

Why Do Computers Crash?

R.L. Feigenbaum
Croton on Hudson, NY

Clay Shields, a professor of computer science at Georgetown University, explains. Computers crash because of errors in the operating system (OS) software or errors in the computer hardware. Software errors are probably more common, but hardware errors can be devastating and harder to diagnose.

A variety of hardware components must function correctly in order for a computer to work. These components, like many things, age over time and can develop faults. Unfortunately, these faults are often transient, and can be hard to diagnose because they do not appear consistently. The system power supply can fail in this manner. Normally a computer's power supply converts alternating current to clean direct current. If it starts to fail, the computer can crash accidentally when the power supply generates a noisy signal. The random access memory (RAM) can also fail in an intermittent way, particularly if it gets hot. Because the values RAM stores get corrupted unpredictably, it causes random system crashes. The central processing unit (CPU) can also be the source of crashes due to excessive heat. The (often loud) fans on most common computers are there to prevent this type of crash, though they may eventually fail. The fans that bring cooling air into the case also carry dirt and dust inside. This dirt can accumulate and cause intermittent short circuits as the dirt blows around. Fortunately, compressed air or a vacuum cleaner easily gets rid of the dirt. Still other hardware problems that can cause crashes are trickier to identify and require software tests or sequential replacement of components.

More permanent faults happen with errors on a computer's disk. Each disk stores information in units named sectors. Most new disks come with bad sectors that occur in the manufacturing process and are marked at the factory. Makers expect this and include ample additional sectors to replace the defective ones. Sectors can go bad later, however, and lose the information stored on them. If these sectors happen to hold system information, they can cause a crash. Worse, a disk can fail completely when the computer gets jarred and the head that reads information makes contact with the disk surface. This may cause all data on the disk to be lost.

Although crashes caused by hardware are possible, most computer crashes are caused by errors in the OS software. The OS does more than provide an interface for the user to operate the computer. It also provides a consistent interface between applications and the hardware, and acts to share system resources between different programs. As a result, there are a number of errors that can occur. Perhaps the most common is a glitch that arises when the OS tries to access an incorrect memory address, perhaps as a result of a programming error. In Windows, this can lead to an error known as a General Protection Fault (GPF). Other errors drive the OS into an infinite loop, in which the computer executes the same instructions over and over without hope of escape. In these cases, the computer might seem to "lock up"--the system doesn't crash, but is not longer responsive to input and needs to be reset. Still other problems result when a bug allows information

to be written into a memory buffer that is too small to accept it. The additional data "overflows" out of the buffer and overwrites information in memory, corrupting the OS state. These same errors can occur in application programs. Newer OSs are robust against application crashes, but in older systems application bugs can affect the OS and cause a system-wide crash. Modern operating systems are carefully tested, and tend to be relatively stable, but drivers that are added to the OS to allow the use of additional devices such as printers may not be, and are often the source of crashes. This is why most modern OSs allow for a special boot mode that disables loading drivers. The drivers can then be added one at a time to determine which one causes the error.

The OS can also crash when it fails in its job of managing system resources correctly. It is possible for the OS to reach a state of deadlock, in which multiple programs each have control of some resource another program needs, and each is waiting for the other to relinquish control of the resource. Alternatively, the system might be switching back and forth between a few programs, each of which needs a significant proportion of memory resources. Because the switching takes time (as memory information is stored to and read from the disk), it is possible for the machine to thrash, which means it spends so much time swapping programs back and forth that little or no productive processing occurs. A thrashing machine may be slow or unresponsive, but its disk is still operating and it will generally recover after being left to itself for a few minutes.

Thrashing can occur as a result of the OS failing to allocate and recover memory space properly. As the OS allows programs to run, it allocates memory to them. A memory leak occurs when the OS fails to recover the memory correctly when programs stop. Over time, the OS's internal accounting will show that there is little memory available. Computers can crash as a result of different devices trying to use the same internal ID to operate. These types of crashes are more common after adding new, conflicting hardware to a system.

Finally, an OS can crash if information it needs is corrupted on disk. This often happens when a computer crashes, loses power, or is shut down without having the opportunity to write the contents of memory to the appropriate files. A system crash can therefore lead to later crashes upon rebooting. A virus infecting the system can also cause file corruption.

Given that there are so many ways that a computer can crash, how do you diagnose the problem? It isn't always easy, but there are resources that can provide detailed guidelines about how to approach solving your problem.

A good website is available at: <http://www.pcguides.com/ts/x/sys/crash.htm>