

A Punnet of Berries - Documentation

TeamPi: Adrian Zielonka, Alyssa Biasi, Zach Ryan

CONTENTS

1	Self Assessment	1
1.1	Collaboration Tools	1
1.2	Milestone 01	1
2	Introduction	3
2.1	The Punnet of Berries	3
2.2	The Berry Batch	4
3	Goals and Objectives	5
4	Constraints	6
4.1	Hardware: Raspberry Pi	6
4.2	Software: Architectural Considerations and Code Compilation	6
5	Architecture	8
5.1	Hardware	8
5.2	System Design	8
5.3	The Operating System	9
6	Program Design	11
6.1	Scheduling Algorithms	11
7	Testing	13
7.1	Tests Done	13
8	Roles and Responsibilities	15
9	Work Breakdown by Team Member	16
9.1	Alyssa Biasi's Log	16
9.2	Adrian Zielonka's Log	17
9.3	Zach Ryan's Log	18
10	Summary	19
11	References	20
11.1	Raspberry Pi	20
11.2	Project Research	20
11.3	Arch Linux ARM	20

SELF ASSESSMENT

1.1 Collaboration Tools

The collaboration tools utilised for the development of the Punnet of Berries project reflect the Open Source nature of the project.

Version Control: All code and documentation relating to the Punnet of Berries project is hosted in repositories on [GitHub](#) under the *rmit-teamPi* organisation. GitHub was chosen as it allows individual team members to use either Git or SVN depending on their own preferences.

Project Management: The project and bug-tracking was managed using [Redmine](#), an open source web-based project management tool. The team's Redmine account was hosted on [HostedRedmine](#), a service that hosts standard Redmine projects for free.

1.2 Milestone 01

1.2.1 Reflection

Milestone 1 involved following a cross-compiler recipe and becoming familiar working in a command line Unix environment.

The image was build by one TeamPi member, Alyssa Biasi. Alyssa already has considerable experience working on the command line. After spending some time getting to know the recommended tool, `tmux`, she followed the instructions given in the recipe and performed the sanity check. No problems were encountered during the build.

uClibc vs. glibc: Clibc was developed without consideration for other architectures and thus is a C library targeted directly for embedded Linux. glibc was intended for higher output performance and has additional features not required on a Raspberry Pi. The omitted features allow uClibc to function with a smaller amount of memory.

Cross-Compilers: A cross-compiler is critical in the creation of executables that are able to function on architectures that differ from the platform on which the code was developed.

1.2.2 Result

OSP Check sheet 1 – Cross Compiler

Input	Expected result / output	Verified
cd \$HOME/cross/cross-tools/bin ldd arm-unknown-linux-uclicgnueabi-ar grep 'whoami'	If this returns nothing then the student did not build the compiler as this user.	✓
cd \$HOME/cross/rootfs	Should return 245	✓
find . -type d wc -l	Should return 1815 245	✓
cd \$HOME/cross/images	boot the image the student created.	✓
/share/tools/bin/qemu-system-arm -M versatilepb -cpu arm1176 -m 256 -kernel kernel-qemu -initrd fs.img -append "root=/dev/ram rdinit=/bin/sh console=ttyAMA0" -nographic	Should see a "#" prompt	✓
which find	Returns /usr/bin/find	✓
find . wc -l	Returns 2101 2100	✓
vi -v	Should fail. Returns /bin/vi: invalid option -- v BusyBox v1.21.0 (2013-07-27 16:16:25 EST) multi-call binary. Usage: vi [OPTIONS] [FILE]... Edit FILE -c CMD Initial command to run (\$EXINIT also available) -R Read-only -H List available features	✓
Ctrl-A x	Exit.	✓

Team name: Team P1 Lab time: Friday 14:30-15:30 Date of demonstration: 16/08/13

Result of demonstration: Requirements fulfilled ☒ Resubmit ☐ Demonstrator name: Peter George Signed: [Signature]

INTRODUCTION

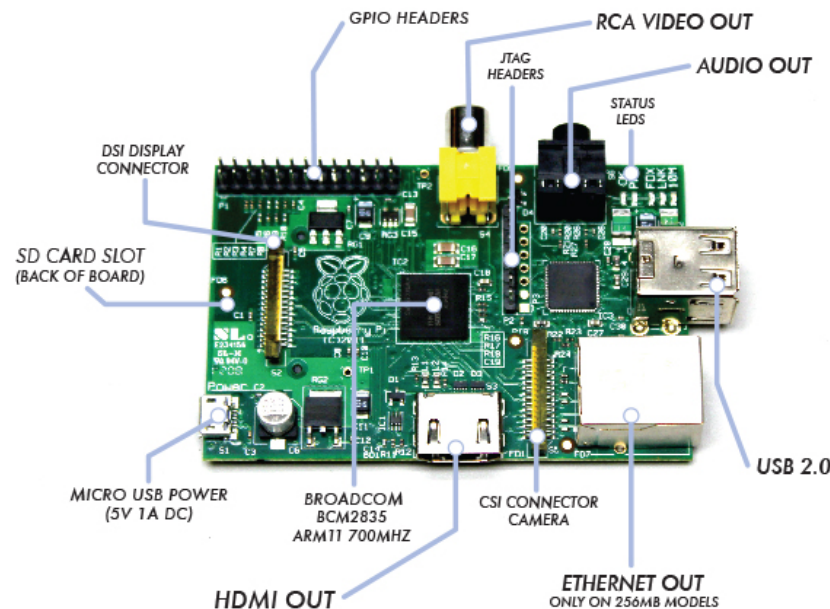
The aim of the **Punnet of Berries** project was to create a financially viable Beowulf cluster, evaluate the power of the Raspberry Pi and demonstrate scheduling concepts.

This document outlines the framework on which the **Punnet of Berries** and its accompanying software were developed. System design, architecture and implementation are all covered in the subsequent sections.

2.1 The Punnet of Berries

A **Beowulf** cluster is a supercomputer built out of a collection of (typically) inexpensive computers. The computers are networked together by a local area network, usually Ethernet, and run a parallel processing software. The individual computers are combined together to form a single system, becoming the nodes of the supercomputer. This concept puts the power of supercomputing into the hands of small research groups and schools. It gives them the ability to access the computational power that normally only large corporations can afford.

The Raspberry Pi is a single-board, credit card sized device. It is capable of running Linux, and other light-weight operating systems, with its ARM processors.



The Raspberry Pi is a powerful, yet, low-cost “mini-computer”. It was developed to educate and inspire the next generation of programmers. **TeamPi** believes that this makes the Raspberry Pi an ideal candidate to create a Beowulf cluster.

The **Punnet of Berries** is a *Raspberry Pi Beowulf cluster*.

2.2 The Berry Batch

In order for a compute cluster to function properly and fairly, its resources must be managed and monitored by a batch system scheduler. Batch systems enforce limits on the number of jobs running at one time and the resources available to them. Compute cluster users request resources by submitting jobs to the batch system, which then determines the scheduling that best utilises the resources available.

The **Berry Batch** is a custom batch system, which manages the Punnet of Berries compute cluster.

As one of the aims of the Punnet of Berries is to demonstrate scheduling concepts, the Berry Batch implements several scheduling algorithms. The user is able to select which algorithm to use when the cluster is initialised.

GOALS AND OBJECTIVES

The key goals of the Punnet of Berries project were to:

- Create a Beowulf compute cluster
- Develop a basic batch system scheduler
- Demonstrate scheduling concepts
- Develop descriptive documentation so that others may recreate the Punnet of Berries project

A financially viable Beowulf cluster was created as a part of the Punnet of Berries project. The cluster is energy efficient and demonstrates the advantages of parallel computing.

During the development of the Punnet of Berries, three key design principles were kept in mind.

- **Scalability** A computer cluster is, by nature, designed to be scalable. It was important that this inherent design characteristic be adhered to in order to avoid any sort of strict node limitations.
- **Usability** The Punnet of Berries project is largely for educational purposes. As such, it was important that the cluster be easily constructed with high cohesion. TeamPi aimed to create an application interface that is readily accessible.
- **Flexibility** In order to best demonstrate scheduling concepts, the Berry Batch was created to allow customisation. This was achieved by offering the user different options for the main features of the Berry Batch.
 - > Scheduling algorithm: First-Come-First-Served (FCFS), Round Robin (RR) or Priority scheduling.
 - > I/O: Blocking or Non-Blocking.

CONSTRAINTS

4.1 Hardware: Raspberry Pi

The Raspberry Pi is a comparatively low end single-board computer. As such, the design process must accommodate for these hardware specifications:

- SoC (System on a Chip) Broadcom BCM2835
- 700 MHz ARM core
- VideoCore IV GPU
- 512 MB of SDRAM.

Aside from an SD card, the Raspberry Pi does not feature any sort of non-volatile storage. Efficiency is a concern when constrained to a low capacity memory card. There is also an interface bottleneck, as both the USB and Ethernet share the same bandwidth.

An additional concern is of power distribution in regards to scalability; this would likely be resolved with the utilization of custom power supplies, but nonetheless it is something to consider.

4.2 Software: Architectural Considerations and Code Compilation

The Raspberry Pi hosts a single-core ARM processor, operating at 700MHz (with overclocking available). Due to this, any software developed for a Raspberry Pi should be in languages that are high-level, architecturally dependent and need to be recompiled on the Pi itself.

As there are several pre-configured Linux operating system images for the Raspberry Pi, it was decided that it was not necessary to create a custom Linux-From-Scratch image. In order to choose an appropriate operating system, it was necessary to evaluate the compute cluster's basic needs. The key factors in the decision making process were:

- Performance
- Size
- Compatibility

With this in mind, each node in the Punnet of Berries runs the **Arch Linux Arm** operating system. Arch Linux is:

- ARM compatible.
- Light-weight
 - The entire image is ~150 MB.
 - The default installation is bare bones, with nothing extra included.

- Boots in around 10 seconds.
 - Allowing the entire cluster to *boot in under 20 seconds*.

ARCHITECTURE

5.1 Hardware

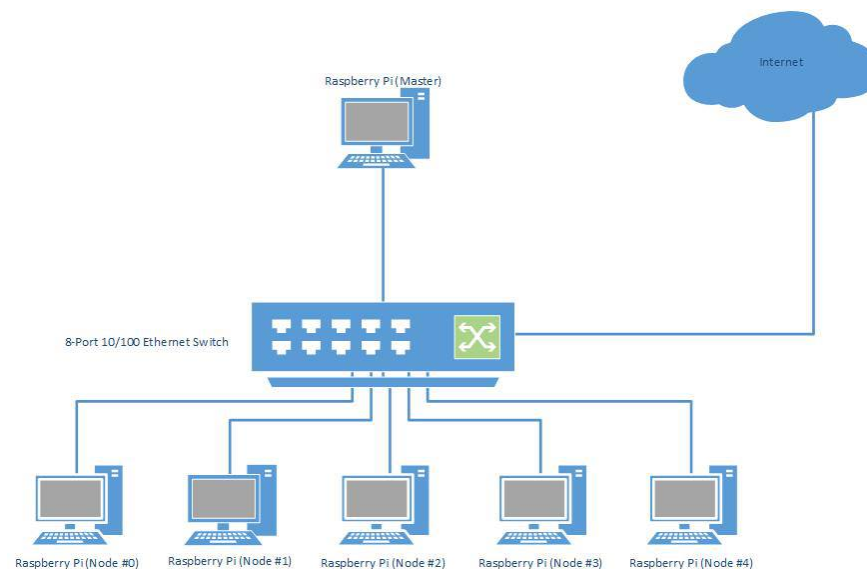
Category	Qty	Description	Supplier	Unit Cost	Total Cost
Computer	6	Raspberry Pi Model B 512MB	element14	\$41.80	\$250.80
Storage	6	Assorted brands 8GB SD card	MULTIPLE	\$12.00	\$72.00
Enclosure/Case	6	Raspberry Pi Case (Clear)	element14	\$9.79	\$58.74
Power Supply	6	Power Supply USB 5V/1.2A	element14	\$22.00	\$132.00
Ethernet Cables	6	1.5m Cat-5e Patch Cable	MULTIPLE	\$3.00	\$18.00
Switch	1	8-Port 10/100 Mbps Ethernet Switch	UNKNOWN	\$34.00	\$34.00

The above table details the hardware components used to create the Punnet of Berries compute cluster.

5.2 System Design

A chain is only as strong as its weakest link.

In a distributed model super computer, this statement cannot be more true. No matter how quickly each individual node can process its allocated tasks, if the communications link between the system's master and each of the slave nodes becomes a bottleneck, then the entire system will suffer in performance.



In order to produce an efficient, flexible and reliable system, the selection of a networking topology which can fulfil those three specific characteristics is crucial. By selecting the star network topology as the basis for the Punnet of Berries super computer, the key requirements of efficiency, flexibility and reliability are met.

5.2.1 Efficiency

By design, star networks utilise a central device, such as a networking switch, to handle the flow of packets in a network. This allows data packets to only travel to those nodes which they are intended for, reducing the overall traffic in the network.

5.2.2 Flexibility

By utilising a central device for communications, a star network can easily be expanded to support additional nodes without affecting the entire network. Furthermore nodes can be removed at any time, if servicing is required.

5.2.3 Redundancy

The likelihood of total system failure is greatly reduced by using a central device for communication. The failure of an individual node wouldn't have any effect on the overall system. However, the central device failing would be catastrophic, bringing the entire system down.

5.3 The Operating System

As discussed earlier, it was decided that **Arch Linux ARM** would be the Punnet of Berries' operating system.

The following section describes the steps that were taken to configure a basic Arch Linux image for use in the compute cluster. The base image was obtained from the Raspberry Pi downloads page (<http://www.raspberrypi.org/downloads>).

5.3.1 Setting up Arch Linux ARM

- Resizing:

```
# fdisk /dev/mmcblk0
>> p
>> d
>> 2
>> n
>> e
>> (return) = accept default partition no
>> (return) = accept default start
>> (return) = accept default end
>> n
>> l
>> (return) = accept default start
```

```
>> (return) = accept default end
```

```
>> p
```

```
>> w
```

```
# sync; reboot
```

```
# resize2fs /dev/mmcblk0p5
```

- Set the timezone:

```
# ln -s /usr/share/zoneinfo/Australia/Melbourne /etc/localtime
```

- Update System:

```
# pacman-key --init
```

Press (Alt+F2) to switch to a 2nd virtual console, then enter the following command:

```
# ls -R / && ls -R / && ls -R /
```

Press (Alt+F1) to switch back 1st virtual console. Check whether the “pacman-key --init” command has finished running.

```
# pacman -Syu
```

- Users and Groups:

Note: Use the following credentials when executing the steps below:

Username: rpcluster

Password: rpcluster

```
# useradd -m <username>
```

```
# passwd <username>
```

```
# groupadd admin
```

```
# gpasswd -a <username> admin
```

- Install the following using Arch Linux’s package manager (pacman).

1. **Sudo:** # pacman -S sudo

– Give “admin” group sudo rights.

```
# visudo
```

Find “#%wheel ALL=(ALL) ALL”. Change it to:

```
%admin ALL=(ALL) ALL
```

2. **Vim:** # pacman -Sy vim
3. **GCC:** # pacman -Sy gcc
4. **Make:** # pacman -Sy make
5. **OpenMPI:** # pacman -Sy openmpi
6. **OpenSSH:** # pacman -Sy openssh
7. **NFS:** # pacman -Sy nfs-utils

PROGRAM DESIGN

The Berry Batch consists of a centralised manager daemon and worker daemons. While the Punnet of Berries' centralised master node runs the manager daemon, each slave node runs a worker daemon.

The Berry Batch manager iteratively pulls submitted jobs from the queue and determines the most appropriate slave to carry out the job. This is done by monitoring the system's resources and the jobs running or waiting. Users interact directly with the Berry Batch manager to manage their job requests. A user is able to:

- Submit jobs.
- View all queued jobs.
- View the status of all of a particular user's jobs.
- View the status of a particular job.
- Cancel their own jobs.

The workers exist to execute the jobs submitted to the system. They operate in a polling fashion. They will:

1. Wait to be assigned a job.
2. Execute the job.
3. Return the job's exit status.
4. Back to 1.

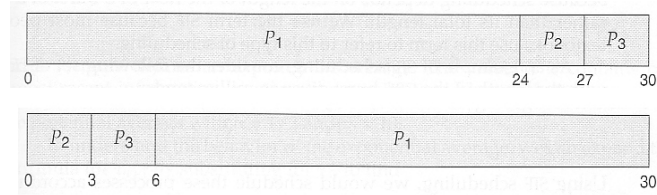
6.1 Scheduling Algorithms

As discussed earlier in this document, the user is able to select which scheduling algorithm the Berry Batch should use. The next few sections will outline the algorithms available.

6.1.1 First-Come-First-Served

The **First-Come-First Served (FCFS)** algorithm is very simple. As the name suggests, jobs are processed in the order that they are submitted. While the FCFS algorithm is very simple and easy to implement, its simplicity can also be its biggest flaw.

As shown in the following image, the waiting time for jobs in queue can vary greatly.



This is due to the running order being determined only by the job arrival time. Ignoring other factors, such as the estimated length of the job, often results in the CPU and device utilisation being lower than it could have been had shorter running jobs been scheduled first.

TESTING

7.1 Tests Done

7.1.1 Cluster Performance Tests

In order to test the performance and functionality of the ‘Punnet of Berries’ cluster, the system will be tested using a ‘HPL (High-Performance Linpack Benchmark)’. This benchmark provides guesstimates of how many GFLOPS (G iga **FL** oating-point **O** perations **P** er **S** econd) of processing performance the Punnet of Berries cluster is able to achieve.

Once the cluster itself has been benchmarked, the ‘HPL’ will then be run on a single Raspberry Pi as well as sepearte system containing a modern x86/x64 based processor. This will provide a meaningful comparison with regards to the performance gain a cluster of six Raspberry Pi’s has over various other systems.

7.1.2 Batch System Tests

To prove that the Berry Batch is functioning correctly, it will undergo a range of tests. These include, but are not limited to, the following:

Test Cases	Expected Result
Submit 1 short job.	Confirmation. Job put in short queue.
Submit 1 medium job.	Confirmation. Job put in medium queue.
Submit 1 long job.	Confirmation. Job put in long queue.
Submit 1 special job.	Job set to wait for permission.
One user submits a short job followed by one medium job.	First job put in short queue. Second job put in medium queue.
One user submit a short job. Second user submits a short job.	First user's job put in short queue. Second user's job put in short queue after the first.
One user submit a short job. Second user submits a medium job.	First user's job put in short queue. Second user's job put in medium queue.
One user submit a short job. Second user submits a medium job. Third user submits a medium job.	First user's job put in short queue. Second user's job put in medium queue. Third user's job put in long queue.
User views all current jobs.	List of jobs currently in the system.
User views a particular user's jobs.	List of user's jobs currently in the system.
User views a specific job.	Details of the given job.
First user views all current jobs. Second user views a specific jobs.	First user sees list of jobs currently in the system. Second user sees the details of the given job.
User cancels a queued job.	Confirmation of job cancellation.
User cancels a running job.	Confirmation of job cancellation.
User cancels a completed job.	Warning of invalid state.
User cancels an already cancelled job.	Warning of invalid state.
One user cancels another user's job.	Permission denied.

ROLES AND RESPONSIBILITIES

The members of TeamPi are:

- **Alyssa Biasi:**
Primary Role: Documentation, Testing and Management
Secondary Role(s): Operating System Configuration
- **Adrian Zielonka**
Primary Role: Hardware and Operating System Configuration
Secondary Role(s): Application Development, Documentation
- **Zach Ryan**
Primary Role: Application Development
Secondary Role(s): Documentation

WORK BREAKDOWN BY TEAM MEMBER

9.1 Alyssa Biasi's Log

9.1.1 Week 1

1. Raspberry PI cross compiler recipe - Milestone 1.
2. Project research.

9.1.2 Week 3

1. Evaluation of project management tools
 - > Trello
 - > Ganttter
 - > Jira
 - > Redmine
2. Setting up Redmine
 - > Hosted by www.hostedredmine.com
3. Setting up repositories
 - > hostedredmine.com only supports SVN
 - > Repository hosted on www.github.com/rmit-teamPi
 - Using GitHub's "Organizations" to allow team access
 - GitHub provides support for SVN allowing individual members to pick their preferred method of version control.

9.1.3 Week 4

1. Arch Linux Arm image setup.
 - > Partition layout changed in July 2013 (<http://davidnelsono.me/?p=218>)

9.1.4 Week 5

1. Setting up docutils, rst2pdf and Sphinx on a VM running Ubuntu for the generation of documentation.
 - > apt-get install python-docutils
 - > apt-get install rst2pdf
 - > apt-get install python-sphinx
 - > apt-get install texlive-latex-recommended
 - > apt-get install texlive-latex-extra
 - > apt-get install texlive-fonts-recommended
2. Work on design specification.

9.1.5 Week 6

1. Work on design specification.
2. Completed Milestone 2 - design specification.

9.1.6 Week 7

1. Setting up Arch Linux Arm images on 8GB SD cards.
2. Set up basic documents and sphinx for the final portfolio.

9.1.7 Week 8

1. Fixing SD cards to cluster specifications.

9.1.8 Week 11

1. Portfolio

9.1.9 Week 12

1. Portfolio

9.2 Adrian Zielonka's Log

9.2.1 Week 1

1. Brainstormed and researched viable project ideas.

9.2.2 Week 2

1. Brainstormed and researched viable project ideas.

9.2.3 Week 5

1. Worked on Design Specification.

9.2.4 Week 6

1. Worked on Design Specification.
2. Complete Milestone 2 - design specification.
3. Installed ArchLinux ARM (for Raspberry Pi) image (8GB SD Card).
4. Setup “masterpi” image for MASTER Raspberry Pi.

9.2.5 Week 8

1. Installed ArchLinux ARM (for Raspberry Pi) image (8GB SD Card).
2. Setup “slavepi0” image for SLAVE #0 Raspberry Pi.
3. Created baseline performance benchmark of single Raspberry Pi.

9.3 Zach Ryan’s Log

9.3.1 Week 1

9.3.2 Week 2

9.3.3 Week 3

9.3.4 Week 4

9.3.5 Week 5

9.3.6 Week 6

9.3.7 Week 7

SUMMARY

REFERENCES

11.1 Raspberry Pi

Murray, M. (2012, July 13). *Raspberry Pi*. Retrieved from <http://www.pcmag.com/article2/0,2817,2407058,00.asp>
RPi Easy SD Card Setup. Retrieved from http://elinux.org/Rpi_Easy_SD_Card_Setup
Build Guide - Linux From Scratch on the Raspberry Pi. Retrieved from <http://www.intestate.com/pilfs/guide.html>
Linux From Scratch. Retrieved from <http://www.linuxfromscratch.org/lfs/view/development>

11.2 Project Research

About Beowulf. Retrieved from <http://yclept.ucdavis.edu/Beowulf/aboutbeowulf.html>
Vaughan-Nichols, S. (2013, May 23). *Build your own supercomputer out of Raspberry Pi boards*. Retrieved from <http://www.zdnet.com/build-your-own-supercomputer-out-of-raspberry-pi-boards-7000015831>
Kiepert, J. (2013, May 22). *RPiCluster*. Retrieved from http://coen.boisestate.edu/ece/files/2013/05/Creating.a.Raspberry.Pi-Based.Beowulf.Cluster_v2.pdf
The Annual Unnamed UD Internet Contest - Problem 6: Supercomputer Job Scheduling. Retrieved from <http://www.eecis.udel.edu/~breech/contest.inet.fall.09/problems/sc-sched.html>
Simple Linux Utility for Resource Management. Retrieved from <https://computing.llnl.gov/linux/slurm/overview.html>
Sample PBS Script for Serial Job. Retrieved from http://qcd.phys.cmu.edu/QCDcluster/pbs/run_serial.html
Bell, J. *Operating Systems: CPU Scheduling*. Retrieved from http://www.cs.uic.edu/~jbell/CourseNotes/OperatingSystems/5_CPU_Sche

11.3 Arch Linux ARM

Nelson, D. (2013, June 16). *Growing the root filesystem on Arch Linux ARM for the Raspberry Pi*. Retrieved from <http://davidnelson.me/?p=218>
Norman (2013, June 4). *Beginner's Guide to Arch Linux on the Raspberry Pi*. Retrieved from <http://qdosmsq.dunbar-it.co.uk/blog/2013/06/beginners-guide-to-arch-linux-on-the-raspberry-pi/>
Norman (2013, June 12). *Beginner's Guide to Arch Linux on the Raspberry Pi - Part 2*. Retrieved from <http://qdosmsq.dunbar-it.co.uk/blog/2013/06/beginners-guide-to-arch-linux-on-the-raspberry-pi-part-2/>
Beginners' Guide - ArchWiki. Retrieved from https://wiki.archlinux.org/index.php/Beginners'_Guide

Arch Linux Install Guide. Retrieved from http://elinux.org/ArchLinux_Install_Guide