Spencer Phippen

Diff: Does it Work?

The engineering department here at Diff-o-Matic™ has been investigating the use of the diff[1] command line utility as a back-end, and I have been assigned the task of vetting the tool. Given our clients in the defense industry, the reliability of this tool could be a matter of national security – we certainly don't want to end up like the folks at Doff-o-Matic™ after last year's scandal.

There were two broad categories of functionality that I attempted to test for and analyze:

1. Exceptional behavior
2. Non-exceptional, but incorrect behavior

The first category includes things like hanging, segmentation faults, deleting the current subdirectory, granting a root shell, etc. The second includes ostensibly more benign behavior: generating output—usually in the form of a diff report—that seems to have the correct form, but is not in fact an accurate diff report of the files being analyzed. For example, reports where two lines of the input file are swapped or where a differing line was not mentioned are examples of behavior in the second category.

I spent most of my time investigating ways to trigger behavior in the category (1); unfortunately, I was not able to find bugs of this nature in the diff program. I started by looking for the obvious warning signs: unsanitized input used as format strings, code around calls to C string functions, naïve handling of fixed size buffers. However, any suspicious-looking operations ended up being simple to verify the correctness of, and this avenue of analysis revealed no clear errors or vulnerabilities.

Another avenue for detecting category (1) behavior involves runtime checkers and instrumentation tools such as valgrind, Murphy, and others. Valgrind didn't report any errors or memory leaks when diff[2] was run with a variety of options and input files. I spent some time with Murphy but wasn't able to get it working very well, so I wasn't able to test diff's behavior in situations where the kernel is behaving strangely.

Because the version of diff installed on the servers here at Diff-o-Matic™ was released 19 years ago, I decided to look at the development mailing list to see if any serious bugs had been reported since then. I found one bug that affects our platform in the archives since then: diff fails to print the line "No newline at end of file" when certain files without ending newlines are compared. Based on my knowledge of our product, this bug will not affect our software because newline information is not reported directly to the user.

Regarding stability, diff is one of the few utilities in Miller's original fuzzing paper (1990) that never hung or crashed, and the utility has not changed much since then – this is consistent with my findings.

[1] diffutils version 2.7

[2] compiled under gcc with "-O2 -NDEBUG -g -fprofile-arcs -ftest-coverage -Wall"

Because of the time spent trying to find category (1) bugs, I wasn't able to do very robust testing of category (2). I've run a few manual tests (they all passed), but there is definitely the possibility of improper output in corner cases. If I were to continue testing the software, I would definitely focus my efforts on some sort of automatic, random diff correctness-checker.

Final Verdict: Maybe (but it won't crash on you!)