Rapport de stage Ingénieur

Implémentation d'un ordonnanceur temps réel sur plateforme multi-cœur hétérogène

BELPOIS Vincent 2023





Table des matières

Table des figures			
1	Intr	$\mathbf{roduction}$	3
2	os	compatibles avec la carte	4
	2.1	Présentation de la carte de développement	4
	2.2	Installation d'un système d'exploitation	4
		2.2.1 Installation d'une image pré-compilée	5
		2.2.2 Compilation de Linux depuis le code source	Ę
		2.2.3 Compilation croisée	7
	2.3	Etude des versions de Linux compatibles	8
		2.3.1 Comment Linux gère le suport d'un processeur	8
		2.3.2 Essais de différentes versions	8
3	LIT	$^{ m rMUS^{RT}}$	10
	3.1	Présentation de LITMUS ^{RT}	10
	3.2	Présentation de feather-trace	10
	3.3	Implémentation d'un ordonnanceur EDF partitionné	11
		3.3.1 Algorithme considéré	11
		3.3.2 Implémentation	11
	3.4	Implémentation d'un ordonnanceur RM partitionné	12
C	onclu	ısion	13
Annexe			1 4
Bi	bilio	graphie	2 4
Glossaire			24





1 Introduction

J'ai pu réaliser mon stage ingénieur, durant ma deuxième année d'étude à l'ISAE-ENSMA, au laboratoire du LIAS de Chasseneuil-du-Poitou. Le LIAS, ou Laboratoire d'Informatique et d'Automatique pour les Systèmes, regroupe plusieurs dizaines d'enseignants chercheurs dans les domaines de l'automatique, le génie électrique et l'informatique. Le site de Chasseneuil-du-Poitou regroupe deux équipes, l'équipe Ingénierie des Données et des moDèles (IDD) et l'équipe Systèmes Embarqués Temps Réel (SETR).

J'ai été accueilli au sein de cette dernière afin de travailler avec Antoine BERTOUT et Thomas GASPARD sur le projet SHRIMP. Mon stage s'intéresse à l'implémentation d'un ordonnanceur sur plateforme hétérogène[1], tandis que le reste du projet s'intéresse entre autre à la conception d'un ordonnanceur temps réel global et dynamique pour des plateformes unrelated.

Mon stage aura donc pour objectif d'identifier une solution permettant de programmer un ordonnanceur temps réel sur une carte de développement particulière. Il me faudra ensuite étudier les mécanismes les plus adaptés pour la migration tâches sur cette plateforme. L'objectif final sera alors d'implémenter une politique d'ordonnancement temps réel globale hétérogène naïve, voir plus évoluée en fonction de l'avancement des travaux de thèse.





2 Systèmes d'exploitation compatibles avec la carte ROCK960

Affin d'étudier les systèmes d'exploitations temps réel compatibles avec la carte de développement qui m'a été fourni, je me suis d'abord familiariser avec celle-ci en installant des OS fournis par le fabriquant avant d'étudier la compatibilité avec des systèmes plus complexes.

2.1 Présentation de la carte de développement

Le stage s'intéressant à l'implémentation d'un algorithme d'ordonnacement sur une plateforme hétérogène, une carte possédant un tel processeur est mis à ma disposition. Cette carte se nomme ROCK960 et est fabriquée par l'entreprise 96Boards. Cette carte de développement contient de nombreuses interfaces mais nous nous contenteront d'utiliser l'interface Série TTL à laquelle nous nous connecterons via un convertisseur USB vers TTL. Cela me permettra d'interfacer via un terminal qui fonctionnera avec une liaison série.

Au centre de la carte est un SOC Rockchip RK3399. Ce processeur contient deux type de cœurs, ou processeurs. Deux d'entre eux sont des processeurs Cortex-A72 et les quatre autres sont des processeurs Cortex-A53. Ces 6 processeurs utilisent le même jeu d'instruction : ARMv8-A 64-bit. Cela sera important par la suite afin de faciliter la migration de tache entre les processeurs, en effet si les jeux d'instructions des processeurs étaient différents, plusieurs copies du code compilé devrait exister tout en maintenant un lien d'équivalence entre les deux codes. Cela est bien au delà de la portée de mon stage mais sera un point intéressant à explorer.

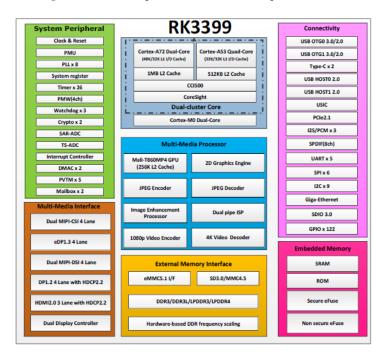


FIGURE 1 – Architecture du processeur RK3399

Le SOC RK3399 contient bien d'autres composants et peut interfacer avec de nombreux périphériques (écran HDMI, USB, caméra MPI-CSI, SPI, UART, I2C, etc...) comme le montre la figure 1. Ce diagramme nous montre aussi que les deux cluster de processeurs ne partagent pas les cache L1 ni L2 mais sont interconnectés par une interface CCI-500 qui, selon le site des développeurs ARM, permet la cohérence des caches des deux clusters.

2.2 Installation d'un système d'exploitation

Le code que l'on veut exécuter sur la carte de développement doit être compilé pour celle-ci puis placé sur un support de stockage. Ce code peut être un simple programme ou bien un système d'exploitation complet. J'ai premièrement installé des images pré-compilées de Linux, puis j'ai compilé moi-même le noyau Linux depuis son code source en réalisant une compilation croisée.





2.2.1 Installation d'une image pré-compilée

Pour premier tester l'Installation de linux sur la carte de développement, j'ai utilisé une image de la distribution Ubuntu fournie par le fabriquant 96Boards disponible sur leur site. Cette image se présente sous la forme d'une archive au format .tar.gz. Elle contient à la fois le bootloader, le noyau Linux, et le système de fichier. Cette image (system.img) peut alors être gravée (ou flashée) sur une carte micro SD.

Depuis un terminal, en se déplaçant dans le dossier de l'archive extraite, on exécute la commande suivante :

\$ sudo dd if=system.img of=/dev/XXX bs=4M oflag=sync status=noxfer

Listing 1 – Linux Command

EXPLIQUER CE QUE FAIT CETTE COMMANDE

Aussi dire en quoi on s'en servira dans des scripts afin d'accélérer le développement.

2.2.2 Compilation de Linux depuis le code source

Afin d'utiliser une version de Linux différente de la version précompilé par le fabriquant de la carte de développement, il faut se premièrement se procurer le code source du noyau Linux. Celui-ci est disponible sur un dépôt de code git hébergé par GitHub. Il est disponible a l'adresse https://github.com/torvalds/linux, sous le profile du créateur de Linux : Linus Torvalds. Durant mon stage j'étais libre d'utiliser le logiciel de gestion de version de mon choix, j'ai donc principalement utilisé git en ligne de commande et j'ai parfois utilisé un client git nommé GitKraken afin plus facilement explorer les anciens commits de certains projets comme LITMUS^{RT}.

Une fois le code source du noyau téléchargé, et en se déplaçant dans le dossier linux depuis un terminal, on peu alors procédé a la compilation. Pour mes premiers essais j'ai premièrement décidé de compiler Linux pour une machine virtuelle que je ferai tourner sur ma machine de travail.

Le noyau Linux est un programme ayant une compilation basée sur la configuration : cela signifie que certaines parties du code peuvent rajouter ou omises par le simple biais d'une fichier de configuration. Ce configuration se présente sous la forme d'un fichier .config qui doit être créé à la racine du noyau. Pour créer ce fichier, des utilitaires sont mise a notre disposition dans le noyau :

- make defconfig: cet outil est utilisé pour générer une configuration pas défaut. Ici l'architecture de la machine qui réalise la compilation sera sélectionnée. Dans mon cas, éxécuter cette commande créer un fichier de configuration basé sur la config 'x86_64_defconfig'.
- make menuconfig : cet outil permet d'éditer la configuration actuelle du fichier .config via une interface graphique. Ce menu permet aussi de rechercher des paramètres, de voir leur description et d'enregistrer différentes configurations.





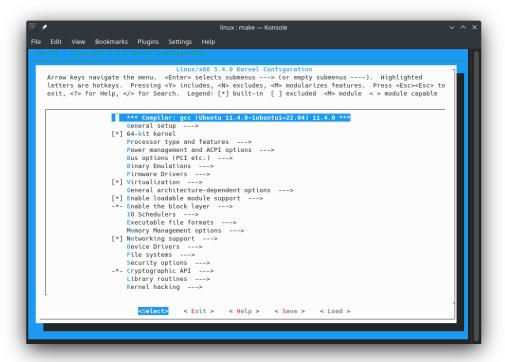


Figure 2 – Interface de configuration du noyau

EXPLIQUER ou est exactement stockée cette config expliquer que l'on peut stocker ce fichier manuellement.

Une fois la configuration crée, nous pouvons passer à la compilation du noyau. Linux utilise l'utilitaire de compilation *GNU Make*: il permet l'automatisation de la compilation, la gestion des dépendances et gère la personnalisation de la compilation de chaque dossier. Ces règles sont alors dictés par des fichiers MakeFile présent dans chaque dossier contenant des fichiers à compiler du projet.

Note : chaque distribution Linux possède un ensemble différent de programmes préinstallés, il faudra alors peut-être installer des programmes nécessaires a la compilation. Par exemple, installer libelf-dev "une bibliothèque partagée qui permet de lire et écrire des fichiers ELF à un niveau élevé" ¹.

```
$ make -j16
```

Listing 2 – Compilation sur plusieurs processeurs

Le paramètre -j 16 signifie que l'on veux exécuter la compilation avec 16 tâches en parallèles. Il est recommandé sur internet par beaucoup d'utiliser comme nombre de tâches, le double du nombre de processeurs dans l'ordinateur qui réalise la compilation.

Par la suite il me sera parfois nécessaires de changer la version de Linux que je compile afin de tester si celle si fonctionne. Sur le dépôt de code de linux, les différentes versions sont stockés sous forme d'un certain commit qui à été "tagé" afin de le retrouver. On peut alors changer de version en revenant à ce commit grâce à la commande checkout de git :

```
$ git checkout v5.4
```

Listing 3 – Retour sur un commit tagé

On peut avoir la list de ces commits tagés de la manière suivante :

```
$ git tag -l
```

Listing 4 – Comment lister les tags





 $^{1.\} d'après\ la\ description\ sur\ packages. debian.org/fr/sid/libelf-devent approximation and approximation of the development of the property of the development of the development$

On aura alors l'ensemble des tags de tout le dépôt de code, et on peut filtrer ces résultats avec grep par exemple si l'on veut retrouver une version particulière.

Bien que j'étais déjà familier avec git, cela m'a pris un certain temps de comprendre comment ce changement de version s'effectuait. La nuance que l'on ne changeait pas de branche dans le dépôt, mais que l'on revenait simplement au commit correspondant à la version étais était la plus compliquée a comprendre. Tout au long de ce stage j'ai pu utiliser git afin d'explorer comment certains projet ont été construits en remontant l'historique de leurs commits, mais j'ai pu aussi utilisé git pour géré le stockage du code que j'ai développé, que ce soit des outils ou des modifications du noyau.

2.2.3 Compilation croisée

Nous ne pouvons malheureusement pas compiler directement le noyau linux pour la carte de développement dû à la différence de jeu d'instruction. Il faut donc réaliser une *cross compilation* ou compilation croisée : c'est le fait de compiler un programme sur une architecture qui n'est pas celle de l'architecture cible.

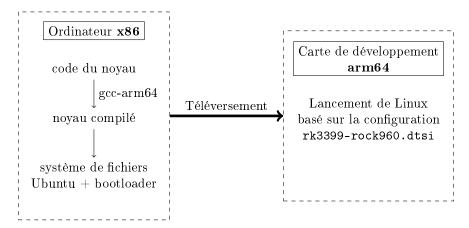


Figure 3 – Compilation croisée du noyau Linux

Dans notre cas, l'ordinateur qui réalisera la compilation a un jeu d'instruction x86 tandis que la carte de développement a un jeu d'instruction arm64. Lors de la compilation du noyau Linux, on doit alors fixer des variables d'environnement comme l'architecture cible et le chemin vers la toolchain à utiliser :

```
ARCH="arm64"
CROSS_COMPILE="../toolchain/bin/aarch64-linux-gnu-"
```

Listing 5 – Variables pour la compilation croisée du noyau Linux

Ces variables seront lues par le fichier Makefile principal du noyau. Par exemple, le compilateur qui sera utilisé est aarch64-linux-gnu-gcc et fait partie de la toolchain que j'ai pu me procurer sur le site de Linaro². Cette toolchain contient entre autre un compilateur qui s'exécutera sur une machine x86 et compilera vers une architecture arm64. Cette toolchain contient aussi des bibliothèques propres a l'architecture cible, comme arm_neon.h par exemple (librairie qui permet d'utiliser les unités de calcul NEON sous les architecture ARM compatibles qui permettent le calcul en parallèle grâce aux instructions SIMD).

Le noyau linux est ainsi compilé par la commande :

```
make Image dtbs -j16
```

Listing 6 – Linux Command

Cela produit entre autre un fichier Image dans le dossier litmus-rt/arch/arm64/boot. Ce fichier est le noyau Linux compilé, non compressé, contenant les modules compilés en tant que built-in-modules mais ne contenant pas les modules externes. Ces derniers peuvent être compilés séparément et installés par une simple copie sur le système de notre choix.

 $^{2.\} releases. linaro.org/components/toolchain/binaries/latest-7/aarch64-linux-gnu/gcc-*-x86_64_aarch64-linux-gnu.tar.xz$





Une fois le noyau étant compilé, on peut le flasher sur la carte microSD que lira la carte de développement à son démarrage. Cependant, il faut joindre ce noyau a plusieurs autres programmes afin d'avoir un système d'exploitation utilisable :

- un bootloader, dans notre cas u-boot
- un système de fichiers et des programmes utilitaires, dans mon cas j'ai choisi une image minimale de Ubuntu³

Le bootloader est un court programme qui est chargé d'amorcer le système d'exploitation principal. Il est stocké dans une mémoire non volatile et est exécuté au démarrage de la carte de développement. Il est alors chargé de charger le noyau Linux et de lui passer la main. Dans notre cas, le fabriquant de la carte de développement fourni un bootloader nommé u-boot qui est déjà compilé et qui est disponible sur leur dépôt de code. Il est alors possible de le télécharger et de le flasher sur la carte microSD.

Toolchain : qu'est ce que c'est, de quoi elle est constituée? Expliquer que l'on compile sur du x86 mais qu'on veut compiler pour du ARMv8xxx.

Variables d'environement? Qu'est ce que c'est sous linux, comparer a des

Réalisation de scripts linux pour accélérer le développement. Que doit on charger pour charger le nouveau code compilé?

Copie de l'image du noyau pour faire encore plus rapide.

2.3 Etude des versions de Linux compatibles

2.3.1 Comment Linux gère le suport d'un processeur

Pour gérer la compatibilité avec un processeur, le bootloader charge au démarrage de Linux le fichier Device Tree qui contient les informations sur le matériel présent sur la carte. Ce fichier est ensuite utilisé par le noyau Linux pour initialiser le matériel. Dans notre cas, le fichier rock960-rk3399.dts charge le fichier rk3399.dtsi qui contient les informations sur le processeur. On peut y trouver les informations sur la connectique, les périphériques, les contrôleurs, les bus, etc. La partie nous intéressant est celle sur la structure des processeurs qui se trouve dans le fichier rk3399.dtsi. On y trouve les informations sur les différents cœurs du processeur, leur fréquence, leur cache. C'est vers la fin de mon stage que j'ai pu me rendre compte d'un oubli dans le fichier décrivant ce processeur : les caches ne sont pas décrits. Cela a des conséquences sur les algorithmes d'ordonnancement utilisant cette information pour concevoir les clusters de processeurs les plus adaptés. En effet, les caches sont utilisés pour déterminer les coûts de migration d'une tâche d'un cœur à un autre. Sans cette information, les algorithmes d'ordonnancement ne peuvent pas déterminer les coûts de migration et ne peuvent donc pas déterminer les clusters les plus adaptés.

Dans le listing 10 de l'annexe, on peut voir les modifications que j'ai apporté au fichier rk3399.dtsi pour ajouter les informations sur les caches. Je me suis appuyés sur les informations du datasheet du processeur pour ajouter ces informations. La manière d'ajouter ces informations n'étais pas documentée mais j'ai pu trouver des exemples d'autres processeurs pour m'aider à ajouter ces informations.

Ces modifications ne sont toujours pas présentes dans la version actuelle du noyau Linux, il y a alors ici une possibilité de soumettre une pull request pour ajouter ces informations au noyau Linux. Je me suis renseigné sur la procédure à suivre pour soumettre une pull request au noyau Linux et j'ai pu trouver un guide[4] expliquant la procédure à suivre. Cependant je n'ai pas eu de réponses au patch que j'ai soumis 4 .

De plus, ces informations ne sont pas indispensables pour le fonctionnement de Linux sur ce processeur et ne sont donc pas une priorité pour les développeurs du noyau Linux. Il est donc possible que cette *pull request* ne soit pas acceptée : j'ai pu trouver une pull request similaire datant de plusieurs année qui a mis du temsp à être acceptée ⁵ alors qu'elle s'intéresse à un processeur plus répandu que le RK3399.

2.3.2 Essais de différentes versions

Le problème principal que j'ai rencontré est que le noyau LITMUS^{RT}, dont plus de détails sont donné dans la partie 3, n'est pas compatible avec les versions récentes du noyau Linux. En effet, la dernière version officielle du noyau LITMUS^{RT} est basé sur la version 4.9.30 du noyau Linux et





 $^{3.\} ubuntu_server_16.04_arm64_rootfs_20171108.ext4$

 $^{4. \ \} Patch \ \ \underline{envoy} \\ \hat{a} : < \underline{linux-rockchip@lists.infradead.org} > \ \underline{et} \\ < \underline{linux-arm-kernel@lists.infradead.org} > \ \underline{et} \\ < \underline{linux-arm$

date de Mai 2017. Cependant, la carte de développement ROCK960 n'est supportée que depuis la version 6.6 du noyau Linux comme en témoigne un commit de septembre 2018 ⁶. Il est donc nécessaire de trouver une version du noyau Linux compatible avec la carte de développement et avec le noyau LITMUS^{RT}.

Comme recommandé par Antoine Bertout, les mailling lists de LITMUS^{RT}m'ont permis de trouver d'autres chercheurs ayant patché des versions plus récentes de Linux. Après plusieurs essaais, je me suis arrêté sur la version 5.4 du noyau Linux pour laquelle Joshua Bakita à fait le travail de patcher linux pour le rendre compatible avec LITMUS^{RT}. J'ai pu trouver son travail sur son $github^7$.

J'ai moi même essayer de réaliser ce travail pour une autre version de Linux, celle fournie par le fabriquant de la carte de développement. En effet, je n'ai jamais pu faire fonctionner le WiFi de la carte avec une autre version de Linux que celle faite par 96Boards, c'est pourquoi j'ai tenter de patcher cette version de Linux avec les commits nécessaires pour faire fonctionner LITMUS^{RT}. Cependant j'ai rencontré de nombreux problèmes, à la fois lors de la compilation, et lors du fonctionnement de ma version patché de Linux. C'est donc pourquoi j'ai arrêté ce développement pour me concentrer sur la version de Joshua Bakita. Ma compréhension du noyau Linux n'étais pas suffisante pour comprendre les problèmes que je rencontrais et j'ai donc préféré me concentrer sur la version du noyau patché qui fonctionnait.

Je n'ai pas testé le bon fonctionnement de toutes les fonctionnalités de la carte sous cette version de Linux (HDMI, USB, GPU, etc.) car je n'en avait pas besoin pour mon développement et que je n'avais pas le temps de tester toutes les fonctionnalités de la carte. Dans cette version le WiFi ne fonctionne pas, cela aurait été utile afin d'accélérer le développement lors de mon stage : téléverser les fichiers sur la carte de développement aurait été plus rapide. Cependant, le WiFi n'est pas indispensable pour le fonctionnement de la carte et je n'ai donc continuer ainsi.

Cette étape d'essais de versions et de drivers pour le WiFi étais fastidieuse et malgré l'existence d'outils pour accélérer cela comme Yocto ou Buildroot, je n'ai pas eu le temps de les utiliser. En effet, ces outils permettent de compiler un système d'exploitation Linux complet pour une carte de développement donnée. Cependant, il faut alors configurer ces outils pour qu'ils utilisent les bons drivers et les bonnes versions de Linux. Cela aurait donc nécessité de comprendre comment ces outils fonctionnent et comment les configurer pour qu'ils utilisent les bons drivers et les bonnes versions de Linux. Cela aurait été utile pour accélérer le développement mais j'ai préféré me concentrer sur le développement de l'ordonnanceur plutôt que sur la configuration de ces outils.

^{7.} Depôt de code de linux-5.4-lit mus





3 LITMUS^{RT}

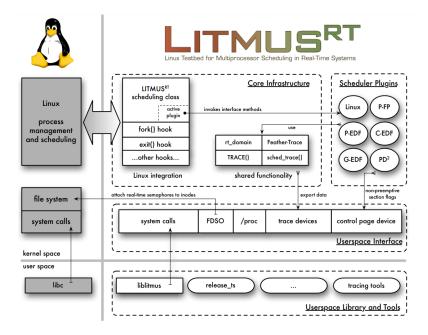


Figure 4 – Architecture de LITMUS RT

LITMUS^{RT}est un patch au noyau Linux constitué de quatre parties :

- LITMUS^{RT} Core : le patch au noyau Linux qui permet d'ajouter les fonctionnalités temps réel à Linux
- plusieurs ordonnanceur
- une interface dans l'espace utilisateur
- des outils dans l'espace utilisateur tel que feather-trace et liblitmus

3.1 Présentation de LITMUS^{RT}

LITMUS^{RT}, qui signifie Linux Testbed for Multiprocessor Scheduling in Real-Time Systems est un moyen de développer des applications temps réel sur le noyau Linux. Il contient des modifications au noyau habituel de Linux, des interfaces utilisateurs permettant d'interagir a bas niveau avec l'ordonnancement des taches sous Linux, ainsi qu'une infrastructure de traçage de l'exécution de l'ordonnanceur. LITMUS^{RT}à été développé par Björn B Brandenburg [3] afin de faciliter la recherche et la comparaison des algorithmes d'ordonnancement. Actuellement, beaucoup de publications utilisent LITMUS^{RT}afin de comparer différents protocoles de gestion de resources partagées par plusieurs processeurs. Mais LITMUS^{RT}est aussi utilisé pour sa facilité à être implémenté sur des plateformes récente dù au fait qu'il est construit par dessus le noyau Linux et que ce dernier est le système d'exploitation qui supporte le plus de plateformes.

Ce dernier point est principalement pourquoi nous avons choisis LITMUS^{RT}comme système d'exploitation sur lequel nous implémenterons des algorithmes d'ordonnancement pour la carte de développement ROCK960. Les autres candidats, comme FreeRTOS, étaient souvent dirigés vers les microcontrôleurs ou bien n'étaient simplement pas compatibles avec la carte de développement.

3.2 Présentation de feather-trace

Feather-trace [2] est outil de suivi d'événements léger conçu pour être intégré dans des applications, systèmes d'exploitation ou systèmes embarqués. Il est dans notre cas, à la fois intégré dans le noyau modifié LITMUS^{RT}, mais aussi dans les algorithmes d'ordonnancement que nous implémenterons. Il a été choisi pour sa simplicité et sa légèreté. Il permet d'enregistrer sous forme de fichier de log de multiples données de l'ordonnancement, par exemple l'arrivée d'une nouvelle tache, le début d'un nouveau job de cette tache, la date de la fin d'exécution, et bien d'autres événements. De multiples wrapper des fonctions de base de feather-trace sont fournie dans LITMUS^{RT} afin de pouvoir log des informations supplémentaires, comme le processeur depuis lequel l'exécution du log est effectuée ou encore depuis quelle fonction l'appel est fait.





Cela a été très utile lorsque j'ai développé des nouveaux ordonnanceur sous LITMUS^{RT}afin de corriger des erreurs. Mais cet outil m'a aussi été essentiel afin de comprendre comment fonctionnait les algorithmes d'ordonnancement fournis avec LITMUS^{RT}. J'ai aussi pu comprendre comment le noyau Linux communiquait avec les ordonnanceur en activant des sorties de debug additionnels dans la configuration du noyau.

Enfin, des outils permettant d'extraire, de synthétiser ou de tracer des graphique de certaines données de ces fichiers de logs sont mis à notre disposition sur un dépôt de code présent sur github nommé feather-trace-tools. Voici un exemple du tracé de l'ordonnancement réel de INSERER ORDONBANCEUR:

METTRE UNE IMAGE D'UN DES TRACES

Pour cela, les temps d'exécution, les débuts et les fins de chaque jobs s'exécutant sur chaque processeur ont été enregistrés sur la carte de développement avec l'outil st-trace-schedule du dépôt de code mentionnée précédemment. Cet outil générer autant de fichiers que de processeurs sont présents. On peut alors tracer l'exécution réel avec cette fois ci l'outil st-draw en lui fournissant les fichiers générés au préalable. Ici une durée de XXXXX a aussi été donnée en argument afin de limiter la durée du tracé.

3.3 Implémentation d'un ordonnanceur EDF partitionné

Le but du stage étant l'implémentation d'algorithmes d'ordonnancement sur platforme hétérogène avec migration de taches et de jobs entre les différent processeur, il faudra être capable de réalise des préemptions de jobs (une exécution de tache), les migrer, assurer le traitement d'égalités et bien d'autre problèmes.

3.3.1 Algorithme considéré

On cherche alors, pour commencer, à implémenter un algorithme d'ordonnancement simple afin de se familiariser avec les méthodes et fonctions fourni par LITMUS^{RT}. J'ai donc choisi un algorithme partitionné pour la simplicité d'ordonnancement par processeur que cela offre. Un algorithme EDF (*Earliest Deadline First*) est alors choisi pour la simplicité du choix de la tache a exécuter. Comme son nom l'indique, on choisi à chaque instant la tache ayant l'échéance la plus proche. On nommera par la suite cet algorithme P-EDF (*Partitionned Earliest Deadline First*).

Pour montrer le fonctionnement de cet algorithme, si l'on se place sur un même processeur, on peut visualiser l'exécution de deux tache periodiques :

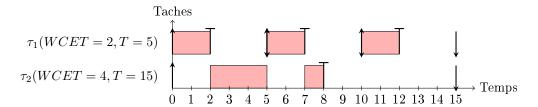


FIGURE 5 – Exemple de EDF à 2 taches

On a ici une première tache τ_1 avec un pire temps d'éxécution (Worst Case Execution Time) de 2 et une période de 5, et une seconde tache τ_2 avec un pire temps d'éxécution de 4 et une période de 15. On a alors préemption de la τ_2 à t=5 afin d'éxécuter τ_1 . Cela est dû au réveil de la tâche τ_1 (représenté par la flêche montante) et à la date d'échéance plus proche de cette dernière.

3.3.2 Implémentation

La construction d'un plugin d'ordonnancement nécessite la déclaration d'un module au sens de Linux. Pour Linux un module est un élément de code qui peut être chargé dynamiquement lors de l'exécution du système d'exploitation. Un module permet alors d'étendre les fonctionnalités du noyau, il a donc ont accès aux fonctions du noyau, à ses resources et peut aussi réaliser des appels systèmes.

Pour que notre nouvel ordonnanceur soit reconnu par le noyau Linux modifié (LITMUS^{RT}); il faut alors déclarer une fonction d'initialisation :





```
#include #include #include #include init()

static int __init init_p_edf(void)
{
    return 0; // indicates a successful initialisation
}

module_init(init_p_edf); // specify the entry point of the module
```

On peut alors enregistrer ce fichier sous le nom sched_p_edf.c pour suivre la nomenclature des autres ordonnanceurs fournis avec avec LITMUS^{RT}. Ce fichier est enregistré dans le dossier llinux /litmus. On peut alors modifier le fichier Makefile de ce dossier afin de l'ajouter au fichier à compiler :

```
obj-y = sched_p_edf.o
```

On place notre fichier à compiler sous le mot clé obj-y pour signifier que l'on veut ce module compilé et inclus lors de la compilation du noyau Linux.

Une fois le makefile modifié, la compilation de notre module sera exécutée lors de la compilation du noyau Linux à l'aide de make. La compilation du noyau est discuté dans la partie 2.2.2.

3.4 Implémentation d'un ordonnanceur RM partitionné





Conclusion

Il faut une petite conclusion qui fait le taf





Annexe

Listing 7 – linux/litmus/Makefile

```
# Makefile for LITMUS^RT
   #
3
   obj-y = sched_plugin.o litmus.o \
          preempt.o \
          litmus_proc.o \
          budget.o \
          clustered.o \
          jobs.o \
          sync.o \
          rt_domain.o \
          edf_common.o \
13
          fp_common.o \
14
          fdso.o \
          locking.o \
          srp.o \
          bheap.o \
          binheap.o \
          ctrldev.o \
          uncachedev.o \
          sched_gsn_edf.o \
          sched_psn_edf.o \
          sched_pfp.o \
           sched_p_edf.o
26
   obj-$(CONFIG_PLUGIN_CEDF) += sched_cedf.o
   obj-$(CONFIG_PLUGIN_PFAIR) += sched_pfair.o
   obj-$(CONFIG_FEATHER_TRACE) += ft_event.o ftdev.o
30
   obj-$(CONFIG_SCHED_TASK_TRACE) += sched_task_trace.o
   obj-$(CONFIG_SCHED_DEBUG_TRACE) += sched_trace.o
   obj-$(CONFIG_SCHED_OVERHEAD_TRACE) += trace.o
   obj-y += sched_pres.o
36
   obj-y += reservations/
```

Listing 8 – linux/litmus/sched_p_edf.c

```
#include <linux/module.h>
   #include <linux/percpu.h>
   #include <linux/sched.h>
   #include <litmus/litmus.h>
   #include <litmus/budget.h>
   #include <litmus/edf_common.h>
   #include <litmus/jobs.h>
   #include <litmus/litmus_proc.h>
   #include <litmus/debug_trace.h>
   #include <litmus/preempt.h>
   #include <litmus/rt_domain.h>
   #include <litmus/sched_plugin.h>
12
   #include <litmus/sched_trace.h>
1.3
   struct p_edf_cpu_state {
15
          rt_domain_t local_queues;
16
          int cpu;
```





```
struct task_struct* scheduled;
   };
   static DEFINE_PER_CPU(struct p_edf_cpu_state, p_edf_cpu_state);
   #define cpu_state_for(cpu_id) (&per_cpu(p_edf_cpu_state, cpu_id))
   #define local_cpu_state()
                                 (this_cpu_ptr(&p_edf_cpu_state))
24
                             (&per_cpu(p_edf_cpu_state, cpu).local_queues)
   #define remote_edf(cpu)
   #define remote_pedf(cpu) (&per_cpu(p_edf_cpu_state, cpu))
   #define task_edf(task)
                             remote_edf(get_partition(task))
   static struct domain_proc_info p_edf_domain_proc_info;
   static long p_edf_get_domain_proc_info(struct domain_proc_info **ret)
31
32
33
          *ret = &p_edf_domain_proc_info;
          return 0;
34
   }
36
   static void p_edf_setup_domain_proc(void)
38
          int i, cpu;
          int num_rt_cpus = num_online_cpus();
40
41
          struct cd_mapping *cpu_map, *domain_map;
42
          memset(&p_edf_domain_proc_info, 0, sizeof(p_edf_domain_proc_info));
          init_domain_proc_info(&p_edf_domain_proc_info, num_rt_cpus, num_rt_cpus);
45
          p_edf_domain_proc_info.num_cpus = num_rt_cpus;
46
          p_edf_domain_proc_info.num_domains = num_rt_cpus;
47
          i = 0;
49
          for_each_online_cpu(cpu) {
                  cpu_map = &p_edf_domain_proc_info.cpu_to_domains[i];
                  domain_map = &p_edf_domain_proc_info.domain_to_cpus[i];
                  cpu_map->id = cpu;
54
                  domain_map->id = i;
                  cpumask_set_cpu(i, cpu_map->mask);
                  cpumask_set_cpu(cpu, domain_map->mask);
                  ++i;
58
          }
59
   /* This helper is called when task 'prev' exhausted its budget or when
   * it signaled a job completion. */
   static void p_edf_job_completion(struct task_struct *prev, int budget_exhausted)
64
          sched_trace_task_completion(prev, budget_exhausted);
      TRACE_TASK(prev, "job_completion(forced=%d).\n", budget_exhausted);
       tsk_rt(prev)->completed = 0;
          /* Call common helper code to compute the next release time, deadline,
          * etc. */
71
          prepare_for_next_period(prev);
   }
73
7.4
   /* Add the task 'tsk' to the appropriate queue. Assumes the caller holds the
       ready lock.
   static void p_edf_requeue(struct task_struct *tsk, struct p_edf_cpu_state *
```





```
cpu_state)
    {
           if (is_released(tsk, litmus_clock())) {
7.9
                   /* Uses __add_ready() instead of add_ready() because we already
                      * hold the ready lock. */
                   __add_ready(&cpu_state->local_queues, tsk);
82
                  TRACE_TASK(tsk, "added to ready queue on reschedule\n");
83
           } else {
                   /* Uses add_release() because we DON'T have the release lock. */
                   add_release(&cpu_state->local_queues, tsk);
86
                  TRACE_TASK(tsk, "added to release queue on reschedule\n");
           }
89
    static int p_edf_check_for_preemption_on_release(rt_domain_t *local_queues)
           struct p_edf_cpu_state *state = container_of(local_queues,
                                                struct p_edf_cpu_state,
                                                    local_queues);
           /* Because this is a callback from rt_domain_t we already hold
               * the necessary lock for the ready queue. */
9.8
           if (edf_preemption_needed(local_queues, state->scheduled)) {
                  preempt_if_preemptable(state->scheduled, state->cpu);
                  return 1:
           }
           return 0;
104
    static long p_edf_activate_plugin(void)
108
           int cpu;
           struct p_edf_cpu_state *state;
           for_each_online_cpu(cpu) {
                  TRACE("Initializing CPU%d...\n", cpu);
                   state = cpu_state_for(cpu);
                   state->cpu = cpu;
114
                   state->scheduled = NULL;
                   edf_domain_init(&state->local_queues,
                                 p_edf_check_for_preemption_on_release,
                                 NULL);
118
           }
           p_edf_setup_domain_proc();
           return 0;
    }
123
124
    static long p_edf_deactivate_plugin(void)
           destroy_domain_proc_info(&p_edf_domain_proc_info);
           return 0;
128
    }
    static struct task_struct* p_edf_schedule(struct task_struct * prev)
           struct p_edf_cpu_state *local_state = local_cpu_state();
           /* next == NULL means "schedule background work". */
137
```





```
struct task_struct *next = NULL;
           /* prev's task state */
140
           int exists, out_of_time, job_completed, self_suspends, preempt, resched;
           raw_spin_lock(&local_state->local_queues.ready_lock);
143
144
           BUG_ON(local_state->scheduled && local_state->scheduled != prev);
           BUG_ON(local_state->scheduled && !is_realtime(prev));
146
147
           exists = local_state->scheduled != NULL;
           self_suspends = exists && !is_current_running();
149
           out_of_time = exists && budget_enforced(prev) && budget_exhausted(prev);
           job_completed = exists && is_completed(prev);
           /* preempt is true if task 'prev' has lower priority than something on
153
           * the ready queue. */
154
           preempt = edf_preemption_needed(&local_state->local_queues, prev);
           /* check all conditions that make us reschedule */
           resched = preempt;
           /* if 'prev' suspends, it CANNOT be scheduled anymore => reschedule */
              (self_suspends) {
                  resched = 1;
           }
           /* also check for (in-)voluntary job completions */
              (out_of_time || job_completed) {
                   p_edf_job_completion(prev, out_of_time);
                  resched = 1;
168
           }
169
           if (resched) {
                   /* First check if the previous task goes back onto the ready
                   * queue, which it does if it did not self_suspend.
                   */
174
                   if (exists && !self_suspends) {
                          p_edf_requeue(prev, local_state);
                   }
                  next = __take_ready(&local_state->local_queues);
178
           } else {
                   /* No preemption is required. */
                  next = local_state->scheduled;
           }
182
183
           local_state->scheduled = next;
184
           if (exists && prev != next) {
185
                  TRACE_TASK(prev, "descheduled.\n");
           }
           if (next) {
188
                   TRACE_TASK(next, "scheduled.\n");
189
           /* This mandatory. It triggers a transition in the LITMUS^RT remote
           * preemption state machine. Call this AFTER the plugin has made a
           * local scheduling decision.
194
           */
           sched_state_task_picked();
           raw_spin_unlock(&local_state->local_queues.ready_lock);
198
```





```
return next;
    }
200
201
    static long p_edf_admit_task(struct task_struct *tsk)
203
           if (task_cpu(tsk) == get_partition(tsk)) {
204
                   TRACE_TASK(tsk, "accepted by p_edf plugin.\n");
205
           }
207
           return -EINVAL;
208
    }
    static void p_edf_task_new(struct task_struct *tsk, int on_runqueue,
211
                              int is_running)
212
    {
213
214
           /* We'll use this to store IRQ flags. */
           unsigned long flags;
215
           struct p_edf_cpu_state *state = cpu_state_for(get_partition(tsk));
216
           lt_t now;
           TRACE_TASK(tsk, "is a new RT task %11u (on runqueue: %d, running: %d) \n",
219
                       litmus_clock(), on_runqueue, is_running);
221
           /* Acquire the lock protecting the state and disable interrupts. */
222
           raw_spin_lock_irqsave(&state->local_queues.ready_lock, flags);
224
           now = litmus_clock();
           /* Release the first job now. */
227
           release_at(tsk, now);
228
           if (is_running) {
                   /* If tsk is running, then no other task can be running
231
                       * on the local CPU. */
                   BUG_ON(state->scheduled != NULL);
                   state->scheduled = tsk;
234
           } else if (on_runqueue) {
235
236
                   p_edf_requeue(tsk, state);
           }
238
           if (edf_preemption_needed(&state->local_queues, state->scheduled))
239
                   preempt_if_preemptable(state->scheduled, state->cpu);
240
           raw_spin_unlock_irqrestore(&state->local_queues.ready_lock, flags);
242
244
    static void p_edf_task_exit(struct task_struct *tsk)
245
246
           unsigned long flags;
           struct p_edf_cpu_state *state = cpu_state_for(get_partition(tsk));
           raw_spin_lock_irqsave(&state->local_queues.ready_lock, flags);
249
           rt_domain_t*
                              edf;
250
251
           /* For simplicity, we assume here that the task is no longer queued
252
               anywhere else. This
               * is the case when tasks exit by themselves; additional queue
253
                   management is
               * is required if tasks are forced out of real-time mode by other tasks
                   . */
255
           if (is_queued(tsk)){
256
```





```
edf = task_edf(tsk);
                   remove(edf,tsk);
258
           }
              (state->scheduled == tsk) {
261
                   state->scheduled = NULL;
262
263
           preempt_if_preemptable(state->scheduled, state->cpu);
265
           raw_spin_unlock_irqrestore(&state->local_queues.ready_lock, flags);
266
    }
267
    /* Called when the state of tsk changes back to TASK_RUNNING.
269
       * We need to requeue the task.
270
271
       * NOTE: If a sporadic task is suspended for a long time,
        * this might actually be an event-driven release of a new job.
273
274
    static void p_edf_task_resume(struct task_struct *tsk)
           unsigned long flags;
277
           struct p_edf_cpu_state *state = cpu_state_for(get_partition(tsk));
278
           lt t now;
279
           TRACE_TASK(tsk, "wake_up at %llu\n", litmus_clock());
           raw_spin_lock_irqsave(&state->local_queues.ready_lock, flags);
281
           now = litmus_clock();
           if (is_sporadic(tsk) && is_tardy(tsk, now)) {
285
                   /* This sporadic task was gone for a "long" time and woke up past
286
                       * its deadline. Give it a new budget by triggering a job
                       * release. */
288
                   inferred_sporadic_job_release_at(tsk, now);
289
                   TRACE_TASK(tsk, "woke up too late.\n");
           }
           /* This check is required to avoid races with tasks that resume before
293
               \ast the scheduler "noticed" that it resumed. That is, the wake up may
294
               * race with the call to schedule(). */
           if (state->scheduled != tsk) {
296
                   TRACE_TASK(tsk, "is being requed\n");
                   p_edf_requeue(tsk, state);
                   if (edf_preemption_needed(&state->local_queues, state->scheduled))
                          preempt_if_preemptable(state->scheduled, state->cpu);
300
                   }
301
           }
302
303
           raw_spin_unlock_irqrestore(&state->local_queues.ready_lock, flags);
304
    }
306
307
    static struct sched_plugin p_edf_plugin = {
308
                                  = "P-EDF",
           .plugin_name
           .schedule
                                  = p_edf_schedule,
           .task_wake_up
                                  = p_edf_task_resume,
311
           .admit_task
                                  = p_edf_admit_task,
312
           .task_new
                                  = p_edf_task_new,
           .task_exit
                                  = p_edf_task_exit,
314
           .get_domain_proc_info = p_edf_get_domain_proc_info,
315
           .activate_plugin
                                  = p_edf_activate_plugin,
316
```





```
.deactivate_plugin = p_edf_deactivate_plugin,
.complete_job = complete_job,
};

static int __init init_p_edf(void)
{
    return register_sched_plugin(&p_edf_plugin);
}

module_init(init_p_edf);
```

Listing 9 – Partie du fichier .config liée a LITMUS $^{\mathrm{RT}}$

```
# LITMUS^RT
   #
   #
   # Scheduling
   CONFIG_PLUGIN_PFAIR=y
   # CONFIG_RELEASE_MASTER is not set
   CONFIG_PREFER_LOCAL_LINKING=y
   CONFIG_LITMUS_QUANTUM_LENGTH_US=1000
   CONFIG_BUG_ON_MIGRATION_DEADLOCK=y
   # end of Scheduling
14
   # Real-Time Synchronization
15
   CONFIG_NP_SECTION=y
   CONFIG_LITMUS_LOCKING=y
18
   # end of Real-Time Synchronization
   # Performance Enhancements
   CONFIG_ALLOW_EARLY_RELEASE=y
   # CONFIG_EDF_TIE_BREAK_LATENESS is not set
   CONFIG_EDF_TIE_BREAK_LATENESS_NORM=y
   # CONFIG_EDF_TIE_BREAK_HASH is not set
   # CONFIG_EDF_PID_TIE_BREAK is not set
   # end of Performance Enhancements
30
   # Tracing
33
   CONFIG_FEATHER_TRACE=y
34
   CONFIG_SCHED_TASK_TRACE=y
   CONFIG_SCHED_TASK_TRACE_SHIFT=9
   CONFIG_SCHED_OVERHEAD_TRACE=y
   CONFIG_SCHED_OVERHEAD_TRACE_SHIFT=22
   CONFIG_SCHED_DEBUG_TRACE=y
   CONFIG_SCHED_DEBUG_TRACE_SHIFT=18
   CONFIG_SCHED_DEBUG_TRACE_CALLER=y
   # CONFIG_PREEMPT_STATE_TRACE is not set
   # CONFIG_REPORT_TIMER_LATENCY is not set
   # end of Tracing
   # end of LITMUS^RT
```

Listing 10 – Modifications apportées au fichier rk3399.dtsi





```
cpus {
           #address-cells = <2>;
2
           \#size-cells = <0>;
           cpu-map {
                    cluster0 {
                           core0 {
                                   cpu = <&cpu_10>;
                            core1 {
                                    cpu = <&cpu_l1>;
                           };
12
                           core2 {
                                    cpu = <&cpu_12>;
14
                           };
                           core3 {
                                    cpu = \langle \&cpu\_13 \rangle;
17
                           };
18
                   };
                    cluster1 {
                           core0 {
                                    cpu = \langle \&cpu_b0 \rangle;
                           };
                            core1 {
                                    cpu = \langle \&cpu_b1 \rangle;
                           };
                   };
28
           };
3.0
           cpu_10: cpu@0 {
                   device_type = "cpu";
                    compatible = "arm,cortex-a53";
                   reg = <0x0 0x0>;
                    enable-method = "psci";
                   next-level-cache = <&12_0>;
36
                    capacity-dmips-mhz = <485>;
37
                    clocks = <&cru ARMCLKL>;
3.8
                    #cooling-cells = <2>; /* min followed by max */
                    dynamic-power-coefficient = <100>;
40
                    cpu-idle-states = <&CPU_SLEEP &CLUSTER_SLEEP>;
41
42
                    12_0: 12-cache {
                    compatible = "cache,arm,arch-cache";
44
           };
45
           };
47
           cpu_l1: cpu@1 {
48
                   device_type = "cpu";
                    compatible = "arm,cortex-a53";
                    reg = <0x0 0x1>;
51
                    enable-method = "psci";
                   next-level-cache = <&12_0>;
                    capacity-dmips-mhz = <485>;
                    clocks = <&cru ARMCLKL>;
                    \#cooling-cells = <2>; /* min followed by max */
                   dynamic-power-coefficient = <100>;
                    cpu-idle-states = <&CPU_SLEEP &CLUSTER_SLEEP>;
           };
59
           cpu_12: cpu@2 {
61
```





```
device_type = "cpu";
                   compatible = "arm,cortex-a53";
                   reg = <0x0 0x2>;
                   enable-method = "psci";
                   next-level-cache = <&12_0>;
                   capacity-dmips-mhz = <485>;
                   clocks = <&cru ARMCLKL>;
                   #cooling-cells = <2>; /* min followed by max */
                   dynamic-power-coefficient = <100>;
                   cpu-idle-states = <&CPU_SLEEP &CLUSTER_SLEEP>;
           };
           cpu_13: cpu@3 {
                   device_type = "cpu";
                   compatible = "arm, cortex-a53";
                   reg = <0x0 0x3>;
                   enable-method = "psci";
                   next-level-cache = <&12_0>;
                   capacity-dmips-mhz = <485>;
                   clocks = <&cru ARMCLKL>;
                   #cooling-cells = <2>; /* min followed by max */
82
                  dynamic-power-coefficient = <100>;
83
                   cpu-idle-states = <&CPU_SLEEP &CLUSTER_SLEEP>;
           };
           cpu_b0: cpu@100 {
                   device_type = "cpu";
                   compatible = "arm,cortex-a72";
                   reg = <0x0 0x100>;
                   enable-method = "psci";
91
                  next-level-cache = <&12_1>;
                   capacity-dmips-mhz = <1024>;
                   clocks = <&cru ARMCLKB>;
                   #cooling-cells = <2>; /* min followed by max */
                   dynamic-power-coefficient = <436>;
                   cpu-idle-states = <&CPU_SLEEP &CLUSTER_SLEEP>;
98
                   12_1: 12-cache {
                   compatible = "cache,arm,arch-cache";
           };
           };
           cpu_b1: cpu@101 {
                   device_type = "cpu";
                   compatible = "arm,cortex-a72";
                   reg = <0x0 0x101>;
107
                   enable-method = "psci";
108
                   next-level-cache = <&12_1>;
109
                   capacity-dmips-mhz = <1024>;
                   clocks = <&cru ARMCLKB>;
                   #cooling-cells = <2>; /* min followed by max */
                   dynamic-power-coefficient = <436>;
                   cpu-idle-states = <&CPU_SLEEP &CLUSTER_SLEEP>;
114
           };
   }
```





Table des figures

1	Architecture du processeur RK3399	4
2	Interface de configuration du noyau	6
	Compilation croisée du noyau Linux	
4	Architecture de LITMUS ^{RT}	10
5	Exemple de EDF à 2 taches	11





RÉFÉRENCES Glossaire

Références

[1] Antoine Bertout, Joël Goossens, Emmanuel Grolleau, and Xavier Poczekajlo. Workload assignment for global real-time scheduling on unrelated multicore platforms. In *Proceedings of the 28th International Conference on Real-Time Networks and Systems*, pages 139–148, 2020.

- [2] B Brandenburg and J Anderson. Feather-trace: A lightweight event tracing toolkit. In *Proceedings of the third international workshop on operating systems platforms for embedded real-time applications*, pages 19–28. Citeseer, 2007.
- [3] Bjorn B Brandenburg. Scheduling and locking in multiprocessor real-time operating systems. PhD thesis, The University of North Carolina at Chapel Hill, 2011.
- [4] Nick Desaulniers. Submitting your first patch to the linux kernel and responding to feedback. https://nickdesaulniers.github.io/blog/2017/05/16/submitting-your-first-patch-to-the-linux-kernel-and-responding-to-feedback/, 2021. Accessed: 2021-05-01.

Glossaire

bootloader court programme chargé au démarrage de l'ordinateur initialisant le système d'exploitation. 5

cluster ensemble interconnecté de plusieurs processeurs. 4

git système de gestion de versions décentralisé, utilisé pour suivre les modifications apportées à des fichiers sources dans un projet de développement logiciel. 5, 6, 7

plateforme hétérogène système formé d'un ensemble de processeurs différents. 3

préemption processus par lequel un système d'exploitation interrompt temporairement l'exécution d'une tâche en cours pour donner la priorité à une autre tâche de plus haute priorité. 11

processeur Ca c'est la définition. 11

SHRIMP Scheduling of Real-Time Heterogeneous Multiprocessor Platform ou Ordonnancement Temps réel de Plateforme Multiprocesseur Hétérogène. 3

SOC ou Système On a Chip est . 4



