The OMNI Thread Abstraction

Version 4.3

1 Introduction

The OMNI thread abstraction is designed to provide a common set of thread operations for use in programs written in C++. Programs written using the abstraction should be much easier to port between different architectures with different underlying threads primitives.

The programming interface is designed to be similar to the C language interface to POSIX threads (IEEE draft standard 1003.1c — previously 1003.4a, often known as “pthreads” [POSIX94]).

Much of the abstraction consists of simple C++ object wrappers around pthread calls. However for some features such as thread-specific data, a better interface can be offered because of the use of C++.

Some of the more complex features of pthreads are not supported because of the difficulty of ensuring the same features can be offered on top of other thread systems. Such features include thread cancellation and complex scheduling control (though simple thread priorities are supported).

See the `omnithread.h` header file for full details of the API. The descriptions below assume you have some previous knowledge of threads, mutexes, condition variables and semaphores. Also refer to other documentation ([Birrell89], [POSIX94]) for further explanation of these ideas (particularly condition variables, the use of which may not be particularly intuitive when first encountered).

2 Synchronisation objects

Synchronisation objects are used to synchronise threads within the same process. There is no inter-process synchronisation provided. The synchronisation objects provided are mutexes, condition variables and counting semaphores.

2.1 Mutex

An object of type `omni_mutex` is used for mutual exclusion. It provides two operations, `lock()` and `unlock()`. The alternative names `acquire()` and `release()`

can be used if preferred. Behaviour is undefined when a thread attempts to lock the same mutex again or when a mutex is locked by one thread and unlocked by a different thread.

2.2 Condition Variable

A condition variable is represented by an `omni_condition` and is used for signalling between threads. A call to `wait()` causes a thread to wait on the condition variable. A call to `signal()` wakes up at least one thread if any are waiting. A call to `broadcast()` wakes up all threads waiting on the condition variable.

When constructed, a pointer to an `omni_mutex` must be given. A condition variable `wait()` has an implicit mutex `unlock()` and `lock()` around it. The link between condition variable and mutex lasts for the lifetime of the condition variable (unlike pthreads where the link is only for the duration of the wait). The same mutex may be used with several condition variables.

A wait with a timeout can be achieved by calling `timedwait()`. This is given an absolute time to wait until. The routine `omni_thread::get_time()` can be used to turn a relative time into an absolute time. `timedwait()` returns `true` if the condition was signalled, `false` if the time expired before the condition variable was signalled.

2.3 Counting semaphores

An `omni_semaphore` is a counting semaphore. When created it is given an initial unsigned integer value. When `wait()` is called, the value is decremented if nonzero. If the value is zero then the thread blocks instead. When `post()` is called, if any threads are blocked in `wait()`, exactly one thread is woken. If no threads were blocked then the value of the semaphore is incremented.

If a thread calls `trywait()`, then the thread won’t block if the semaphore’s value is 0, returning `false` instead.

There is no way of querying the value of the semaphore.

3 Thread object

A thread is represented by an `omni_thread` object. There are broadly two different ways in which it can be used.

The first way is simply to create an `omni_thread` object, giving a particular function which the thread should execute. This is like the POSIX (or any other) C language interface.

The second method of use is to create a new class which inherits from `omni_thread`. In this case the thread will execute the `run()` member function of the new class. One advantage of this scheme is that thread-specific data can be implemented simply by having data members of the new class.

When constructed a thread is in the "new" state and has not actually started. A call to `start()` causes the thread to begin executing. A static member function `create()` is provided to construct and start a thread in a single call. A thread exits by calling `exit()` or by returning from the thread function.

Threads can be either detached or undetached. Detached threads are threads for which all state will be lost upon exit. Other threads cannot determine when a detached thread will disappear, and therefore should not attempt to access the thread object unless some explicit synchronisation with the detached thread guarantees that it still exists.

Undetached threads are threads for which storage is not reclaimed until another thread waits for its termination by calling `join()`. An exit value can be passed from an undetached thread to the thread which joins it.

Detached / undetached threads are distinguished on creation by the type of function they execute. Undetached threads execute a function which has a `void*` return type, whereas detached threads execute a function which has a `void` return type. Unfortunately C++ member functions are not allowed to be distinguished simply by their return type. Thus in the case of a derived class of `omni_thread` which needs an undetached thread, the member function executed by the thread is called `run_undetached()` rather than `run()`, and it is started by calling `start_undetached()` instead of `start()`.

The abstraction currently supports three priorities of thread, but no guarantee is made of how this will affect underlying thread scheduling. The three priorities are `PRIORITY_LOW`, `PRIORITY_NORMAL` and `PRIORITY_HIGH`. By default all threads run at `PRIORITY_NORMAL`. A different priority can be specified on thread creation, or while the thread is running using `set_priority().` A thread's current priority is returned by `priority()`.

Other functions provided are `self()` which returns the calling thread's `omni_thread` object, `yield()` which requests that other threads be allowed to run, `id()` which returns an integer id for the thread for use in debugging, `state()`, `sleep()` and `get_time()`.

4 Per-thread data

omnithread supports per-thread data, via member functions of the `omni_thread` object.

First, you must allocate a key for with the `omni_thread::allocate_key()` function. Then, any object whose class is derived from `omni_thread::value_t` can be stored using the `set_value()` function. Values are retrieved or removed with `get_value()` and `remove_value()` respectively.

When the thread exits, all per-thread data is deleted (hence the base class with virtual destructor).

Note that the per-thread data functions are **not** thread safe, so although you can access one thread's storage from another thread, there is no concurrency control. Unless you really know what you are doing, it is best to only access per-thread data from the thread it is attached to.

References

[POSIX94] *Portable Operating System Interface (POSIX) Threads Extension*, P1003.1c Draft 10, IEEE, September 1994.

[Birrell89] *An Introduction to Programming with Threads*, Research Report 35, DEC Systems Research Center, Palo Alto, CA, January 1989.