CS 4414 Operating Systems – Fall 2017

Homework #1

Maurice Wong

Problem Description:

The goal of this assignment was to implement a simple UNIX shell capable of running programs and supporting pipes and file redirection. The assignment was successfully completed.

Approach:

The shell was written as a continuously running while loop that reads in a line of input at a time. The body of the shell was divided into four main phases:

1. Reading in a line into an array of characters.
2. Lexing the lines into tokens (words and operators).
3. Parsing the list of tokens into groups that represent each command and its arguments.
4. Interpreting each token group and executing the command.

Initially, the shell begins with the *readLine()* function that reads in individual characters from stdin into an array of characters until a newline character appears. Error checks for exceeding the maximum number of characters in a line occur here.

For the lexing phase, a *tok_list* struct was created to store an array of strings and the size of the array for ease of iterating through. This struct was used to store all the tokens within a line. In the *tokenizeLine()* function, the stdlib function *strtok()* is utilized to divide the line of characters into tokens separated by spaces. Each token is then validated to ensure it is either an operator or a word and that it does not have invalid characters.

In the parsing phase, tokens are grouped together into structs called *tok_group*. This struct has a number of fields to represent the possible options that a token group has for an instruction:

- *command* – the command string
- *args* – a *tok_list* of arguments for that instruction
- *input_redirect & output_redirect* – boolean fields indicating if the instruction performs a file input or output redirection, respectively
- *input_file & output_file* – string fields with the name of the file used for input/output redirection, respectively

By first iterating through the tokens and counting the number of pipe operators "|", the number of groups can be determined. The *parseTokens* function loops through the tokens and forms these *tok_group* structs, setting the fields of the structs as appropriate based on what operators and words are encountered. Any parsing errors are detected within this function, which are improper orderings of the tokens. The results are then stored in a created struct called *tok_group_list* which contains an array of *tok_groups* and the size of the array for ease of iterating.

Once the line is divided into token groups, the token groups can be iterated over and executed in separate processes. This is done in the *runCommands()* function. Based on the number of token groups *n*, it creates a set of *n-1* pipes to direct input/ouput between the list of token groups so that all the forked processes will have access to the pipes. Each token group is then iterated over and a new child process is forked to execute each instruction. In the child process, the appropriate files are opened for input/output file redirection as well as setting up pipe redirection using *dup2()*. The commands are then run with *execve* and the appropriate

arguments. The parent process stores the child process id's into an array and iterates over them, waiting for them to finish and printing their exit codes to stderr.

The main shell loop concludes with freeing allocated memory used for storing tokens and groups. The while-loop continues to process the next line of input.

Results:

The resulting simple shell successfully works. Commands were tested that involved combinations of input/output file redirection and pipes were all used successfully. Commands that took up the terminal (like *emacs*) were also tested and were successful. Invalid commands that had too many characters per line and improper syntax were all tested and failed correctly. Invalid programs were attempted to be exec'd and invalid files were attempted to be opened for redirection, failing as intended.

Problems Encountered:

The most difficult part of the shell was figuring out how to setup the pipes. Initially, only the ends of the pipes that were not used in each process were closed, but child processes hung when this code was tested. It was later deduced after some experimentation that all the file descriptors in both parent and child processes pointing to the write-end of the pipe needed to be closed. This is because the child process continues to read from the write-end of the pipe until all file descriptors pointing to it are closed.

Some strange behavior was observed when printing to the terminal. When redirecting command lines into the shell executable on a local VM. The result of the commands printed first, and then all the shell prompts ">" that should have printed between the commands were all printed at the end.

In addition, the output from the exec'd processes sometimes came after the printed exit codes, even though the exit codes were explicitly to be printed after the child processes had terminated.

Analysis:

Some experimentation with the code was performed in attempts to debug this strange behavior shell prompt printing behavior for file input redirection of commands. It was found that when printing text with a newline, the behavior went away and shell prompts were printed between output from lines of command. This can be explained by I/O buffering, where the OS saves input to be printed to the terminal in a buffer and prints it when ready. In the case of the shell prompt, using *fflush(stdout)* flushed out the buffer to the screen, fixing the strange order.

Also it was discovered that redirecting the output of *./msh* to another file would correct the order of the child processes' output and the exit codes, which is also likely an issue that can be attributed to I/O buffering.

Conclusion:

This was an informative assignment on understanding how to create and handle multiple processes, as well as pipe communication and file redirection for processes. It was challenging to get pipe communication and other input/output redirections to work, but some experimentation with file descriptors yielded successful results.

On my honor, I have neither given nor received aid on this assignment: Maurice Wong