

Xcrypt Manual

E-Science Group, Nakashima Laboratory, Kyoto University

January 23, 2010

Contents

I	General	3
1	Introduction	4
1.1	Overview	4
1.2	Environment	4
2	Script	5
2.1	Module	5
2.2	Template	6
2.3	Procedure	6
2.4	Example	6
3	Flow	8
3.1	Model	8
3.2	Execution	8
3.3	Product	9
II	Details	11
4	Module	12
4.1	core	12
4.2	limit	12
4.3	successor	13
4.4	convergence	13
4.5	n_section_method	13
4.6	dry	14
4.7	minimax	14
5	Template	15
5.1	id	15
5.2	exe	15
5.3	arg <i>i</i>	15
5.4	linkedfile <i>i</i>	15
5.5	copiedfile <i>i</i>	15
5.6	copieddir <i>i</i>	16
5.7	stdofile	16
5.8	stdefile	16
5.9	queue	16

5.10	<code>cpu</code>	16
5.11	<code>proc</code>	16
5.12	<code>option</code>	16
5.13	<code>successor</code>	16
6	Built-in Function	17
6.1	<code>prepare</code>	17
6.2	<code>submit</code>	20
6.3	<code>sync</code>	20
6.4	<code>addkeys</code>	21
6.5	<code>addperiodic</code>	21
6.6	<code>prepare_submit</code>	21
6.7	<code>submit_sync</code>	21
6.8	<code>prepare_submit_sync</code>	22
A	How to Implement Job Class Extension Modules	23
A.1	How to Define and Use Extension Modules	23
A.2	Scripts of Extension Modules	23
A.3	Special Methods	25
A.3.1	<code>new</code>	25
A.3.2	<code>before_isready</code>	25
A.3.3	<code>before</code>	26
A.3.4	<code>start</code>	26
A.3.5	<code>after_isready</code>	26
A.3.6	<code>after</code>	26

Requirement Xcrypt requires

- Marc Lehmann’s Coro (where `conftest.c` is not contained), EV,
- Joshua Nathaniel Pritikin’s Event, and
- Daniel Muey’s Recursive,

and wants Marc Lehmann’s AnyEvent, `common::sense`, and Guard (warns if none).

Part I

General

Chapter 1

Introduction

1.1 Overview

In using a high-performance computer, we usually commit job processing to a job scheduler. At this time, we often go through the following procedures:

- to create a script in its writing style depending on the job scheduler,
- to pass the script to the job scheduler, and
- to extract data from its result, create another script from the data, and pass it to the job scheduler.

However, such procedures require manual intervention cost. It therefore seems better to remove manual intervention in mid-processing by using an appropriate script language. Xcrypt is a script language for job parallelization. We can deal with jobs as objects (called *job objects*) in Xcrypt and manipulate the jobs as well as objects in an object-oriented language. Xcrypt provides some functions and modules for facilitating job generation, submission, synchronization, etc. Xcrypt makes it easy to write scripts to process job, and supports users to process jobs easily.

1.2 Environment

Xcrypt requires a superset of Bourne or C shell, Perl 5.10.0 or any later version, and Perl/Tk 8.4 for GUI.

Chapter 2

Script

Xcrypt is a script language, and an extension of Perl. Xcrypt provides some functions and modules (not in Perl) which support how to deal with *jobs*.

An Xcrypt script consists of descriptions of

1. module,
2. template, and
3. procedure.

2.1 Module

Modules for job objects are used as follows,



```
use base qw(core);
```

When you use multiple modules, it is enough to write



```
use base qw(my module core);
```

Every module should be used in order. The details of the modules are described in Chapter 4.

Commonly-used modules can be loaded as follows,

```
use builtin;
```

similarly to how to use modules in Perl.

Some modules define variables. In order to use such variables, it is enough to write e.g.,

```
$separator = '-';
```

2.2 Template

Xcrypt's templates are implemented as Perl's hashes. For example,

```
%myjob = (  
  'id' => 'myjob',  
  'exe' => './myexe',  
  'arg0' => '100',  
  'arg1' => 'myinput',  
  'linkedfile0' => 'myexe',  
  'copiedfile0' => 'myinput',  
  'stdofile' => 'myout',  
  'stdefile' => 'myerr',  
  'queue' => 'myqueue',  
  'option' => 'myoption'  
);
```

Keys in templates are described in Chapter 5 in detail.

2.3 Procedure

Procedures of job processing are described in Xcrypt (and Perl) instead of manually carried out. Xcrypt's functions are described in Chapter 6.

2.4 Example

An example script is as follows,

```
use base qw(limit core);

&limit::initialize(10);
$separator = '-';

%myjob = (
    'id' => 'myjob',
    'exe' => './myexe',
    'arg0' => '100',
    'arg1' => 'myinput',
    'linkedfile0' => 'myexe',
    'copiedfile0' => 'myinput',
    'stdofile' => 'myout',
    'stdefile' => 'myerr',
    'queue' => 'myqueue',
    'option' => 'myoption'
);
&prepare_submit_sync(%myjob, 'arg0@' => [2,4]);
```

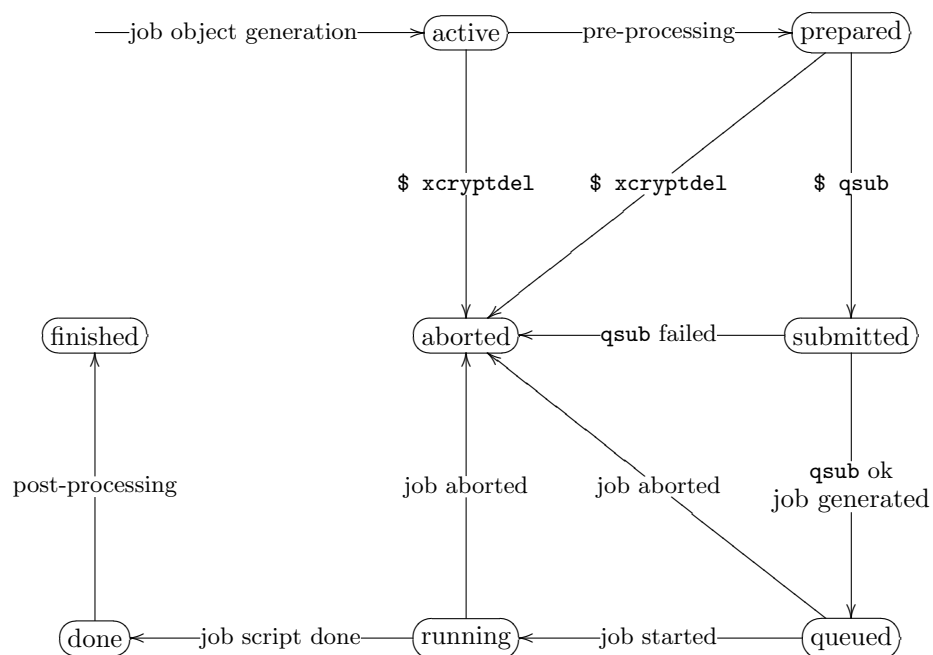

Chapter 3

Flow

In this chapter, we introduce how jobs are processed.

3.1 Model

Any job transits to *states* as follows,



3.2 Execution

Set the following environment variables appropriately. **XCRYPT** should be set to the installation directory of Xcrypt (e.g., `/usr/share/xcrypt`). For example, it is enough to type

```
$ XCRYPT=/usr/share/xcrypt; export XCRYPT
```

when you use Bourne shell. In addition, execute the following:

```
$ cd $XCRYPT; source source-me.sh
$ cd $XCRYPT/cpan; ./do-install.sh
```

If necessary, `XCRJOBSCHED`¹ should be set to an appropriate job scheduler.
Next, move to the working directory (e.g., `$HOME/wd`)

```
$ cd $HOME/wd
```

and write an Xcrypt script (e.g., `sample.xcr`). See Section 2.4 in order to know how to write.
Finally, execute Xcrypt with the script:

```
$ $XCRYPT/bin/xcrypt sample.xcr
```

3.3 Product

Xcrypt creates the following in the working directory during and after its execution.

directory

is created for each job (called a *job working directory*). It is named by the value of the job hash's key `id`. Job-processing is done in the job working directory.

myscript.pl

is a Perl script created from the Xcrypt script *myscript*.

file and soft link

of the names (which are the values of `copiedfilei` and `linkedfilei`) are copied and created in the job working directory, respectively.

`$XCRJOBSCHED.sh`

is a job script, which is passed to a job scheduler.

`sh.sh`

is a Bourne shell script, which is executed, regarding OS as a job scheduler.

`request_id`

is a file storing the job's request ID (which the job scheduler returns).

¹SGE, TSUKUBA, TOKYO, KYOTO, and `sh` are available. In the case of `sh`, jobs are dealt with as processes in OS. The default is `sh`

stdout

is a file storing the job's standard output. When **stdofile** is defined, the file is renamed as its value.

stderr

is a file storing the job's standard error. When **stdefile** is defined, the file is renamed as its value.

Part II

Details

Chapter 4

Module

In this chapter, we introduce some modules available in Xcrypt scripts.

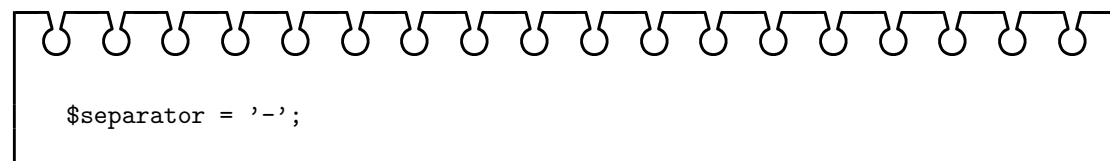
4.1 core

This module is the Xcrypt core module, and required to be read in order to use anything particular to Xcrypt.

It creates a directory of the name

```
$myjob->{'id'} $separator $myjob->{'arg0'} $separator ...
```

under the job working directory, where %myjob is a template. The default of \$separator is .. In order to redefine \$separator to be -, for example, it is enough to write as follows,



```
$separator = '-';
```

Any word composed of ASCII printable characters except

```
@_["$%&'/:;<=>?[\\]`{|}
```

is available.

This module also makes the variable \$separator_noccheck, which becomes available in Xcrypt scripts. When the value of \$separator_noccheck is 1, Xcrypt skips a check of whether \$separator consists of only available characters or not. The default is 0.

4.2 limit

This module limits the number of jobs submitted simultaneously. In order to limit the number of jobs to 10, for example, it is enough to write as follows,

```
&limit::initialize(10);
```

4.3 successor

This module indicates job objects which can be defined declaratively. For example, in order to define job objects of the name `%x`, `%y`, write:

```
...
'successors' => ['x', 'y'],
...
```

using the key `successors` in the template.

4.4 convergence

This module provides a function for a Plan-Do-Check-Action (PDCA) cycle, to deal with convergence of difference of job's results. The keys `initialvalue`, `isConvergent`, `inputfile`, `sweepname`, `outputfile`, and `extractrules` can be used in templates.

4.5 n_section_method

This module provides *n*-section method, a root-finding algorithm. The only difference from bisection method¹ is the number of sections.

The values `partition` and `epsilon` denote a partition number and an error, respectively. An interval is expressed by `x_left` and `x_right`. The values `y_left` and `y_right` are values on `x_left` and `x_right`. Typically, we can call the function `n_section_method` with these keys, e.g.,

```
&n_section_method::n_section_method(%job,
  'partition' => 12, 'epsilon' => 0.01,
  'x_left'    => -1, 'x_right' => 10,
  'y_left'    => 0.5, 'y_right' => -5
);
```

¹http://en.wikipedia.org/wiki/Bisection_method/

4.6 dry

This module provides job-processing in dry mode (skipping any command execution). Description in a template

```
...  
'dry' => 1,  
...
```

makes any job (derived from this hash) to be processed in dry mode.

4.7 minimax

This module provides a function of a tree algorithm *minimax*. This function takes a tree and a static function on nodes.

The following is a sample script:

```
%myjob = (  
  'id'           => 'job10',  
  'linkedfile0' => 'getchildren',  
  'linkedfile1' => 'komadoku',  
  'linkedfile2' => 'kuraidori',  
  'linkedfile3' => 'ikiataribattari',  
  'arg0'         => '9',           # depth of lookahead  
  'arg1'         => '0',           # position  
  'arg2'         => 'getchildren'  # get next positions  
);  
  
&prepare_submit_sync(%myjob, 'arg3@' => ['komadoku', 'kuraidori', 'ikiataribattari']);
```

where komadoku, kuraidori, and ikiataribattari are static functions.

Chapter 5

Template

In this chapter, we introduce keys and values available in templates by default.

5.1 id

Its value is a word. The value is used for creating job objects and identifying the job objects as their prefixes. Any word of ASCII printable characters except

@_ "\$%&' / : ; < = > ? [\] ' { | }

is available.

5.2 exe

Its value denotes a command. The command is executed as follows,

```
$ myexe myarg0 myarg1 ... myarg255
```

with `argi` explained below.

5.3 arg_{*i*}

Its values are arguments of a command. The command is executed as follows,

```
$ myexe myarg0 myarg1 ... myarg255
```

5.4 linkedfile_{*i*}

A soft link of the file (whose name is its value) is created in the job working directory.

5.5 copiedfile_{*i*}

The indicated file is copied to the job working directory.

5.6 copieddir*i*

All files in the indicated directory are copied in the job working directory.

5.7 stdofile

The standard output is stored in the indicated file. The default is **stdout**.

5.8 stdefile

The standard error is stored in the indicated file. The default is **stderr**.

5.9 queue

Its value is a name of a queue.

5.10 cpu

Its value is the number of cores used exclusively.

5.11 proc

Its value is the number of processes used exclusively.

5.12 option

Its value is an option of a job scheduler.

5.13 successor

Its values' jobs are generated after the job is done.

Chapter 6

Built-in Function

In this chapter, we introduce built-in functions.

6.1 prepare

This function takes a job definition hash and parameters of references¹, and returns an array of job objects. The ids of job objects are generated by `RANGEi` as described later.

Format

```
prepare(%template
    [, 'RANGE0' => \@myparam0s]...[, 'RANGE255' => \@myparam255s]
    [, 'key@' => \@myparams]...);
```

where *key@* denotes the one whose postfix is the character @ (e.g., `arg0@`).

Any word of ASCII printable characters except

`@_"$%&'/:;<=>?[\\'{}|}`

is available for `RANGEi`'s values.

Example

```
@jobs = prepare('id' => 'myjob', 'exe' => './myexe', 'arg0@' => [10,20]);
```

¹In this manual, references do not denote type globs.

This is almost the same as

```
@jobs = ();  
push(@jobs, {'id' => 'myjob_0', 'exe' => './myexe', 'arg0' => '10'});  
push(@jobs, {'id' => 'myjob_1', 'exe' => './myexe', 'arg0' => '20'});
```

Declarative description is also available as follows,

```
%template = ('id' => 'myjob', 'exe' => './myexe', 'arg0@' => [10,20]);  
prepare(%template);
```

Description of a template can be separated from parameters:

```
%template = ('id' => 'myjob', 'exe' => './myexe');  
prepare(%template, 'arg0@' => [10,20]);
```

Multiple job objects with various parameter values can be generated by using `RANGE`'s. For example,

```
@jobs = prepare(%template, 'RANGE0' => [0..99], 'arg0@' => '2 * $R0');
```

is the same as a sequence of `prepare(%template, 'arg0' => 0)`, `prepare(%template, 'arg0' => 2)`, ..., `prepare(%template, 'arg0' => 198)`.

Advanced

It is possible to generate job objects by using multiple parameters. For example,

```
%template = ('id' => 'myjob', 'exe' => './myexe');
@jobs = prepare(%template, 'arg0@' => [0,1], 'arg1@' => [2,3]);
```

is almost the same as

```
%template = ('id' => 'myjob', 'exe' => './myexe');
@jobs = ();
@job0 = prepare(%template, 'arg0' => '0', 'arg1' => '2');
push(@jobs, $job0[0]);
@job1 = prepare(%template, 'arg0' => '1', 'arg1' => '3');
push(@jobs, $job1[0]);
```

To use `RANGE`is makes it to generate job products by multiple parameters, e.g.,

```
%template = ('id' => 'myjob', 'exe' => './myexe');
@jobs = prepare(%template, 'RANGE0' => [0,1], 'RANGE1' => [2,4],
                'arg0@' => '$R0 + $R1');
```

is almost the same as

```
%template = ('id' => 'myjob', 'exe' => './myexe');
@jobs = ();
@job0 = prepare(%template, 'arg0' => '0', 'arg1' => '2');
push(@jobs, $job0[0]);
@job1 = prepare(%template, 'arg0' => '0', 'arg1' => '4');
push(@jobs, $job1[0]);
@job2 = prepare(%template, 'arg0' => '1', 'arg1' => '2');
push(@jobs, $job2[0]);
@job3 = prepare(%template, 'arg0' => '1', 'arg1' => '4');
push(@jobs, $job3[0]);
```

6.2 submit

This function takes an array of job objects and passes the jobs (corresponding to the job objects) to a job scheduler. Its return value is also the array of job objects.

Format

```
submit(@myjobs);
```

Example

Typically, this function takes a return value of `prepare`.

```
@jobs = prepare('id' => 'myjob', 'exe' => './myexe', 'arg0' => [10,20]);  
submit(@jobs);
```

It is possible to define job references without using `prepare` (although not recommended).

```
submit('id' => 'myjob_0', 'exe' => './myexe', 'arg0' => '10',  
      'id' => 'myjob_1', 'exe' => './myexe', 'arg0' => '20');
```

6.3 sync

This function takes an array of job objects and synchronizes the job objects. Its return value is also the array of job objects.

Format

```
sync(@myjobs);
```

Example

Typically, this function takes a return value of `prepare` (same as `submit`).

```
@jobs = prepare('id' => 'myjob', 'exe' => './myexe', 'arg0@' => [10,20]);  
submit(@jobs);  
sync(@jobs);
```

6.4 addkeys

This function takes an array of words and makes it available as keys in job definition hashes.

Format

```
addkeys(@words);
```

6.5 addperiodic

This function takes an Xcrypt's script code (denoted as *mystring*) and an integer *i*, and evaluates it each *i* seconds.

Format

```
addperiodic(mystring, i);
```

6.6 prepare_submit

This function is an abbreviation of `prepare` and `submit`. Its format follows `prepare`.

6.7 submit_sync

This function is an abbreviation of `submit` and `sync`. Its format follows `submit`.

6.8 `prepare_submit_sync`

This function is an abbreviation of `prepare_submit` and `sync`. Its format follows `prepare`.

Appendix A

How to Implement Job Class Extension Modules

Any job object generated by the Xcrypt's function `prepare` belongs to the class `core`, defined by `$XCRYPT/lib/core.pm`. Xcrypt users and developers can extend the class `core` by defining modules, consequently expand the function of Xcrypt. In this chapter, we introduce how to implement such extension modules.

A.1 How to Define and Use Extension Modules

In order to define an extension module of the name *mymodule*, it is enough for Xcrypt developers to put it into any directory designated by `$XCRYPT/lib/` (or `$PERL5LIB`).

Xcrypt users write the following:

```
use base (... mymodule ... core);
```

on the header of his/her script and use the extension modules.

A.2 Scripts of Extension Modules

A typical definition script of an extension modules is as follows,


```

package mymodule;

use strict;
use ...;

&addkeys('my_instance_member', ...);

my $my_class_member;

# special methods
sub new {
    my $class = shift;
    my $self = $class->NEXT::new(@_);
    ...
    return bless $self, $class;
}

sub before_isready { ... }
sub before { ... }

sub start
{
    my $self = shift;
    ...
    $self->NEXT::start();
    ...
}

sub after { ... }
sub after_isready { ... }

# general methods
sub another_method
{
    ...
}

```

In the following, we introduce components of scripts.

1. Definition of the module name: is designated by **package**. The module name must coincide with the file name's extension(**.pm**)-dropped part.
2. Use of Perl modules: is done in the usual way, i.e., by using **use**.
3. Addition of instance variables: is done by the function **addkeys**. The added instance variables are accessed as properties of job objects by writing, e.g.,

`$job->{my_instance_member}`

in Xcrypt scripts and modules. Also, by writing, e.g.,

`%template = { ..., my_instance_member=>value, ...}`

users can set values to them.

4. Definition of class variables: is done in the usual way in object-oriented programming, i.e., class variables are defined as global variables in packages. The variables can be accessed, e.g.,

`$mymodule::my_class_member`

5. Definition of methods: is defined in the usual way, i.e., methods added and extended in modules are defined as top-level functions in packages. However, note that some methods of particular names have special meanings as explained in the next section.

A.3 Special Methods

Xcrypt gives special meanings to the following class methods.

A.3.1 new

The method **new** is a class method, the so-called *constructor*. The method **new** in the most specialized class (the left-most module declared on the script header) is called.

The method **new** takes the following as arguments:

1. the package name (= **user**) to which an Xcrypt script belongs,
2. a reference to a job object¹.

Note that each object that **prepare** generates is applied to **new**.

In the body of a method, the method **new** in the parent class is called as like

`$class->NEXT::new($self,$obj)`

where **\$class** and **\$obj** are the class name and reference to the object, the arguments of **new**, respectively.

Typically, each **new** calls **new** in his parent class with the same two arguments, processes its return value (a object), and returns `bless \textit{reference to the object}, the class name` as return values.

In the module **core**, **new** is defined. The **new** creates a job directory, soft links, and copies of files (explained in Section 3.3). Note that this required procedure is skipped unless **news** in children classes call the **new** in the **core**.

A.3.2 before_isready

In Xcrypt, application of the function **submit** (cf. Section 6.2) makes a job object's state **prepared**. The methods **before_isready**s are applied to a job object of the state **prepared** (cf. Section 3.1). Its argument is a reference to the job object. The order of calling **before_isready**s is the one from children classes to parents classes. Each method returns a boolean value. When some **before_isready**s called in the object returns false, **before_isready**s are called after a little interval time.

¹The object members has values in the template passed to the function **prepare**.

A.3.3 before

The methods **before**s are applied to a job object when all the methods **before_isreadys** applied to the job object return true. Its argument is a reference to the job object. The order of calling **before**s is the one from children classes to parents classes. Return values of the methods are abandoned.

A.3.4 start

The methods **starts** are applied to a job object after **before**s to the job objects are applied. Its argument is a reference of the job object. The method **start** in the most specialized class (the left-most module declared on the script header) is called.

In the body of a method, the method **new** in the parent class is called as like

```
$obj->NEXT::start()
```

where **\$obj** is the reference to the object.

In the module **core**, **start** is defined. The **start** creates a job script and submit the job to a job scheduler. Note that this required procedure is skipped unless **starts** in children classes call the **start** in the **core**.

A.3.5 after_isready

In Xcrypt, a completion notice of a job submitted by the method **core::start** makes the job object's state **done**. The methods **after_isreadys** are applied to a job object of the state **done** (cf. Section 3.1). Its argument is a reference to the job object. The order of calling **after_isreadys** is the one from parents classes to children classes. Each method returns a boolean value. When some **after_isreadys** called in the object returns false, **after_isreadys** are called after a little interval time.

A.3.6 after

The methods **before**s are applied to a job object when all the methods **after_isreadys** applied to the job object return true. Its argument is a reference to the job object. The order of calling **after**s is the one from parents classes to children classes. Return values of the methods are abandoned.

Remark on Concurrency of before_isready, before, start, after_isready, after. Each method is called in concurrently with Xcrypt main thread. However, it is not necessary that each method is concurrently called at each job object. Xcrypt guarantee the following concurrency:

- each of **before_isreadys**, **before**s, **starts**, **after_isreadys**, **after**s method is called in concurrent with Xcrypt main thread,
- at most one method is simultanously called in **before_isreadys**, **before**s, and **starts**,
- at most one method is simultanously called in **after_isreadys** and **after**s,
- just after any **before_isready** of a job object returns true, the **before**s of the job object are applied, i.e., no method is applied to the **other** job objects,

- just after **before**s are applied to a job object, **start**s are applied to the job object, i.e., no method is applied to the other job objects,
- just after any **after_isready** of a job object returns true, the **after**s of the job object are applied , i.e., no method is applied to the **other** job objects,