ident86 manual

This document has been compiled on 2025-10-12.

Contents

Section 1: Overview and purpose

ident86 operates on two binary files, attempting to show differences between the files and disassembling different parts as 8086 code. It is intended for detecting code that is encoded or disassembles differently but has the same semantic meaning.

At least two and up to four input files can be specified. Traditionally, the first input is an "original" binary and the second is a "replicated" binary that's to be adjusted to match the "original". The two files may be flat binaries or MZ executables, but should both be of the same type.

The third input file is a .tls (trace listing) file that's used to detect instruction boundaries and data items (non-code parts) in the second file. Instruction boundaries cannot be readily detected without a .tls file.

The fourth file is a .map file, created by WarpLink with `/mx` switch or by NASM with `-f bin` format. It is used only to detected "map ranges" that indicate what bytes in the second binary file are alignment inserted by the linker that do not correspond to any data recorded in the .tls file. (NASM's `-f bin` output format is considered a linker that's internal to the assembler for these purposes.)

Section 2: Switches

Switches may be specified anywhere on the command line. Switches are cap-sensitive.

2.1 `-m` and `-M` - Minimum and maximum offset switches

The switch `-m` or `--minimum-offset` specifies the minimum offset within the file to process. The subsequent number may be in decimal, or in hexadecimal with a leading `0x`. Default is `0`.

The switch `-M` (capital M) or `--maximum-offset` specifies the maximum offset within the file to process. Before the subsequent number, a plus sign may be specified, in that case the maximum is calculated from the number taken as an amount of bytes to process starting at the minimum offset. The number may be in decimal, or in hexadecimal with a leading `0x`. Default is `0xFFFFffff`.

2.2 `-z` - Skip header switch

The switch `-z` or `--skip-header` enables to skip listing of byte differences in the MZ executable header.

2.3 `-Z` - Apply relocation switch

The switch `-Z` (capital Z) or `--apply-reloc` accepts a subsequent 16-bit number. For an MZ executable, the number is applied as a relocation factor to the relocations listed in the MZ header.

2.4 `-x` - Only second pass switch

The switch `-x` or `--only-second` skips the first pass over the .tls file in the find source line function. May be needed to pick up a label properly.

2.5 `-a` - Auto length switch

The switch `-a` or `--auto-length` is followed by a number. If this many different bytes occur in a row, the automatic difference length is considered reached. Internally, the maximum offset to process is set so that processing stops after the different byte that reached the automatic difference length.

2.6 `-A` - Auto display length switch

The switch `-A` (capital A) or `--auto-display` is followed by a number. If the automatic difference length is reached, then by default disassembly occurs for that length of the last run. By soecifying an auto display length, a shorter disassembly may be used for this last run.

2.7 -s - Side by side switch

The switch -s or --side-by-side is essential for some aspects of ident86. It enables a two-column display where disassembled lines from the first file are in the first column and disassembled lines from the second file afe in the second column.

2.8 -o - Offset switch

The switch -o or --offset is followed by a number. A plus sign may occur before the number. The number may be negative. The number is applied as an offset to match the trace listing (.tls) file. Default is +0.

If there is a leading plus sign, then the MZ executable header size, if any, is added to the following number to calculate the offset.

If the MZ executable header size is 512 bytes then the calculated offset should be 512.

If the binaries are flat-format executables created using exe2bin or x2b2 from an MZ executable with an entrypoint of 0:256 (that is, 256 bytes of padding that is cut out) then the calculated offset should be -256.

2.9 -v - Version switch

The switch -v or --version makes ident86 only display its own source control revision ID and the lDebug version that it runs. The program will terminate immediately after this.

2.10 -V - No version switch

The switch -V (capital V) or --no-version avoids the version display usually done at the beginning of the report.

2.11 -d - Difference length switch

The switch -d or --difference-length is followed by a number. After the earliest difference (not no difference and not fuzzy same), the maximum offset is forced to the offset of that difference plus the difference length number.

2.12 -X - Trailer size switch

The switch -X (capital X) or --trailer-size is followed by a number. After one of the files is at EOF but the other isn't yet, only list as many bytes as this number indicates. Default is 16.

2.13 -e - Specify edit file switch

The switch -e or --edit-file specifies which file is to be edited. It is followed by a number in the range 0 to 2, inclusive. Traditionally, the second file is to be edited, which corresponds to using -e 2 as a switch.

2.14 -E - Do edit switch

The switch -E (capital E) or --do-edit signals to ident86 that it should attempt to edit the corresponding source text file to make the two binaries match more closely. Only one edit is

done per ident86 iteration. The `--edit-file` or `-e` switch must also be specified in order to use the `-E` switch.

2.15 `-S` - Show edit switch

The switch `-S` (capital S) or `--show-edit` signals to ident86 that it should display a part of the corresponding source text file that would be or is edited using the `-E` switch. The display includes line numbers and the line content. The `--edit-file` or `-e` switch must also be specified in order to use the `-S` switch.

2.16 `-j` - Show diff switch

The switch `-j` or `--show-diff` makes ident86 display the .tls file lines that correspond to differences.

2.17 `-J` - Show diff colour switch

The switch `-J` (capital J) or `--show-diff-colour` causes the .tls lines displayed by `-j` to highlight the hex dump parts corresponding to different bytes. The highlighting is done using reverse video escape codes. The `-J` switch has no effect if the `-j` switch is not specified.

2.18 `-c` - Cookie switch

The switch `-c` or `--cookie` is followed by a filename. If the file exists at the beginning of the ident86 iteration, its last line is read and the resulting number is used as a minimum offset. If an earliest difference (not no difference and not fuzzy same) is found in an iteration, its offset is appended to the cookie file unless the last line in the cookie file already matches the same offset.

The point of the cookie file is to continue the ident86 operation after an assumed good prefix on subsequent iterations, saving a lot of time.

2.19 `-b` - Build scriptlet switch

The switch `-b` or `--build` is followed by a list of shell commands. At the beginning of an ident86 iteration, ident86 will call the shell to run the specified commands. If the shell command returns a nonzero return code, ident86 is aborted.

This is particularly useful in combination with the `-E` and `-r` switches.

2.20 `-p` - Pattern switch

The switch `-p` or `--pattern` is followed by a filename modification pattern. This is used to run a regular expression. The pattern is split into up to three parts with double-colon `::` as a separator. The first part is the search string. The second part is the replacement string, it defaults to the empty string. The third part is how often to apply the search and replace, it defaults to 1. The search string may be '`<None>`' to match an empty filename.

Every given pattern is applied to the current .tls trace listing source filename to produce a source text filename, as needed if either or both of the `-E` or `-S` switches are used. If no `-p` switch is given, a single default patterh that corresponds to '`\.lst$::.nas::0`' is used.

2.21 `-I` - Include switch

The switch `-I` (capital I) or `--include` is followed by an include directory pathname. A trailing pathhame separator, such as a slash `/`, should be included. Source text filenames generated from the .tls trace listing source using the `-p` patterns are searched for in the current directory first, then in every include directory in order. One or more `-I` switches may be needed if either or both of the `-E` or `-S` switches are used.

2.22 `-P` - Checksum switch

The switch `-P` (capital P) or `--pruef` indicates to show a checksum of each binary at the beginning of each ident86 iteration. The name comes from the german word for "checksum". The checksums are used internally if the `-r` switch is specified. The `-P` switch however determines whether they are displayed.

2.23 `-t` - Time switch

The switch `-t` or `--time` indicates to show the current date and time at the beginning of every ident86 iteration.

2.24 `-T` - Time UTC switch

The switch `-T` (capital T) or `--time-utc` indicates to show the current date and time in UTC at the beginning of every ident86 iteration. If both `-t` and `-T` are specified, the UTC time takes precedence.

2.25 `-r` - Repeat switch

The switch `-r` or `--repeat` indicates to run multiple ident86 iterations. It requires the `-E` switch be used as well. The `-b` switch generally needs to be specified too. For every iteration, the build scriptlet runs to regenerate the file to edit and the trace listing .tls file and the .map file. If an autonomous edit is possible, then it is applied and ident86 runs for another iteration. At the beginning of a subsequent iteration, the checksums are compared to ensure that the "original" file did not change and the "replicated" file did change.

The cookie file mechanism (`-c` switch) can be used to speed up subsequent iterations, with the assumption that the autonomous edit won't introduce changes before the prior earliest difference.

2.26 `-f` - Fuzzy comparison switch

The switch `-f` or `--fuzzy` is followed by a number. The default is 32. Fuzzy comparison calculates the absolute values of numeric differences in otherwise matching disassembled instructions, such as for branch destination addresses, memory operand offsets, and immediate operands. If all the absolute values do not exceed the `-f` switch number, then the instruction is considered fuzzy same.

2.27 `-D` - Dump all switch

The switch `-D` (capital D) or `--dump-all` indicates to dump disassembly of all instructions in differing ranges. It disables the matching that skips exact disassembly matches at the beginning of a differing range, and consequently also makes it impossible for the "no difference" message to be written for a differing range.

2.28 `-Y` - Show dump colour switch

The switch `-Y` (capital Y) or `--show-dump-colour` indicates to highlight with reverse video the length indicator of every disassembled line displayed that contains at least one different byte value.

2.29 `-F` - Forced disassembly switch

The switch `-F` (capital F) or `--forced` is followed by one or two numbers. If a second number is specified, it is separated from the first number using a double-colon `::`.

If a single number is specified, the byte at this offset is considered a mismatch for the purposes of detecting differing ranges (even if it in fact is a match). The byte is not considered a mismatch for the purpose of `-Y` or `-J` switch processing however.

If two numbers are specified, then the first number specifies the start of a range and the second number specifies the end of a range (inclusive). All bytes in this range are considered a mismatch.

Source Control Revision ID

hg c60af0786e18, from commit on at 2025-10-12 13:17:28 +0200

If this is in ecm's repository, you can find it at https://hg.pushbx.org/ecm/ldebug/rev/c60af0786e18