Why the "volatile" type class should not be used

C programmers have often taken volatile to mean that the variable could be changed outside of the current thread of execution; as a result, they are sometimes tempted to use it in kernel code when shared data structures are being used. In other words, they have been known to treat volatile types as a sort of easy atomic variable, which they are not. The use of volatile in kernel code is almost never correct; this document describes why.

The key point to understand with regard to volatile is that its purpose is to suppress optimization, which is almost never what one really wants to do. In the kernel, one must protect shared data structures against unwanted concurrent access, which is very much a different task. The process of protecting against unwanted concurrency will also avoid almost all optimization-related problems in a more efficient way.

Like volatile, the kernel primitives which make concurrent access to data safe (spinlocks, mutexes, memory barriers, etc.) are designed to prevent unwanted optimization. If they are being used properly, there will be no need to use volatile as well. If volatile is still necessary, there is almost certainly a bug in the code somewhere. In properly-written kernel code, volatile can only serve to slow things down.

Consider a typical block of kernel code:

```
    spin_lock(&the_lock);
    do_something_on(&shared_data);
    do_something_else_with(&shared_data);
    spin_unlock(&the_lock);
```

If all the code follows the locking rules, the value of shared_data cannot change unexpectedly while the_lock is held. Any other code which might want to play with that data will be waiting on the lock. The spinlock primitives act as memory barriers - they are explicitly written to do so - meaning that data accesses will not be optimized across them. So the compiler might think it knows what will be in shared_data, but the spin_lock() call, since it acts as a memory barrier, will force it to forget anything it knows. There will be no optimization problems with accesses to that data.

If shared_data were declared volatile, the locking would still be necessary. But the compiler would also be prevented from optimizing access to shared_data _within_ the critical section, when we know that nobody else can be working with it. While the lock is held, shared_data is not volatile. When dealing with shared data, proper locking makes volatile unnecessary - and potentially harmful.

The volatile storage class was originally meant for memory-mapped I/O registers. Within the kernel, register accesses, too, should be protected by locks, but one also does not want the compiler "optimizing" register accesses within a critical section. But, within the kernel, I/O memory accesses are always done through accessor functions; accessing I/O memory directly through pointers is frowned upon and does not work on all architectures. Those accessors are written to prevent unwanted optimization, so, once again, volatile is unnecessary.

Another situation where one might be tempted to use volatile is when the processor is busy-waiting on the value of a variable. The right way to perform a busy wait is:

```
    while (my_variable != what_i_want)
        cpu_relax();
```

The cpu_relax() call can lower CPU power consumption or yield to a hyperthreaded twin processor; it also happens to serve as a compiler barrier, so, once again, volatile is unnecessary. Of course, busy-waiting is generally an anti-social act to begin with.

There are still a few rare situations where volatile makes sense in the kernel:

- The above-mentioned accessor functions might use volatile on architectures where direct I/O memory access does work. Essentially, each accessor call becomes a little critical section on its own and ensures that the access happens as expected by the programmer.

- Inline assembly code which changes memory, but which has no other visible side effects, risks being deleted by GCC. Adding the volatile keyword to asm statements will prevent this removal.

- The jiffies variable is special in that it can have a different value every time it is referenced, but it can be read without any special locking. So jiffies can be volatile, but the addition of other variables of this type is strongly frowned upon. Jiffies is considered to be a "stupid legacy" issue (Linus's words) in this regard; fixing it would be more trouble than it is worth.

- Pointers to data structures in coherent memory which might be modified by I/O devices can, sometimes, legitimately be volatile. A ring buffer used by a network adapter, where that adapter changes pointers to indicate which descriptors have been processed, is an example of this type of situation.

For most code, none of the above justifications for volatile apply. As a result, the use of volatile is likely to be seen as a bug and will bring additional scrutiny to the code. Developers who are tempted to use volatile should take a step back and think about what they are truly trying to accomplish.

Patches to remove volatile variables are generally welcome - as long as they come with a justification which shows that the concurrency issues have been properly thought through.

NOTES

[1] http://lwn.net/Articles/233481/
[2] http://lwn.net/Articles/233482/

CREDITS

Original impetus and research by Randy Dunlap
Written by Jonathan Corbet
Improvements via comments from Satyam Sharma, Johannes Stezenbach, Jesper
        Juhl, Heikki Orsila, H. Peter Anvin, Philipp Hahn, and Stefan
        Richter.