Daniel Setser

CS 5959

Reliability of Split

Split is a very simple utility allowing you to split large files into multiple smaller files. You can do this by number of lines per file or number of bytes per file. I looked at the most recent version of split and an older version (which was closer to the very old version that the CADE lab is running). I didn't find any difference in the reliability between versions. The only differences I noticed were that the older version has fewer options for splitting the file, and can't split files into chunks larger than 2GB.

In my reliability testing, I was unable to find anything that caused split to crash, hang, or give incorrect output. I ran several tests using Clang's runtime checks for undefined or suspicious behaviors, all tests performed as expected. I also ran tests with all of the options in the Murphy kernel shim (excluding those noted as untested), and all were successful. I ran tests on different files including: large files over 2GB, text files run through a fuzzer with a significant amount of flipped bits, and binary files. I tested splitting these files by different numbers of lines and bytes. All files were split into the correct number of files, with the correct number of lines, the correct filenames, and could be appended together and pass a diff showing they were identical to the original file. I also examined the source code to ensure every system call's return value was checked and that there wasn't anything that looked suspicious.

Based on this evidence and my examination of the source, I believe that the split utility can be relied on for important tasks.