories. The satellites are almost always "push" model, however the central master could be either "push" or "pull" model.
- The snowflake topology is like a hierarchical star topology, with contributors feeding staging posts, which eventually feed the master repository. Common for large Open Source Internet projects. Towards the master repository is almost always "push" model, and away from the master is almost always "pull" model.
- The network topology is your basic anarchic autonomous collective, with change sets flying about peer-to-peer with no particular structure. Often done as a "push" model through an e-mail mailing list.

All of these topologies, and any mixture you can dream up, are supported by Aegis and *aedist*. The choice of the right topology depends on your project and your team.

11.1.4. The Rest of this Chapter

Aegis is the ideal medium for hosting distributed projects, for all the above reasons, and the rest of this chapter describes a number of different ways of doing this:

□ The second section will describe how to perform these actions manually, both send and receive, as this demonstrates the method efficiently, and represents a majority of the use made of the mechanism.

□ The third section will show how to automate e-mail distribution and receipt. Automated e-mail distribution is probably the next most common use.

□ The fourth section will show how to configure distribution and receipt using World Wide Web servers and browsers.

□ The fifth section deals with security issues, such as validating messages and coping with duplicate storms.

11.2. Manual Operation

This section describes how to use *aedist* manually, in order to send and receive change sets.

11.2.1. Manual Send

In order to send a change set to another site, it must be packaged in a form which captures all of the change's attributes and the contents of the change's files. This package must be compressed and encoded in a form which will survive the various mail transport agents it must pass through, and then given to the local mail transport agent. This is done by a single command

```
% aedist -send -c number | \
  mail joe.blow@example.com
%
```

All of the usual Aegis command line options are available, so you could specify the project on the command line if you needed to.

This command will send the sources from the development directory, if the change is not yet completed. This is useful for collaboration between developers, but it isn't reviewed and integrated, so beware.

It is more usual to send a change which has been completed. In this case the version of the file which was committed is sent. If necessary, the history files will be consulted to reconstruct this information. See the "*Automatic Send*" section, below, for more discussion of this.

There are many options for customizing the e-mail message sent to `joe.blow@example.com`, see *aedist*(1) for more information.

11.2.2. Sending Baselines

In order to send the entire contents of the repository to someone, you use a very similar command:

```
% aedist -send -baseline | \
  mail joe.blow@example.com
%
```

This can be a very large change set, because it is all files of the project.

11.2.3. Sending Branches

There are times when remote developers are not interested in a blow-by-blow update of your repository. Instead they want to have updates from time to time. In order to send them the current state of your active development branch, in this example "*example.4.2*", you would use a

command of the form

```
% aedist -send -p example.4 -c 2 | \
  mail joe.blow@example.com
%
```

Notice how the correspondence between branches and change sets is exploited. The baseline of a branch is the development directory of the “super change” is represents.

Branch change sets like this are smaller than whole baselines, because they include only the files altered by this branch, rather then the state of every file in the project.

11.2.4. Manual Receive

The simplest form of receiving a change set is to save it from your e-mail program into a file, and then

```
% aedist -receive -file filename
...lots of information...
%
```

where *filename* is where you saved the e-mail message. If your e-mail program is able to write to a pipe, you can use a simpler form. This example uses the Rand Mail Handler’s *show*(1) command:

```
% show | aedist -receive
...lots of information...
%
```

Each of these examples assumes that you have used the same project name locally as that of the sender (it’s stored in the package). If this isn’t the case, you will need to use the `-project` option to tell *aedist* which project to apply the change to.

The actions performed by *aedist* on receive are not quite a mirror of what it does on send. In particular, send usually extracts its information from the repository, but receive **does not** put the change set directly into the repository.

On receipt of a change set, *aedist* creates a new change with its own development directory, and unpacks the change set into it, in much the same way as a change would normally be performed by a developer. (Indeed, the receiver must be an authorized developer.)

Once the change is unpacked, it goes through the usual development cycle of build, difference and test. If any portion of this fails, *aedist* will stop with a suitable error message. If all goes well, development of the change will end, and it will be left in the *being reviewed* state.

At this point, a local reviewer must examine the change, and it proceeds through the change integration process as normal.

If there is a problem with the change, it can be dealt with as you would with any other defective change − by developing it some more. Or you can email the sender telling them the problem and use *aedbu*(1) and *aencu*(1) to entirely discard the change.

11.2.5. Getting Started

In order to receive a change, you must have a project to receive it into. Also, changes tend to be the *difference* between an existing repository and what it is to become. You need some way to get the starting point of the differences before you can apply any differences. This section describes one way of doing this.

You start by creating a normal Aegis project in the usual way. That is covered earlier in this User Guide. It helps greatly if you give your local project exactly the same name as the remote project. It doesn’t need the same pathname for the project directory, just the same project name.

Once you have this project created, request the remote repository send you a “baseline” change, as described above. Once you have received this, and it is integrated successfully, you are ready to receive and apply change sets. This is an inherently “pull” activity, as the source may never have heard of you before. The initial baseline may arrive by e-mail, or floppy disk, or you may retrieve it from the web, it all depends how the project is being managed.

You will be warned about "potential trojan horse" files in the baseline change set. This is normal, because you are receiving all project configurations file, build files and test files. All of these contain executable commands *that will be executed*. Caveat emptor. Make sure you trust the source.

11.3. Sneaker Net

Another common method of transporting data, sometimes a quite large amount of it, is to write it onto transportable media and carry it.

To write a change set onto a floppy, you would use commands such as

```
% mount /mnt/floppy
% aedist -send -no-base64 \
  -o /mnt/floppy/change.set
% umount /mnt/floppy
%
```

The above command assumes the floppy is pre-formatted, and that there is a user-permitting line in the */etc/fstab* file, as is common for many Linux distributions. The *change.set* can be any filename you like, but is usually project-name and change-number related.

It takes a very sizable change set to fail to fit on a 1.44MB floppy, because they are compressed (and change sets exceeding 8MB of source are rare, even for huge projects). The `-no-base64` option is used to avoid the MIME base 64 encoding, which is necessary for e-mail, not not necessary in this case. The receive side will automatically figure out there is no MIME base 64 encoding.

Reading the change set is just as simple, as it closely follows the other commands for receiving commands sets.

```
% mount /mnt/floppy
% aedist -rec -f /mnt/floppy/change.set
...lots of output...
% umount /mnt/floppy
%
```

This technique will work for any of the disks available these days including floppies, Zip, Jaz, *etc*.

11.4. Automatic Operation

This section describes how to use *aedist* to automatically send change sets via e-mail.

11.4.1. Sending

Change sets can be sent automatically when a change passes integration. You do this by setting the *integrate_pass_notify_command* field of the project attributes.

In this example, the “example” project sends all integrations to all the addresses on the `example-developers` mailing list. (The mailing list is maintained outside of Aegis, *e.g.* by Majordomo.) The relevant attribute (edited by using the *aepa*(1) command) looks like this:

```
integrate_pass_notify_command =
  "aedist -p $project -c $change | \
  mail example-users";
```

Please note that project attributes are inherited by branches when they are created. If you don’t want all branches to broadcast all changes, you need to remember to clear this project attribute *from the branch* once the branch has been created. Alternatively, use the `$version` substitution to decide who to send the change to.

11.4.2. Receiving

write this section

You need to set up an e-mail alias, with a wrapper around it − you probably *don’t* want "daemon" as a registered developer.

While *aedist*(1) makes every attempt to spot potential trojan attacks, you really, *really* want PGP validation (or similar industrial strength digital signatures) before you accept this kind of input.

11.5. World Wide Web

This section describes how to use *aeget*(1) and *aedist*(1) to transport change sets using the World Wide Web. This requires configuration of the web server to package and send the change sets, and configuration of the browser to receive and unpack the change sets. You can also automatically track a remote site, efficiently downloading and applying new change sets as they appear.

11.5.1. Server

Aegis has a read-only web interface to its database, it is a web server CGI interface. If you are running Apache, or similar, all you have to do is copy (or symlink, if you have symlinks enabled) the */usr/local/bin/aeget* file into the web server’s *cgi-bin* directory. For example, the default Apache install would need the following command:

```
ln -s /usr/local/bin/aeget /var/www/cgi-bin/aeget
```

11.5.2. Browser

You need to set the appropriate mailcap entry, so that `application/aegis-change-set` is handled by *aedist −receive*.

Edit the */etc/mailcap* file, and add the lines

```
# Aegis
application/aegis-change-set;/usr/local/bin/aedist -re
```

You may need to restart your web browser for this to take effect.

11.5.3. Hands-Free Tracking

Clients of sites using a web server, such as the various developers in an open sourec project, it is possible to automatically "replay" change sets on the server which have not yet been incorporated at your site.

The command

```
aedist −replay −f name-of-web-server
```

will automatically download any remote change sets not present in the local repository. It downloads them by using the *aedist*(1) command. It uses commands of the form

```
aedist −receive −f url-of-change-set
```

to download the change sets, which have to go through all of the usual Aegis process before vecoming part of your local repository. This includes code review, unless you have configured the *develop_end_action* field of the project configuration to be *goto_awaiting_development*.

If you add this command to a *crontab(1) entry, you can check to see if there are change sets to synchronize with once a day, or however often you set the line to run.*

11.6. Security

This section deals with security issues. Security isn't just "keep the bad guys out", it actually covers *availability*, *integrity* and *confidentiality*.

Availability:
: refers to the system being available for use by authorised users. Denial of service and crashes are examples of bad things in this area.

Integrity:
: refers to the system being in a known good state. Corrupted change sets and un-buildable repositories are examples of bad things in this area.

Confidentiality:
: refers to the system being available *only* to authorised users. For many Open Source projects, this isn't a large concern, but for corporate users of Aegis, non-disclosure of change-sets as they cross the Internet is a serious requirement.

As you can see, a strategy of "keep the bad guys out" is necessary, but not sufficient, to satisfy security.

This section covers the above security issues, as applied to the use of *aedist* to move change sets around.

11.6.1. Trojan Horses

"A Trojan horse is an apparently useful program containing hidden functions that can exploit the privileges of the user [running the program], with a resulting security threat. A Trojan horse does things that the program user did not intend[26]."

In order to forestall this threat, *aedist* will cease development of the change if it detects the potential for a Trojan horse. These include...

- Changing the project *aegis.conf* file. This file contains the build command and the difference commands, both of which would be run before a reviewer had a chance to confirm they were acceptable.
- Changing any of the files named in the *trojan_horse_suspect* field of the project *aegis.conf* file. This lets you cover things like the build tool's configuration file (*e.g.* the Makefile or the cookbook), and any scripts or code generators which would be run by the build.

This isn't exhaustive protection, and it depends on keeping the *trojan horse suspect* list up-to-date. (It accepts patterns, so it's not too onerous.) For better protection, you need to validate the sender and the message.

11.6.2. PGP

PGP can be used to validate that the source of a change set distribution is really someone you trust.

anyone want to advise me what to put here?

11.6.3. Sorcerer's Apprentice

In a push system, with a central master server and a collection of contributors, all of which are using automatic send, as described above, a potential explosion of redundant messages is possible. Viz:

- Contributor integrates a change, which is dispatched to the master server.
- Maintainer integrates the change set into the master repository.
- Master repository automatically dispatches the change set to all of the contributors.

[26] Summers, Rita C., *Secure Computing Threats and Safeguards*, McGraw-Hill, 1997.

- Each of the contributors receives and integrates the change, each of which are dispatched to the master server.
- The master server is inundated with change sets it already has.
- If these change sets were to be integrated, the storm repeats, growing exponentially every time it goes around the loop.

To prevent this, *aedist* does several things...

- Before the change is built, an *aecpu −unchanged* is run. If there is nothing left, the change is abandoned, because you already have it. (This doesn't always work, because propagation delays may try to *reverse* a subsequent local change.)
- When a change set is sent, an RFC 822 style header is added to the description. This includes From and Date. When a change set is received, a Received line is added. Too many Received lines causes the change set to be dropped − for a star topology the maximum is 2. (This doesn't always work, because the description could be edited to rip it off again.) (This doesn't always work, because the maintainer may edit it in some ways before comitting it, and forget to rip off (enough of) the header.) (This doesn't always work, because hierarchical topologies will group change sets together.) (This doesn't always work, because a baseline pull will group change sets together.)
- Set the description to indicate it was received by aedist? Use this to influence the decision to send it off again at integrate pass? How?

11.7. Patches

In the open source community, patches are common way of sharing enhancements to software. This was particularly common before the World Wide Web, and usenet was the more common medium of distribution. Patches also have the advantage of being fairly small and usually tansportable by email with few problems.

11.7.1. Send

If you are participating in an open source project, and using Aegis to manage your development, the *aepatch −send* command may be used to construct a patch to send to the other developers.

It is very similar in operation to the *aedist*(1) command, however it is intended for non-Aegis-using recipients.

To send a change to someone (a completed change, or one in progress) simply use a command such as

```
% aepatch −send −c number | \
  mail joe.blow@example.com
%
```

to send your change as a patch. Note that it will be compressed (using GNU Zip) and encoded (using MIME base 64), which produces small files which are going to survive email transport.

11.7.2. Receive

The simplest way of receiving a patch and turn it into a change set is to save it from your e-mail program into a file, and then

```
% aepatch −receive −file filename
...lots of information...
%
```

where *filename* is where you saved the e-mail message. If your e-mail program is able to write to a pipe, you can use a simpler form. This example uses the Rand Mail Handler's *show*(1) command:

```
% show | aepatch −receive
...lots of information...
%
```

Each of these examples assumes that you have already set the project name, either via *aeuconf*(5) or *ae_p*(1), or you could use the *−project* option.

The actions performed by *aepatch* on receive are not quite a mirror of what it does on send. In particular, send usually extracts its information from the repository, but receive **does not** put the change set directly into the repository.

On receipt of a change set, *aepatch* creates a new change with its own development directory, copies the files into it, and applies the patch to the files. The receiver must be an authorized developer.

Once the patch is applied, it goes through the usual development cycle of build, difference and test. If any portion of this fails, *aepatch* will stop with a suitable error message. If all goes well, development of the change will end, and it will be left in the *being reviewed* state.

At this point, a local reviewer must examine the change, and it proceeds through the change integration process as normal.

If there is a problem with the change, it can be dealt with as you would with any other defective change − by developing it some more. Or you

can email the sender telling them the problem and use *aedbu*(1) and *aencu*(1) to entirely discard the change.

11.7.3. Limitations

Despite a great similarity of command line operations and operation, the *aepatch* command should **not** be thought of as an equivalent for the *aedist* command, or a replacement for it.

The information provided by *aedist −send* is sufficiently complete to recreate the change set at the remote end. No information is lost. In contrast, the *aepatch −send* command is limited to that information a patch file (see the *patch*(1) command, from the GNU Diff utils). There is no guarantee that the *aepatch −send* output will be given to *aepatch −receive*; it must work with *patch*(1), and similar tools.

Conversely, there is no guarantee that the input to *aepatch −receive* came from *aepatch −send*. It can and must be able to cope with the outout of a simple *diff −r −N −c* command, with no additional information.

All this means, use *aedist* wherever possible. The *aepatch* command is to simplify and streamline communication with non-Aegis developers.

12. Further Reading

This chapter contains information about books, articles or web sites relevant to some aspect of Aegis or using Aegis. These references should not be taken as endorements.

If I've missed a good reference, it isn't personal, but I can't and haven't read everything out there. Email me the information and I'll add it to this chapter (no advertising, please).

12.1. Software Configuration Management

Eaton, D. (1995), Configuration Management Frequently Asked Questions, http://www.dav-eeaton.com/scm/CMFAQ.html

This is an essential first-stop for information about Software Configuration Management. It has an excellent book list.

Pool, D., CM Today, http://www.cmtoday.com/

This is a configuration management portal site, with news and other links.

12.2. Reviewing

Baldwin, J. (1992), An Abbreviated C++ Code Inspection Checklist, http://www2.ics.hawaii.edu/~johnson/FTR/Bib/Baldwin92.html

This web page talks about C++ code inspections with some useful suggestions about how to conduct (rather formal) reviews and some for C++ constructs to watch out for.

13. Appendix A: New Project Quick Reference

For those of you too impatient to read a whole great big document about how to use the aegis program, this appendix gives a quick look at how to place a project under aegis.

The style here is an itemized list. It does not try to be exhaustive. For exact details on how to use the various aegis commands, you should see the manual pages, ditto for the formats and contents of some files.

Probably the quickest start of all is to copy an already existing project. The project used in chapter 2 is complete, assuming you use the author's "cook" dependency maintenance tool. The entirety of this example may be found, if slightly obfuscated, in the aegis source file *test/00/t0011a.sh* distributed with aegis.

13.1. Create the Project

The *aenpr* command is used to create a project. You must supply the name on the command line. The name should be ten characters or less, six characters or less if you want version numbers included.

The user who creates the project is the owner of the project, and is set as the administrator. The default group of the user who created the project is used as the project's group.

You may want to have a user account which owns the project. You must create the project as this user, and then use the *aena* and *aera* commands to add an appropriate administrator, and remove the owning user as an administrator. After this, the password for the owning user may be disabled, because the aegis program will, at appropriate times, set file ownership to reflect project ownership or execute commands on behalf of the project owner *as* the project owner.

13.1.1. Add the Staff

The *aend* command is used to add developers. The *aenrv* command is used to add reviewers. The *aeni* command is used to add integrators. These commands may only be performed by a project administrator.

You will still have to do this, even if the person who created the project will be among these people, or even be all of these people.

13.1.2. Project Attributes

The *aepa* command is used to change project attributes. These attributes include the description of the project, and booleans controlling whether, for example, developers may review their own work.

The project attributes file is described in the *aepattr*(5) manual entry.

13.2. Create Change One

The *aenc* command is used to create a new change. You will need to construct a change attributes file with your favorite text editor before running this command.

The change attributes file is described in the *aecattr*(5) manual entry.

13.3. Develop Change One

This is the most grueling step. Indeed, the integration step will probably reveal things you missed, and you may return to the *being developed* state several times.

One of the people you nominated as a developer will have to use the *aedb* command to commence development of the first change. The *aecd* command can be used to change directory into the just-created development directory.

Add files to the change. The *aenf* command is used to create new files. If you don't use *aenf* then the aegis program has no way of knowing whether that file lying there in the development directory is significant to the project, or just a shopping list of the groceries you forgot to buy yesterday.

One particular new file which *must* be created by this change is the project configuration file, usually called *aegis.conf* but can be named something else. This file tells Aegis what history mechanism you wish to use, what dependency maintenance command to use, what file difference tools to use, and much more. The *aepconf*(5) manual entry describes this file.

If you are going to use the "cook" dependency maintenance tool, another new file you will need to create in this change is the "Howto.cook" file. Some other tool will want some other rules file.

You probably have a prototype or some other "seed" you have sort-of working. Create new files for each source file and *then* copy the files

from wherever they are now into the development directory.

Use the *aeb* command to build the change. It will need to build cleanly before it can advance to the next step.

Use the *aed* command to difference the change. It will need to difference cleanly before it can advance to the next step.

Use the *aent* command to add new tests to the command. It will need to have tests before it can advance to the next step.

Most existing projects don't have formal tests. These tests will form a regression test-bed, used to make sure that future changes never compromise existing functionality.

Use the *aet* command to test the change. It will need to test cleanly before it can advance to the next step.

Once the change builds, differences and tests cleanly, use the *aede* command to end development.

13.4. Review The Change

One of the people nominated as reviewers will have to run the *aerpass* command to say that the change passed review.

The aegis program does not mandate any particular review mechanism: you could use a single peer to do the review, you could use a panel, you could set the project so that developers may review their own work effectively eliminating the review step. In projects with as few as two people, it is always beneficial for someone other than the developer to review changes. It is even beneficial for the developer herself to review the next day.

Should a reviewer actually want to *see* the change, the *aecd* command may be used to change directory to the development directory of the change. The difference files all end with a "comma D" suffix, so the

```
more ‘find . -name "*,D" -print |
sort‘
```

command may be used to search them out and see them. This is probably fairly useless for the first change, but is vital for all subsequent changes. There is a supplied alias for this command, it is *aedmore* and there is a similar *aedless* alias if you prefer the *less*(1) command.

There are some facts that a reviewer *knows* because otherwise the change would not be in the "being reviewed" state: • the change compiles cleanly, • the change passes all of its tests. Other information about the change may be obtained using the "change_details" variation of the *ael* command.

The *aerfail* command may also be used by reviewers to fail reviews and return a change to the developer for further work; the reviewer must supply a reason for the change history to record for all time. Similarly, the *aedeu* command may be used by the developer to resume development of a change at any time before it is integrated; no stated reason is required.

13.5. Integrate the Change

A person nominated as an project integrator then integrates the change. This involves making a copy of the integration directory, applying the modifications described by the change to this integration directory, then building and testing all over again.

This re-build and re-test is to ensure that no special aspect of the developers environment influenced the success up to this point, such as a unique environment variable setting. The re-build also ensures that all of the files in the baseline, remembering that this includes source files and all other intermediate files required by the build process, remain consistent with each other, that the baseline is self-consistent. The definition of the baseline is that it passes its own tests, so the tests are run on the baseline.

Use the *aeib* command to begin integration.

The *aeb* command is used to build the integration copy of the change.

The *aet* command is used to test the integration copy of the change.

On later changes, the integration may also require the *aet −bl* command to test the change against the baseline. This tests ensures that the test *fails* against the baseline. This failure is to ensure that bug fixes are accompanied by tests which reproduce the bug initially, and that the change has fixed it. New functionality, naturally, will not be present in the old baseline, and so tests of new functionality will also fail against the old baseline.

Later changes may also have the regression tests run, using the *aet −reg* command. This can be a very time-consuming step for projects with a long history, and thus a large collection of tests. The *aet −suggest* command can also be used to run

"representative" sets of existing tests, but a full regression test run is recommended before a major release, or, say, weekly if it will complete over the weekend. This command is also available to developers, so that they have fewer surprises from irate integrators.

The integrator may use the *aeifail* command to return a change to its developer for further work; a reason must be supplied, and should include relevant excerpts from the build log in the case of a build failure (not the *whole* log!), or a list of the tests which failed for test failures.

The *aeipass* command may be used to pass an integration. When the change passes, the file histories are updated. In the case of the first change, the history is created, and problems with the project configuration file's history commands will be revealed at this point. The integration won't pass, and should be failed, so that the developer may effect repairs. There are rarely problems at this point for subsequent changes, except for disk space problems.

Once the history is successfully updated, aegis renames the integration directory as the baseline, and throws the old baseline away. The development directory is deleted at this time, too.

13.6. What to do Next

There, the first change is completed. The whole cycle may now be repeated, starting at "Create Change," for all subsequent changes, with very few differences.

It is recommended that you read the *Change Development Cycle* chapter for a full worked example of the first four changes of an example project, including some of the twists which occur in real-world use of aegis.

Remember, too, the definition:

> **aegis** (ee.j.iz) *n.* a protection, a defence.

It is not always the case that aegis exists to make life "easier" for the software engineers. The goal is to have a baseline which always "works", where "works" is defined as passing all of its own tests. Wherever possible, the aegis program attempts to be as helpful and as unintrusive as possible, but when the "working" definition is threatened, the aegis program intrudes as necessary. (Example: you can't do an integrate pass without the integration copy building successfully.)

All of the "extra work" of writing tests is a long-term win, where old problems never again reappear. All of the "extra work" of reviewing changes means that another pair of eyes sees the code and finds potential problems before they manifest themselves in shipped product. All of the "extra work" of integration ensures that the baseline always works, and is always self-consistent. All of the "extra work" of having a baseline and separate development directories allows multiple parallel development, with no inter-developer interference; and the baseline always works, it is never in an "in-between" state. In each case, not doing this "extra work" is a false economy.

14. Appendix B: Glossary

The following is an alphabetical list of terms used in this document.

administrator
: Person responsible for administering a *project*.

awaiting_development
: The state a change is in immediately after creation.

awaiting_integration
: The state a change is in after it has passed review and before it is integrated.

awaiting review
: An optional state a change is in after it is developed, but before someone has chosen to review it..

baseline
: The repository; where the project master source is kept.

being developed
: The state a change is in when it is being worked on.

being integrated
: The state a change is in when it is being integrated with the baseline.

being reviewed
: The state a change is in after it is developed.

change
: A collection of files to be applied as a single atomic alteration of the baseline.

change number
: Each *change* has a unique number identifying it.

completed
: The state a change is in after it has been integrated with the baseline.

delta number
: Each time the *aeib*(1) command is used to start integrating a *change* into the *baseline* a unique number is assigned. This number is the delta number. This allows ascending version numbers to be generated for the baseline, independent of change numbers, which are inevitably integrated in a different order to their creation.

dependency maintenance tool
: A program or programs external to aegis which may be given a set of rules for how to efficiently take a set of source files and process them to produce the final product.

DMT
: Abbreviation of Dependency Maintenance Tool.

develop_begin
: The command issued to take a change from the *awaiting development* state to the *being developed* state. The change will be assigned to the user who executed the command.

develop_begin_undo
: The command issued to take a change from the *being developed* state to the *awaiting development* state. Any files associated with the change will be removed from the development directory and their changes lost.

develop_end
: The command issued to take a change from the *being developed* state to the *being reviewed* state, or optionally to the *awaiting reviewed* state. The change must be known to build and test successfully.

develop_end_undo
: The command issued to take a change from the *being reviewed* state back to the *being developed* state. The command must be executed by the original developer.

developer
: A member of staff allowed to develop changes.

development directory
: Each change is given a unique development directory in which to edit files and build and test.

history tool
: A program to save and restore previous versions of a file, usually by storing edits between the versions for efficiency.

integrate_pass
: The command used to take a change from the *being integrated* state to the *completed* state. The change must be known to build and test successfully.

integrate_begin
: The command used to take a change from the *awaiting integration* state to the *being integrated* state.

integrate_begin_undo
: The command used to take a change from the *being integrated* state to the *awaiting integration* state.

integrate_fail
: The command used to take a change from the *being integrated* state back to the *being developed* state.

integration
: The process of merging the *baseline* with the *development directory* to form a new baseline. This includes building and testing the merged directory, before replacing the original *baseline* with the new merged version.

integration directory
: The directory used during *integration* to merge the existing *baseline* with a change's *development directory*.

integrator
: A staff member who performs *integrations.*

new_change
: The command used to create new changes.

new_change_undo
: The command used to destroy changes.

review_begin
: The command used to take a change from the *awaiting review* state to the *being reviewed* state.

review_fail
: The command used to take a change from the *being reviewed* state back to the *being developed* state.

review_pass
: The command used to take a change from the *being reviewed* state to the *awaiting integration* state.

reviewer
: A person who may review *changes* and either pass or fail them (*review_pass* or *review_fail* respectively).

state
: Each *change* is in one of seven states: *awaiting development*, *being developed*, *awaiting review*, *being reviewed*, *awaiting integration*, *being integrated* or *completed*.

state transition
: The event resulting in a *change* changing from one state to another.

15. Appendix D: Why is Aegis Set-Uid-Root?

The goal for aegis is to have a project that "works". There is a fairly long discussion about this earlier in this User Guide. One of the first things that must be done to ensure that a project is not subject to mystery break downs, is to make sure that the master source of the project cannot be in any way altered in an unauthorized fashion. Note this says "cannot", a stronger statement than "should not".

Aegis is more complicated than, say, set-group-id RCS, because of the flaw with set-group-id: the baseline is writable by the entire development team, so if a developer says "this development process stinks" he can always bypass it, and write the baseline directly. This is a *very* common source of project disasters. To prevent this, you must have the baseline read-only, and so the set-group-id trick does not work. (The idea here is that there is *no* way to bypass the QA portions of the process. Sure, set-group-id will prevent accidental edits on the baseline, if the developers are not members of the group, but it does not prevent *deliberate* checkin of unauthorized code. Again, the emphasis is on "cannot" rather than "should not".)

Also, using the set-group-id trick, you need multiple copies of RCS, one for each project. Aegis can handle many projects, each with a different owner and group, with a single set-uid-root executable.

Aegis has no internal model of security, it uses UNIX security, and so becomes each user in turn, so UNIX can determine the permissions.

15.1. Examples

Here are a few examples of the uid changes in common aegis functions. Unix "permission denied" errors are not shown, but it should be clear where they would occur.

new change (aenc):
: become invoking user and read (edit) the change attribute file, validate the attribute file, then become the project owner to write the change state file and the project state file.

develop begin (aedb):
: become the project owner and read the project state file and the change state file, to see if the change exists and is available for development, and if the invoking user is on the developer access control list. Become the invoking user, but set the default group to the project group, and make a development directory. Become the project owner again, and update the change state file to say who is developing it and where.

build (aeb):
: become the project owner to read the project and change state files, check that the invoking user is the developer of the change, and that the change is in the *being developed* state. Become the invoking user, but set the default group to the project group, to invoke the build command. Become the project owner to update the change state to remember the build result (the exit status).

copy file into change (aecp):
: become the project owner to read the project and change state files. Check that the invoking user is the developer and that the change is in the *being developed* state, and that the file is not already in the change, and that the file exists in the baseline. Become the invoking user, but set the default group to the project group, and copy the file from the baseline into the development directory. Become the project owner, and update the change state file to remember that the file is included in the change.

integrate fail (aeifail):
: become the project owner to read the project and change state files. Check that in invoking user is the integrator of the change, and that the change is in the *being integrated* state. Become the integrator to collect the integrate fail comments, then become the project owner to delete the integration directory, then become the developer to make the development directory writable again. Then become the project owner to write the change state file, to remember that the change is back in the *being developed* state.

integrate pass (aeipass):
: become the project owner to read the project and change state files. Check that in invoking user is the integrator of the change, and that the change is in the *being integrated* state. Make the integration directory the new baseline directory and

remove the old baseline directory. Write the change and project states to reflect the new baseline and the change is in the *completed* state. Then become the developer to remove the development directory.

All the mucking about with default groups is to ensure that the reviewers, other members of the same group, have access to the files when it comes time to review the change. The umask is also set (not shown) so that the desired level of "other" access is enforced.

As can be seen, each of the uid change either (a) allows UNIX to enforce appropriate security, or (b) uses UNIX security to ensure that unauthorized tampering of project files cannot occur. Each project has an owner and a group: members of the development team obtain read-only access to the project files by membership to the appropriate group, to actually alter project files requires that the development procedure embodied by aegis is carried out. You could have a single account (not a user's account, usually, for obvious conflicts of interest) which owns all project sources, or you could have one account per project. You can have one group per project, if you don't want your various projects to be able to see each other's work, or you could have a single group for all projects.

15.2. Source Details

For implementation details, see the `os_become*` functions in the *aegis/os.c* file. The `os_become_init` function is called very early in `main`, in the *aegis/main.c* file. After that, all accesses are bracketed by `os_become` and `os_become_undo` function calls, sometimes indirectly as `project_become*` or `user_become*`, etc, functions. You need to actually become each user, because root is not root over NFS, and thus `chown` tricks do not work, and also because duplicating kernel permission checking in aegis is a little non-portable.

Note, also, that most system calls go via the interface described in the *aegis/glue.h* file. This isolates the system calls for UNIX variants which do not have the `seteuid` function, or do not have a correctly working one. The code in the *aegis/glue.c* file spawns "proxy" process which uses the `setuid` function to become the user and stay that way. If the `seteuid` function is available, it is used instead, making aegis more efficient. This isolation, however, makes it possible for a system administrator to audit the aegis code (for trojans) with some degree of confidence.

System calls should be confined to the *aegis/log.c*, *aegis/pager.c*, *aegis/os.c* and *aegis/glue.c* files. System calls anywhere else are probably a Bad Thing.

16. Appendix I: Internationalization and Localization

The Aegis source code has been internationalized, which is the process of modifying the original source code to permit error messages and other text to be presented in a language other than the author's native English. This was a large and often painful task, but it allows a degree of customization of error messages and other behaviours which would not have been otherwise possible. (It also makes the job of running a spell-checker over the error messages significantly easier.)

Localization is the process of translating the error messages and other text into various different languages or nationalities. This appendix is primarily aimed at localizers of Aegis.

16.1. The ".po" Files

The "lib/en/LC_MESSAGES" directory in the source tree contains the various message files needed to localize Aegis. You will find a number of ".po" files in this directory, which translates "programmer cryptic" into English. You will see that each message has a comment attached, describing the message and the context in which it is used. Many messages also have "substitutions" described, which are strings similar to shell variables which may be substituted into the message − such as the file name for messages which have something to do with a specific file.

The substitution mechanism is the same one as is used for the various commands in the project *aegis.conf* file, and so all of the substitutions described in *aesub*(5) are available to the translator. Note frequent use of the *plural* substitution, which allows grammatically correct error messages to be issued when faced with the singular/plural dichotomy. Other substitutions include the login name of the executing user, names of projects, number and state of changes, etc.

Ideally, the task for a translator is to take the *.po* files and translate the `msgstr` lines into the appropriate language. The job will, of course, not be that simple and so references into the code have been included, so that you can read the code should context be necessary to correctly translate the message.

16.2. Checking the Code

There are a number of keywords you need to have for the *xgettext* program when extracting message strings. The `gettext` keyword is not used directly, because of the substitution mechanism wrapped around it.

i18n	error_intl
io_comment_append	fatal_intl
report_error	verbose_intl
report_error	gram_error
rpt_value_error	

In general, the *etc/Howto.cook* file causes the messages to be extracted into *i18n-tmp/*.po* for checking during the build.

16.3. Translators Welcome

If *you* are able to translate the error messages into another language, please contact Peter Miller <pmiller@opensource.org.au> and he will tell you how it is done. (Actually, he'll point you to this part of the User Guide. :−)

To translate the error messages, look up the two-letter abbreviation (http://www.w3.org/WAI/ER/IG/ert/iso639.htm) of the language you are going to translate the error messages to. The rest of these instructions will call it *xx*.

In the source tree, you will see a directory called *lib/en/LC_MESSAGES* which contains some `.po` files. These are the text form of the message catalogues. You can view them with a simple text editor.

Create a new directory for your translations, and copy the English messages into it.

```
mkdir lib/xx/LC_MESSAGES
cp lib/en/LC_MESSAGES/*.po \
        lib/xx/LC_MESSAGES
```

Now you need to edit each of the `lib/xx/LC_MESSAGES/*.po` files, replacing the `msgstr` strings with suitable translations. Leave the `msgid` strings and the comments untranslated. These are text files, you can edit them with a simple text editor. GNU Emacs has a PO mode to make this easier.

The GNU Gettext (http://www.gnu.org/directory/gettext.html) sources have fairly good documentation (http://www.gnu.org/manual/gettext/index.html) about this process.

If you want to test your translations, you need to "compile" the text into the binary form used by

the `gettext()` system call. This is done using the *msgfmt*(1) program from the GNU Gettext package. To see your new translations in action, you create a `/usr/local/lib/aegis/`*xx*`/LC_MESSAGES` directory and arrange for the *msgfmt*(1) output to be placed in it. Some of the messages are hard to trigger, don't expect complete test coverage.

There are almost 600 error messages. If you average 1 message every 2 minutes, this is approximately 20 hours work. The German translation, for example, required around 12 hours.

When you are done translating, email the results to Peter Miller <pmiller@opensource.org.au> and they will be included in the next release of Aegis.

/* vim: set ts=8 sw=4 et : */

.