



docker

MSE

MASTER OF SCIENCE
IN ENGINEERING

Hes·SO

Haute Ecole Spécialisée
de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences and Arts
Western Switzerland

Projet de semestre Docker and embedded systems

Auteur :

Gary MARIGLIANO

Encadrant :

Jean-Roland SCHULER

Contact :

gary.marigliano@master.hes-so.ch

Mandant :

Haute École d'ingénierie et
d'architecture de Fribourg

Version 0.0.3
23 mai 2016

Historique

Version	Date	Auteur(s)	Modifications
0.0.1	02.05.16	Gary MARIGLIANO	Création du document
0.0.2	05.05.16	Gary MARIGLIANO	Ajout des chapitres Matériel utilisé, Présentation de Docker
0.0.3	11.05.16	Gary MARIGLIANO	Rédaction des sections Situation actuelle, Séparation des données Docker dans une partition chiffrée, Planning

1. Résumé du document

TODO

Table des matières

1	Résumé du document	
2	Introduction	2
2.1	Contexte	2
2.2	Objectifs	2
3	Présentation de Docker	3
3.1	Introduction	3
3.2	Containers vs machines virtuelles	3
3.3	Docker images et Docker containers	4
3.4	Système de fichiers en couche	5
3.5	Isolation	6
3.5.1	Les namespaces	7
3.5.2	cgroups - Control Groups	7
3.6	Contraintes liées au monde de l'embarqué	7
4	Matériel utilisé et mise en place de la cible	9
4.1	La carte ODROID-XU3 Lite	9
4.2	Installation du système	9
4.3	Installation de Docker	10
5	Objectif 3 - Docker et la sécurité	11
5.1	Situation actuelle	11
5.2	Structure de la suite du document	11
6	Configuration du système d'exploitation hôte	12
6.1	Passage en revue du benchmark de sécurité : CIS Docker 1.11.0 Benchmark	12
6.1.1	Situation actuelle	12
6.1.2	1.1 - Create a separate partition for containers	15
6.1.3	2.1 - Restrict network traffic between containers	15
6.1.4	2.5 - Do not use the aufs storage driver	15
6.1.5	2.8 - Enable user namespace support	17
6.1.6	2.11 - Use authorization plugin	17
6.1.7	2.12 - Configure centralized and remote logging	19
6.1.8	2.13 - Disable operations on legacy registry (v1)	20
6.1.9	4.1 - Create a user for the container	21
6.1.10	4.5 - Enable Content trust for Docker	23
6.1.11	5.3 - Restrict Linux Kernel Capabilities within containers	24
6.1.12	5.4 - Do not use privileged containers	24
6.1.13	5.6 - Do not run ssh within containers	24
6.1.14	5.10 - Limit memory usage for container	24
6.1.15	5.11 - Set container CPU priority appropriately	24
6.1.16	5.12 - Mount container's root filesystem as read only	24
6.1.17	Situation finale	24
6.2	Séparation des données Docker dans une partition chiffrée	24
6.2.1	Redimensionnement de la partition principale	25
6.2.2	Création d'une partition chiffrée	27
6.2.3	Montage manuel de la partition chiffrée	27
6.2.4	Montage au boot de la partition chiffrée	28
6.2.5	Preuve de bon fonctionnement du montage automatique	28

6.2.6	Binding des données de Docker sur la partition chiffrée	28
6.2.7	Analyse de la solution proposée	29
6.3	Conclusion sur la configuration du système d'exploitation hôte	30
7	Création et utilisation des images Docker	31
8	Utilisation des containers	32
9	Déroulement du projet	33
9.1	Planning initial	33
9.2	Planning final	33
10	Proposition d'améliorations vis à vis du travail précédent	34
11	Conclusion	35
	Appendices	38
A	Installation de Archlinux ARM sur ODROID-XU3 Lite	39
A.1	Micro SD Card Creation	39
A.2	eMMC Module Creation	40

2. Introduction

2.1 Contexte

Ce document est le rapport de fin de projet de semestre Docker and embedded systems. Un des buts de ce projet est de cross compiler Docker à partir de ses sources pour produire un binaire exécutable sur un ODROID-XU3 Lite (ARMv7). De plus, une partie concernant la sécurité de Docker est également traitée.

Lien : https://github.com/krypty/docker_and_embedded_systems

Ce projet de semestre s'inscrit dans une certaine continuité avec les projets de semestre et de bachelor de M. Loic Bassang [26]. Plusieurs pistes intéressantes avaient en effet été mentionnées dans ces projets-là notamment une partie concernant la sécurité et Docker. Ainsi, ce rapport fera parfois des parallèles avec ces documents.

Il est important de noter que la vitesse de développement de Docker est assez hallucinante. En effet, sur Github (<https://github.com/docker/docker>) les commits se succèdent à vitesse grand V. Entre chaque version de Docker, qui sortent environ tous les mois, il est courant d'avoir plus de 3000 commits qui ont été *pushés*. Tout ceci pour dire qu'à la lecture de ce document, il est quasiment sûr que certaines pistes explorées soient définitivement obsolètes ou au contraire deviennent la voie à suivre du à une mise à jour quelconque.

2.2 Objectifs

De manière plus précise, ce projet vise à maîtriser les parties suivantes :

1. Construction d'un système Linux capable de faire tourner Docker et son *daemon* en utilisant Buildroot. Pour générer le dit système, on dispose d'un *repository* Gitlab hébergé à la Haute École d'ingénierie et d'architecture de Fribourg
2. Cross compilation de Docker et de son *daemon*, capable de faire tourner des containers
3. Comprendre, analyser et évaluer l'aspect sécurité de Docker dans le cadre d'une utilisation avec une carte embarquée

Les deux premiers points ont été traités dans un précédent rapport appelé "État de l'art à la mi-projet de semestre Docker and embedded systems - Ou comment ne pas cross compiler Docker sur ARM" qu'il est recommandé de lire.

Ce document se concentre, dès lors, sur le dernier point ainsi que sur le déroulement global du projet.

3. Présentation de Docker

Remarque : Si le lecteur a déjà lu le rapport "État de l'art à la mi-projet de semestre Docker and embedded systems - Ou comment ne pas cross compiler Docker sur ARM", il ne lui est pas nécessaire de relire ce chapitre sachant qu'il s'agit du même contenu.

3.1 Introduction

Le but de ce chapitre est d'apporter un contexte au projet réalisé afin de comprendre ce qui est expliqué dans la suite de ce document. Si le lecteur souhaite connaître les tréfonds de Docker, il est recommandé de lire le rapport "Docker and Internet of Things (travail de semestre)" de M. Loic Bassang [25].

Docker est un outil qui permet d'empaqueter une application et ses dépendances dans un container léger, auto-suffisant, isolé et portable. En intégrant leur application dans un container, les développeurs s'assurent que celle-ci va tourner sur n'importe quel environnement GNU/Linux. Ainsi, le temps passé à configurer les différents environnements (développement, test et production typiquement) est réduit voire même unifié[28][5][20].

Les caractéristiques principales des containers Docker sont les suivantes :

- *légers* : les containers partagent le même kernel et librairies que le système d'exploitation hôte permettant ainsi un démarrage rapide des containers et une utilisation mémoire contenue
- *isolés* : les containers sont isolés grâce à des mécanismes offerts par le kernel. Voir section 3.5 pour plus de détails
- *éphémères et maintenables* : les containers sont conçus pour être créés et détruits régulièrement contrairement à un serveur ou une machine virtuelle pour lesquels un arrêt est souvent critique. Ils sont maintenables dans le sens où il est possible de revenir à une version précédente facilement et qu'il est aisé de déployer une nouvelle version

Les sections qui suivent abordent un peu plus en profondeur certains points clés de Docker qui ont été jugés importants.

3.2 Containers vs machines virtuelles

Suite à la lecture de la section précédente, il serait normal de se dire que ces containers partagent certaines caractéristiques avec les machines virtuelles.

Voici les différences majeures qu'il existe entre les containers et les machines virtuelles[20].

Les machines virtuelles possèdent leur propre OS qui embarque ses propres binaires et librairies. Ceci engendre une perte d'espace disque importante surtout si les binaires ou librairies sont communes à plusieurs machines virtuelles. De plus, démarrer une machine virtuelle prend du temps (jusqu'à plusieurs minutes). En outre, les machines virtuelles doivent installer leurs propres drivers afin de communiquer avec l'hyperviseur (logiciel s'exécutant à l'intérieur d'un OS hôte qui gère les machines virtuelles). Un avantage cependant est l'isolation complète d'une machine virtuelle qui ne peut communiquer avec les autres par défaut.

Les containers s'exécutent de manière isolée par dessus l'OS hôte qui partage ses ressources (kernel, binaires, librairies, périphériques,...). Plus légers, les containers démarrent en quelques secondes seulement. Sur une machine, il est tout à fait possible de lancer de milliers de containers similaires car l'empreinte mémoire est réduite et l'espace disque occupé est partagé si les containers sont semblables. Ceci est expliqué plus en détails à la section 3.4. Les containers sont isolés mais ils peuvent aussi communiquer entre eux si on leur a explicitement donné l'autorisation.

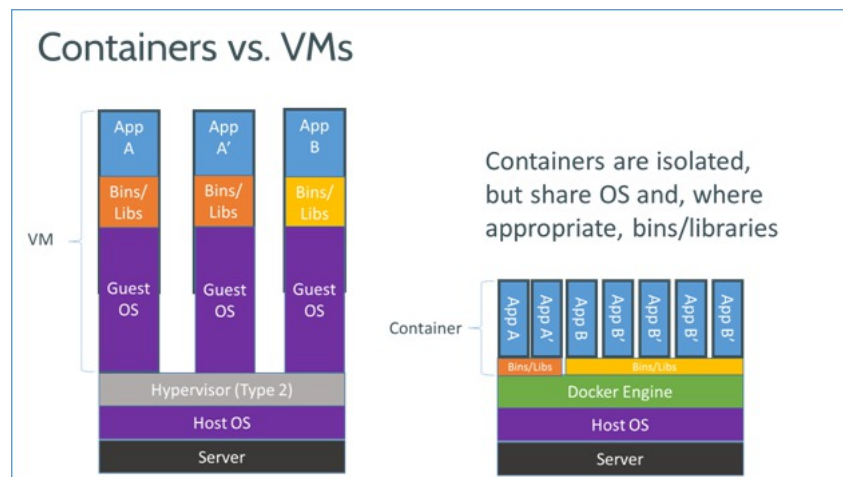


FIGURE 3.1 – Containers vs machines virtuelles

Si le lecteur désire connaître plus de détails concernant la virtualisation, il est conseillé de lire le chapitre 3.2 du rapport "Docker and Internet of Things (travail de semestre)" de M. Loic Bassang [25] qui amène une bonne introduction aux différents types de virtualisation.

3.3 Docker images et Docker containers

Avec Docker, une application est encapsulée avec toutes ses dépendances et sa configuration dans une **image**.

Pour construire cette image, on utilise un Dockerfile. Il s'agit d'un fichier qui décrit les étapes de création et de configuration nécessaires à l'obtention de l'application configurée. C'est dans ce fichier qu'on retrouve l'OS à utiliser, les dépendances à installer et toutes autres configurations utiles au bon fonctionnement de l'application à déployer.

Typiquement un Dockerfile permettant de lancer un serveur web Nginx qui affiche un "hello world" ressemble à ceci :

```
1 FROM alpine # image de départ
2 MAINTAINER support@tutum.co # mainteneur du Dockerfile
3 RUN apk --update add nginx php-fpm && \ # installation des dépendances
4     mkdir -p /var/log/nginx && \
5     touch /var/log/nginx/access.log && \
6     mkdir -p /tmp/nginx && \
7     echo "clear_env = no" >> /etc/php/php-fpm.conf
8 ADD www /www # ajout des sources de l'application
9 ADD nginx.conf /etc/nginx/ # ajout d'un fichier de configuration
10 EXPOSE 80 # ouverture du port 80
11 CMD php-fpm -d variables_order="EGPCS" && (tail -F
    ↪ /var/log/nginx/access.log &) && exec nginx -g "daemon off;" # commande
    ↪ à lancer au lancement du container
```

Source : <https://github.com/tutumcloud/hello-world/blob/master/Dockerfile>

Un Dockerfile est en quelque sorte la recette de cuisine qui permet de construire une image Docker.

Une fois l'image construite, on peut exécuter l'application dans un container. Un container Docker est donc une instance de l'image fraîchement créée. La figure 3.2 montre les relations entre un Dockerfile, une image et un container.

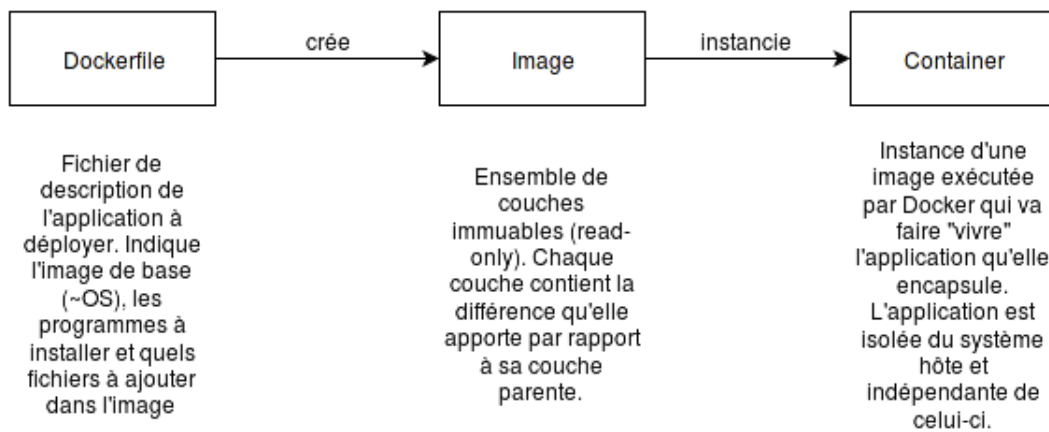


FIGURE 3.2 – Dockerfile, image et container

3.4 Système de fichiers en couche

Chaque image Docker est composée d'une liste de couches (*layers*) superposées en lecture seule[19]. Chaque couche représente la différence du système de fichiers par rapport à la couche précédente. Sur la figure 3.3, on peut voir 4 couches (identifiables avec des ID) et leur taille respective.

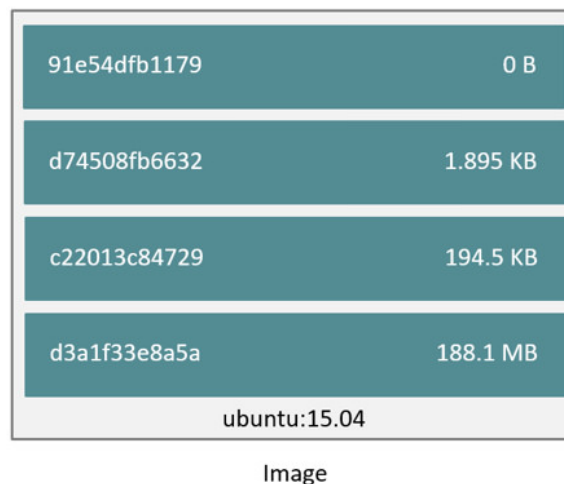


FIGURE 3.3 – Couches d'une image Docker

A la création d'un container, une nouvelle couche fine est ajoutée. Cette couche, appelée "container layer" est accessible en écriture durant l'exécution du container. La figure 3.4 le montre clairement.

Un mot supplémentaire sur une nouvelle caractéristique arrivée avec Docker 1.10 (mars 2016) ; avant cette version, Docker attribuait des UUID¹ générés aléatoirement pour identifier les couches d'une image. Désormais, ces UUID sont remplacés par des hash appelés *secure content hash*.

Les différences principales entre un UUID et un hash sont :

- Un UUID est généré aléatoirement, donc deux images exactement identiques auront un UUID différent alors qu'en utilisant un hash, le résultat sera identique
- Avec les UUID, même si la probabilité est rare², il est possible de générer deux fois le même UUID ce qui peut poser des problèmes lors de la construction des images

1. UUID : https://fr.wikipedia.org/wiki/Universal_Unique_Identifier

2. Probabilité de doublon : https://en.wikipedia.org/wiki/Universally_unique_identifier#Random_UUID_probability_of_duplicates

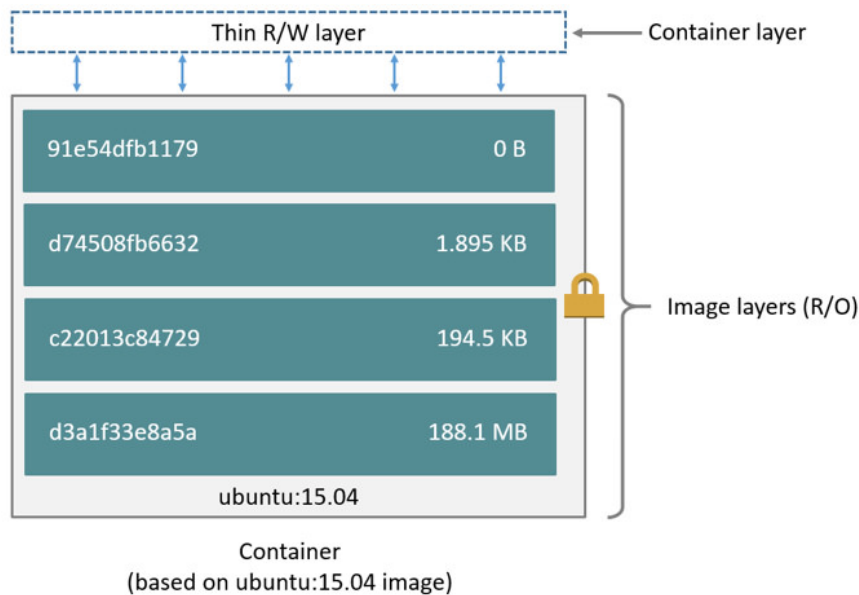


FIGURE 3.4 – Couches d'un container Docker

- Une image téléchargée chez une personne A aura un UUID différent que la même image téléchargée chez une personne B. Impossible de s'assurer de l'intégrité de l'image téléchargée en se basant sur l'UUID. En utilisant le hash, on s'assure du même résultat si l'image est identique

Par conséquent, Docker avance les avantages suivants :

- Intégrité des images téléchargées et envoyées (sur Docker Hub par exemple)
- Évite les collisions lors de l'identification des images et des couches
- Permet de partager des couches identiques qui proviendraient de *build* différents

Le dernier point est relativement intéressant. En effet, si deux images de base (Ubuntu et Debian) sont différentes mais qu'une couche supérieure est identique (par exemple l'ajout d'un même fichier texte) alors cette couche supérieure peut être partagée entre les deux images (puisqu'elle possède le même hash). Ceci peut potentiellement offrir un gain d'espace disque conséquent si plusieurs images partagent plusieurs couches identiques. Ceci n'aurait pas pu être possible avec les UUID car les deux couches auraient produit deux UUID différents. Un exemple est visible à la figure 3.5

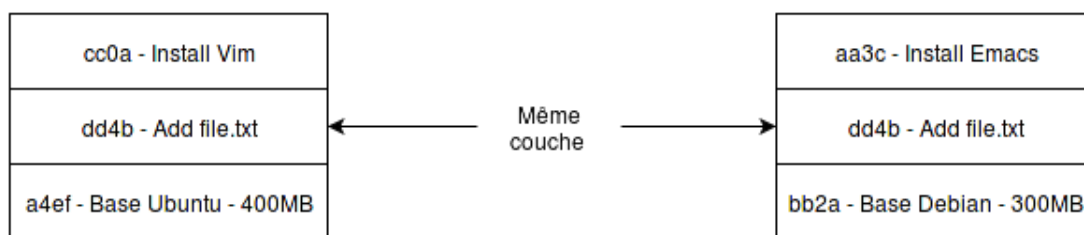


FIGURE 3.5 – Couche partagée entre deux images Docker

D'autres explications plus détaillées sur le système de fichiers en couches ont fait l'objet du chapitre 5.4.1 du rapport "Docker and Internet of Things (travail de semestre)" de M. Loïc Bassang.

3.5 Isolation

Docker met en avant le fait que ses containers soient isolés du système hôte. Pour ce faire, Docker utilise des mécanismes fournis par le kernel. On peut citer parmi ces mécanismes les *namespaces* et *cgroups*.

3.5.1 Les namespaces

Les namespaces permettent d'isoler certaines fonctionnalités d'un système d'exploitation utilisant Linux. Comme chroot permet aux processus de voir comme racine un dossier isolé du système et non pas la "vraie" racine, les namespaces isolent certains aspects du système comme les processus, les interfaces réseaux, les points de montage, etc.

Jusqu'à très récemment (docker < 1.10.0), Docker supportait les namespaces suivants[4] :

- PID namespace, chaque conteneur a ses propres id de processus
- UTS namespace, pour avoir son propre hostname
- IPC namespace, qui permet d'isoler les Communications Inter-Processus
- Network namespace, chaque conteneur peut avoir sa propre interface réseau, son ip, ses règles de filtrage

Docker a désormais ajouté le support d'un nouveau namespace : user namespace. Celui-ci permet à un processus d'avoir les droits root au sein d'un namespace mais pas en dehors. Avant, Docker lançait les containers en root ce qui pouvait poser des problèmes de sécurité si un processus dans le container venait à en sortir ; il se retrouverait root sur le système hôte. Avec la prise en charge de ce namespace, un container Docker a l'impression d'être root alors qu'il n'est, en réalité, qu'un utilisateur normal sur le système hôte.

3.5.2 cgroups - Control Groups

Cgroups (control groups) est une fonctionnalité du kernel pour limiter, prioriser, isoler et contrôler l'utilisation des ressources (CPU, RAM, utilisation disque, utilisation réseau,...). Pour limiter les ressources, cgroups propose de créer un groupe (profil) qui décrit les limitations à respecter. Par exemple, Si on crée un groupe appelé "groupe 1" et qu'on exige de lui qu'il n'utilise qu'au maximum 25% de la charge CPU et n'utilise qu'au maximum 100 MB de RAM. Alors, il devient possible de lancer des programmes qui appartiennent à ce groupe et qui respectent les limites fixées.

Lorsqu'on utilise la commande `docker run` de Docker pour lancer un container, Docker peut utiliser cgroups et ainsi limiter les ressources du container³.

3.6 Contraintes liées au monde de l'embarqué

On entend par système embarqué, un système qui est/peut être léger, autonome, à puissance limitée, à stockage réduit, avec un OS minimal et souvent connecté.

Dans le monde de l'embarqué, il existe plusieurs problèmes récurrents lorsqu'on développe, déploie et maintient une application sur une cible. On peut citer les problèmes suivants :

- Cross-compilation souvent obligatoire
- Aucune interface graphique
- Installation et configuration des dépendances sur la cible
- Mises à jour de l'application et de ses dépendances
- Tests et journalisation (logs)
- Limiter l'utilisation en CPU, RAM et disque

Bien que ces problèmes peuvent être aussi présents dans le cas d'un développement desktop, ils ne sont pas aussi préoccupants.

3. Docker - Runtime constraints on resources : <https://docs.docker.com/engine/reference/run/#runtime-constraints-on-resources>

Docker peut être utile dans le cadre d'une application embarquée car il permet une maintenance plus aisée de l'application car on peut **versionner son installation et sa configuration ainsi que celle de ses dépendances**. Ceci permet de plus facilement mettre à jour une application mais également de pouvoir revenir à une version précédente. De plus, si un accès réseau est disponible, il est même possible d'administrer Docker à distance depuis un poste de développement en se connectant au *daemon* Docker de la cible.

Cependant, il reste quelques freins et pré-requis pour pouvoir utiliser Docker sur une carte embarquée :

GNU/Linux : Il faut un système GNU/Linux et si possible une distribution qui intègre Docker dans ses packages

Images compatibles : Les images doivent être compatibles avec la plateforme de la cible. Les images x64 ne fonctionnent pas sur ARM. Actuellement, la majorité des images sont basées sur une image de base x64. Dans la plupart des cas, il suffit de trouver l'équivalent ARM de l'image de base. Par exemple, à la lecture d'un Dockerfile, il suffit de remplacer `FROM ubuntu` par `FROM armhf/ubuntu` pour que l'image arrive à se construire. Dans les autres cas, il faudra adapter les instructions du Dockerfile.

Espace disque limité : Il faut veiller à l'utilisation de l'espace disque. En effet, la plupart des images se basent sur des Ubuntu (400 MB) ou des Debian (300 MB) ce qui peut être trop volumineux pour un système embarqué. De plus plusieurs versions de ces images peuvent être téléchargées si les Dockerfiles le spécifient. Par exemple, `FROM ubuntu:14.04` pour l'image 1 et `FROM ubuntu:15.10` pour l'image 2. On favorisera l'utilisation d'une image de base légère, comme Alpine Linux⁴, commune à plusieurs applications ou processus tournant sur la cible.

4. Alpine Linux : <http://www.alpinelinux.org/>

4. Matériel utilisé et mise en place de la cible

Ce chapitre présente le matériel utilisé dans le projet ainsi que son installation et sa configuration de base.

Afin de réaliser ce projet, une carte embarquée ODROID-XU3 Lite a été mise à disposition afin d'y faire tourner Docker.

4.1 La carte ODROID-XU3 Lite

Cette carte possède les caractéristiques suivantes [8] :

- Samsung Exynos5422 Cortex™-A15 1.8Ghz quad core and Cortex™-A7 quad core CPUs
- Mali-T628 MP6(OpenGL ES 3.0/2.0/1.1 and OpenCL 1.1 Full profile)
- 2Gbyte LPDDR3 RAM at 933MHz (14.9GB/s memory bandwidth) PoP stacked
- eMMC5.0 HS400 Flash Storage
- USB 3.0 Host x 1, USB 3.0 OTG x 1, USB 2.0 Host x 4
- HDMI 1.4a for display

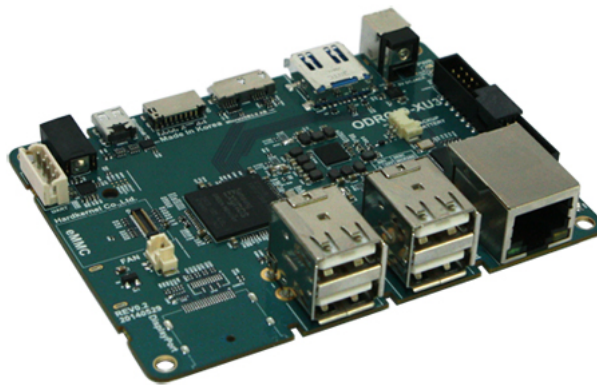


FIGURE 4.1 – ODROID-XU3 Lite

Lien vers le constructeur : http://www.hardkernel.com/main/products/prdt_info.php?g_code=G141351880955.

4.2 Installation du système

Initialement, il était prévu de générer un système d'exploitation minimal qui aurait été capable de faire tourner Docker et des containers. Malheureusement, il n'a pas été possible de cross compiler Docker *et son daemon* afin de lancer des containers sur ce système minimal. Plus d'informations sont disponibles dans le rapport "État de l'art à la mi-projet de semestre Docker and embedded systems - Ou comment ne pas cross compiler Docker sur ARM".

Ainsi, il a été décidé, de la même manière que pour le travail de bachelor précédent, d'utiliser une distribution GNU/Linux proposant Docker dans ses packages. Le choix s'est donc porté sur **Archlinux ARM**¹.

Sur la page wiki de la distribution (<https://archlinuxarm.org/platforms/armv7/samsung/odroid-xu3>), on peut suivre un guide de génération de la carte SD qui contient le système d'exploitation. Une copie de ce guide est disponible à l'appendice A.

1. Arch Linux ARM : <https://archlinuxarm.org/>

4.3 Installation de Docker

Une fois le système installé et démarré, il convient d'installer Docker depuis les dépôts de la distribution.

Ceci peut être accompli avec les commandes suivantes :

```
1  pacman -Syuu
2  reboot
3  pacman -S docker
4  systemctl enable docker
5  systemctl start docker
6  docker version
```

Sortie :

```
1  Client:
2    Version:      1.10.3
3    API version:   1.22
4    Go version:    go1.6
5    Git commit:    20f81dd
6    Built:         Sat Mar 12 12:49:56 2016
7    OS/Arch:       linux/arm
8
9  Server:
10   Version:      1.10.3
11   API version:   1.22
12   Go version:    go1.6
13   Git commit:    20f81dd
14   Built:         Sat Mar 12 12:49:56 2016
15   OS/Arch:       linux/arm
```

5. Objectif 3 - Docker et la sécurité

5.1 Situation actuelle

Pour cette partie du projet, l'objectif est le suivant : Analyser et évaluer l'aspect sécurité de Docker dans le cadre d'une utilisation avec une carte embarquée.

Pour rappel, on utilise la carte ODROID-XU3 Lite . Un des objectifs de ce projet était de générer un système GNU/Linux minimal contenant un binaire Docker cross- compilé. Malheureusement, cet objectif n'a pas abouti comme expliqué dans le rapport "État de l'art à la mi-projet de semestre Docker and embedded systems - Ou comment ne pas cross compiler Docker sur ARM".

Pour continuer sur la partie sécurité, il a été choisi d'utiliser une distribution GNU/Linux compatible avec la carte et qui propose dans ces dépôts un package Docker précompilé. Tout comme le travail précédent de M. Loïc Bassang, Arch Linux ARM a été sélectionné.

5.2 Structure de la suite du document

Pour ce projet, il a été décidé d'étudier la question de la sécurité avec Docker avec une approche en couches. A peu à la manière du modèle OSI¹ en réseau, chaque couche représente un ensemble de fonctionnalités qui, dans le cas de ce projet, doit faire l'objet d'une évaluation de la sécurité.

L'étude de la sécurité de Docker a donc été séparée avec les couches *arbitraires* suivantes :

- Compilation et installation de Docker : en particulier les options de compilation
- Configuration du système d'exploitation hôte : configuration du kernel, configuration des options de lancement de Docker, etc.
- Création et utilisation des images Docker : Bonnes pratiques et contraintes liées au monde de l'embarqué
- Utilisation des containers : options de lancement

Remarque : Chacune de ces couches fait l'objet d'un chapitre dans ce rapport excepté le premier point : Compilation et installation de Docker. En effet, celui-ci n'est pas traité car, comme annoncé précédemment, la cross compilation de Docker sur un système ARM n'a pas aboutie. L'effort est alors concentré sur les autres points.

1. Modèle OSI : https://fr.wikipedia.org/wiki/Mod%C3%A8le_OSI

6. Configuration du système d'exploitation hôte

Dans ce chapitre, on présente diverses bonnes pratiques et configurations dans le but de sécuriser Docker et/ou le système l'hébergeant.

Parmi ces techniques, on peut citer :

- Création d'une partition chiffrée pour stocker les données de Docker (containers et images)
- Modification des options de lancement du *daemon* Docker

6.1 Passage en revue du benchmark de sécurité : CIS Docker 1.11.0 Benchmark

Il s'agit d'un document à l'intention d'administrateurs systèmes ou spécialistes de la sécurité qui souhaitent intégrer, développer ou sécuriser des solutions qui utilisent Docker.

Le but de ce document est de proposer un éventail de recommandations concernant la sécurité de Docker sur plusieurs points. Voici les points évalués :

1. *Host Configuration* : Recommandations pour la configuration de la machine hôte
2. *Docker daemon configuration* : Recommandations concernant le comportement du *daemon* Docker
3. *Docker daemon configuration files* : Recommandations concernant la permission des fichiers et dossiers de configuration du *daemon* Docker
4. *Container Images and Build File* : Recommandations concernant les images de base et les fichiers de build
5. *Container runtime* : Recommandations sur les options de lancement des containers
6. *Docker Security Operations* : Rappels des bonnes pratiques concernant la sécurité de Docker

Lien : https://benchmarks.cisecurity.org/tools2/docker/CIS_Docker_1.11.0_Benchmark_v1.0.0.pdf

Dans cette partie du rapport, quelques points de ce benchmark seront analysés et appliqués à la cible dans des conditions les plus réelles possibles.

6.1.1 Situation actuelle

Avant de chercher à sécuriser la cible, il faut connaître son état de sécurisation actuel. Pour ce faire, il suffit de lancer le script du benchmark qui va exécuter toutes les recommandations proposées et indiquer quels points sont à corriger.

Avant de commencer, il faut lancer un container en arrière plan afin de pouvoir évaluer la partie 5 - *Container runtime* qui évalue les containers actuellement exécutés. Pour ce faire, on choisit de lancer en tâche de fond un serveur web sur la cible.

A faire sur la cible :

```
1 docker run --name=my_container -d -p 8080:80 hypriot/rpi-busybox-httpd
```

Lancement du benchmark :

```
1 git clone https://github.com/docker/docker-bench-security.git
2 cd docker-bench-security/
3 sudo sh docker-bench-security.sh
```


On obtient la sortie suivante :

```

1   Initializing Sat May 14 14:12:18 UTC 2016
2
3
4   [INFO] 1 - Host Configuration
5   [WARN] 1.1 - Create a separate partition for containers
6   [PASS] 1.2 - Use an updated Linux Kernel
7   [PASS] 1.4 - Remove all non-essential services from the host - Network
8   [PASS] 1.5 - Keep Docker up to date
9   [INFO]      * Using 1.11.1 which is current as of 2016-04-27
10  [INFO]      * Check with your operating system vendor for support and
    ↪ security maintenance for docker
11  [INFO] 1.6 - Only allow trusted users to control Docker daemon
12  [INFO]      * docker:x:996:alarm
13  [WARN] 1.7 - Failed to inspect: auditctl command not found.
14  [WARN] 1.8 - Failed to inspect: auditctl command not found.
15  [WARN] 1.9 - Failed to inspect: auditctl command not found.
16  [WARN] 1.10 - Failed to inspect: auditctl command not found.
17  [WARN] 1.11 - Failed to inspect: auditctl command not found.
18  [INFO] 1.12 - Audit Docker files and directories - /etc/default/docker
19  [INFO]      * File not found
20  [INFO] 1.13 - Audit Docker files and directories - /etc/docker/daemon.json
21  [INFO]      * File not found
22  [WARN] 1.14 - Failed to inspect: auditctl command not found.
23  [WARN] 1.15 - Failed to inspect: auditctl command not found.
24
25
26  [INFO] 2 - Docker Daemon Configuration
27  [WARN] 2.1 - Restrict network traffic between containers
28  [PASS] 2.2 - Set the logging level
29  [PASS] 2.3 - Allow Docker to make changes to iptables
30  [PASS] 2.4 - Do not use insecure registries
31  [WARN] 2.5 - Do not use the aufs storage driver
32  [INFO] 2.6 - Configure TLS authentication for Docker daemon
33  [INFO]      * Docker daemon not listening on TCP
34  [INFO] 2.7 - Set default ulimit as appropriate
35  [INFO]      * Default ulimit doesn't appear to be set
36  [WARN] 2.8 - Enable user namespace support
37  [PASS] 2.9 - Confirm default cgroup usage
38  [PASS] 2.10 - Do not change base device size until needed
39  [WARN] 2.11 - Use authorization plugin
40  [WARN] 2.12 - Configure centralized and remote logging
41  [WARN] 2.13 - Disable operations on legacy registry (v1)
42
43
44  [INFO] 3 - Docker Daemon Configuration Files
45  [PASS] 3.1 - Verify that docker.service file ownership is set to root:root
46  [PASS] 3.2 - Verify that docker.service file permissions are set to 644
47  [PASS] 3.3 - Verify that docker.socket file ownership is set to root:root
48  [PASS] 3.4 - Verify that docker.socket file permissions are set to 644
49  [PASS] 3.5 - Verify that /etc/docker directory ownership is set to
    ↪ root:root
50  [PASS] 3.6 - Verify that /etc/docker directory permissions are set to 755
51  [INFO] 3.7 - Verify that registry certificate file ownership is set to
    ↪ root:root
52  [INFO]      * Directory not found
53  [INFO] 3.8 - Verify that registry certificate file permissions are set to
    ↪ 444
54  [INFO]      * Directory not found

```

```

55 [INFO] 3.9 - Verify that TLS CA certificate file ownership is set to
    ↳ root:root
56 [INFO]      * No TLS CA certificate found
57 [INFO] 3.10 - Verify that TLS CA certificate file permissions are set to
    ↳ 444
58 [INFO]      * No TLS CA certificate found
59 [INFO] 3.11 - Verify that Docker server certificate file ownership is set
    ↳ to root:root
60 [INFO]      * No TLS Server certificate found
61 [INFO] 3.12 - Verify that Docker server certificate file permissions are
    ↳ set to 444
62 [INFO]      * No TLS Server certificate found
63 [INFO] 3.13 - Verify that Docker server key file ownership is set to
    ↳ root:root
64 [INFO]      * No TLS Key found
65 [INFO] 3.14 - Verify that Docker server key file permissions are set to 400
66 [INFO]      * No TLS Key found
67 [PASS] 3.15 - Verify that Docker socket file ownership is set to
    ↳ root:docker
68 [PASS] 3.16 - Verify that Docker socket file permissions are set to 660
69 [INFO] 3.17 - Verify that daemon.json file ownership is set to root:root
70 [INFO]      * File not found
71 [INFO] 3.18 - Verify that daemon.json file permissions are set to 644
72 [INFO]      * File not found
73 [INFO] 3.19 - Verify that /etc/default/docker file ownership is set to
    ↳ root:root
74 [INFO]      * File not found
75 [INFO] 3.20 - Verify that /etc/default/docker file permissions are set to
    ↳ 644
76 [INFO]      * File not found
77
78
79 [INFO] 4 - Container Images and Build Files
80 [WARN] 4.1 - Create a user for the container
81 [WARN]      * Running as root: my_container
82 [WARN] 4.5 - Enable Content trust for Docker
83
84
85 [INFO] 5 - Container Runtime
86 [WARN] 5.1 - Verify AppArmor Profile, if applicable
87 [WARN]      * No AppArmorProfile Found: my_container
88 [WARN] 5.2 - Verify SELinux security options, if applicable
89 [WARN]      * No SecurityOptions Found: my_container
90 [PASS] 5.3 - Restrict Linux Kernel Capabilities within containers
91 [PASS] 5.4 - Do not use privileged containers
92 [PASS] 5.5 - Do not mount sensitive host system directories on containers
93 [PASS] 5.6 - Do not run ssh within containers
94 [PASS] 5.7 - Do not map privileged ports within containers
95 [PASS] 5.9 - Do not share the host's network namespace
96 [WARN] 5.10 - Limit memory usage for container
97 [WARN]      * Container running without memory restrictions: my_container
98 [WARN] 5.11 - Set container CPU priority appropriately
99 [WARN]      * Container running without CPU restrictions: my_container
100 [WARN] 5.12 - Mount container's root filesystem as read only
101 [WARN]      * Container running with root FS mounted R/W: my_container
102 [WARN] 5.13 - Bind incoming container traffic to a specific host interface
103 [WARN]      * Port being bound to wildcard IP: 0.0.0.0 in my_container
104 [WARN] 5.14 - Set the 'on-failure' container restart policy to 5
105 [WARN]      * MaximumRetryCount is not set to 5: my_container
106 [PASS] 5.15 - Do not share the host's process namespace
107 [PASS] 5.16 - Do not share the host's IPC namespace

```

```

108 [PASS] 5.17 - Do not directly expose host devices to containers
109 [INFO] 5.18 - Override default ulimit at runtime only if needed
110 [INFO]      * Container no default ulimit override: my_container
111 [PASS] 5.19 - Do not set mount propagation mode to shared
112 [PASS] 5.20 - Do not share the host's UTS namespace
113 [PASS] 5.21 - Do not disable default seccomp profile
114 [PASS] 5.24 - Confirm cgroup usage
115 [WARN] 5.25 - Restrict container from acquiring additional privileges
116 [WARN]      * Privileges not restricted: my_container
117
118
119 [INFO] 6 - Docker Security Operations
120 [INFO] 6.4 - Avoid image sprawl
121 [INFO]      * There are currently: 2 images
122 [INFO] 6.5 - Avoid container sprawl
123 [INFO]      * There are currently a total of 8 containers, with 1 of them
    ↪ currently running

```

Comme on peut le voir, il y a plusieurs points qui sont en *WARN*. Pour la suite du document, plusieurs de ces points vont être analysés et mitigés.

6.1.2 1.1 - Create a separate partition for containers

On recommande pour ce point de créer une partition séparée pour les containers. Ce point est traité à la section 6.2. Dans cette section, une étape supplémentaire est proposée ; la partition est chiffrée.

6.1.3 2.1 - Restrict network traffic between containers

Par défaut, le trafic réseau est autorisé entre containers du même hôte. On devrait désactiver ce comportement par défaut et explicitement autoriser une communication entre containers grâce au "linking".

Pour remédier à ceci, il faut lancer le *daemon* Docker avec cette option : `docker daemon -icc=false`

A faire sur la cible :

```
1 sudo vim /usr/lib/systemd/system/docker.service
```

Ajouter l'option `icc=false` à l'instruction `ExecStart`. Le résultat devrait ressembler à ceci :

```
1 ExecStart=/usr/bin/docker daemon -H fd:// --icc=false
```

Redémarrage du daemon :

```
1 sudo systemctl daemon-reload # recharge la configuration des services
2 sudo systemctl restart docker # relance le service docker
```

Résultat du benchmark après cette modification : **[PASS] 2.1 - Restrict network traffic between containers.**

6.1.4 2.5 - Do not use the aufs storage driver

Pour stocker les images et les différentes couches qui les composent, Docker utilise un driver de stockage [19], [16], [24]. Un driver de stockage (*driver storage*) est l'implémentation d'une abstraction de services liés à la gestion des couches des images et de la couche inscriptible (*writable*) des containers.

L'utilisation d'un driver ou d'un autre est transparent pour l'utilisateur. Cependant, une certaine implémentation, aufs, est considérée peu stable et peut causer des crashes au niveau du kernel, voir [18] page 45. Ce driver est donc à remplacer par un autre driver de stockage comme devicemapper par exemple.

Inconvénient : aufs est le seul driver de stockage qui permet aux containers de partager la mémoire des bibliothèques partagées. Cela peut être utile si la cible fait tourner des milliers de containers utilisant les mêmes programmes ou bibliothèques [18].

Dans le cadre d'une carte embarquée, on ne va pas faire tourner des milliers de containers en même temps et par conséquent changer de driver storage n'est pas un problème.

Par défaut Docker utilise devicemapper mais suivant la distribution ou la plateforme utilisée, un autre storage driver est utilisé. Dans le cas présent, aufs est malheureusement utilisé.

Vérification que device-mapper est présent sur le système :

```

1  [alarm@alarm docker-bench-security]# pacman -Qi device-mapper
2  Name           : device-mapper
3  Version        : 2.02.149-1
4  Description     : Device mapper userspace library and tools
5  Architecture   : armv7h
6  URL            : http://sourceware.org/dm/
7  Licenses       : GPL2  LGPL2.1
8  Groups         : base
9  Provides       : None
10 Depends On    : glibc  systemd
11 Optional Deps : None
12 Required By   : cryptsetup  docker  lvm2
13 Optional For  : None
14 Conflicts With : None
15 Replaces     : None
16 Installed Size : 871.00 KiB
17 Packager      : Arch Linux ARM Build System
   ↳ <builder+xu2@archlinuxarm.org>
18 Build Date    : Sun Apr 17 12:52:54 2016
19 Install Date  : Mon Apr 18 14:12:44 2016
20 Install Reason : Explicitly installed
21 Install Script : No
22 Validated By  : SHA-256 Sum

```

Comme avant, on change les paramètres de lancement du *daemon* Docker :

```

1  sudo vim /usr/lib/systemd/system/docker.service

```

Ajouter l'option **-storage-driver=devicemapper** à l'instruction ExecStart. Le résultat devrait ressembler à ceci :

```

1  ExecStart=/usr/bin/docker daemon -H fd:// --icc=false
   ↳ --storage-driver=devicemapper

```

```

1  sudo systemctl daemon-reload # recharge la configuration des services
2  sudo systemctl restart docker # relance le service docker

```

6.1.5 2.8 - Enable user namespace support

Les namespaces permettent d'isoler certaines fonctionnalités d'un système d'exploitation utilisant Linux. Comme chroot permet aux processus de voir comme racine un dossier isolé du système et non pas la "vraie" racine, les namespaces isolent certains aspects du système comme les processus, les interfaces réseaux, les points de montage, etc.

Jusqu'à très récemment (docker < 1.10.0), Docker supportait les namespaces suivants :

- PID namespace, chaque conteneur a ses propres id de processus
- UTS namespace, pour avoir son propre hostname
- IPC namespace, qui permet d'isoler les Communications Inter-Processus
- Network namespace, chaque conteneur peut avoir sa propre interface réseau, son ip, ses règles de filtrage

Pour plus de détails, voir section 3.5.

Il y a peu de temps, Docker ajouté le support d'un nouveau namespace[4] : user namespace. Celui-ci permet à un processus d'avoir les droits root au sein d'un namespace mais pas en dehors. Avant, Docker lançait les containers en root ce qui pouvait poser des problèmes de sécurité si un processus dans le container venait à en sortir ; il se retrouverait root sur le système hôte. Avec la prise en charge de ce namespace, un container Docker a l'impression d'être root alors qu'il n'est, en réalité, qu'un utilisateur normal sur le système hôte.

Pour mettre en place ce mécanisme, il faut réaliser les opérations suivantes :

```
1 touch /etc/subuid /etc/subgid
```

Et ajouter l'option `-userns-remap=default` au démarrage du *daemon* Docker.

Malheureusement, il a été impossible de démarrer le *daemon* avec cette option que ce soit sur la cible (Archlinux, kernel 3.10.96), sur un portable Archlinux (kernel 4.5.4) ou sur un autre portable Xubuntu 16.04 (kernel 4.4.0). Le *daemon* refusait de démarrer ou se bloquait. Il semblerait que cela ne soit pas un cas isolé. En effet plusieurs issues (#20751, #21130¹) ont été ouvertes sur le *repository* Github de Docker.

6.1.6 2.11 - Use authorization plugin

Le modèle de permission Docker par défaut est tout ou rien. Tout utilisateur qui a accès au *daemon* Docker peut effectuer n'importe quelles opérations sur tous les containers. Dans le cas où l'on a besoin d'un contrôle plus fin, il est possible de créer des plugins d'autorisation (*authorization plugins*) et les ajouter à la configuration du *daemon* Docker. Il est, par exemple, possible de définir que l'utilisateur Bob ne puisse pas arrêter les containers mais uniquement les lister.

Un plugin d'autorisation approuve ou refuse les requêtes du client (un utilisateur) vers le *daemon* en se basant sur le contexte actuel d'authentification et le contexte de commande. Le contexte d'authentification contient toutes les détails de l'utilisateur et de la méthode d'authentification et le contexte de commande contient toutes les données utiles à la requête de l'utilisateur.

Remarque : La communication entre le *daemon* et le client[17] se fait à travers un socket Unix (`/var/run/docker.sock`) mais peut aussi se faire via un socket TCP (`tcp://0.0.0.0:2375`) pour accéder au *daemon* de manière distante. Grâce à ce mécanisme, Docker propose *Docker Remote API*[14] une API *REST-like* qui permet d'accéder à la liste des containers, des images, de créer des containers, etc.

1. Docker issues : <https://github.com/docker/docker/issues>

Pour créer un plugin d'autorisation, il faut respecter le contrat décrit par la *Docker Plugin API*. De manière similaire à la *Docker Remote API*, cette API définit un cadre afin d'étendre les possibilités de Docker comme par exemple autoriser ou refuser l'accès à certaines fonctionnalités de Docker pour un utilisateur ou un groupe donné.

Maintenant les notions de communication client *daemon* et API REST éclaircies, on peut enfin voir l'architecture de base des plugins d'autorisation.

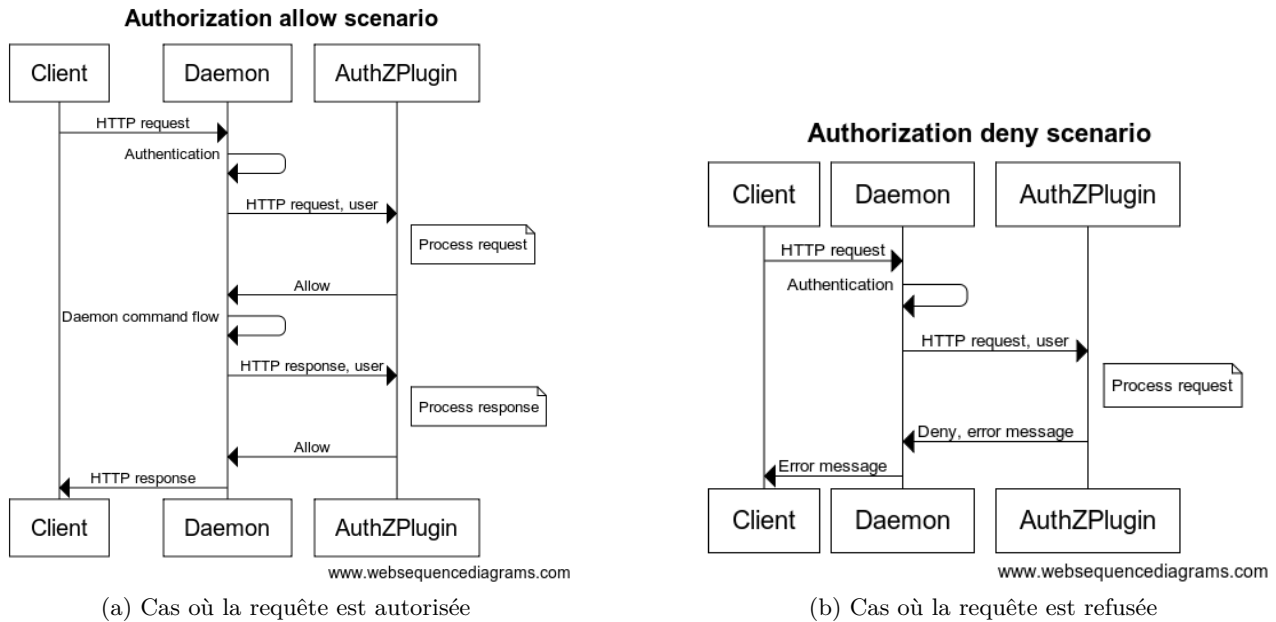


FIGURE 6.1 – Scénarios d'autorisation Docker²

Comme on peut le voir sur les figures ci-dessus, les requêtes sont des requêtes HTTP JSON comme on peut le trouver fréquemment dans avec les API REST.

Il est possible d'utiliser plusieurs plugins en même temps. Dans ce cas, pour qu'une requête soit acceptée, il faut qu'elle puisse traverser tous les plugins (donc être approuvée par tous les plugins). Si un seul plugin refuse la requête alors elle est refusée.

La création et mise en place de plugin n'est pas présenté dans ce rapport. Néanmoins, voici un exemple de plugin : <https://github.com/twistlock/authz>.

Analyse : Dans le cas d'une utilisation embarquée, il n'est pas indispensable d'avoir recours à un contrôle plus fin de l'utilisation de Docker car :

- la cible est souvent mono-utilisateur
- les containers appartiennent tous au même utilisateur. Dans le cas d'une plateforme offrant un service de type CaaS (Containers as a Service), on peut comprendre que chaque utilisateur n'ait accès qu'aux containers qu'il lance.

Par conséquent, on décide de ne pas mettre en place ce mécanisme. On rappelle que ce benchmark a surtout pour but de montrer un ensemble de bonnes pratiques et de rendre attentif à certains aspects de la sécurité avec Docker. Il n'est néanmoins pas toujours indispensable d'appliquer toutes les mesures de sécurités si on en exprime pas le besoin.

2. Source des images : https://docs.docker.com/engine/extend/plugins_authorization/

6.1.7 2.12 - Configure centralized and remote logging

Docker supporte plusieurs *log drivers* (manière de journaliser). Par défaut, les logs des containers sont produits dans des fichiers JSON sur la machine hôte. Mais il est recommandé de centraliser les logs sur une machine distante pour voir et traiter les logs globalement.

Le benchmark CIS recommande d'externaliser les logs en optant pour le log driver *syslog*. *Syslog*[29][10] est un protocole définissant un service de journaux d'événements d'un système informatique. Son but est de transporter par réseau les messages de journalisation générés par une application (dans ce cas Docker) vers un serveur hébergeant un serveur *Syslog*. Le but étant de regrouper tous les journaux au même endroit.

Pour la suite, on va mettre en place un petit exemple de démonstration qui n'a absolument pas la prétention d'être exhaustif ni sécurisé. Cet exemple a uniquement pour but de montrer un cas simple de *remote logging*.

Pour cet exemple, il y a à disposition un ordinateur portable qui fait office de serveur *Syslog* et la cible Odroid qui va lancer un container dont les logs sont envoyés à l'ordinateur portable. On peut voir cette configuration ci-dessous.

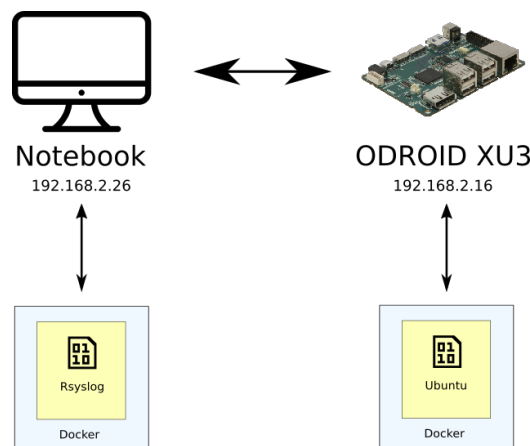


FIGURE 6.2 – Structure de démonstration Syslog

Le container (ici Ubuntu) va être lancé en indiquant que les logs doivent être envoyés au PC portable disposant d'un serveur syslog dans un container appelé rsyslog.

Pour le serveur *Syslog*, un container Docker prêt à l'emploi va être utilisé. Ce container utilise une image Rsyslog dont la source se trouve ici : <https://hub.docker.com/r/panoptix/rsyslog/>.

Lancement du container Rsyslog sur le portable et affichage des logs distants :

```
1  # démarrage du container
2  docker run --name rsyslog -t -d -p 514:514/udp -p 514:514 panoptix/rsyslog
3
4  # affichage des logs en temps réel
5  docker exec -it rsyslog bash
6  tail -f /var/log/syslog
```

Lancement du container Ubuntu (producteur de logs) sur la cible :

```
1  docker run -d --log-driver=syslog --log-opt
   ↪ syslog-address=tcp://192.168.2.26 --log-opt tag="remote" armhf/ubuntu
   ↪ echo "hello from Odroid"
```


Ce container va lancer la commande `echo` et se stopper immédiatement après. Évidemment dans un cas réel, le container effectuera des opérations plus longues et produirait une plus grande quantité de logs mais ici le but est de voir le principe général.

Sur l'ordinateur portable, on peut observer l'affichage suivant :

```
1 May 22 18:23:42 fc6dca22947a rsyslogd: [origin software="rsyslogd"
   ↪ swVersion="8.11.0" x-pid="1" x-info="http://www.rsyslog.com"] start
2 May 22 18:38:09 192.168.2.16 docker/remote[284]: hello from Odroid
```

Comme on peut le voir, on est capable de récupérer les logs de la cible sur l'ordinateur portable.

Remarque : Pour cet exemple, les options de logging ont été précisées au niveau du lancement du container (`-log-driver=XXX -log-opt syslog-address=tcp://aaa.bbb.ccc.ddd -log-opt YYY`). Idéalement, il faudrait définir par défaut toutes ces options au démarrage du *daemon*. Pour ce faire, il suffit de modifier la configuration de lancement du *daemon* comme il a été montré aux sections précédentes. Par exemple, on pourrait configurer le *daemon* ainsi :

```
1 ExecStart=/usr/bin/docker daemon -H fd:// --icc=false
   ↪ --storage-driver=devicemapper --log-driver=XXX --log-opt
   ↪ syslog-address=tcp://aaa.bbb.ccc.ddd --log-opt YYY
```

Analyse : On l'a vu les containers produisent des logs qu'il est recommandé de centraliser sur une machine distante. Pour ce faire, Docker propose plusieurs *log drivers* qui permettent ce besoin. Dans cet exemple, *Syslog* a été utilisé mais il existe d'autres *log drivers*³.

L'exemple présenté est sommaire mais il montre le principe général. Dans un cas plus pratique, par exemple un ensemble de cibles qui transmettent la température ambiante à un serveur (ou autre application IoT), on configurerait les cibles pour émettre leurs logs sur un serveur distant. Ce serveur disposerait ensuite d'une interface web affichant l'ensemble des logs des cibles et il serait ainsi possible de suivre le fonctionnement de toutes les cibles. De plus, dans un cas pratique, il faudrait aussi veiller à regarder le contenu de l'image rsyslog téléchargée et vérifier qu'elle fasse bien uniquement ce qu'on lui demande et rien de plus ou alors créer une image rsyslog de toutes pièces.

6.1.8 2.13 - Disable operations on legacy registry (v1)

Un *registry* est un service responsable de l'hébergement et de la distribution d'images Docker. C'est une sorte d'App Store pour image Docker. Le *registry* par défaut est Docker Hub⁴. Il est également possible d'utiliser son propre *registry*. Dans tous les cas, il est recommandé de ne plus utiliser la version 1 de l'API de communication avec les *registries* car la version 2 permet notamment le transfert sécurisé de données en utilisant TLS[11].

Pour forcer Docker à n'utiliser que des *registries* v2, il suffit d'ajouter l'option `-disable-legacy-registry` au démarrage du *daemon* Docker.

A faire sur la cible :

```
1 sudo vim /usr/lib/systemd/system/docker.service
```

Ajouter l'option `icc=false` à l'instruction `ExecStart`. Le résultat devrait ressembler à ceci :

```
1 ExecStart=/usr/bin/docker daemon -H fd:// --icc=false
   ↪ --storage-driver=devicemapper --disable-legacy-registry
```

3. Docker log drivers : <https://docs.docker.com/engine/admin/logging/overview/>

4. Docker Hub : <http://hub.docker.com/>

Redémarrage du daemon :

```
1 sudo systemctl daemon-reload # recharge la configuration des services
2 sudo systemctl restart docker # relance le service docker
```

Résultat du benchmark après cette modification : [PASS] 2.13 - Disable operations on legacy registry (v1).

6.1.9 4.1 - Create a user for the container

Par défaut, lorsqu'on lance un container, l'utilisateur à l'intérieur de ce container est root. Pour des raisons évidentes de sécurité, root n'est pas toujours l'utilisateur le plus adapté quand on lance un container. De plus, si par une quelconque raison, on arrive à sortir du container, on se retrouve root sur la machine hôte!

Pour mitiger ces effets, il est possible de :

- Créer un utilisateur avec des droits limités dans l'image du container (USER toto dans le Dockerfile)
- Spécifier l'utilisateur à utiliser lors du lancement du container (docker run -user toto)
- Faire croire au container qu'il est root alors qu'en réalité il s'agit d'un utilisateur à droits limités (user namespace)

La suite a pour but de montrer les différentes techniques pour changer l'utilisateur au sein du container.

État initial

On utilise l'image Alpine sans modifications et on observe les résultats suivants :

```
1 docker run -it alpine /bin/sh
2 / # whoami
3 root
4 / # touch /root.txt
5 / # ls /root.txt
6 /root.txt
7 / # ls -l /root.txt
8 -rw-r--r-- 1 root root 0 May 23 13:41 /root.txt
```

On voit que par défaut l'utilisateur dans le container est root. Cet utilisateur peut donc écrire un fichier à la racine.

Dans le Dockerfile

Pour l'exemple qui suit, on va créer une image qui utilise Alpine Linux comme base. Cette image est modifiée pour qu'elle se lance par défaut avec un utilisateur toto qui a des droits limités.

```
1 mkdir alpine-user
2 vim Dockerfile
```

Dans le Dockerfile, on écrit :

```
1 FROM alpine
2
3 # tâches nécessitant un accès root
4 # par exemple, installer des packages
5 RUN apk update && apk add vim
```

```
6
7  # fin des tâches demandant un accès root
8  # passage à un user à droits limités
9  RUN adduser -h /home/toto -s /bin/sh -D toto
10 USER toto
11
12 CMD ["/bin/sh"]
```

On utilise donc l'image de base d'Alpine sur laquelle on effectue des tâches administratives nécessitant root. Une fois ces tâches effectuées, il n'est plus nécessaire d'être root. On crée alors un utilisateur à droit limité appelé toto. C'est cet utilisateur qui sera utilisé par défaut au lancement d'un container utilisant cette image.

```
1  docker run -it alpine-user
2  / whoami
3  toto
4  / touch /toto.txt
5  touch: /toto.txt: Permission denied
```

Ici, on voit que l'utilisateur en cours est toto et qu'il n'a pas les droits root.

Au lancement du container

Il est également possible de spécifier l'utilisateur à utiliser au lancement du container grâce à l'option `-user`.

Les cas suivants sont testés :

- root : on peut spécifier root si l'image a défini un autre utilisateur par défaut. Utilisation déconseillée.
- nobody : l'utilisateur nobody est un utilisateur aux droits très limités et la plupart des distributions l'intègre de base donc pas besoin le créer expressément.
- gary : cas où l'utilisateur existe sur l'hôte mais pas dans le container
- toto : cas où l'utilisateur n'existe pas sur l'hôte mais dans le container

Utilisateur root :

```
1  > docker run --user root -it alpine-user
2  / # whoami
3  root
4  / # touch /root.txt
5  / # ls -al /root.txt
6  -rw-r--r--    1 root    root          0 May 23 19:35 /root.txt
```

Le container est lancé en root. On peut écrire à la racine sans souci.

Utilisateur root :

```
1  > \ $ docker run --user nobody -it alpine-user
2  ~ \ $ whoami
3  nobody
4  ~ \ $ touch /nobody.txt
5  touch: /nobody.txt: Permission denied
```

Le container est lancé avec l'utilisateur nobody (présent sur l'OS hôte et surtout dans le container). L'écriture sur / est refusée.

Utilisateur gary :

```
1 > docker run --user gary -it alpine-user
2 docker: Error response from daemon: linux spec user: Unable to find user
   ↳ gary.
```

Lancer un container avec un utilisateur qui n'existe pas dans le container n'est pas possible.

Utilisateur toto :

```
1 > \ $ docker run --user toto -it alpine-user
2 / \ $ whoami
3 toto
4 / \ $ touch /toto.txt
5 touch: /toto.txt: Permission denied
```

Le container est lancé avec l'utilisateur toto (absent sur l'OS hôte et mais présent dans le container). L'écriture sur / est refusée.

Analyse : on remarque que par défaut, les containers sont lancés en root. Si le container n'a pas besoin d'avoir des accès root, il est fortement recommandé de construire une image mettant en place un utilisateur à droits limités. Dans le cas où l'on utilise une image écrite par quelqu'un d'autre, on peut essayer de lancer avec l'utilisateur nobody si on estime qu'elle n'a pas besoin des droits root pour fonctionner. Évidemment, cela ne fonctionnera pas si l'image a réellement besoin des droits root comme par exemple pour installer des packages.

Le dernier cas concerne l'utilisation du user namespace. Ici c'est différent, on mappe un utilisateur standard de la machine hôte à l'utilisateur root du container. Ainsi, on peut effectuer des tâches nécessitant les droits root dans le container. Cependant, si le processus au sein du container vient à s'en échapper, il n'est pas root mais l'utilisateur mappé aux droits limités.

En utilisant le user namespace

TODO voir chapitre dédié

```
1 gary@gary-portable ~/alpine-hello-world
2 > \ $ docker run -it alpine /bin/sh
3 / # whoami
4 root
5 / # exit
6
7 gary@gary-portable ~/alpine-hello-world
8 > \ $ docker run --user nobody -it alpine /bin/sh
9 ~ \ $ whoami
10 nobody
11 ~ \ $ whoami
12 ~ \ $ exit
13
14 gary@gary-portable ~/alpine-hello-world
15 ↳ [11:21:48]
16 > \ $ docker run --user gary -it alpine /bin/sh
17 docker: Error response from daemon: linux spec user: Unable to find user
   ↳ gary.
```

6.1.10 4.5 - Enable Content trust for Docker

TODO

6.1.11 5.3 - Restrict Linux Kernel Capabilities within containers

TODO

6.1.12 5.4 - Do not use privileged containers

TODO explain only

6.1.13 5.6 - Do not run ssh within containers

TODO explain only

6.1.14 5.10 - Limit memory usage for container

TODO

6.1.15 5.11 - Set container CPU priority appropriately

TODO

6.1.16 5.12 - Mount container's root filesystem as read only

TODO explain only

6.1.17 Situation finale

TODO

6.2 Séparation des données Docker dans une partition chiffrée

Dans cette section, une partition chiffrée sera créée sur la cible afin d'isoler les données (images et containers principalement) de Docker du reste du système. La partition sera chiffrée afin d'accroître la sécurité des données.

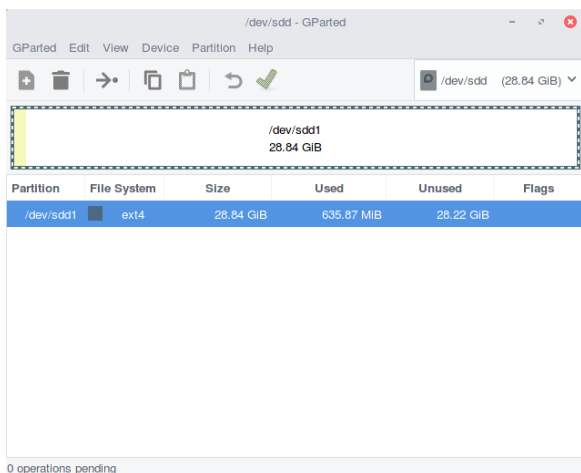
Les étapes suivantes seront réalisées :

1. Redimensionnement de la partition principale
2. Création d'une partition chiffrée
3. Montage manuel de la partition chiffrée
4. Montage au boot de la partition chiffrée
5. Preuve de bon fonctionnement du montage automatique
6. Binding des données de Docker sur la partition chiffrée
7. Analyse de la solution proposée

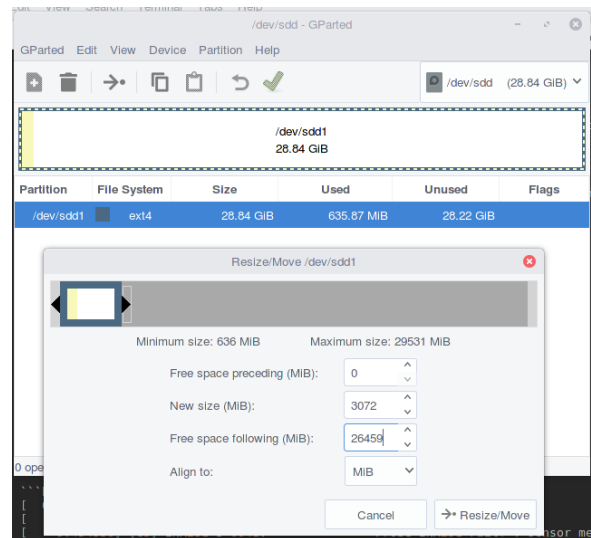
6.2.1 Redimensionnement de la partition principale

Actuellement, il n'y a qu'une partition sur la cible ; la partition principale. Il faut réduire la partition principale et créer une partition dans le nouvel espace libre. Pour ce faire, l'outil GParted ⁵ a été utilisé car il permet de voir visuellement ce qui est réalisé.

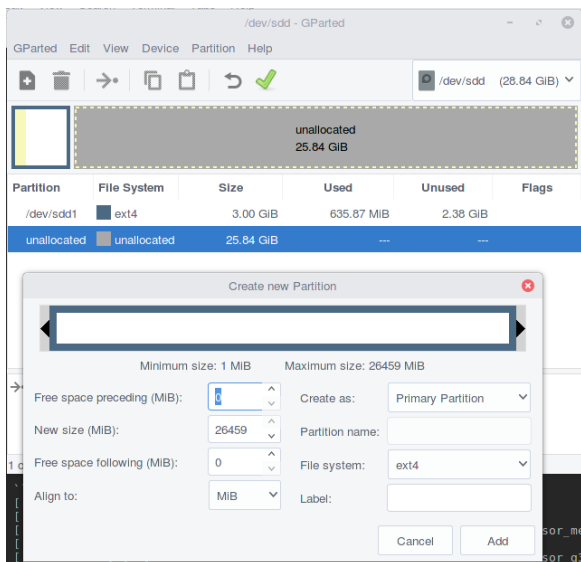
5. GParted : <http://gparted.org/>



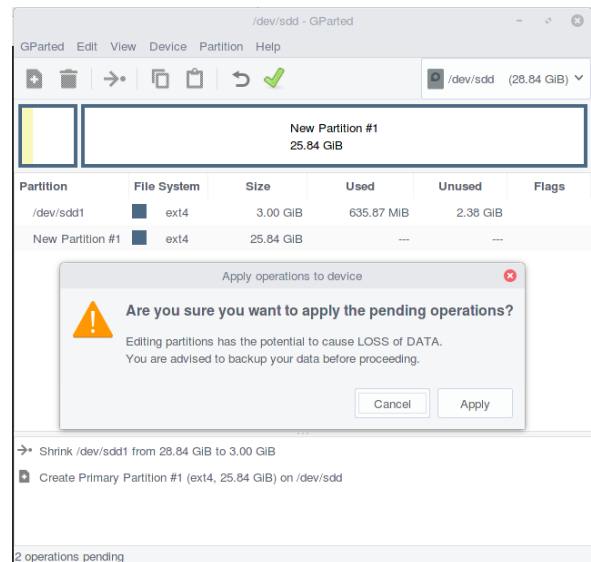
(a) Situation initiale - Une seule partition



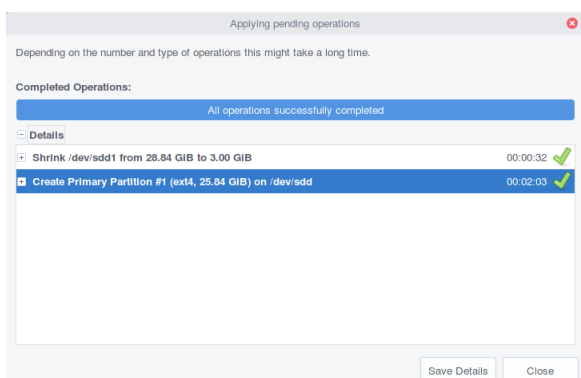
(b) Redimensionnement de la partition principale



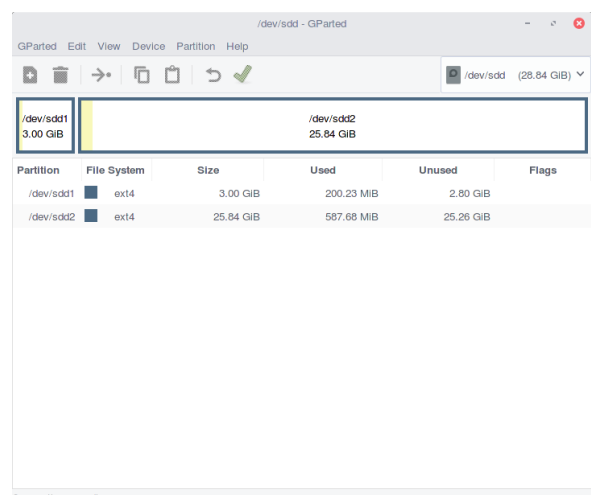
(c) Création de la partition pour les données Docker



(d) Confirmation



(e) Partitionnement en cours...

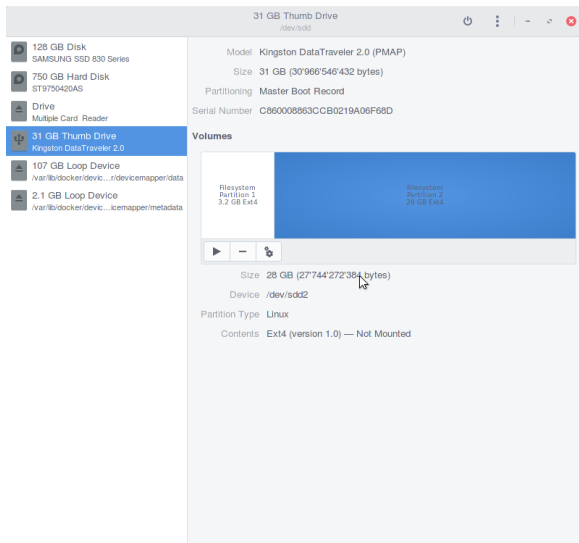


(f) Situation finale - Deux partitions, la principale et celle pour Docker

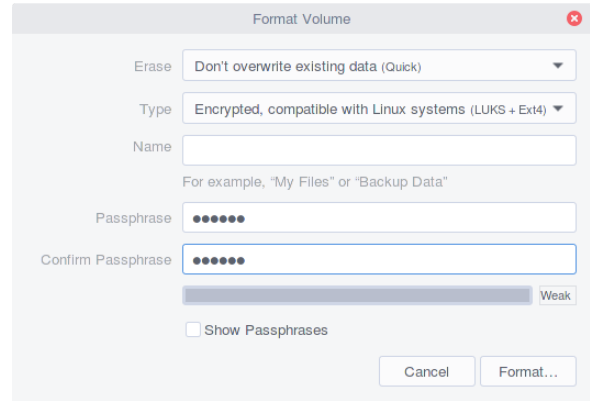
FIGURE 6.3 – Redimensionnement de la partition principale

6.2.2 Création d'une partition chiffrée

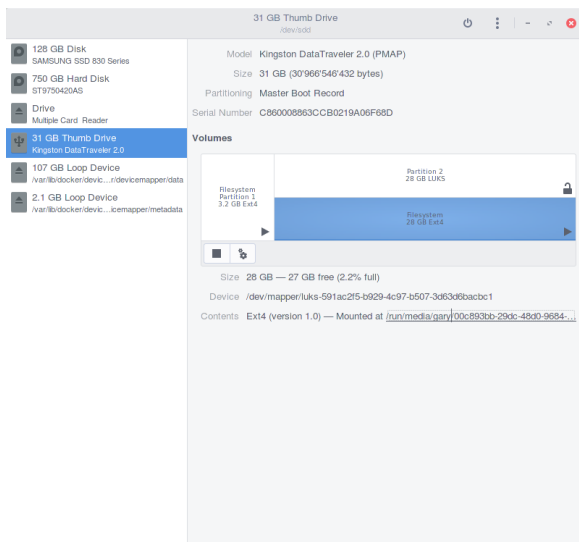
Maintenant que la carte SD correctement partitionnée, il faut créer la partition chiffrée avec LUKS. Pour ce faire, Gnome Disks Utility est utilisé⁶.



(a) Sélection de la partition à chiffrer. Cliquer sur la roue crantée et sélectionner "Format Partition..."



(b) Sélectionner "Encrypted, compatible with Linux systems (LUKS + ext4)" et saisir une passphrase (mot de passe)



(c) La partition chiffrée est créée

FIGURE 6.4 – Création de la partition chiffrée avec Gnome Disk Utility

6.2.3 Montage manuel de la partition chiffrée

Maintenant que la partition chiffrée a été créée, on peut replacer la carte SD dans la cible et la démarrer. Pour monter manuellement la partition, il faut réaliser les commandes suivantes :

```
1 sudo mkdir /mnt/docker-data
2 sudo cryptsetup luksOpen /dev/mmcblk0p2 docker-data
3 sudo mount /dev/mapper/docker-data /mnt/docker-data/
```

6. Gnome Disk Utility : <https://doc.ubuntu-fr.org/gnome-disk-utility>

`/dev/mmcblk0p2` représente la partition chiffrée de la carte SD.

6.2.4 Montage au boot de la partition chiffrée

Monter manuellement la partition n'est pas une solution possible si l'on souhaite y mettre les données de Docker dessus. En effet, pour que Docker stocke ces données (rappel : les images et les containers) dans la partition chiffrée, il faut monter cette partition au démarrage de la cible et créer un *bind mount* sur `/var/lib/docker`.

Avant d'effectuer le *bind mount*, il faut faire en sorte de monter la partition chiffrée au boot.

Pour monter la partition chiffrée au démarrage du système, il faut ouvrir le fichier `/etc/crypttab` sur la cible, et ajouter la ligne suivante :

```
1  docker-data /dev/mmcblk0p2  none luks
```

De manière similaire, il faut ajouter cette ligne dans le fichier `/etc/fstab` :

```
1  /dev/mapper/docker-data /mnt/docker-data ext4 rw 0 0
```

6.2.5 Preuve de bon fonctionnement du montage automatique

Pour vérifier que tout ceci fonctionne bien, il suffit de redémarrer la cible et on devrait observer les lignes suivantes au boot :

```
1  [ OK ] Found device /dev/mmcblk0p2.
2      Starting Cryptography Setup for docker-data...
3  Please enter passphrase for disk docker-data on /mnt/docker-data! *****
4  [*** ] (1 of 2) A start job is running for...for docker-data (13s / no
   ↪ limit)[ 22.871246] [c5] NET: Registered pr8
5  [** ] (1 of 2) A start job is running for...for docker-data (13s / no
   ↪ limit)[ 23.925487] [c6] bio: create slab <1
6  [ OK ] Found device /dev/mapper/docker-data.
7  [ OK ] Started Cryptography Setup for docker-data.
8  [ OK ] Reached target Encrypted Volumes.
9      Mounting /mnt/docker-data...
10 [ 24.117386] [c5] EXT4-fs (dm-0): mounted filesystem with ordered data
   ↪ mode. Opts: (null)
11 [ OK ] Mounted /mnt/docker-data.
```

Le système va se bloquer pour que la saisie de la passphrase soit effectuée :

```
1  Please enter passphrase for disk docker-data on /mnt/docker-data! *****
```

Une fois connecté, il est tout à fait possible de d'écrire des données sur la partition chiffrée :

```
1  echo "LUKS je suis ton père !" > /mnt/docker-data/starwars.txt
```

6.2.6 Binding des données de Docker sur la partition chiffrée

Maintenant, on veut faire en sorte que les données de Docker soient dans cette partition chiffrée. Pour ce faire, on va effectuer un *bind mount* entre un dossier dans la partition chiffrée et le dossier de Docker.

Un *bind mount* n'est rien d'autre qu'un montage de dossier dans un autre dossier. On va donc monter `/mnt/docker-data/docker` dans `/var/lib/docker`. De cette manière, aucune configuration supplémentaire de Docker n'est nécessaire et ceci est donc complètement transparent pour lui. Cette opération est persistante après un redémarrage. Les *bind mount* sont également utilisés comme redirection pour ne pas avoir de points de montage introuvables s'ils ont été déplacés.

Attention : Les opérations suivantes impliquent la perte de toutes les images et containers précédemment créés.

Création du dossier **docker** sur la partition chiffrée, à faire sur la cible :

```
1  sudo systemctl stop docker # on stoppe le service Docker avant tout
2  cd /mnt/docker-data
3  mkdir -p var/lib/docker
4  sudo rm -rf /var/lib/docker/ # supprime toutes les données de Docker !!!
5  sudo mkdir /var/lib/docker
```

Sur la cible, on fait ensuite un *bind mount* entre la partition chiffrée et le dossier Docker de base

```
1  sudo vim /etc/fstab
2
3  # dans ce fichier, on ajoute la ligne suivante
4  /mnt/docker-data/var/lib/docker /var/lib/docker none bind 0 0
```

A la suite de ces modifications, le fichier `/etc/fstab` ressemble à ceci :

```
1  #
2  # /etc/fstab: static file system information
3  #
4  # <file system> <dir> <type> <options> <dump> <pass>
5  /dev/mapper/docker-data /mnt/docker-data ext4 rw 0 0
6  /mnt/docker-data/var/lib/docker /var/lib/docker none bind 0 0
```

Désormais les données de Docker sont stockées sur une partition séparée et chiffrée. Pour prouver que c'est bien le cas, il suffit de relancer le benchmark de sécurité :

```
1  cd /home/alarm/docker/docker-bench-security
2  sudo bash docker-bench-security.sh
3
4  # la sortie indique:
5  [PASS] 1.1 - Create a separate partition for containers
```

6.2.7 Analyse de la solution proposée

Stocker les données de Docker dans une partition chiffrée est une bonne initiative mais il faut garder à l'esprit les implications suivantes :

- La partition chiffrée est montée au boot. Ceci implique quand cas de reboot de la machine, il faut avoir un accès physique à celle-ci car le service ssh n'est pas encore lancé à ce moment là. En cas de dépannage à distance, il n'est pas donc possible de reboot la cible.
- Les performances quand on utilise Docker devraient être diminuées dû au chiffrement
- L'utilisation d'une partition séparée permet de réinstaller le système sans perdre les données liées à Docker et ses containers.
- Le chiffrement des containers Docker apporte une plus value en terme de sécurité

6.3 Conclusion sur la configuration du système d'exploitation hôte

TODO dire ce qui s'est passé et ce qui serait bien de faire en plus.

7. Création et utilisation des images Docker

TODO : ce chapitre montre comment et pourquoi créer et utiliser des images Docker qui soient sécurisées et adaptées au monde de l'embarqué (taille principalement). Utilisation des hash cryptographiques pour l'OS utilisé et les versions des logiciels installés. Montrer un cas pratique avec une image commune basée sur alpine linux dire ce que ca apporte de travailler ainsi et pourquoi il faut le faire.

8. Utilisation des containers

TODO : parler de la philosophie des containers légers et éphémères, comment les exécuter de manière sûre et sécurisée, comment limiter le pouvoir des containers (conso ram, ddos, communication avec le monde extérieur ou entre containers, ...)

9. Déroulement du projet

9.1 Planning initial

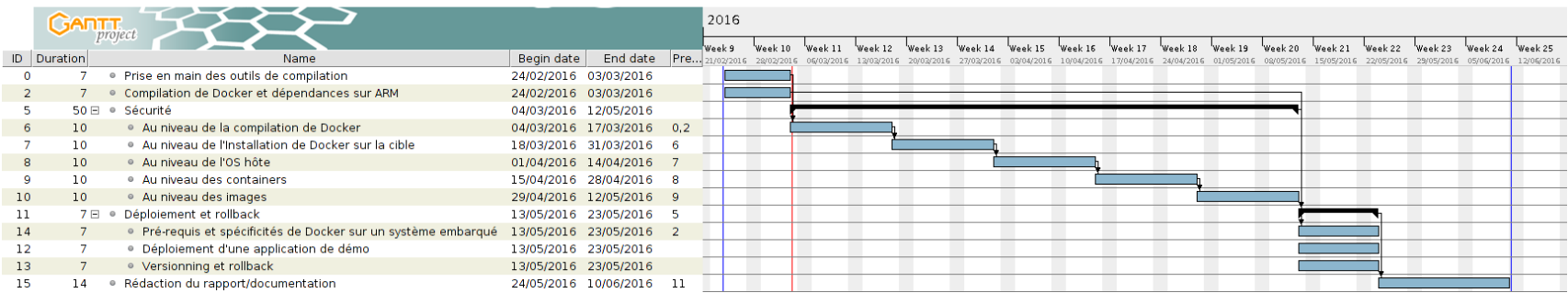


FIGURE 9.1 – Planning initial - 04.03.2016

9.2 Planning final

img/initial_planning.png

FIGURE 9.2 – Planning final - TODO date

10. Proposition d'améliorations vis à vis du travail précédent

TODO : Pas prioritaire, à voir si on a le temps ! passer en revue et critique positivement le travail de Bachelor précédent. Dire que ce n'est pas une critique négative mais apporter un avis supplémentaire et plus récent (Docker évoluant beaucoup)

11. Conclusion

TODO... ce qu'on voulait faire, ce qui a été fait, ce qu'il reste à faire/pourrait faire

Bibliographie

- [1] 10.04 - Mount a LUKS partition at boot - Ask Ubuntu, mai 2016. <http://askubuntu.com/questions/21025/mount-a-luks-partition-at-boot>.
- [2] Cgroups, mai 2016. <https://fr.wikipedia.org/wiki/Cgroups>.
- [3] Creating containers - Part 1, mai 2016. <http://crosbymichael.com/creating-containers-part-1.html>.
- [4] Docker 1.10 et les user namespace, mai 2016. <http://linuxfr.org/users/w3blogfr/journaux/docker-1-10-et-les-user-namespace>.
- [5] Docker (software), mai 2016. https://en.wikipedia.org/wiki/Docker_%28software%29.
- [6] Introduction to Linux Control Groups (Cgroups), mai 2016. <https://sysadmincasts.com/episodes/14-introduction-to-linux-control-groups-cgroups>.
- [7] ODROID-XU3 | Arch Linux ARM, avril 2016. <https://archlinuxarm.org/platforms/armv7/samsung/odroid-xu3>.
- [8] ODROID-XU3 Lite | Hardkernel, mai 2016. http://www.hardkernel.com/main/products/prdt_info.php?g_code=G141351880955.
- [9] Separation Anxiety : A Tutorial for Isolating Your System with Linux Namespaces, mai 2016. <https://www.toptal.com/linux/separation-anxiety-isolating-your-system-with-linux-namespaces>.
- [10] Syslog, mai 2016. <https://fr.wikipedia.org/wiki/Syslog>.
- [11] Ben Firshman. Faster and Better Image Distribution with Registry 2.0 and Engine 1.6, mai 2016. <https://blog.docker.com/2015/04/faster-and-better-image-distribution-with-registry-2-0-and-engine-1-6/>.
- [12] Brand Pending. Shipping logs to rsyslog from a Docker container using the syslog Logging Driver, mai 2016. <http://www.brandpending.com/blog/2015/11/9/shipping-logs-to-rsyslog-from-a-docker-container-using-the-syslog-logging-driver>.
- [13] Colin Barschel. Encrypt Partitions, mai 2016. <http://sleepyhead.de/howto/?href=cryptpart>.
- [14] Docker Inc. API Documentation for Docker, mai 2016. https://docs.docker.com/engine/reference/api/docker_remote_api/.
- [15] Docker Inc. Docker Documentation, mai 2016. <https://docs.docker.com/>.
- [16] Docker Inc. Select a storage driver, mai 2016. <https://docs.docker.com/engine/userguide/storagedriver/selectadriver/>.
- [17] Docker Inc. The daemon command description and usage, mai 2016. <https://docs.docker.com/engine/reference/commandline/daemon/>.
- [18] Center for Internet Security. CIS Docker 1.11.0 Benchmark, avril 2016. https://benchmarks.cisecurity.org/tools2/docker/cis_docker_1.11.0_benchmark_v1.0.0.pdf.
- [19] Docker Inc. Understand images, containers, and storage drivers, mai 2016. <https://docs.docker.com/engine/userguide/storagedriver/imagesandcontainers/>.
- [20] Docker Inc. What is Docker ?, mai 2016. <https://www.docker.com/what-docker>.
- [21] Jaroslav. Docker 1.10, mai 2016. <http://www.jrslv.com/docker-1-10/>.
- [22] Jean-Roland Schuler. Filesystems - Secured Embedded Systems, mai 2016. Cours HES-SO Master.
- [23] Jean-Tiare Le Bigot. Docker for your users - Introducing user namespace, mai 2016. <https://blog.yadutaf.fr/2016/04/14/docker-for-your-users-introducing-user-namespace/>.
- [24] Jérôme Petazzoni. Deep dive into Docker storage drivers, mai 2016. <https://jpetazzo.github.io/assets/2015-03-03-not-so-deep-dive-into-docker-storage-drivers.html>.

- [25] Loïc Bassang. Docker and Internet of Things. Technical report, Haute École d'ingénierie et d'architecture de Fribourg, 2015.
- [26] Loïc Bassang. Docker and Internet of Things - Travail de Bachelor. Technical report, Haute École d'ingénierie et d'architecture de Fribourg, 2015.
- [27] Adrian Mouat. *Using Docker*. O'Reilly Media, 2005.
- [28] Katherine Noyes. Docker : A 'Shipping Container' for Linux Code, mai 2016. <https://www.linux.com/news/docker-shipping-container-linux-code>.
- [29] ram-0000. Présentation du protocole Syslog, mai 2016. <http://ram-0000.developpez.com/tutoriels/reseau/Syslog/>.
- [30] Sehrope Sarkuni. Encrypting Docker containers on a Virtual Server, mai 2016. <https://launchbylunch.com/posts/2014/Jan/13/encrypting-docker-on-digitalocean/>.
- [31] Prakhar Srivastav. Docker for beginners, avril 2016. <http://prakhar.me/docker-curriculum/>.
- [32] Viktor Petersson. Manage Docker resources with Cgroups, mai 2016. <https://www.cloudsigma.com/manage-docker-resources-with-cgroups/>.

Appendices

A. Installation de Archlinux ARM sur ODROID-XU3 Lite

Remarque : Ce guide requiert l'utilisation d'un ordinateur sous GNU/Linux.

Source : <https://archlinuxarm.org/platforms/armv7/samsung/odroid-xu3>

A.1 Micro SD Card Creation

Replace sdX in the following instructions with the device name for the SD card as it appears on your computer.

1. Zero the beginning of the SD card :

```
1 dd if=/dev/zero of=/dev/sdX bs=1M count=8
```

2. Start fdisk to partition the SD card :

```
1 fdisk /dev/sdX
```

3. At the fdisk prompt, create the new partitions :

- a. Type o. This will clear out any partitions on the drive.
- b. Type p to list partitions. There should be no partitions left.
- c. Type n, then p for primary, 1 for the first partition on the drive, and enter twice to accept the default starting and ending sectors.
- d. Write the partition table and exit by typing w.

4. Create and mount the ext4 filesystem :

```
1 mkfs.ext4 /dev/sdX1
2 mkdir root
3 mount /dev/sdX1 root
```

5. Download and extract the root filesystem (as root, not via sudo) :

```
1 wget
  ↳ http://os.archlinuxarm.org/os/ArchLinuxARM-odroid-xu3-latest.tar.gz
2 bsdtar -xpf ArchLinuxARM-odroid-xu3-latest.tar.gz -C root
```

6. Flash the bootloader files :

```
1 cd root/boot
2 sh sd_fusing.sh /dev/sdX
3 cd ../../
```

7. (Optional) Set the MAC address for the onboard ethernet controller :

- a. Open the file root/boot/boot.ini with a text editor.
- b. Change the MAC address being set by the setenv macaddr command to the desired address.

- c. Save and close the file.
8. Unmount the partition :
`umount root`
9. Set the boot switches on the ODROID-XU3 board to boot from SD :
 - a. With the board oriented so you can read the ODROID-XU3 on the silkscreen, locate the two tiny switches to the left of the ethernet jack.
 - b. The first switch (left) should be in the off position, which is down.
 - c. The second switch (right) should be in the on position, which is up.
10. Insert the micro SD card into the XU3, connect ethernet, and apply 5V power.
11. Use the serial console (with a null-modem adapter if needed) or SSH to the IP address given to the board by your router.
 - Login as the default user alarm with the password alarm.
 - The default root password is root.

A.2 eMMC Module Creation

1. Attach the eMMC module to the micro SD adapter, and plug that into your computer.
2. Follow the above steps to install Arch Linux ARM, and boot the board with the eMMC still attached to micro SD adapter, plugged into the SD slot in the board.
3. Re-flash the bootloader to the protected boot area of the eMMC module :

```
1  cd /boot
2  ./sd_fusing.sh /dev/mmcblk0
```

4. Power off the board :

```
1  poweroff
```

5. Remove the micro SD adapter, and detach the eMMC module.
6. Set the boot switches on the ODROID-XU3 board to boot from eMMC :
 - a. With the board oriented so you can read the ODROID-XU3 on the silkscreen, locate the two tiny switches to the left of the ethernet jack.
 - b. The first switch (left) should be in the on position, which is up.
 - c. The second switch (right) should be in the on position, which is up.
7. Connect the eMMC module to the XU3, ensuring you hear a click when doing so, connect ethernet, and apply 5V power.
8. Use the serial console (with a null-modem adapter if needed) or SSH to the IP address given to the board by your router.
 - Login as the default user alarm with the password alarm.
 - The default root password is root.