



An Educational Kernel for the Raspberry Pi

Thomas Archbold

1602581

Department of Computer Science

University of Warwick

April 29, 2019

CS310 Third-year Project
supervised by Adam Chester

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Keywords – operating systems, kernel, Raspberry Pi, process scheduling, inter-process communication

Contents

1	Introduction	4
1.1	Motivation	4
2	Background	4
2.1	Relevant Material	4
2.2	Why is this project worthwhile?	5
2.3	Useful concepts	5
3	Design	6
3.1	Hardware	6
3.1.1	The Raspberry Pi's boot process	6
3.2	Resources & Environment	7
3.2.1	Technical Documentation	7
3.2.2	System V ABI	8
3.2.3	ARM environment	8
3.2.4	Tools	9
3.3	System Overview	9
3.4	Project management	10
3.4.1	Development methodology	10
3.4.2	Organisation	11
3.4.3	Testing & Debugging	11
3.4.4	Legal, Ethical, and Social Issues	12
4	Implementation	12
4.1	System initialisation	12
4.1.1	Linker Script	12
4.1.2	Initialising the C runtime	12
4.2	Organising memory	13
4.2.1	Atags	13
4.2.2	Pages	14
4.2.3	Dynamic memory allocator	14
4.3	Serial Output	14
4.3.1	Peripherals	15
4.4	Interacting with the GPU	15
4.4.1	Mailbox Peripheral	16
4.5	Interrupts and Exceptions	17
4.6	System Timer	17
4.7	Processes	17
4.7.1	Handlers	17
4.7.2	Reapers	17
4.7.3	Initialising Processes	17
4.8	Scheduling	17
4.9	Interacting with the Memory Management Unit	17
4.10	Synchronisation	17
4.11	Inter-process Communication	17
4.12	User interaction	17
5	Testing and Issues	17
5.1	Loading onto real hardware	17
5.2	Switch to the Raspberry Pi 1	17
5.3	ACT debugging	17
5.4	Static libraries	17
6	Evaluation	17
6.1	Achievements	17
6.2	Limitations	17
6.3	Further work	17

A	Source code	18
B	Timetable	18

1 Introduction

Operating systems are some of the most pervasive pieces of software around, but due to their inherent complexity, their inner workings are often impenetrable to understand without specialist knowledge. While widespread access to a personal computer is nothing new, the introduction of the Raspberry Pi in recent years has rendered experimentation with computers much more affordable and hence readily available, inviting tinkering at all levels with less concern of economic loss. The Pi therefore provides an ideal platform for operating systems education – novice developers looking to get involved in writing such systems have access to a standardised set of hardware that is inexpensive both to maintain and replace, if and when things go wrong. Now seven years since its initial release, the Raspberry Pi has several official operating systems to offer, each addressing its own issue such as ease-of-installation, Internet of Things integration, or classroom management [1], with many more unofficial ones. However, there is less in the way of those written to teach concepts of the operating system itself – this project attempts to fill this gap by providing a configurable, educational operating system for the Raspberry Pi 1 Model B+. The focus is on presenting code which is simple to understand and providing clear interfaces that encourage ease of extensibility, and hence a practical, software-driven approach to learning about operating systems.

1.1 Motivation

An operating system draws together aspects from all over the field of computer science, whose development requires intimate knowledge of low-level concepts such as the computer’s organisation and architecture, up to an understanding for the more abstract in designing how processes communicate or implementing filesystems. The opportunity to write one as part of this project was therefore appealing as it served not only to provide the author with experience in an area of computer science of great interest, but also the chance to unite and put into practice many of the topics learned throughout the undergraduate computer science course. One of the key motivations of this project was therefore to gain experience in low-level systems programming and interacting with real-world hardware, all the while creating a useful and entirely self-contained piece of software.

The main goal of this project is to provide an operating system for the Raspberry Pi that is capable of booting on real hardware, for both educational and hobbyist use. An important aspect of this is gaining practical experience in this unique area of software development, and so in addition to being configurable at compile-time, offering multiple approaches to process scheduling and inter-process communication, it also aims to be easily-understandable and open to extension. In order to achieve this it must also implement a basic interface to manage this compile-time configuration. Also key to the project’s success is the ability to boot on real hardware - while it would largely be possible to implement in a solely emulated environment with the help of a virtual machine, this project attempts to provide the additional valuable experience of taking real-world design considerations and observing their effects on live hardware, the latter of which is easily ignored if working only through an emulator.

2 Background

2.1 Relevant Material

To this end, there are a handful of modern resources for getting involved with operating systems development – a particularly useful one at the time of first carrying out research for the project’s proposal was wiki.osdev.org, which contains information about the creation of operating systems and acts as a community for hobbyist operating system developers. However, much of the focus is on the x86 platform and past providing a brief overview of the idiosyncrasies of the Raspberry Pi as well as the code to get a barebones kernel to boot, there is little material on the specifics to get core systems working on the platform. Cambridge’s *Baking Pi* [2] provides more help in this regard, with Alex Chadwick’s comprehensive tutorials proving an invaluable resource for information such as accessing registers and peripherals specific to the Raspberry Pi. The project can, however, be much further extended to guide through the implementation of core operating system concepts such as memory management, the process model, inter-process communication, and filesystems. Another aspect in which this series of tutorials diverges with the goal of this project is the language in which it has been implemented – while assembly is an undeniably language in which to be competent, it is not the most easily-understandable, in stark contrast to what this project hopes to achieve. The resource which aligns most tightly with the aim of

this project is [3], whose tutorials have served as an outline to how many key features of the project have been implemented.

Finally, other notable resources which are in place to teach general operating systems development are Stanford's *Pintos* [4] and Tanenbaum's MINIX operating system [5]; the former was written to accompany the university's CS140 Operating Systems course, while the latter is an illustrative operating system written alongside the book *Operating Systems: Design and Implementation*, as a means of providing concrete examples of how operating system features are implemented in practice. Helin and Renberg's *The Little Book About OS Development* [6] also serves as a guide to writing one's own operating system. The only drawback to these three is their focus on the x86 architecture, and while they are useful resources it is in concept only, given the gap which was found to quickly form from focusing on a different processor.

2.2 Why is this project worthwhile?

The project is worthwhile firstly as it provides an accessible gateway into systems programming and operating systems development. Given the relative difficulty and additional effort required to get involved with this area of software development as opposed to others, for example by reading technical reference manuals and building an intimate knowledge of the hardware with which you are working before even starting, it therefore finds its use in easing this transition and making the learning curve associated with its involvement less intimidating and more approachable. In doing this, the project is also worthwhile in that it demystifies some of the key considerations that go into operating system implementations, not only in high-level concepts such as processes, but also the low-level with notions such as memory-mapped I/O and the processor's registers. In providing this opportunity to see theory in practice, it further opens up the opportunity for experimentation and invites practical self-learning, and hopefully clarifying why existing operating systems work the way they do.

While there are similarities to be drawn between the aims of this project and those of the current background material, both looking to create a more accessible way in to operating systems development, this project addresses the gap that they leave unfilled by tackling a different architecture, as well as in approaching feature implementation in a more modular manner. Thus the project forms one more part in the ecosystem of introductory and instructional operating systems.

2.3 Useful concepts

Operating System - a program that manages a computer's hardware that acts as an intermediary between the user and said hardware [7], providing an environment in which a user can execute programs conveniently and efficiently.

Kernel - the one program running at all times throughout an operating system's execution, the kernel is often tasked with managing the most vital/recurring tasks. Other key types of programs include system programs and application programs.

Freestanding environment - typical programs are written to run in a hosted environment, meaning they have access to a C standard library and other useful runtime features. Conversely, a freestanding environment is one which uses no such pre-supplied standard, meaning a bespoke one must be supplied. Any functions we wish to use as part of the operating system we must define ourselves.

Cross-compiler - a regular compiler will generate machine-code which is specific to that on which the code has been compiled. By contrast, a cross-compiler allows us to write code on any machine and compile it for our target architecture (the architecture on which we design our code to run).

Linker - responsible for linking all object (.o) files generated by a compiler/assembler into a single executable (or 'library file', see [?]). It also defines the entry point and the location of the various sections (detailed in Section 3.2.2) in the final ELF file.

Exception - an event triggered when something exceptional happens during normal execution, for example hardware providing the CPU with data, a privileged action, or bad instruction.

Process - a program that has been loaded into memory and is executing.

Context Switch - the act of saving the currently executing process into memory (**state save**) followed by loading the saved state of a different process (**state restore**). It allows the CPU to switch from executing one process to executing another.

Concurrency - the ability for multiple processes to make progress seemingly simultaneously, as a result of being rapidly brought into and out of memory with respect to a policy enforced by some scheduling algorithm.

Synchronisation - the prevention of **race conditions**: when the outcome of two concurrently executing processes depends on the order in which data access took place.

Inter-process Communication - the mechanism by which processes may exchange data and information.

Peripheral - a device with a specific memory address which it may write data to and read data from. All peripherals may be described by an offset from some base address, covered in Section 4.3.

Note: Throughout this report the units kiB and MiB are used (in the context of memory only) to denote $2^{10} = 1024$ bytes and $(2^{10})^2 = 1,048,576$ bytes, respectively. This is to ensure clarity and avoid confusion with their SI-prefixed counterparts, namely kB and MB, which instead correspond to 10^3 and 10^6 bytes.

3 Design

3.1 Hardware

The project has been developed for the Raspberry Pi 1 Model B+. Some of the relevant hardware onboard includes:

- System-on-Chip: Broadcom BCM2835
- CPU: 700MHz ARM1176JZF-S
- GPU: 250MHz Broadcom VideoCore IV
- SDRAM: 512MiB, shared with the GPU
- Video output: HDMI and DSI
- Storage: MicroSDHC slot
- 4x USB 2.0 ports
- 40 General Purpose Input/Output (GPIO) pins

Development was initially planned for the Raspberry Pi 2 Model B+ simply due to its availability, having been received as a gift some years prior. However, focus was switched to target the Raspberry Pi 1 Model B+ as a result of the difficulties encountered with interacting with the GPU via the mailbox interface, with the processing differing slightly and being more complex on the 2. The underlying architecture of the 1's BCM2835 chip is, however, identical to that of the BCM2836 and BCM2837 [8], used by the Raspberry Pi 2 and 3, respectively. They only differ in that the 1 uses the ARM1176JZF-S processor, as opposed to the quad-core Cortex-A7 and quad-core Cortex-A53 cluster used by these later boards, in addition to the 512MiB extra available to them. Therefore, the choice between specific models for which to develop would have made little difference to the outcome of the project, and indeed much of the code is transferable. As has been discussed, this standard set of hardware was desirable for the project's aims, rendering the operating system more widely accessible and for less effort.

3.1.1 The Raspberry Pi's boot process

The decision to work with the Raspberry Pi in particular, as opposed to a more open-ended PC setup, is not only due to the relative lack of material available for operating systems development on the ARM architecture, but also as a result of the much simpler boot process in contrast to other hardware, details of which were found from [9]. Booting is handled almost entirely by the Pi's system-on-chip, thus does not require the writing of a custom bootloader. Instead, it relies on closed-source proprietary firmware

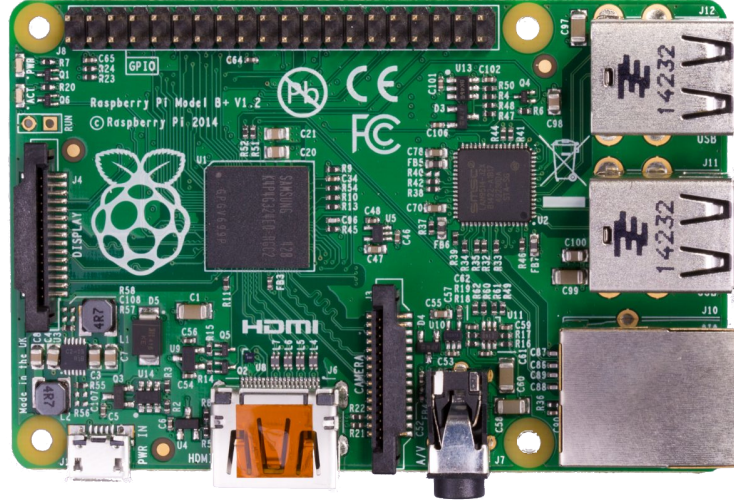


Figure 1: Raspberry Pi 1 Model B+

programmed into the SoC processor which may not be modified. The necessary files can be acquired by either downloading them from [10], or by downloading an existing operating system for the Pi and using the files that it provides (since it still requires the same firmware).

During system boot, the ARM CPU does not act as the main CPU, but rather as a coprocessor to the VideoCore GPU [11]. When the system is first powered on, the ARM CPU is halted and SDRAM is disabled. Control is passed to the GPU, whose responsibility it is to execute the bootloader. The bootloader itself is divided into three stages: the first stage, stored in ROM on the system-on-chip; the second stage, `bootcode.bin`, and the third stage, `start.elf`. The first mounts the FAT32 boot partition on the SD card to enable execution of the second-stage bootloader, and then loads this into the L2 cache to run it. Control is then passed to the second-stage bootloader, which enables SDRAM and loads `start.elf` for execution. This final stage allows the GPU to start up the CPU. An additional file, `fixup.dat`, is used to configure the shared SDRAM partition between the two processors. The GPU firmware reads the files `config.txt` and `cmdline.txt` to load the kernel image, then releases the CPU from reset and transfers control to it to begin executing the kernel.

After an operating system is loaded, the code on the GPU is not unloaded, but rather runs its own simple operating system, the VideoCore Operating System (VCOS) [12]. This can be used to communicate with the services provided by the GPU (for example, providing a framebuffer), using the mailbox peripheral and ARM CPU interrupts, which the GPU is capable of producing. The GPU is not only in charge of graphical functions as, for example, it also controls the system timer and audio – in this way it is therefore more akin to a regular PC's BIOS.

3.2 Resources & Environment

The project was developed on various x86 machines running the Linux kernel version 4.16 onwards. Since the target environment, the Raspberry Pi 1 Model B+, runs on an ARM CPU, the target architecture is therefore different to that of the development machines. Therefore, a cross-compiler is required in order to compile the code for the target machine. Available on the ARM developer website is the GNU Embedded Toolchain [13]. Conveniently, this suite of tools is available from Arch Linux's package manager, Pacman [14], and this is the version that has been used to develop the project.

3.2.1 Technical Documentation

With no prior experience working in assembly and, in particular, how to use ARM assembly in conjunction with the C programming language, it was necessary to become better acquainted with this

environment, including becoming familiar with the instructions available and the calling conventions that must be followed (for example, returning the result of an assembly procedure call in register 0). This information was gathered by reading through both official and unofficial documentation on the ARM environment in general and the specifics of working with the Pi. In particular, the following were the main resources of technical documentation used throughout the project:

- *ARM1176JZF-S Technical Reference Manual* [15]
- *ARM Developer Suite Assembler Guide* [16]
- *ARM Cortex-A Series: Programmer's Guide* [17]
- *Broadcom BCM2835 ARM Peripherals Manual* [18]

The technical reference manual provides important, in-depth architectural information upon which the processor operates, and, for example, details concepts such as the programmer's model or interaction with the system control coprocessor and the memory management unit. The online guide at [16] provides comprehensive coverage of the instructions to use when using assembly to program on the ARM CPU, and used in conjunction with the programmer's guide provides an understanding of when and how to use certain instructions. Although not written for the processor used by the Raspberry Pi 1, instead written for the CPU used by the 2 and 3, the programmer's guide provided further examples of various instructions' use-cases, and its utility continued even when focus shifted from the Raspberry Pi 2 to the Raspberry Pi 1.

On the Raspberry Pi are various peripherals, such as the Universal Asynchronous Receiver/Transmitter (UART), system and ARM timers, or the interrupt controller. Information regarding the layout of their registers and the memory addresses from and to which to read and write was obtained from the peripherals manual. Since the underlying architecture behind the BCM2835 and BCM2836/7 is largely the same, the manual was helpful both in development for the Raspberry 1 and 2, with the BCM2836 [19] providing some extra processor specifics in the case of the latter.

3.2.2 System V ABI

Of particular use within GNU's suite of tools is the cross-compiler for `arm-non-eabi`, which provides a toolchain to target the System V ABI (Application Binary Interface). This is a set of specifications that detail the calling conventions, object and executable file formats, dynamic linking semantics, and more, for systems complying with the System V Interface Definition, of which the Raspberry Pi is one. For example, it defines the Executable and Linkable Format, or ELF, which is a format for storing programs or fragments of programs on disk that are generated as a result of compiling and linking. Each ELF file is divided into sections, specifically:

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

These are covered in more depth in the discussion of the linking process in section 4.1.1. Additionally the SysV ABI defines the `.comment`, `.note`, `.stab`, and `.stabstr` sections for compiler and linker toolchain comments and debugging information.

3.2.3 ARM environment

The ARM1176JZF-S processor implements the ARM11 ARMv6 architecture [15], supporting both the ARM and Thumb instruction sets¹. The processor contains 33 general-purpose 32-bit registers and 7 dedicated 32-bit registers, 16 of which are accessible for general use at any one time in the ARM state. These are as follows:

- `r15` - Program Counter
- `r14` - Link Register

¹A subset of the most commonly-used 32-bit ARM instructions. Each Thumb instruction is 16 bits long, and has a corresponding 32-bit ARM instruction with an equivalent effect on the processor model. While not useful and out-of-scope for this project, the instruction sets can be easily switched between to enable the programmer to optimize for either performance or code density as they see fit.

- **r13** - Stack Pointer
- **r12** - Intra-Procedure-Call Scratch Register
- **r4-r11** - local variables to a function
- **r0-r3** - arguments passed to a function, and the returned result

The Current Program Status Register (CPSR) contains code flags, status bits, and current mode bits. Another register, the Saved Program Status Register (SPSR), is similar to the CPSR but is only available in privileged modes. This contains the condition code flags, status bits, and current mode bits saved as a result of the exception which prompted the processor to enter the current mode.

The architecture asserts a full descending stack: full, meaning the stack pointer points to the topmost entry in the stack²; and descending, meaning the stack grows downwards, starting from a high memory address and progressing to lower addresses as items are pushed.

As a final note on correctly interacting with the ARM environment, any procedure calls must preserve the contents of registers 4 to 11 and the stack pointer. Further, subroutines calling other subroutines must save the return address (found in the link register) to the stack before calling that subroutine.

3.2.4 Tools

The following is a summary of the tools used to develop the project:

Languages	C, ARM assembly
Cross-compiler toolchain	GNU Embedded Toolchain (arm-none-eabi-*)
Build automation	GNU Make
Version control	Git, hosted remotely on Github
Emulation	QEMU
Documentation	Doxygen

Table 1: Tools used by the project

The project uses the C language both due to its familiarity as well as its ease-of-setup on the embedded environment, however there are times in such an environment, and especially for operating systems development, that assembly is more suitable [20], and was thus opted for instead. As already discussed, the GNU Embedded Toolchain has been used to target the ARM architecture for which the project is built.

The GNU Embedded Toolchain consists of several utilities, all of which are used at various stages in the project’s development: **arm-none-eabi-gcc** is responsible for compiling the C and ARM assembly files into object files; **arm-none-eabi-ld** is an embedded platform-independent linker, used to link multiple object files into a single **.elf** file; **arm-none-eabi-objcopy** converts the resultant **.elf** file into a binary (system executable, or **.img**) file; **arm-none-eabi-objdump** provides helpful debugging information about the final executable file (for example, used as a disassembler it presents the kernel image in assembly form); and **arm-none-eabi-gdb** provides the familiar interface of the GNU debugger.

As the project increased in complexity, especially with the addition of multiple source files, GNU’s Make utility was used to automate the build process. This involved becoming familiar with concepts regarding rules, with GNU’s online manual page [21] and [22] being of particular help.

Throughout its early stages, the project operated solely in an emulated environment, simply as it provided quicker feedback with respect to testing, which is already limited and slow in such an environment. For this purpose QEMU was consequently used. However, due to its lack of simulation of a system timer, an important aspect when programming interrupts and process scheduling, focus had to eventually be shifted towards operating on real hardware, limiting QEMU’s influence on the project as a whole.

3.3 System Overview

The project presents an operating which is capable of booting on a Raspberry Pi 1 Model B+ and running (with reduced functionality) on an environment emulating the Raspberry Pi 2 Model B, and initialises a C runtime environment. It provides a mechanism for memory management, first in the form

²Contrasted to empty, which points to the next free location, i.e. the address at which the next item will be stored.

of determining the total memory available to the system, and then in dividing this up into 4kiB pages. A dynamic memory allocator is set up in the form of a 1MiB portion of “heap” memory, and provides an interface similar to the C standard library’s `malloc()` and `free()` in order to manage this.

Visual feedback is provided in the form of output to HDMI and, in the virtual environment, output to serial via the UART peripheral. The mechanism for communication with the GPU is set up and managed using the mailbox peripheral, and output via HDMI is achieved by requesting and providing relevant data (dimensions, bits per pixel, etc.) to the framebuffer. User interaction is so far only achieved in the virtual environment, again doing so via UART, with real-world interaction via USB proving to be a significant source of challenge throughout development.

The project sets up interrupts and exceptions by providing various exception handlers. This is with the main aim of interacting with the system timer, a peripheral capable of generating interrupts every set amount of time, for use in process scheduling. It also provides a somewhat intuitive interface for registering new “types” of interrupts (further covered in Section ??). Processes are implemented and stored using the standard job- and ready-queues, and a simple interface for creating new threads of execution is provided.

With process scheduling being one of the main systems designed for ease of understanding and extensibility, the project provides a simple interface for using and programming custom process schedulers, with all code required contained entirely in one file dedicated to such a task. Taking into account the added complexity brought on by concurrency, synchronisation is achieved by implementing both spinlocks and mutual exclusion locks.

A simple and somewhat rudimentary form of inter-process communication is provided, making use of the “shared memory” model. User interaction via USB keyboard is attempted and provided as a compile-time option, and makes use of a statically-linked library provided by Rene Stange’s project *USPi* [23]. While not an initial requirement, the system provides the ability to enable the Memory Management Unit, allowing the programmer to define their own mapping from virtual to physical memory and control which memory is bufferable and cacheable.

The project also provides a means of configuring various options at compile-time through sending directives via the Makefile, resulting in a clean and simple interface to change between various systems. At present, the system may be configured for running on the Raspberry Pi 1 or 2, using Round-Robin or First Come First Served Scheduling, and which form of inter-process communication to use (for which only shared memory is currently provided). While there is not much choice currently, this system is relatively simple and easy to extend as more features are developed.

Finally, since the ability to demystify the codebase of an operating system is an important aspect of the project, it has been designed to be compatible with Doxygen [24] in order to generate documentation for the project in a simple collection of HTML files. It provides similar functionality to Oracle’s Javadoc [25].

Future work will be concentrated on achieving user-interaction via USB keyboard and using this to implement a shell/command interpreter. A final important system to implement is filesystems and permanent storage, for which an interface to the External Mass Media Controller (EMMC) must be written, to enable interaction with the SD card.

3.4 Project management

3.4.1 Development methodology

The project has been developed in a waterfall-style approach – this has been well-suited to the rigidity of system dependencies, especially in early versions of the operating system. For example, it was necessary to write the boot code and set up the C environment before writing a interrupt handlers before implementing process scheduling, and any deviation from this order would have made little sense. This inflexibility allowed for such core features to be planned and designed in advance, and meant that focus only progressed to another feature once they had been implemented successfully. Once these lower-level systems were in place, the project opened up more, particularly after having completed development of processes; from this point focus at any point could be concentrated on features such as process scheduling, inter-process communication, or user interaction. As such, development became more agile, and the decision to pursue development on certain features was taken by assessing their importance to the project’s success in relation to the other possible features. Throughout the course of the project, goals have been set and achieved in an incremental manner; this involved, once deciding upon the system to implement, breaking it down into appropriate subtasks and completing each in turn. An example of this has been the development of memory management, first requiring the parsing of atags, followed by

creating and maintaining a list of page metadata, and finally using these to write the dynamic memory allocator.

3.4.2 Organisation

From the very start of the project, the effective use of version control software has been vital to its success. Integrating with Git and hosting it remotely on Github occurred early on and has been a constant consideration throughout. Since testing on real hardware is so important to determining whether or not an embedded system truly works, the ability to keep track of all previous builds has been useful. In particular, when features do not work, and especially if their failure interferes with other systems, being able to revert to the previous working version has helped not only in getting the project back to a functional state, but also in investigating and understanding precisely why certain attempts broke the build. Github also provides several useful features to track the evolution of the codebase, from detailing code frequency to monitoring branches. Figure 2 highlights Github’s consistent aid throughout the project, in the form of weekly commits over the span of its development.

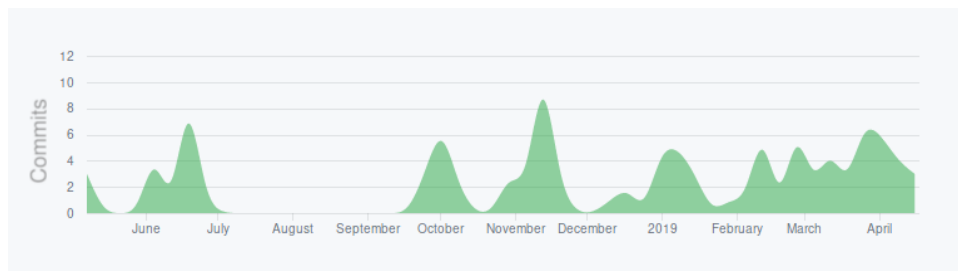


Figure 2: Weekly commits to the Github repository (source: repository’s “Insights” page)

Progress was initially considered in the project specification, for which a preliminary schedule was devised. After two months of development, this was then revised in the progress report. A copy of this version is provided in Appendix B, and an evaluation of how effective and accurate these timetabling estimates were is discussed in Section 6. In order to monitor progress, regular meetings were held with the project supervisor, Adam Chester, during which recent achievements and current issues were discussed, with guidance given in the case of the latter. These further aided in consolidating lessons learned and milestones passed, when otherwise it might be easy to lose track – not vital, but helpful in tracking the project’s progression. During Term 1, these meetings were held fortnightly as work on the project was less of a focus in the light of the higher course load regarding other modules. The frequency of meetings increased to weekly the following term as more time became available and more challenging and abstract aspects of the project presented themselves.

3.4.3 Testing & Debugging

The project was tested throughout development, and with particular frequency once able to boot on real hardware. This means that, once a subtask had been completed, the entire system was compiled and loaded onto hardware, with specific unit tests written each time to ensure the feature’s correct operation. Also important was integration testing, so as to ensure a working implementation of one feature did not compromise the correct functioning of another. This was the case when enabling the MMU, having worked in isolation from other components, but interfering with the return addresses of other important functions when used in the wrong place or memory addresses in-use and to which it should not have had access – this only became clear as a result of integration testing.

There is little infrastructure to debug low-level software of this type, so much had to be done manually. Specifically, this involved inspecting the `.map` and `.list` files and using the Linux command-line utility `nm` [26] to display information about the generated executable files. Furthermore, prior to successfully sending output to a real screen via HDMI, there was no way to gather textual information about the system during execution. The only other useful way to provide debugging information during execution was interacting with the two programmable LEDs on the Pi, namely the ACT and PWR LEDs, for example causing them to blink at different points and in different patterns. Soon after successfully displaying to HDMI, it was important to develop a working version of `printf` to display debugging information in a flexible manner.

3.4.4 Legal, Ethical, and Social Issues

All software being used to develop is available under the GNU General Public License, which grants permission for commercial and private use, as well as modification and distribution. Most software used as a study resource is similarly available under the GNU General Public License, while several others are available under the MIT license, the main difference being software distributed under the former must be made open-source, while this is not true of the latter. Throughout the project's development, informal feedback has been sought from friends and colleagues, particularly for non-functional aspects of the software, such as usability and overall polish. Due to the informal nature of the feedback, however, ethical, social, and legal considerations need not apply.

4 Implementation

The following section largely documents the system as it was developed chronologically, and is not representative of the order in which systems are initialised in the final piece of software. This better highlights the dependencies between each discrete stage in development and reinforces the plan-driven approach to the project as a whole.

4.1 System initialisation

There are two main files which handle the entirety of system startup, namely `linker.ld` and `boot.S`, the former responsible for defining the layout of the final executable we will be producing, and the latter handling the setup of the C runtime environment and passing control to the main kernel function.

4.1.1 Linker Script

In order to create a kernel, or any program for that matter, we must link all compiled object files into a single executable. For user-space programs, that is, programs written to run in an established operating system, there are default scripts provided by the compiler which do this. Since the operating system on which we will run the kernel does not yet exist, we must create a linker script ourselves.

We begin by declaring that the symbol `_start` is the entry point for our entire program (this is usually `main` for user-space programs), meaning execution will jump to this address when the system starts. We first declare the symbols `__start` and `__text.start` at address `0x8000`, which informs the bootloader where to place our kernel image. Addresses up to `0x8000` will be reserved for the stack. After this we must declare one by one each of the major sections which will be used in the ELF file, the first of which is the `.text` section, containing executable code. The code from `boot.S` is placed in the `.text.boot` section, and we use the `KEEP` directive to inform the linker not to optimise the code in that section – we want it running exactly what is written there. We declare the sections `.rodata`, `.data`, and `.bss` similarly, each time declaring global symbols for them should the need to use them as variables arise (as is the case, for example, when placing the metadata for pages directly after the kernel image, see Section 4.2.2). We also use the directive `ALIGN` to define a page size for the kernel image – specifically, this sets the current address to the next available address that is divisible by the page size we set, 4096 bytes.

4.1.2 Initialising the C runtime

With our kernel image sections defined, we can actually write the code that will boot into our kernel, contained in the file `boot.S`. We begin by declaring that this is to be placed in the `.text.boot` section (as used by the bootloader) and we make our start symbol visible from any file using the `.global` directive – this is required for the linker to see which location to jump to at system startup.

The project was initially written for the Raspberry Pi 2, containing the quad-core Cortex-A7 CPU. However, since multicore programming entails extra difficulty even in a familiar user-space environment, it was always designed to only use one of these cores. The following code remains included in the project, wrapped in a test to determine which processor we have compiled for, and sends three of the four cores to hang. We will discuss it now to introduce the idea of the ARM coprocessor:

```
_start:
    mrc p15, #0, r1, c0, c0, #5
    and r1, r1, #3
```

```

cmp r1, #0
bne halt

```

Listing 1: Code to halt three of the four cores

The `mrc` and `mcr` instructions allow for interaction with the system control processor, and denote “move data from coprocessor to ARM register” and “move data from ARM register to coprocessor” respectively. The purpose of this is to control and provide status information for the functions implemented on the ARM1176JZF-S processor [15, pg. 3-2], and exposes optional³ additional functionality that is not provided by the core instruction set. Among others, these functions include:

- System control and configuration
- MMU control and configuration
- Cache control and configuration
- Direct Memory Access (DMA) control
- System performance monitoring

As it stands, the project only makes use of the first two. The code block itself reads the Multiprocessor Affinity Register [27] for the identifier of the current CPU it is testing, and if neither of the two lower bits returned are 1, then the core is sent to halt indefinitely in a low-power state.

With just a single core to develop for, the next step is to set up the stack pointer at address `0x8000`. The program counter for the kernel starts at address `0x8000` and grows upwards, and, since the stack grows downwards, it is safe to place it at this address without it interfering with the kernel. Next the start and end of the Basic Service Set (BSS) is loaded into registers – recall that this is where statically allocated global variables are stored. Since the C standard requires uninitialised global variables to be zeroed, we must zero out (that is, store the value 0) in the registers between the addresses of `__bss_start` and `__bss_end`.

The final responsibility of `boot.S` is to transfer control to our kernel entry point, `kernel_main`, by loading the address of this symbol into the program counter. Overall, this small portion of the codebase initialises a minimum C environment, meaning the stack is initialised and the BSS segment zeroed before we pass control to our custom kernel. Note that the use of registers 0, 1, and 2 is avoided as the bootloader uses these to pass system parameters and hardware information to the main function⁴ at runtime.

4.2 Organising memory

4.2.1 Atags

One important piece of data passed by the bootloader to the main kernel function is the address of the “atags” – this is a list of tags which describe various system parameters such as total system memory, the initial ramdisk, and command-line parameters to pass, and is passed through address `0x100` on the Raspberry Pi 1. Throughout the execution of the operating system, computations will require memory and since the project functions entirely in kernel-space, any and all memory is available to use. To impose some structure on the memory and manage it effectively, however, we organise it into 4kiB pages, allowing for equal-sized blocks of memory, which are neither insignificantly nor prohibitively large, to be allocated and tracked. Therefore, the most important piece of information among the atags is the total system memory, which the project uses in its implementation of a dynamic memory allocator.

In order to parse this list of tags, we need the kernel to understand the layout of the data it is being passed. In particular, each tag consists of two values: an unsigned 32-bit integer denoting the length of the tag (in 32-bit words), and the tag itself, details of which are found at [28]. This provides information concerning the memory tag itself, namely that it contains two unsigned 32-bit integers describing memory size and start address, in this order. Therefore, we define the C structure `atag_mem` to easily access these two fields once we come across the memory tag. Thus, to determine the total memory available we iterate over the atags list until we find the memory tag, cast this to our `atag_mem` structure, and consequently return the field denoting the size of system memory.

³ARM only provides specifications for processors; it is down to hardware manufacturers to implement their designs.

⁴Namely, the device from which the system was booted, the ARM Linux Machine Type (`0xc42` for the Pi), and the address of the atags, which contain more important information such as available memory.

4.2.2 Pages

The total number of pages is the total size of memory divided by the page size, which we have set to 4kiB. In order to effectively manage each page, we create a list containing metadata for each page: this metadata includes, for example, the virtual address to which it maps, along with bit-field flags detailing whether the page is allocated, if it is available for sharing, if it is a kernel page (for when user-space is implemented), and whether it is a regular page or a page on the heap. We use the global symbol `__end`, declared in the linker script, to place the page metadata list directly after the end of the kernel image.

To allocate a page, we need knowledge of which ones are free to do so, for which we create a linked list of all such pages. Then all that is needed is to return a pointer to its location in memory and zero it (for security), modify the appropriate flags in the page array, and remove it from our free pages list. When a page's use is no longer required, it makes sense to make it available for future allocations, for example when a process has finished executing (see Section 4.7.2). Thus, to free a page we use its physical address to index into the array containing all pages, set the `allocated` to 0, and add it back to the list of free pages.

4.2.3 Dynamic memory allocator

To allocate memory dynamically, and less restrictively than entire pages at a time, we reserve a 1MiB portion of memory directly after the page metadata for heap memory, which is used to allocate segments of memory instead, with a similar interface to the C standard library's `malloc` and `free`. This choice of 1MiB is fairly arbitrary, but is not too restrictive for most uses of dynamic memory and does not use too large a portion of memory that will eventually be used by user-space.

Memory allocation is implemented by first defining for each allocation a header containing the allocation size, in bytes, and whether it has been allocated, and then using these headers to form a linked list. Thus, in order to grant an allocation we must find a header that is at least the requested number of bytes in size, and not currently in use. If the best-fitting segment, that is, the segment with the smallest size that satisfies the request, is relatively large compared to what was requested, we split it so that we may use what is effectively “left over” to satisfy some future request.

When freeing memory allocations, we mark the segment as unallocated by modifying the bit in the allocation header and merge adjacent free segments in the linked list into one, thus decreasing the likelihood of internal fragmentation – unused or unusable memory within the segment [7, pg. 363]. Coalescing adjacent free segments to the left is shown below; the process is identical for coalescing to the right, except that it modifies the pointer to the `next` segment header instead.

```
/* merge segments to the left */
while (seg->prev != NULL && !seg->prev->allocated) {
    seg->prev->next = seg->next;
    seg->prev->segment_size += seg->segment_size;
    seg = seg->prev;
}
```

Listing 2: Coalescing heap segments to the left

When initialising the heap at system startup, we must first allocate the appropriate number of pages and mark them as kernel heap pages, then initialise the linked list of segments by declaring a single segment 1MiB in size. Although the total heap size remains the same, it will become split into increasingly many segments, all forming a linked list, as allocation requests are satisfied.

4.3 Serial Output

An operating system would be useless without an efficient means of interacting with it during execution, so the next task to present itself was that of interfacing with the Universal Asynchronous Receiver/-Transmitter, a peripheral on the Pi, which is capable of serial⁵ communication. While output would evolve to be via HDMI, it was simpler to initialise this when working within QEMU, and also prompted for an intuitive interface for GPIO to be written, which would again become useful when debugging with HDMI output.

⁵One bit at a time

Offset	Register	Size
0x00	Data Register	32
0x04	RSRECR	32
0x18	Flag Register	32
0x20	Unused	32
0x24	Integer Baud Rate Divisor	32
0x28	Fractional Baud Rate Divisor	32
0x2c	Line Control Register	32
0x30	Control Register	32
0x34	Interrupt FIFO Level Select Register	32
0x38	Interrupt Mask Set Clear Register	32
0x3c	Raw Interrupt Status Register	32
0x40	Masked Interrupt Status Register	32
0x44	Interrupt Clear Register	32
0x48	DMA Control Register	32
0x80	Test Control Register	32
0x84	Integration Test Input Register	32
0x88	Integration Test Output Register	32
0x8c	Test Data Register	32

Table 2: Each register is described by an offset from the UART base address, 0x20201000.

4.3.1 Peripherals

As mentioned, a peripheral is a device with a specific address from and to which it may read and write data, and all interactions with peripherals make use of Memory Mapped I/O (MMIO) to do so on the Raspberry Pi – the process of performing I/O by reading from and writing to predefined memory address. Moreover, each peripheral may be described by an offset from the Raspberry Pi’s Peripheral Base Address. This varies for different models of the Pi, but on the Raspberry Pi 1 Model B+, this is located at address 0x20000000, and the physical address for peripherals span from this base address up to 0x20ffffff. The peripheral itself contains a collection of registers which may be read and written, and these are defined at offsets from that specific peripheral’s base address (see Table 2 for clarification).

There are several steps to initialise the UART for use on the Pi, and involves setting various configuration flags. The following steps are performed in order in `uart_init()` in `gpio.c`:

1. disable all aspects of the UART
2. disable all GPIO pins
3. disable pins 14 and 15
4. clear all pending interrupts
5. set up the baud rate of the connection
6. enable FIFOs and set word length to 8 bits
7. disable all interrupts
8. enable the UART hardware, and the ability to transmit and receive data

To send a byte to the data register, we wait until the FIFO (the data structure by which the UART sends and stores information) is not full and write the byte to the data register. Receiving a byte is done by waiting until the FIFO is non-empty and returns whatever is in the data register. Thus we have serial input and output, and may use this to implement `putc` and `getc` respectively. When running this build on QEMU, passing the `-serial` flag enables the sending and receiving of serial output via the host computer, thus we can specify `stdin` to be able to send character input via the host system’s keyboard. This enables us complete interactivity with the kernel in the emulated environment.

4.4 Interacting with the GPU

While QEMU was a powerful tool to allow for quick development and testing of core low-level features, as discussed its lack of simulation of a system timer required that the project eventually be moved to real

hardware to tackle core features such as processes and interrupts. Furthermore, the Pi comes without a dedicated serial port, therefore without a TTL-to-USB adapter we are unable to use the UART to output to a real screen. Instead, it is much more natural to use the onboard HDMI port to output to a display, and for this we require communicating with the GPU.

Displaying anything to the screen via the HDMI requires the use of a framebuffer; this is a piece of memory which is shared between the CPU and GPU. The CPU writes RGB pixels to the buffer and the GPU reads from it to render the pixels to whichever output device is connected. We must first request a framebuffer from the GPU, and this interaction takes place over the mailbox peripheral.

4.4.1 Mailbox Peripheral

The mailbox peripheral is a peripheral, just like the UART, which facilitates communication between the CPU and GPU, more specifically the VideoCore Operating System, and starts at offset 0xb880 [29]. It contains a read register, located at offset 0x0 from the base, which holds messages sent by the GPU; a status register, located at offset 0x18, which is used to signify whether the read register is empty or full; and a write register, located at offset 0x20, which the CPU can use to send data to the GPU. Since the mailbox peripheral can also be used to send and receive information about multiple systems, such as power management, audio, and the on-board camera, the peripheral requires communication through certain channels. This is simply a number that specifies the meaning behind the data being sent; for the framebuffer, we communicate via channel 1.

Crucially, the process of communicating with the GPU via the mailbox peripheral differs depending on the model of Raspberry Pi being used; since the project was initially intended to run on the Pi 2 Model B, an interface was written for this version, however since it is a much simpler process on the Pi 1 Model B+, after some weeks debugging and making little progress, the decision to switch the target platform was switched at this point. Within the source code are both methods, and both are guarded by macros (see Section ?? on Makefile directives) which will prevent compilation of the wrong one.

- 4.5 Interrupts and Exceptions
- 4.6 System Timer
- 4.7 Processes
 - 4.7.1 Handlers
 - 4.7.2 Reapers
 - 4.7.3 Initialising Processes
- 4.8 Scheduling
- 4.9 Interacting with the Memory Management Unit
- 4.10 Synchronisation
- 4.11 Inter-process Communication
- 4.12 User interaction

5 Testing and Issues

- 5.1 Loading onto real hardware
- 5.2 Switch to the Raspberry Pi 1
- 5.3 ACT debugging
- 5.4 Static libraries

6 Evaluation

- 6.1 Achievements
- 6.2 Limitations
- 6.3 Further work

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] “Building an operating system for the raspberry pi.” <https://jsandler18.github.io/>. Page accessed: 2018-05-17.
- [4] B. Pfaff, “Pintos.” https://web.stanford.edu/class/cs140/projects/pintos/pintos.html#SEC_Top, December 2009. Document accessed: 2018-08-24.
- [5] A. S. Tanenbaum and A. S. Woodhull, *Operating Systems: Design and Implementation*. Pearson, 3 ed.
- [6] E. Helin and A. Renberg, “The little book about os development.” <https://littleosbook.github.io/>, January 2015. Document accessed: 2018-11-22.
- [7] A. Silberschatz, P. B. Galvin, and G. Gagne, *Operating System Concepts*. Wiley, 9 ed., 2014.
- [8] “Raspberry pi documentation: Bcm2836.” <https://www.raspberrypi.org/documentation/hardware/raspberrypi/bcm2836/README.md>. Page accessed: 2018-10-08.
- [9] “Understanding the raspberry pi boot process.” https://wiki.beyondlogic.org/index.php?title=Understanding_RaspberryPi_Boot_Process. Page accessed: 2019-04-13.

- [10] “Github: raspbrerypi/firmware.” <https://github.com/raspberrypi/firmware>. Page accessed: 2019-04-13.
- [11] “Raspberry pi: Level of hackability of raspberry pi.” <https://raspberrypi.stackexchange.com/questions/7122/level-of-hackability-of-raspberry-pi/7126#7126>. Page accessed: 2018-11-15.
- [12] “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.
- [13] “Gnu-rm downloads.” <https://developer.arm.com/open-source/gnu-toolchain/gnu-rm/downloads>. Page accessed: 2018-11-18.
- [14] “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.
- [15] ARM, *ARM1176JZF-S Technical Reference Manual*, 2009. Revision r0p7.
- [16] “Arm developer suite.” http://infocenter.arm.com/help/index.jsp?topic=/com.arm.doc_subset.swdev.ads/index.html. Page accessed: 2019-04-20.
- [17] ARM, *ARM Cortex-A Series: Programmer’s Guide*, 2013.
- [18] Broadcom Corporation, Broadcom Europe Ltd., 406 Science Park, Milton Road, Cambridge, *Broadcom BCM 2835 ARM Peripherals*, 2012.
- [19] G. van Loo, *Broadcom BCM 2836 ARM Peripherals*. Broadcom Corporation, Broadcom Europe Ltd., 406 Science Park, Milton Road, Cambridge, 2014.
- [20] “C: Things c can’t do.” https://wiki.osdev.org/C#Things_C_can.27t_do. Page accessed: 2018-08-30.
- [21] “Gnu make: Writing recipes in rules.” <https://www.gnu.org/software/make/manual/make.html#toc-Writing-Recipes-in-Rules>. Page accessed: 2019-04-20.
- [22] “Make: How to use variables.” https://ftp.gnu.org/old-gnu/Manuals/make-3.79.1/html_chapter/make_6.html. Page accessed: 2019-04-20.
- [23] “rsta2: A bare metal usb driver for raspberry pi written in c.” <https://github.com/rsta2/uspi>. Page accessed: 2019-04-22.
- [24] “Doxygen.” <http://doxygen.nl/>. Page accessed: 2019-04-24.
- [25] “Javadoc tool.” <https://www.oracle.com/technetwork/java/javase/documentation/index-jsp-135444.html>. Page accessed: 2019-04-24.
- [26] “nm(1) - linux man page.” <https://linux.die.net/man/1/nm>. Page accessed: 2019-04-24.
- [27] “Trying bare metal on raspberry pi 2.” <https://www.raspberrypi.org/forums/viewtopic.php?p=693724>. Page accessed: 2019-04-24.
- [28] “Bootling arm linux: Tag reference.” http://www.simtec.co.uk/products/SWLINUX/files/bootling_article.html#appendix_tag_reference. Page accessed: 2018-11-20.
- [29] “Raspberry pi firmware: Mailboxes.” <https://github.com/raspberrypi/firmware/wiki/Mailboxes>. Page accessed: 2018-11-21.

Appendix A Source code

Appendix B Timetable