

Windows file system compression had to be dumbed down

devblogs.microsoft.com/oldnewthing/20161101-00

November 1, 2016



Raymond Chen

Note: [This story has been retracted.](#)

I noted some time ago that when you ask Windows to use file system compression, [you get worse compression than WinZip or some other dedicated file compression program](#), for various reasons, one of which is that file system compression is under soft real-time constraints.

The soft real-time constraint means that one of the performance targets for file system compression included limits like “Degrades I/O throughput by at most $M\%$ ” and “Consumes at most $M\%$ of CPU time,” for some values of M and N I am not privy to.

But there’s another constraint that is perhaps less obvious: The compression algorithm must be system-independent. In other words, you cannot change the compression algorithm depending on what machine you are running on. Well, okay, you can compress differently depending on the system, but every system has to be able to decompress every compression algorithm. In other words, you might say, “I’ll use compression algorithm X if the system is slower than K megahertz, but algorithm Y if the system is faster,” but that means that everybody needs to be able to decompress both algorithm X and algorithm Y, and in particular that everybody needs to be able to decompress both algorithm X and algorithm Y *and still hit the performance targets*.

The requirement that a file compressed on one system be readable by any other system allows a hard drive to be moved from one computer to another. Without that requirement, a hard drive might be usable only on the system that created it, which would create a major obstacle for data centers (not to mention data recovery).

And one of the limiting factors on how fancy the compression algorithm could be was the Alpha AXP.

One of my now-retired colleagues worked on real-time compression, and he told me that the Alpha AXP processor was very weak on bit-twiddling instructions. For the algorithm that was ultimately chosen, the smallest unit of encoding in the compressed stream was the nibble; anything smaller would slow things down by too much. This severely hampers your ability to get good compression ratios.

Now, Windows dropped support for the Alpha AXP quite a long time ago, so in theory the compression algorithm could be made more fancy and still hit the performance targets. However, we also live in a world where you can buy a 5TB hard drive from Newegg for just \$120. Not only that, but many (most?) popular file formats are already compressed, so file system compression wouldn't accomplish anything anyway.

We live in a post-file-system-compression world.

Raymond Chen

Follow

