

A modular kernel for the Raspberry Pi: Progress Report

November 26, 2018

Thomas Archbold
1602581
University of Warwick

Introduction

While the personal computer has been widely available to the general public for some time now, the introduction of the Raspberry Pi - a cheap but otherwise fully capable computer - has made experimenting with computers much more accessible, and invites tinkering at all levels. Be it simply getting experience in a new operating system, or interacting with GPIO pins to turn LEDs on and off, the Pi provides a simple platform for this to take place. There are a handful of official operating systems available for the Pi [1], addressing issues such as ease-of-installation, Internet of Things integration, or classroom management, there is little in the way of experimenting with the operating system itself. This project attempts to address this gap by providing customisation at the lowest levels of software that manages the computer; this involves altering its approaches to process scheduling, inter-process communication, and permanent storage, so that different options can be explored, and their impact can be felt, rather than dictated. By its end, the project will provide an operating system for the Raspberry Pi 2 Model B, whose approaches to CPU scheduling, inter-process communication, and filesystems may be altered by providing different compilation options at build-time.

Background & Motivation

An operating system is a piece of software which draws together a wide range of aspects within the field of computer science, from applying knowledge about a computer's organisation and architecture, to applying efficient algorithms on specifically crafted data structures, such as filesystem trees or resource allocation graphs. As such, a good operating system is one that successfully marries much of the theory behind computer science with its application, to produce a piece of software that is a microcosm of the field itself. Thus the main motivation behind the project was to not only gain experience in systems programming and in interacting with real-world hardware, but also to draw on the many areas of computer science to create one useful, entirely self-contained piece of software. This is with the goal in mind that by its end, the project will produce an operating system that offers configuration to the most fundamental degree, and which can serve as a solid foundation on which more useful features of an operating system can be added.

Inspiration has been drawn from a number of sources, not only for the project's conception, but also in guiding some early design decisions. In particular, Cambridge's *Baking Pi* [2], Stanford's *Pintos* [3], and [4] have so far been helpful in providing guidance to achieving a basic functioning operating system, albeit it need of translating for the Raspberry Pi 2. Some guidance from [5] has also been taken, and its use is likely only to increase. As for technical references, the BCM2835/6 Peripherals Manual [6, 7], Cortex-A7 MPCore Technical Reference Manual [8], and ARM Cortex-A Series Programmer's Guide [9] have all been useful in finding specific information about various attributes of the hardware and how to interact with it on the Pi. The official Github repositories, in particular [10], have also helped in this regard.

Current progress

The project is currently at a point at which the operating system is able to successfully boot in the emulated environment provided by QEMU. Since it has been capable of doing this since `boot.s` was written, early on in Term 1, it is important to note the specific stages of initialisation performed by the kernel at this point, as well as discuss the environment setup that has enabled this point in development to be reached.

Development environment

The project is being developed on a machine running Linux kernel version 4.16 onwards. Since the target environment, the Raspberry Pi 2 Model B, is different to that on which it is being developed, a cross-compiler is required to compile code that will run on the target machine, as opposed to the host. In particular, available for download on the ARM developer website [11] is the GNU Embedded Toolchain, which provides tools to target ARM Cortex family of processors, including the GNU Compiler Collection (GCC). Conveniently this suite of tools is available from Arch Linux's package manager, Pacman [12], and this is the version of the cross-compiler used in the makefile.

Before writing any code, as the author had little-to-no prior experience in systems programming, research had to be undertaken in order to get acquainted with this environment. In particular, this involved skimming over the various peripherals manuals, technical reference manuals, and programmer's guides for programming on the Raspberry Pi to learn more about its hardware and how to interface with it using ARM assembly. [6] and [7] detail the peripherals on board the Raspberry Pi 2 Model B, the layout of their related registers, and how to read and write them to do meaningful things with the hardware, while [8] and [9] provided help on ARM assembly's syntax, and how and when to use specific instructions.

Particularly important so far have been the sections of the peripherals manual on GPIO and UART, as until the implementation for the mailbox interface is working, all input and output is done through the serial connection provided by the UART. Since there were issues with getting the Pi to run on real hardware, information about the GPIO peripheral was needed in order to write basic low-level debugging functions, mainly in the form of getting the green ACT LED to blink for various return values of functions.

`boot.s`

The first piece of code written was `boot.s`, which is responsible for providing the basic setup of the entire system, which includes initialising a minimum C environment. In particular, it sends three of the four cores on the CPU to shutdown (to decrease overall complexity of the system, as discussed in the specification), initialises the stack pointer at address `0x8000`, sets up the BSS segment (where statically-allocated variables that are not explicitly initialised are stored) and zeroes it out (as required by the C standard), and then loads our C kernel entry point, `kernel_main`, into memory to begin its execution. Note that the Program Counter for the kernel starts at address `0x8000` and grows upwards, so the stack can safely start at `0x8000` without interfering with the kernel (as it grows downwards).

`linker.ld`

The code in `linker.ld` is responsible for linking all of the compiled object files into one final executable. There are scripts which do this for user-space programs, but since we are our own user-space, being the kernel, we have to create one for ourselves. The various sections that the script defines are as follows:

- `.text` - contains executable code
- `.rodata` - read-only data i.e. global constants
- `.data` - global variables initialised at compile-time
- `.bss` - uninitialised global variables

We also define the entry point of our entire operating system in this script, namely the `_start` routine from `boot.s`, and also set the symbols `__start` and `__text_start` to be at `0x8000`, which is where the bootloader will put the kernel image - the code from `boot.s` will be put in the first part of this section, `.text.boot`. We use `KEEP` to tell the compiler not to try to optimise the code in `.text.boot`, and set

the page size as 4KiB using `ALIGN`. The `.rodata`, `.data`, and `.bss` sections are then declared in similar ways.

Makefile

The makefile was written to speed up the build process, and there are only a few features to note. First is that here we specify that we are using the `arm-none-eabi` toolchain, for the compiler to target the Raspberry Pi's architecture as opposed to our own, in particular the Cortex-A7 processor. The `-fpic` compilation flag is currently being used while we are ignoring the Memory Management Unit, and in particular virtual memory, on the Pi, as it creates position-independent code, and will keep separate applications from interfering with each other within this single address space when processes are implemented. The `-ffreestanding` compiler and linker flag, and the `-nostdlib` linker flag, specifies that we are writing code in a freestanding environment, and as such do not expect much of the C standard library to exist, or for program startup to necessarily be at `main()`. Specifically, we only have access to the following header files: `<float.h>`, `<iso646.h>`, `<limits.h>`, `<stdalign.h>`, `<stdarg.h>`, `<stdbool.h>`, `<stddef.h>`, `<stdint.h>`, and `<stdnoreturn.h>`. The rest of the standard library must be implemented ourselves.

Atags

The first piece of setup to be done was discovering the amount of memory available to the system. On the Raspberry Pi, the bootloader creates a list of information about the hardware called Atags, places it at address `0x100`, and passes it as the third parameter to `kernel_main` in register 2. Each tag in the list contains a header consisting of two unsigned 32-bit values: the size of the tag (in 32-bit words), and the tag value. Each header is then followed by information specific to that tag. To access the information in each of the tags when we come across them, the layout of each of the tags [13] was matched by defining appropriate C structures in `atag.h`.

To find the amount of memory on the device, we skip through the Atags list (using pointer arithmetic and information about the tag's size in its header) until we come across the `ATAG_MEM` tag. The function `get_total_mem()` then simply returns the value of the `size` field of the `atag_mem` struct.

Organising memory

Throughout the operating system's execution, different processes will require memory to perform computations. Although the only thing doing anything meaningful at the moment is the kernel, which can theoretically use any memory it wants, in order to impose some order for later implementations of processes and user space, memory is split into 4KiB pages.

To organise the pages, each page has been given a header containing information about whether the page has been allocated, whether it is a kernel page, and whether this page is part of the heap (used later when dynamically allocating memory). The headers have been organised into a linked list directly after the end of the kernel image, using the `__end` variable from the linker script. Each page also stores the virtual address to which it maps; as virtual memory has not yet been implemented, all pages are simply identity mapped. The pages are then iterated over and their headers initialised, and each is added to a linked list of free pages.

Next, page allocation and deallocation was implemented. Allocating a page is done by simply popping the head of the free page list, setting the appropriate flags, and returning the address of the page. Freeing, meanwhile, is done by passing the address of the page to free, again setting the appropriate flags, and appending this page back to the free page list.

Allocating memory

A 1MiB portion of memory located directly after the page headers is reserved for the heap. The struct `heap_segment` is defined for keeping track of heap allocations, such as the segment size and whether it has been allocated (useful for avoiding external fragmentation later), and the heap is initialised by declaring a single heap segment whose size is equal to that of the heap - as more memory is allocated, this segment will be split into smaller ones, which may be split further to satisfy requests for differing amounts of memory.

Segment allocation is done by iterating over segments in the heap and finding one which best satisfies the number of bytes requested, and which is not in use. If one is found, the address directly after the

segment's header is returned. To free an allocation, we set the appropriate flags, and then attempt to merge consecutive free segments, checking both to the left and the right of the current segment.

Serial output

Initial output was done using the UART on board the Pi, meaning text was sent and received through serial ports. This was mainly due to simplicity in early builds of the system, and while this can be achieved in the final version of the project by using a USB-to-TTL cable, ideally it will use the HDMI port on the Pi instead, which requires interfacing with the Mailbox peripheral (discussed later). This was all done by interacting with the GPIO pins on the Pi, which is done entirely through Memory Mapped I/O (MMIO) - that is, by reading from and writing to predefined memory addresses.

A peripheral on the Raspberry Pi is simply an address to and from which you may read and write data, and all may be described by an offset from the Peripheral Base Address; this is `0x3f200000` on the Raspberry Pi 2 Model B. Moreover, a register is a 32-bit chunk of memory that a peripheral may read from or write to. The BCM2835 Peripherals Manual gives the UART base address as `0x7e201000`. Thus, in `gpio.c`, after setting all the required flags for using the UART, we can implement a serial `putc()` by checking that the FIFO is not full and writing our data to the Data Register, and `getc()` by checking that the FIFO is non-empty and reading from it.¹

HDMI output

The final version, however, will send output through the HDMI port on the Pi. To do so, the Mailbox peripheral has been used. The Mailbox is a peripheral that facilitates communication between the ARM CPU and the VideoCore GPU [14], and starts at offset `0xb880`. We may get data from the GPU via the read register, pass data to the GPU using the write register, and check if either of these are empty or full using the status register, at offsets `0x00`, `0x20`, and `0x18` respectively. Furthermore, a channel is a number that gives meaning to the data being sent to and received from the GPU - for interacting with HDMI, we need the Property channel, channel 8. This provides a means to get and set data about various hardware devices, one of which is the framebuffer.

In order to ask the GPU for a framebuffer, we communicate with it by sending messages and parsing the received response. The messages sent set the framebuffer's physical and virtual dimensions, and its colour depth (bits per pixel). Once all the parameters are set, we ask the GPU for a framebuffer, using the `FB_ALLOCATE_BUFFER` tag (defined in `mailbox.h` as `0x00040001`). The returned values are a pointer to this framebuffer and its size.

Booting on real hardware

This is the first point at which the project has required testing on real hardware, in order to verify that a framebuffer is being correctly requested and supplied by the GPU. This required the kernel to be installed on an SD card and for it to be run on the physical board of the Pi. It is now helpful to detail the unique boot process of Pi, which makes this stage much easier.

The boot process of the Raspberry Pi

The boot process relies on closed-source proprietary code programmed into the SoC processor [15] which cannot be modified. Importantly, the ARM CPU is not the main CPU - it is a coprocessor to the VideoCore GPU. Upon powerup, the ARM CPU is halted and the GPU is run. The firmware then loads the bootloader from ROM to the L2 cache and executes it. This first stage bootloader mounts the FAT32 boot partition on the SD card so that the second stage bootloader may be accessed. This is its only responsibility, to load `bootcode.bin`. The first stage bootloader is programmed into the SoC itself during manufacture and cannot be reprogrammed by the user. Next, the second stage bootloader (`bootcode.bin`) then retrieves the GPU firmware from the SD card, programs the firmware, then starts the GPU. The GPU firmware (`start.elf`) is loaded, and allows the GPU to start up the CPU. An additional file `fixup.dat` is used to configure the SDRAM partition between the GPU and CPU. Here the kernel image is loaded, the CPU is released from reset, and control is transferred to it to execute the kernel.

¹From the manual: "Physical addresses range from `0x20000000` to `0x20ffffff` for peripherals. The bus addresses for peripherals are set up to map onto the peripheral bus address range starting at `0x7e000000`. Thus a peripheral listed at `0x7ennnnnn` will be available at physical address `0x20nnnnnn`."

After the operating system is loaded, the code on the GPU is not unloaded; instead, it runs its own simple operating system, called Video Core Operating System (VCOS). The kernel can then use this to communicate with the services it provides (e.g. providing a framebuffer, as above) using the Mailbox Peripheral and interrupts (the GPU is able to produce ARM interrupts). The GPU is not only in charge of graphical functions - it also controls clocks and audio, for example. In this way the GPU firmware is similar to a normal PC's BIOS (Basic Input/Output System) [16, 17].

Loading using Raspbian

To install the project's kernel onto an SD card, it was copied to an SD card already containing an operating system, in this case Raspbian [18], and its `kernel17.img` was replaced with that produced by the Makefile (after renaming it to `kernel17.img`). Now, when the bootloader would look to load Raspbian's kernel image, it loads the project's instead.

Debugging on real hardware

On the first attempt, printing via HDMI did not work, despite functioning in the emulated environment provided by QEMU. In particular, the screen displayed the "rainbow screen", signifying that somehow the framebuffer is not being initialised properly. In an attempt to debug this, assembly code was written, under the guidance of [2], to make the green ACT LED blink at certain points in the code, to verify that these points were being reached successfully.

Project management

Development

The project has been developed incrementally - this has involved deciding on the next feature to implement, breaking it down into its own set of requirements, and tackling each of these in turn. An example of this is memory management, for which it was broken down into getting information about memory through the Atags, implementing paging, then implementing segmentation and a dynamic memory allocator. Up to this point in the project there has been little flexibility in the order in which things are completed, as some of the systems are still being implemented to even allow "higher-level" features to exist. An example of this is getting the project to work on real hardware before interrupts and interrupts before scheduling: the former because of QEMU's lack of simulation of a system timer, and the latter because processes cannot be interrupted if there is no way to interrupt anything. Once the base systems are implemented, development can open up more, with more flexibility as to deciding on which features to develop next. This will of course need to be managed to ensure all areas that the project aims to address are getting attention.

Version control has been a constant focus during development, and has been important. Due to the size of this project already, with the amount of code from a number of different files all contributing to the system's functionality, any time a change is made that has broken the project where a previous version has worked has been much less of an issue. In particular, using Git has allowed for specific differences between commits to be presented clearly, helping to track down the source of various problems much more quickly.

Progress has been kept in mind by timetabling fortnightly meetings with the project supervisor. This has allowed progress to be regularly considered, when otherwise it may be easy to forget about the project in favour of other modules. Regularly discussing the project has also consolidated things learned over the course of the fortnight, and help identify exactly what progress has been made.

Testing

As the focus of the project so far has been to set up the foundations upon which more meaningful work can later on be built, testing has been done entirely manually, simply as there does not exist the platform to do otherwise at this stage. In particular, nothing much more can be done when testing if the project will boot on a physical system other than loading it onto an SD card and trying it. One aspect that has been able to be checked by unit-testing are implementations of functions in the C standard library. In this case, for example when testing `printf()`, the function was called with a range of different values, for example very large, negative, and zero, to test that such different cases could be handled, and if not,

it was then explored manually why not. In this way, testing has been a slow process, but equally in a way that demands the necessary attention to detail to be paid to writing a stable operating system.

Ethical consent

All software being used to develop the project is available under the GNU Public License. At points in the project's development, informal feedback may be requested from friends and colleagues for more of the non-functional requirements, such as usability or how the project is presented. Due to the informal nature of this feedback, however, any associated ethical, social, or professional issues are insignificant.

Progress issues

Comparison with original timetable

Progress has largely been in line with that which was originally scheduled. Early on, progress was even slightly ahead of the timetable, particularly when writing `boot.s` and `linker.ld`, but this is largely due to there being little need to stray from the example code provided by [19]. Overall, however, what was achieved compared to what was timetabled differed little in the first five weeks of the project. It was attempting to move to working and testing on real hardware where the two began to diverge - although it could boot on real hardware (evidenced by flashing the ACT at different times), problems arose when printing output to a real HDMI screen. Due to the limited facility to debug at this level, it has been particularly difficult to work towards a solution to this problem, and the project has fallen behind by a few weeks on this goal. As a result of this, at the time of writing, progress should be underway on loading programs into memory for execution, whereas in reality processes have not even begun development yet. Interrupts and processes should also have been completed by this point, and while progress has started on the former, the system will need to operate on real hardware for it to be finished, due to QEMU's lack of system timer. Therefore, interrupts and processes will not be able to be completed until the issue of outputting to HDMI is resolved.

A reason for this oversight in the timetable is related to underestimating the problems that could arise in the transition from an emulated environment to a fully real one, as well as the time required to fix them. Furthermore, while course load was considered when writing the initial schedule, unforeseen issues with other pieces of coursework have interfered with progress on this project, more than previously accounted for. Time was however set aside during the winter holidays specifically to catch up in the case of delays, and as such the revised timetable, given in Appendix A, only differs slightly at the start of Term 2.

Areas for timetable revision

The first change made to the timetable is the addition of several tasks, and merging of others into one. For example, booting has been given its own section, and memory management has been broken into organising memory by accessing Atags, and dynamic memory allocation. HDMI support has also been split into sections on mailboxes, framebuffers, and HDMI output as the task to get them working was larger than previously thought and deserving of more separate goals. Booting on real hardware has been added as a section to signify the time spent debugging the problems with transitioning on the real device, and is currently active.

Virtual memory has been introduced, and where previously filesystems was split into persistent and load-on-request, it is now one block. Userspace is now a larger section, and encompasses the fork and execute and threading goals. The shell section has been extended to highlight its ongoing nature, while the stretch goals of networking and a text editor have been added in similarly extended blocks of time, as it is unlikely they will be completely done before the project's end.

One section that has been dropped entirely from the timetable is the development of a C standard library. That is not to say that it is not being implemented, but that it is only an auxiliary task - it will be continually developed as needed by other parts of the operating system, and not, as previously it might have implied, for its own sake. As an example, `itoa()` and `printf()` have been implemented, not because they show up in Linux, but because they provide meaningful functionality for tasks such as debugging. Therefore, this section is still present, albeit implicitly, as it is so closely tied to all other tasks.

Further, tasks have been shifted around in the revised timetable compared to the original. Firstly, up to 25 November, 2018, the timetable now reflects the actual progress which was made, showing that booting and displaying on real hardware is still in progress, and interrupts will be started this week. Most of the tasks following this in the original timetable have been pushed back to highlight this delay. Scheduling now carries into Term 2, and scheduled time to implement more schedulers has also been moved forward where previously it was to be started later. Filesystems have also been moved earlier in the term, as has message passing. This is to allow enough time for compilation configuration to be developed, which has also been moved forward, to allow enough time to get it working well for the presentation. The endpoint for the presentation has also moved forwards to reflect the true deadline, while work on the final report will also now start much earlier.

Next steps

For the rest of Term 1, most efforts will be dedicated to debugging the issue with printing on HDMI on real hardware. The nature of the problem, that nothing but a rainbow displays on a real screen, makes the process more difficult in that debugging will rely on tweaking the flash of an LED. With the winter holidays approaching, this issue should find resolution, if not before the end of term, then certainly before the start of next. Furthermore, to make up the deficit caused by this problem, more attention will need to be paid towards interrupts, exceptions, and processes, in order to finish them in a timely manner so that work can be at least started on scheduling, inter-process communication, and synchronisation before the calendar year's end.

Some further research will also need to be undertaken before the start of Term 2 to decide on a filesystem to implement, and assess the options of implementing a custom one, or one that is already established, such as SFS, FAT, BMFS, or something else [20]. It is true that at this point, there is little to distinguish the project from any other hobbyist operating system around, given that there is nothing it does that is special or particularly unique at this stage. Over next term is where the project will begin to diverge from other operating systems and take on its own form, that is, in granting the user a degree of customisation at a low level which others do not provide. Going into the start of Term 2, focus will be on developing the features that will begin to differentiate this operating system from others, namely different approaches to CPU scheduling and inter-process communication, and thought will need to be put into designing the interface for customising the operating system at compile-time.

Reflection

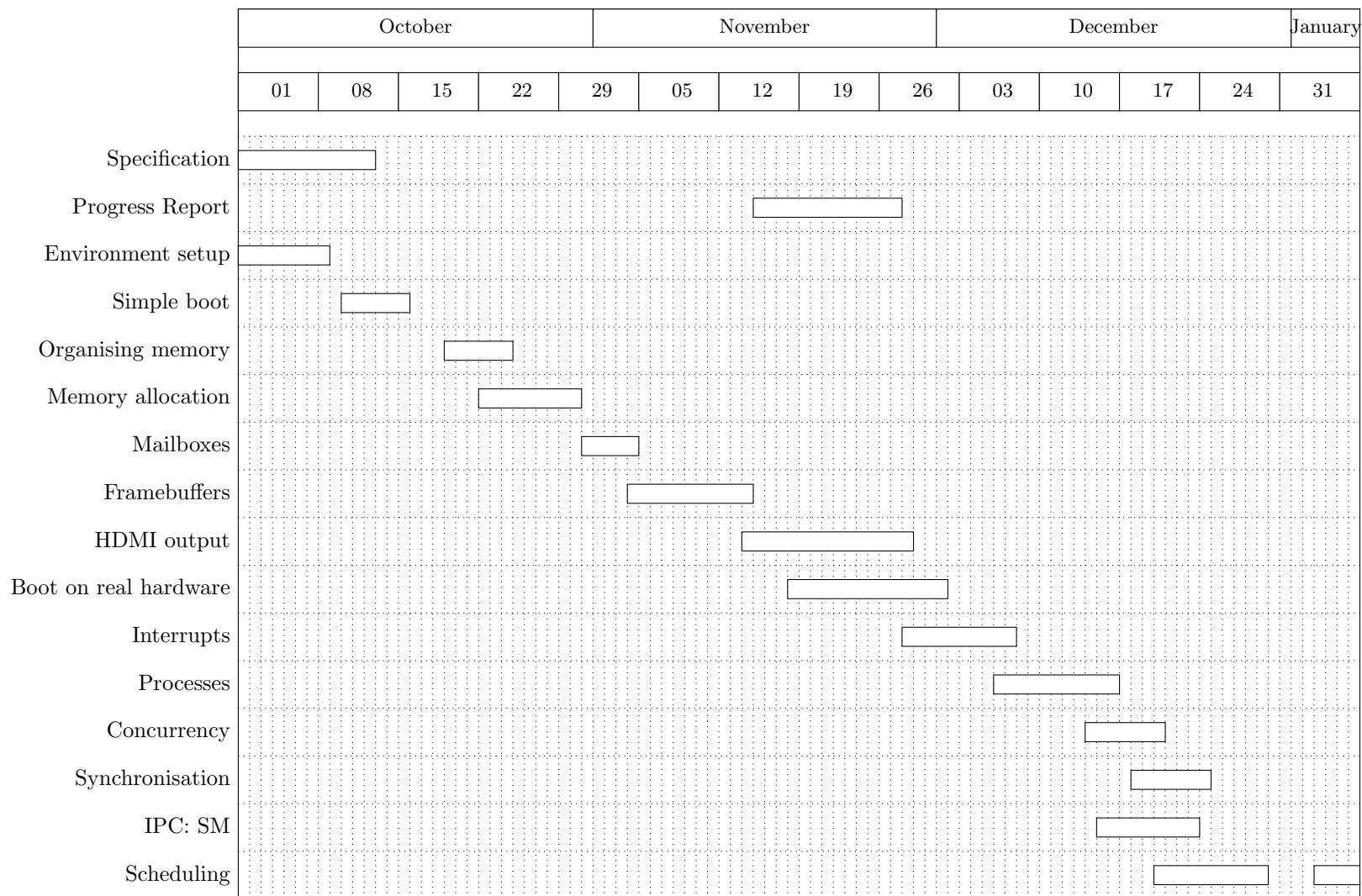
The progress made with the project so far has been pleasing; having little experience in systems programming prior to starting this project, it has been fulfilling learning about every aspect of the Raspberry Pi in minute detail, which is indeed required to make things work at a low level. While it has been noted that at this point there is little to distinguish this piece of work from any other hobbyist operating system, it is nonetheless satisfying to have set this ground work which is, all things considered, still a custom operating system that is capable of running on a Raspberry Pi. Furthermore, despite the issue caused by the framebuffer, it has been a fun challenge to develop a debugging solution for this, and while a blinking green LED may not be a core feature of the final project, learning how to use the GPIO and doing so in assembly has been another valuable, and fun, learning experience. Overall, despite the setbacks caused by running on real hardware, the project is progressing well, and while difficulty is expected to pick up into next term, so to is time spent working on it, hopefully allowing for a smooth transition in between terms and for steady progress to continue being made.

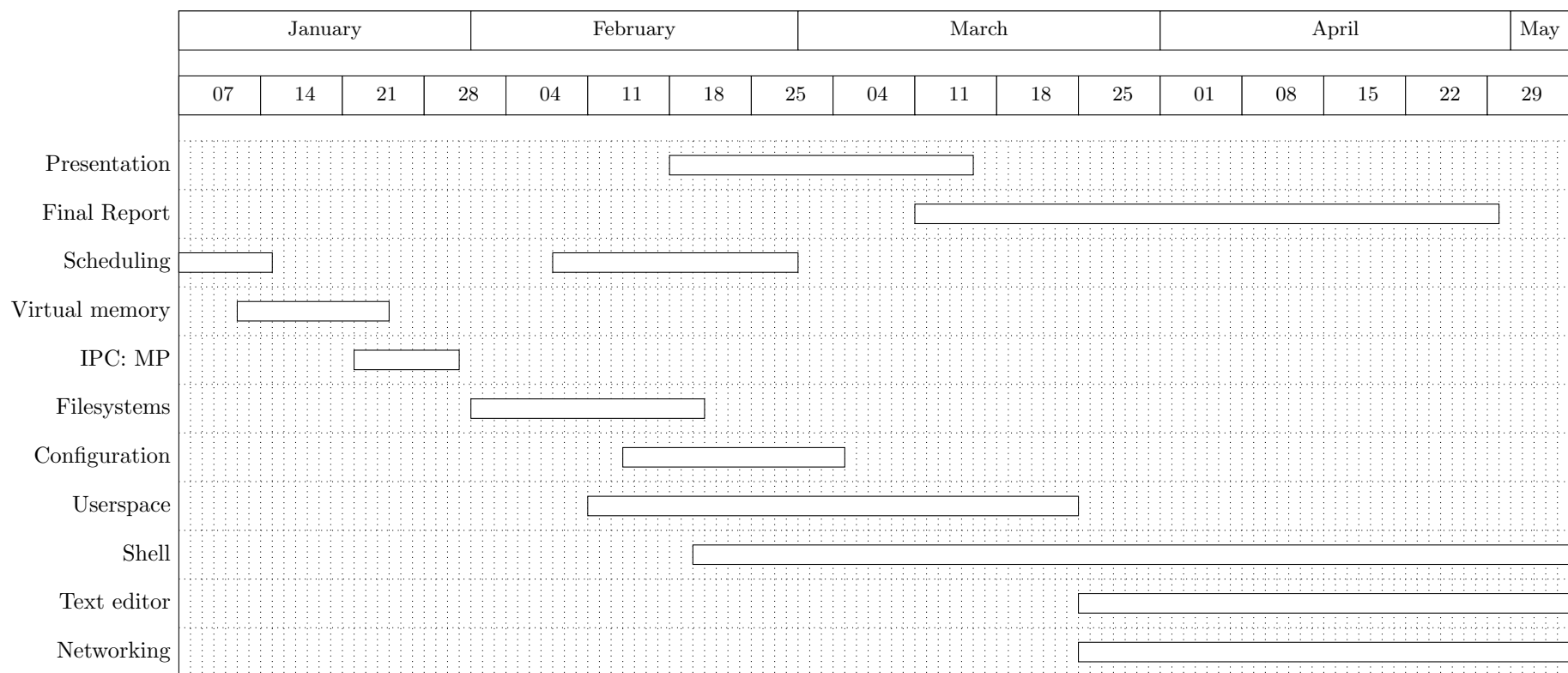
References

- [1] "Downloads." <https://www.raspberrypi.org/downloads/>. Page accessed: 2018-11-22.
- [2] "Baking pi - operating systems development." <https://www.cl.cam.ac.uk/projects/raspberrypi/tutorials/os/index.html>. Page accessed: 2018-11-22.
- [3] B. Pfaff, "Pintos." https://web.stanford.edu/class/cs140/projects/pintos/pintos.html#SEC_Top, December 2009. Document accessed: 2018-08-24.

- [4] "Building an operating system for the raspberry pi." <https://jsandler18.github.io/>. Page accessed: 2018-05-17.
- [5] E. Helin and A. Renberg, "The little book about os development." <https://littleosbook.github.io/>, January 2015. Document accessed: 2018-11-22.
- [6] Broadcom Corporation, Broadcom Europe Ltd., 406 Science Park, Milton Road, Cambridge, *Broadcom BCM 2835 ARM Peripherals*, 2012.
- [7] G. van Loo, *Broadcom BCM 2836 ARM Peripherals*. Broadcom Corporation, Broadcom Europe Ltd., 406 Science Park, Milton Road, Cambridge, 2014.
- [8] ARM, *Cortex-A7 MPCore: Technical Reference Manual*, 2013.
- [9] ARM, *ARM Cortex-A Series: Programmer's Guide*, 2013.
- [10] "Raspberry pi: Documentation." <https://github.com/raspberrypi/documentation>. Page accessed: 2018-11-22.
- [11] "Gnu-rm downloads." <https://developer.arm.com/open-source/gnu-toolchain/gnu-rm/downloads>. Page accessed: 2018-11-18.
- [12] "arm-none-eabi-gcc 8.2.0-1." https://www.archlinux.org/packages/community/x86_64/arm-none-eabi-gcc/. Page accessed: 2018-11-18.
- [13] "Booting arm linux: Tag reference." http://www.simtec.co.uk/products/SWLINUX/files/booting_article.html#appendix_tag_reference. Page accessed: 2018-11-20.
- [14] "Accessing mailboxes." <https://github.com/raspberrypi/firmware/wiki/Accessing-mailboxes>. Page accessed: 2018-11-21.
- [15] "Raspberry pi firmware." <https://github.com/raspberrypi/firmware>. Page accessed: 2018-11-21.
- [16] "Raspberry pi: What bios does raspberry pi use?." <https://raspberrypi.stackexchange.com/questions/8475/what-bios-does-raspberry-pi-use>. Page accessed: 2018-11-15.
- [17] "Raspberry pi: Level of hackability of raspberry pi." . Page accessed: 2018-11-15.
- [18] "Raspbian." <https://www.raspberrypi.org/downloads/raspbian/>. Page accessed: 2018-11-22.
- [19] "Raspberry pi bare bones." https://wiki.osdev.org/Raspberry_Pi_Bare_Bones. Page accessed: 2018-11-26.
- [20] "File systems." https://wiki.osdev.org/File_Systems#.22Beginners.22_filesystems. Page accessed: 2018-11-22.

Appendix A Revised Timetable





Appendix B Project Specification

A modular kernel for the Raspberry Pi: Project Specification

12 October, 2018

Thomas Archbold
1602581
University of Warwick

Background

In most operating systems, many design decisions are made in order to keep things simple for the user, by keeping most of the technical details hidden. In most cases, this is an appropriate approach: needlessly offering more choices for low-level tasks that are usually handled by the operating system, such as CPU scheduling algorithms, would only serve to confuse the average user. It may actually be detrimental to the security and the stability of the system by opening up more opportunities for errors to be introduced. This more insulated approach does mean, however, that the user never really knows what is going on “under the hood”, and indeed whether greater performance can be achieved by making *different* fundamental decisions. Furthermore, a number of operating systems exist for the Raspberry Pi, some focusing on ease-of-installation, with Linux’s NOOBS [1] distribution, others on Internet of Things integration, such as the Windows 10 IoT Core distribution [2]. Yet, none exist to serve as an experimental operating system, designed as a testbed for making and changing these low-level behaviours. This project aims to fill this gap for the operating systems enthusiast, one who wishes to test for themselves the different approaches to CPU scheduling, interprocess communication, and filesystems. It will give the user the ability to alter the fundamental ways in which their machine operates by compiling different modules to handle different tasks, enabling for a more flexible operating system where such things can be tweaked at any point.

Main goal

The goal of this project is to create a modular operating system for the Raspberry Pi 2 Model B that is capable of loading different modules at compilation time to tackle CPU scheduling, interprocess communication, and filesystems in a variety of ways. Specifically, it must have some way to run and switch between multiple processes using a CPU scheduler; to use both shared memory and message passing for interprocess communication; to create, read, update, and delete files and directories using a custom filesystem; and to interface with the SD card for permanent/mass storage. To achieve this, it must implement an interface for compiling different modules, similar to Linux’s `insmod`, `rmmmod`, and `Kbuild` system [3, 4]. Furthermore, as executing processes forms a key functional requirement for the project, there must be a convenient way to load programs into memory and begin their execution. A solution to this is to implement a basic shell/command interpreter.

Finally, a key objective of this project will be to get the operating system to work entirely on real hardware, and not solely in an emulated environment. This includes booting from the SD card installed in the Raspberry Pi. As the boot process is handled by the Pi’s System on Chip (SoC), booting will be possible without writing a custom bootloader. On top of booting from it, the operating system must interact with the SD card in conjunction with a filesystem for permanent/mass storage. Finally, it must be capable of taking input from a keyboard connected via USB, and printing output to a physical screen via its HDMI port.

The kernel

The kernel will be built using the cross-compiler from GCC for `arm-none-eabi`, which provides a toolchain to target the System V Application Binary Interface (ABI). As a result, programs and frag-

ments of programs on disk, and by extension the kernel itself, will be in the Executable and Linkable Format (ELF) after compilation and linking. The kernel will use just a single core of the four available to the BCM2836, but will support multithreading, both at the kernel and user levels, with appropriate interfaces being written in both cases.

The memory available to the operating system will be organised into pages, and furthermore it will use a dynamic memory allocator, similar to the C standard library's `malloc()` and `free()`, to further split the available memory into segments. Processes will need to be loaded into and out of memory, and as such will need an appropriate representation as a Process Control Block (PCB), and will need to be stored in a Process Table to facilitate context switching. On top of this, the kernel must also be able to handle interrupts and exceptions to safely halt processes and bring them out of memory. At this point, CPU scheduling will need to be tackled, and a long-term and multiple short-term schedulers implemented. As QEMU does not simulate a system timer, the move to working on real hardware will coincide with the introduction of multitasking, involving taking input via USB keyboard and printing output via HDMI.

With the possibility of multiple running processes, synchronisation will need to be tackled, most likely using semaphores, and the issues of deadlock avoidance, detection, and correction will need to be considered. Furthermore, solutions for interprocess communication will then be developed, most likely starting with shared memory due to its simplicity. Message passing will follow as a configurable module. Beyond this, a filesystem can then be implemented and development can move to focus more on user space, including a command interpreter, actually accessing mass storage, and implementing `fork()` and `execute()`. At this time the notion of syscalls and operating system traps will also need to be developed. As an operating system needs to be written in a freestanding (as opposed to hosted) environment, a standard library will be continually developed over the course of the project.

Configurable modules

The project must implement the following as modules, which may be configured at compilation time by the user:

- CPU Scheduling:
 - First Come First Served
 - Round Robin
 - Shortest Job First
 - Shortest Remaining Time First
 - Priority Scheduling (preemptive and non-preemptive)
 - Lottery Scheduling
- Interprocess Communication
 - Message passing
 - Shared memory
- Filesystem
 - persistent
 - load-on-request

Stretch goals

Some stretch goals which should be implemented to show understanding of more complex structures would be some more intricate scheduling algorithms, including the following [5, 6, 7, 8]:

- Completely Fair Scheduler
- Multiple Queue Skiplist Scheduler, MuQSS
- Multilevel Queue and Multilevel Feedback Queue
- $\mathcal{O}(n)$ Scheduler
- $\mathcal{O}(1)$ Scheduler

In order to give the operating system more purpose and to increase usability, the collection of relatively simple programs on offer should be extended, including a mix of long running CPU- and I/O-bound programs. This will mean that the relative performance of the schedulers may be seen more easily. While the Not Recently Used (NRU) algorithm will be used for page replacement due to its low overhead and decent performance, other algorithms could be explored and implemented as modules. These may include: First-In-First-Out (to highlight its poor performance), the Clock Page Replacement algorithm, and the Least Recently Used algorithm [9].

Further extensions

Beyond these goals, further extensions would focus on increasing the usability of the system, and start to shape it into one which someone might actually use to get things done. One of the simpler ways to achieve this would be to write a text editor. Additionally, implementing networking into the operating system would vastly increase its usability and general usefulness. Such goals are rather far-fetched given the time frame of the project, but would form meaningful projects later in the life of the operating system.

Out-of-scope

Features which will not be implemented in the project include graphical user interfaces and any form of security. Graphics would increase the complexity of the project too much, and provide too little reward, to be considered a worthwhile goal. While security would be easier to implement, for example by following suit of Linux's permissions interface [10], it would again detract attention from features more in line with the project's goals. After all, the operating system produced will only be experimental and designed for use by one user, and as such security will be an unnecessary feature.

Hardware

Compared to the Raspberry Pi 1, the Raspberry Pi 2 Model B has:

- 900 MHz quad-core ARM Cortex-A7 CPU
- 1GB RAM

Like the Pi 1 Model B+, it has:

- 40 GPIO pins
- 4 USB ports
- Full HDMI port
- MicroSD card slot
- 100 Base Ethernet
- VideoCore IV graphics Core
- Combined 3.5mm audio jack and composite video
- Camera interface (CSI)
- Display interface (DSI)

The main reason for choosing to work with the Raspberry Pi was due to its simple boot process, details of which can be found here [11]. In particular, as it is handled entirely by its SoC, it means a custom bootloader to load the kernel into memory and transfer control to it will need not be written. The Raspberry Pi 2 Model B in particular uses the BCM2836 processor, whose underlying architecture is identical to the 1's BCM2835 and the 3's BCM2837 chips. The only difference is that the 2 uses the quad-core Cortex-A7 cluster as opposed to the ARM1176JZF-S or the quad-core ARM Cortex-A53 cluster, as used by the 1 and 3 respectively [12, 13]. Therefore, as the choice between specific models of the Pi would make little significant difference to the outcome of the project, it made sense to opt for the one already available to the author at the time the project was conceived, namely the Raspberry Pi 2 Model B.

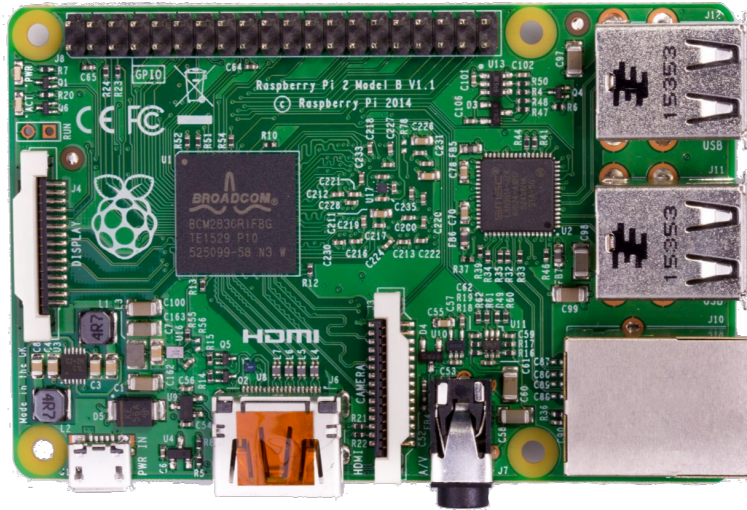


Figure 1: Raspberry Pi 2 Model B [14]

Methodology

The methodology best suited to the project will be a mix between plan-driven and agile approaches. The early stages of the operating system's development will benefit from the former, as the requirements, such as booting before memory management before writing scheduling algorithms, will abide by a rigid structure. An incremental approach will likely be used as opposed to a waterfall methodology, however, due to its less restrictive nature, and to offer choice when it is appropriate about what to implement next. After the foundations are laid, the project will likely move to a more agile approach, where scrum cycles will be useful both for their flexibility and choice, and their focus on finishing one aspect of the project at a time.

Throughout the project, regular meetings will be taken with the supervisor to discuss progress, current problems, and ideas for solutions when necessary. Organisation will of course be a key aspect to the success of the project, and the project timetable will be updated to reflect the project's progress. The meetings will start at once every fortnight in Term 1, and increase to once weekly in Term 2, simply due to timetabling and course load for other modules.

Testing

The project will be tested incrementally. In its early stages, progress will simply not be able to be made until some systems operate correctly, so thorough manual testing of such areas will be vital as there will simply not be the platform to write dedicated unit tests. As it progresses, and unit tests become more viable, they will be written to cover most likely paths of execution to identify shortcomings of the system, and dealt with accordingly. Since the project's aim is to create a configurable operating system, manual testing will again need to be undertaken in order to verify whether it works under the various combinations of modules. This will test both the correctness and the stability of the system rigorously, two of the most important goals of any piece of software.

Timetable

See Appendix.

Technologies

The following technologies will be used by the project:

- Git - version control

- Github - to access the project from multiple sources, as well as to back it up
- C - the language in which most of the operating system will be implemented
- ARM assembly - used when C is unavailable/inappropriate [15]
- GCC cross compiler for ARM EABI - for cross compiling for the target processor, the Cortex-A7
- QEMU - for emulating the Pi to allow quicker and safer testing ¹
- Make - automate the build process

Resources

The following documentation will be used throughout for reference to the architecture of the Cortex-A7 processor and its instruction set, and the peripherals on the Pi:

- Cortex-A7 MPCore Technical Reference Manual
- ARM Cortex-A Series Programmer's Guide
- Broadcom BCM2835 ARM Peripherals Manual

Additional guidance will be taken from the MINIX book [16], Stanford's Pintos [17], and [18].

Legal, social, ethical, and professional considerations

All software used to build the project is available to use under the GNU Public License. Throughout the project's development, some testing will be required from people other than the creator, to gain informal feedback especially with regards to usability; these people are likely to be friends and colleagues, hence the social, ethical, and professional issues are insignificant.

References

- [1] "Noobs." <https://www.raspberrypi.org/downloads/noobs/>. Page accessed: 2018-10-08.
- [2] "An overview of windows 10 iot core." <https://docs.microsoft.com/en-us/windows/iot-core/windows-iot-core>. Page accessed: 2018-10-08.
- [3] "insmod(8) - linux man page." [modprobe://linux.die.net/man/8/insmod](http://linux.die.net/man/8/insmod). Page accessed: 2018-10-03.
- [4] "Kbuild: the linux kernel build system." <https://www.linuxjournal.com/content/kbuild-linux-kernel-build-system>. Page accessed: 2018-10-09.
- [5] "Cfs scheduler." <https://www.kernel.org/doc/Documentation/scheduler/sched-design-CFS.txt>. Page accessed: 2018-10-03.
- [6] "Kernel patch homepage of con kolivas." www.users.on.net/~ckolivas/kernel/. Page accessed: 2018-10-03.
- [7] A. Silberschatz, P. B. Galvin, and G. Gagne, *Operating System Concepts*, ch. 5: Process Scheduling. Wiley, 9 ed., 2014.
- [8] "Scheduling: Linux." [https://en.wikipedia.org/wiki/Scheduling_\(computing\)](https://en.wikipedia.org/wiki/Scheduling_(computing)). Page accessed: 2018-08-30.
- [9] A. S. Tanenbaum and A. S. Woodhull, *Operating Systems: Design and Implementation*, ch. 4: Memory Management. Pearson, 3 ed., 2009.
- [10] "Ownership and permissions." https://access.redhat.com/documentation/en-US/Red_Hat_Enterprise_Linux/4/html/Step_by_Step_Guide/s1-navigating-ownership.html. Page accessed: 2018-10-09.

¹QEMU does not simulate a system timer (at least for the Raspberry Pi 2), so some testing will eventually need to be done on real hardware

- [11] "Raspberry pi boot process." <https://www.raspberrypi.org/forums/viewtopic.php?f=63&t=6685>. Page accessed: 2018-10-11.
- [12] "Bcm2836." <https://www.raspberrypi.org/documentation/hardware/raspberrypi/bcm2836/README.md>. Page accessed: 2018-10-08.
- [13] "Bcm2837." <https://www.raspberrypi.org/documentation/hardware/raspberrypi/bcm2837/README.md>. Page accessed: 2018-10-08.
- [14] "Raspberry pi 2 model b." <https://www.raspberrypi.org/products/raspberry-pi-2-model-b/>. Page accessed: 2018-10-11.
- [15] "C: Things c can't do." https://wiki.osdev.org/C#Things_C_can.27t_do. Page accessed: 2018-08-30.
- [16] A. S. Tanenbaum and A. S. Woodhull, *Operating Systems: Design and Implementation*. Pearson, 3 ed.
- [17] "Pintos." https://web.stanford.edu/class/cs140/projects/pintos/pintos_1.html. Page accessed: 2018-10-09.
- [18] "Building an operating system for the raspberry pi." <https://jsandler18.github.io/>. Page accessed: 2018-10-09.

