# Xcrypt Manual

E-Science Group, Nakashima Laboratory, Kyoto University  ${\rm January~22,~2010}$ 

# Contents

Ι	General	3
1	Introduction           1.1 Overview            1.2 Environment	<b>4</b> 4
2	Script         2.1 Module          2.2 Template          2.3 Procedure          2.4 Example	5 6 6 6
3	Flow         3.1 Model          3.2 Execution          3.3 Product	8 8 8 9
ΙΙ	Details	11
4	4.1 core          4.2 limit          4.3 successor          4.4 n_section_method	12 12 13 13 13
5	5.1 id	15 15 15 15 15 16 16 16 16 16

	5.12	option	16
6	Buil	t-in Function	17
	6.1	prepare	17
	6.2	submit	20
	6.3	sync	20
		addkeys	
		addperiodic	
		prepare_submit	
		submit_sync	
		prepare_submit_sync	
	0.0	proparo_babmro_bjac	
A	ジョ	ブクラス拡張モジュール実装の手引	<b>23</b>
	A.1	はじめに・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	23
	A.2	位作工ジューリの完美と利用	
		孤旅モンユールの足我と利用・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	23
	A.3	拡張モジュールの定義と利用	
		拡張モジュールスクリプトの構成	23
	A.4	拡張モジュールスクリプトの構成	23 25
	A.4	拡張モジュールスクリプトの構成	23 25 25
	A.4	拡張モジュールスクリプトの構成	23 25 25 25
	A.4	拡張モジュールスクリプトの構成	23 25 25 25 26
	A.4	拡張モジュールスクリプトの構成	23 25 25 25 26 26
	A.4	拡張モジュールスクリプトの構成	23 25 25 25 26 26 26
	A.4	拡張モジュールスクリプトの構成	23 25 25 25 26 26 26

# Memo Xcrypt requires

- Marc Lehmann's Coro (Perl Artistic License, Coro/libcoro under GPL), EV (Perl Artistic License, libev under the 2-clause BSD),
- Joshua Nathaniel Pritikin's Event (?), and
- Daniel Muey's Recursive (Perl Artistic License),

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

# Part I General

# 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.

# 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,



When you use multiple modules, it is enough to write

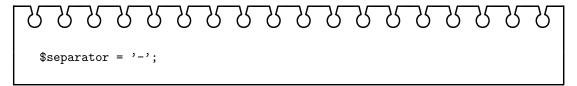


Every module should be used in order. The details of the modules are described in Chapter 4. Commonly-used modules can be loaded as follows,

```
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.,



# 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' => 'myexe',
    '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 = '-';

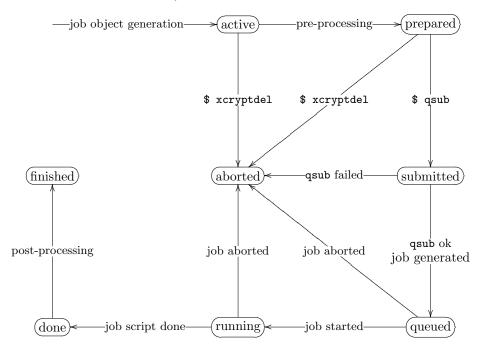
%myjob = (
    'id' => 'myjob',
    'exe' => './myexe',
    'arg0' => '100',
    'arg1' => 'myinput',
    'linkedfile0' => 'myexe',
    'copiedfile0' => 'myinput',
    'stdofile' => 'myout',
    'stdefile' => 'myerr',
    'queue' => 'myqueue',
    'option' => 'myoption'
);
&prepare_submit_sync(%myjob,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

# 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. XCRJOBSCHED<sup>1</sup> should be set to a job scheduler. PERL5LIB should be set to \$XCRYPT/lib and \$XCRYPT/lib/algo/lib.

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 copiedfile i and linkedfile i) 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).

### stdout

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

<sup>&</sup>lt;sup>1</sup>NQS, hNQS, SGE, and tSGE are available. In addition, sh-mode is supported (by being set to sh) for environments without any job scheduler. In that case, jobs are dealt with as processes in OS.

# stderr

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

Part II

Details

# Module

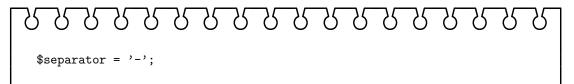
In this chapter, we introduce some modules available in Xcrypt scripts.

## **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

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,



Any word composed of ASCII printable characters except

is available.

This module also makes the variable \$separator\_nocheck, which becomes available in Xcrypt scripts. When the value of \$separator\_nocheck 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 n\_section\_method

This module provides n-section method, a root-finding algorithm. The only difference from bisection method<sup>1</sup> 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.,

# **4.5** dry

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

<sup>1</sup>http://en.wikipedia.org/wiki/Bisection\_method/

```
<u>გგგგგგგგგგგგგგგგ</u>
...
'dry' ⇒ 1,
...
```

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

# **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.

# $5.3 \quad \arg i$

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

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

## 5.4 linkedfilei

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

# 5.5 copiedfilei

The indicated file is copied to the job working directory.

# 5.6 copieddiri

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 values 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.

# **Built-in Function**

In this chapter, we introduce built-in functions.

# 6.1 prepare

This function takes a job definition hash and parameters of references<sup>1</sup>, and returns an array of job objects. The ids of job objects are generated by RANGEi as described later.

#### **Format**

where key@ denotes the one whose postfix is the character @ (e.g., argo@). Any word of ASCII printable characters except

is available for  $\mathtt{RANGE}i$ 's values.

# Example

<sup>&</sup>lt;sup>1</sup>In 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'});
```

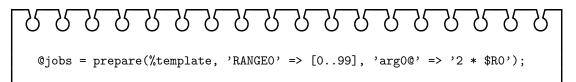
Declarative description is also available as follows,

```
%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  $\mathtt{RANGE}is$ . For example,



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.,

# 

is almost the same as

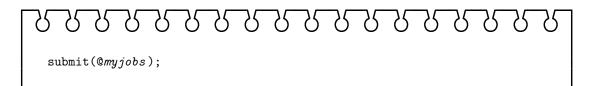
# 

```
%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

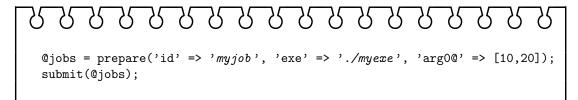
This function takes an array of job objects and passes the jobs (corresponding to the job objects) to a job scheduler. Its return value is also the array of job objects.

## Format

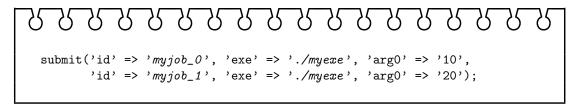


#### Example

Typically, this function takes a return value of prepare.



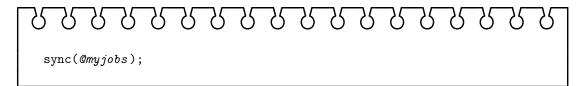
It is possible to define job references without using prepare (although not recommended).



# 6.3 sync

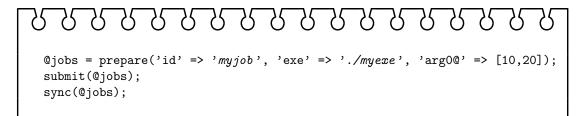
This function takes an array of job objects and synchronizes the job objects. Its return value is also the array of job objects.

#### **Format**



#### Example

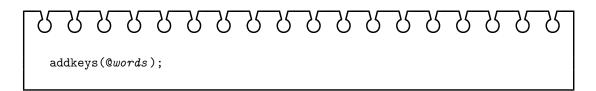
Typically, this function takes a return value of prepare (same as submit).



# 6.4 addkeys

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

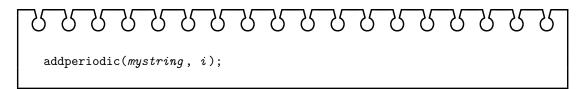
#### **Format**



# 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**



# 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

# ジョブクラス拡張モジュール実装の 手引

# A.1 はじめに

Xcrypt の prepare 関数によって生成された全てのジョブオブジェクトは\$XCRYPT/lib/core.pm で定義された core クラスに属する. Xcrypt のユーザや開発者は,拡張モジュールを定義して core クラスを拡張することで Xcrypt の機能を拡張することが可能である. 本章では,この拡張モジュールを開発するために必要な情報を示す.

# A.2 拡張モジュールの定義と利用

mymodule という名前の拡張モジュールを定義するためには,mymodule.pm という名前のファイルを\$XCRYPT/1ib/(または\$PERL5LIB で指定されているどこかのディレクトリ)に置く.Xcrypt ユーザは,スクリプトの先頭で,

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

のようにモジュール名を指定することで, 当該モジュールの機能を使用できる.

# A.3 拡張モジュールスクリプトの構成

典型的な拡張モジュールの定義スクリプトは以下のように記述される.

# package mymodule:

```
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
{
}
```

以下,スクリプトの各構成要素について説明する.

- 1. モジュール名の定義: package で指定する.ここで指定する名前は,モジュールファイル名の拡張子(.pm)を除いた部分と同一でなければならない.
- 2. Perl モジュールの取り込み: 通常の Perl プログラムと同様に, 必要な Perl モジュールを use で取り込む.
- 3. 追加するインスタンス変数の定義:ジョブクラスに新たに追加したNインスタンス変数の名前を addkeys 関数により定義する.ここで定義したインスタンス変数は,ジョブオブジェクトの属

性として、Xcrypt スクリプトやモジュール内のメソッドから\$job->{my\_instance\_member} などとしてアクセスできる.また、Xcrypt ユーザがジョブ定義ハッシュの記述時に、 %template = { ..., my\_instance\_member=>value, ...} のようにして値を設定することもできる.

- 4. クラス変数の定義: 通常の Perl のオブジェクト指向プログラミングと同様 , クラス変数はパッケージ内のグローバル変数として定義する . ここで定義した変数は ,\$mymodule::my\_class\_memberとしてアクセスできる
- 5. メソッドの定義:このモジュールにおいて追加,拡張するメソッドをパッケージ内のトップレベル関数として定義する.通常の Perl オブジェクト指向プログラミングと同様であるが,特定の名前を持つメソッドは特別な意味を持つので注意する(次節で説明)

# A.4 特別な意味を持つメソッド

### A.4.1 new メソッド

コンストラクタに相当するクラスメソッドであり, Xcrypt の prepare 関数 ( $6.1\,$ 節) の処理中に, 最も specialized なクラス (Xcrypt スクリプトのヘッダで宣言されているモジュール列のうち最も 左に書かれているもの)の new メソッドが呼び出される.

この new メソッドに渡される引数は以下の通りである.

- 1. Xcrypt スクリプトが属するパッケージ名(= user)
- 2. ジョブオブジェクトへの参照.オブジェクトのメンバには, prepare 関数に渡されたジョブ 定義ハッシュに対応する値がセットされている.

ジョブ定義ハッシュの RANGE 等の指定により prepare 関数が複数のジョブオブジェクトを生成した場合は,その各オブジェクトに対して new が適用される.

メソッドの本体では,class-NEXT::new(sself,sobj)(class,sobj はそれぞれ引数として渡されたクラス名およびオブジェクトへの参照)のようにして親クラスの new メソッドを呼び出すことができる.典型的には,各 new メソッドはまず親クラスの new メソッドを,与えられた class つの引数をそのまま引き渡して呼び出し,その返り値のオブジェクトにアクセスしつつ必要な処理を行った後,class カブジェクトへの参照,渡されたクラス名の値をメソッドの返り値として class return すべきである.

core モジュールでも new メソッドが定義されており,ここでは 3.3 節で説明したジョブの作業ディレクトリおよびその中へのファイルのコピー・シンボリックリンクの生成処理を行う.子クラスの new メソッドにより core モジュールの new メソッドが呼び出されないと,この処理が行われなくなるので注意すること.

## A.4.2 before\_isready メソッド

状態(3.1 節)が prepared になったジョブオブジェクトに対して適用される.一般には Xcrypt スクリプト中の submit 関数(6.2 節)の適用により,ジョブの状態が prepared になる.引数は ジョブオブジェクトへの参照である.複数のモジュールで before\_isready メソッドが定義されていた場合は,子クラス 親クラスの順で全ての before\_isready メソッドが呼び出される.

各メソッドは,真偽値を返す.呼び出された before\_isready メソッド列のうち,1 つでも偽を返したメソッドがあれば,しばらくの時間待ち合わせた後,もう一度 before\_isready メソッド列の実行が行われる.

# A.4.3 before メソッド

before\_isready メソッド列が適用され,その全てが真を返したジョブオブジェクトに対して適用される.引数はジョブオブジェクトへの参照である.複数のモジュールで before メソッドが定義されていた場合は,子クラス 親クラスの順で全ての before メソッドが呼び出される.

メソッドの返り値は破棄される.

#### A.4.4 start メソッド

before メソッドの処理が終わったジョブオブジェクトに対して適用される.引数はジョブオブジェクトへの参照である.複数のモジュールで start メソッドが定義されていた場合は,最も specialized なクラスの start メソッドが呼び出される.

メソッドの本体では, \$obj->NEXT::start() (\$obj は引数として渡されたジョブオブジェクトへの参照)のようにして親クラスの start メソッドを呼び出すことができる.

core モジュールでも start メソッドが定義されており,ここではジョブスクリプトの生成およびバッチスケジューラへのジョブ投入処理を行う.子クラスの start メソッドにより core モジュールの start メソッドが呼び出されないと,この処理が行われなくなるので注意すること.

## A.4.5 after\_isready メソッド

状態(3.1 節)が done になったジョブオブジェクトに対して適用される.一般には,core::start メソッドによってバッチスケジューラに投入したジョブからのジョブ完了通知により,ジョブの状態が done になる.引数はジョブオブジェクトへの参照である.複数のモジュールで after\_isready メソッドが定義されていた場合は,親クラス 子クラスの順で全ての after\_isready メソッドが呼び出される.

各メソッドは,真偽値を返す.呼び出された after\_isready メソッド列のうち,1 つでも偽を返したメソッドがあれば,しばらくの時間待ち合わせた後,もう一度 after\_isready メソッド列の実行が行われる.

#### A.4.6 after メソッド

after\_isready メソッド列が適用され,その全てが真を返したジョブオブジェクトに対して適用される.引数はジョブオブジェクトへの参照である.複数のモジュールで after メソッドが定義されていた場合は,親クラス 子クラスの順で全ての after メソッドが呼び出される. メソッドの返り値は破棄される.

#### A.4.7 before\_isready , before , start , after\_isready , after の並行性

before, before\_isready, start, after, after\_isreadyの各メソッドは Xcrypt のユーザスレッドとは並行に実行される. ただし, 各ジョブオブジェクト間についてのこれらのメソッドの実行は完全に並行に実行されるわけではない. Xcrypt 処理系が保証する並行性は以下の通りである.

- どのジョブオブジェクトの before\_isready , before , start , after\_isready , after メソッドも Xcrypt のユーザスレッドとは並行に実行される .
- どのジョブオブジェクトの before\_isready , before , start メソッドも 2 つ以上並行に実行されることはない .
- どのジョブオブジェクトの after\_isready , after , メソッドも 2 つ以上並行に実行されることはない .

- あるオブジェクトに対して before\_isready メソッド列が全て真を返した際 , それに続くその オブジェクトへの before メソッド列の実行の前に ,他のどのオブジェクトへの before\_isready , before , start メソッド適用も行われることはない .
- あるオブジェクトへの before メソッド列の適用と start メソッドの適用の間に,他のどのオブジェクトへの before\_isready, before, start メソッド適用も行われることはない.
- あるオブジェクトに対して after\_isready メソッド列が全て真を返した際 , それに続くそのオブジェクトへの after メソッド列の実行の前に , 他のどのオブジェクトへの after\_isready , after メソッド適用も行われることはない .