



UNIVERSITÀ
DEGLI STUDI
FIRENZE

Scuola di Scienze Matematiche, Fisiche e Naturali
Corso di Laurea in Informatica

Tesi di Laurea

TITOLO ITALIANO

TITOLO INGLESE

NOME CANDIDATO

Relatore: *Relatore*
Correlatore: *Correlatore*

Anno Accademico 2014-2015

Nome Candidato: *Titolo italiano*, Corso di Laurea in Informatica, © Anno
Accademico 2014-2015

INDICE

1	Introduzione	5
1.1	Cos'è un sistema operativo	5
1.2	Microkernel e kernel monolitici	6
1.3	Scheduling	8
1.4	Hypervisor	8
2	SeL4	9
2.1	Capability	10
2.1.1	Proprietà delle capability	11
2.2	Hard Real-Time Systems	11
2.2.1	Mixed-criticality systems	12
2.3	Sicurezza e performance	13
3	Impostazione di seL4	15
3.1	Prerequisiti	15
3.2	Configurazione	17
3.3	Avvio di SeL4	18
3.4	Programmazione con le API livello kernel di seL4	19
3.4.1	Capability	19
3.4.2	Gestione delle memoria	22
3.4.3	Virtual memory management	25
3.4.4	Thread	28
3.4.5	IPC	32

"Inserire citazione"
— *Inserire autore citazione*

INTRODUZIONE

In questa introduzione sarà presente una prima parte che andrà a dare le conoscenze di base minime per comprendere cosa sia un sistema operativo e una piccola classificazione, dopodiché seguiranno la descrizione di alcuni concetti che sono fondamentali per comprendere il resto dell'elaborato.

1.1 COS'È UN SISTEMA OPERATIVO

Un sistema operativo (SO) è un insieme di software che gestisce le risorse hardware e software di un sistema di elaborazione fornendo servizi agli applicativi utente.

In un computer quindi il sistema operativo fornisce l'unica interfaccia diretta con l'hardware e in quanto tale ha un accesso esclusivo con il massimo dei privilegi detto *kernel mode*. Questo comporta che una vulnerabilità all'interno del sistema operativo può portare a gravi conseguenze per l'integrità e la sicurezza del sistema, inoltre qualche malintenzionato potrebbe approfittare di questo bug per trarne profitto. Uno degli obiettivi principali di un SO è quindi quello di garantire la sicurezza, ulteriore scopo è l'efficienza: un buon sistema operativo deve saper sfruttare al meglio tutte le risorse che ha a disposizione, dalla gestione della memoria per sfruttare al meglio lo spazio alla schedulazione dei processi per ottimizzare i tempi di esecuzione. Come ultimo obiettivo, ma non per questo meno rilevante, deve rendere il più semplice possibile l'utilizzo del dispositivo su cui è installato. Dalla definizione di SO data a inizio capitolo possiamo isolare una specifica parte di codice che è quella che permette al software di interfacciarsi con l'hardware, quindi l'accesso e la gestione delle risorse di un dispositivo, questa specifica parte si chiama *kernel* che come suggerisci il nome (nocciolo dall'inglese) rappresenta la parte centrale di un sistema operativo su cui tutto il resto si appoggia.

1.2 MICROKERNEL E KERNEL MONOLITICI

Esistono vari modelli strutturali per i sistemi operativi: monolitici, modulari, a livelli, microkernel ed ibridi, ad oggi i più diffusi sono gli ibridi, che combinano i vari modelli tra di loro, ma che in gran parte si basano su sistemi monolitici i quali consistono di un unico file binario statico al cui interno sono definite tutte le funzionalità del kernel e che viene eseguito in un unico spazio di indirizzi, questo comporta dei vantaggi:

- efficienza → motivo principale per cui la maggior parte dei sistemi operativi ancora si basano su kernel in gran parte monolitici, lavorando nello stesso spazio di indirizzi e gestendo tutto attraverso chiamate a procedura il SO risulterà molto reattivo e performante
- semplicità → in quanto non ha una vera e propria strutturazione ma il codice è tutto in un unico file binario risulta chiaramente più semplice da progettare anche se poi l'implementazione risulta difficile

d'altra parte ha anche degli svantaggi:

- inserimento di un nuovo servizio → questo richiede la ricompilazione del kernel, quindi non permette l'inserimento di un nuovo servizio a runtime (problema risolto nei modelli ibridi)
- dimensione → dovendo gestire tutte le principali funzionalità del sistema operativo il kernel sarà composto da milioni di righe di codice sorgente (MSLOC - linux ha circa 20MSLOC) e questo porta direttamente al successivo grosso svantaggio
- sicurezza → maggiore è il numero di righe di codice maggiore sarà il numero di possibili bug, essendo tutto il codice eseguito nello stesso spazio di indirizzi un bug rischia di far bloccare l'intero sistema anche se il problema è molto piccolo e isolato a una minima funzione del kernel.

All'estremo opposto troviamo i *microkernel* che sono composti da un kernel ridotto al minimo indispensabile, che comprende la gestione della memoria, dei processi e della CPU, le comunicazioni tra processi (IPC) e l'hardware di basso livello, mentre tutto il resto deve essere gestito da server (daemon) che operano sopra al kernel, quindi in spazi di indirizzi separati.

I microkernel sono spesso usati in sistemi embedded in applicazioni

mission critical di automazione robotica o di medicina, a causa del fatto che i componenti del sistema risiedono in aree di memoria separate, private e protette.

Anche questo modello ha dei vantaggi:

- flessibilità → l'inserimento di un nuovo servizio avviene al di sopra del kernel quindi in qualsiasi momento è possibile aggiungere o togliere servizi senza dover modificare il kernel.
- sicurezza → minore quantità di codice eseguita in kernel mode (quindi minore quantità di bug e minore superficie attaccabile) maggiore è la sicurezza del sistema, inoltre i servizi sono lavorano in uno spazio di indirizzi differente da quello del kernel di conseguenza se un server (su cui viene eseguito un servizio) smette di funzionare tutto il resto del sistema continua a funzionare normalmente e si potrà procedere a riavviare quel singolo servizio
- semplicità → essendo il codice composto da giusto qualche decina di migliaia di righe di codice (KSLOC) risulta molto più facile da scrivere

e dall'altro lato ha un grande svantaggio:

- efficienza → dato che ogni servizio gira a livello utente l'utilizzo di uno qualsiasi di questi richiede il ricorso a chiamate di sistema che rallentano fortemente l'esecuzione di ogni operazione, motivo principale per cui ancora oggi i sistemi operativi si basano in gran parte su sistemi monolitici.

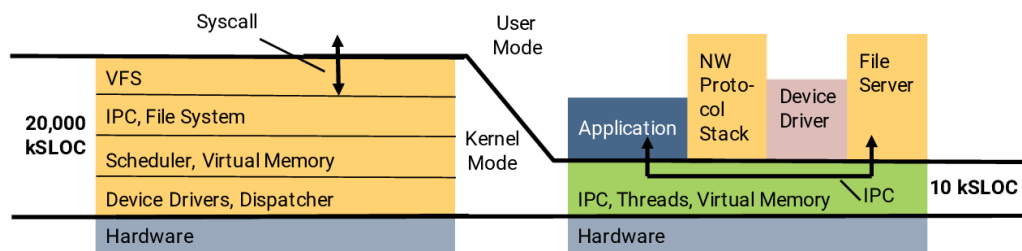


Figura 1: Kernel monolitici (sinistra) VS microkernel (destra)

1.3 SCHEDULING

Proseguendo la lettura verrà nominato lo scheduling: solitamente con il solo termine scheduling si intende quello a breve termine della CPU, cioè la funzionalità che determina quale tra i processi (thread) in attesa della CPU la otterranno, ci sono vari metodi che si differenziano per modalità e prestazioni, gli algoritmi che traducono questi metodi si chiamano politiche di scheduling. Una particolare politica di scheduling rilevante per questo testo è Round Robin o scheduling circolare: consiste nel determinare una quantità di tempo (time slice) nella quale il processo ottiene la CPU al termine del quale il processo viene interrotto e inserito infondo alla lista dei processi in attesa, in questo modo tutti i processi ottengono la CPU per al più un tempo massimo stabilito ed è anche possibile stabilire un tempo massimo di attesa che un processo dovrà attendere in base a quanti processi lo precedono.

1.4 HYPERVISOR

Un *hypervisor*, chiamato anche virtual machine monitor (VMM), è un tipo di software/firmware che permette di creare ed eseguire macchine virtuali. Un computer sul quale un hypervisor esegue una o più macchine virtuali prende il nome di host machine mentre le singole macchina virtuali prenderanno il nome di guest machine, su ognuna di queste è possibile far girare un sistema operativo diverso che eseguirà la maggior parte delle istruzioni direttamente sulle risorse hardware virtualizzate rese disponibili dall'hypervisor.

SEL4

seL4 fa parte della famiglia dei microkernel L4 che risalgono alla prima metà degli anni '90 creato da Jochen Liedtke per sopperire alle scarse performance dei primi sistemi operativi basati su microkernel, ad oggi fa parte del Trustworthy System.

Come descritto poco sopra nell'introduzione, seL4 essendo un microkernel, ha un numero di righe di codice sorgente estremamente piccolo e questo è sufficiente per determinare che non è un sistema operativo ma soltanto un microkernel, infatti non fornisce nessun dei servizi che siamo solitamente abituati a trovare su un comune SO, "è solo un sottile involucro attorno all'hardware" [4], tutti i servizi devono essere eseguiti in modalità utente e questi dovranno essere importati ad esempio da sistemi operativi open-source come Linux (oppure scritti da zero). Data questa sua "incapacità" nel fornire servizi all'utente seL4 è anche un *hypervisor*, quindi è possibile eseguire macchine virtuali sulle quali far girare un comune SO che fornirà i servizi non presenti in seL4.

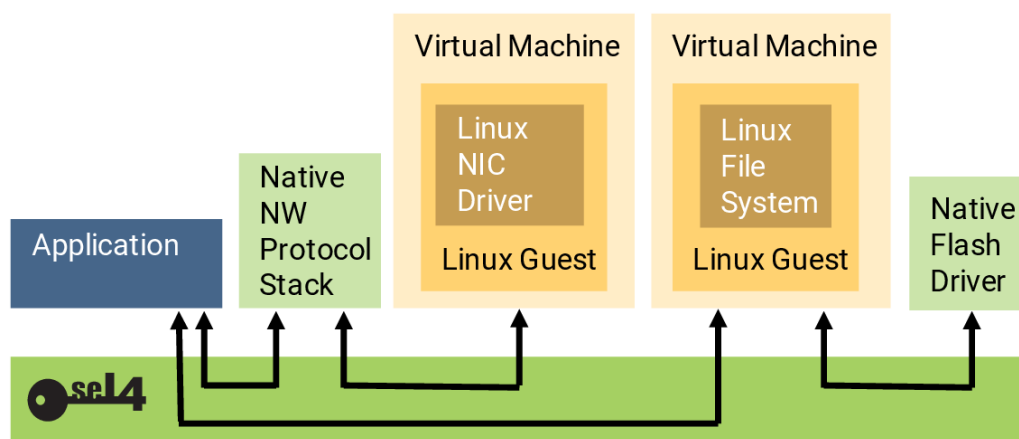


Figura 2: Virtualizzazione del SO Linux per l'integrazione dei servizi di networking e file system

Un esempio pratico può essere quello mostrato in Figura 1, in cui è raffigurato seL4, una generica applicazione e due macchine virtuali (VM) sulle quali viene eseguita una versione ridotta al minimo di Linux, che quindi avrà poco più oltre al servizio che dovrà eseguire. Queste due VM forniranno all'applicazione il servizio di networking ed il file system per la gestione della memoria secondaria (hard disk, supporti rimovibili ecc.), le comunicazioni tra le parti saranno gestite da un canale fornito dal microkernel ma le due macchine virtuali non avranno modo di comunicare tra di loro, non solo questo ma come si vede in figura anche le comunicazioni tra le varie parti e l'applicazione sono ben delineate e precise, nessun'altra comunicazione tra le varie parti è possibile al di fuori di quelle indicate dalle frecce.

2.1 CAPABILITY

Una *capability* è definita formalmente come un riferimento ad oggetto, nel nostro caso specifico possiamo definirla anche come un puntatore immutabile, cioè una capability farà sempre riferimento allo stesso oggetto. SeL4 è un sistema capability-based (basato sulle capability) questo significa che l'unico modo per eseguire un'operazione è attraverso l'invocazione di una capability. Ad ognuna di esse, inoltre, sono associati dei diritti di accesso, quindi una capability è un incapsulamento di un riferimento ad oggetto con i diritti ad esso conferiti. Per dare una definizione meno formale possiamo pensare alle capability come a delle chiavi di accesso estremamente specifiche riguardo quale entità può accedere ad una particolare risorsa del sistema. Inoltre permettono di supportare il *principle of least privilege*, principio del privilegio minimo chiamato anche *principle of least authority* PoLA, questo principio implica che ogni modulo deve avere accesso solo ed esclusivamente alle risorse strettamente necessarie al suo scopo. In seL4 quindi i diritti dati ad un componente possono essere ristretti al minimo indispensabile per svolgere il loro lavoro come richiesto dal PoLA e chiaramente questo è un grosso punto a favore per quanto riguarda la sicurezza.

Nei sistemi operativi più comuni tipo Windows o Linux l'accesso alle risorse è gestito dalle *access-control list* (ACL), quindi, nel caso specifico di Linux, ad ogni file viene associato un set di bit che determinano quali operazioni (lettura, scrittura, esecuzione) possono essere eseguite su di esso dai vari utenti (proprietario, gruppo, altri) questo però implica che ogni file del sistema con lo stesso set di permessi è a disposizione di quello specifico utente, quindi se ci mettiamo nello scenario di voler

avviare un programma, di cui non siamo sicuri della sua attendibilità, su uno specifico file questo non è possibile perchè come può accedere a quel file può accedere anche a tutti gli altri che hanno lo stesso ACL.

Con le capability questo scenario non si può presentare perchè il kernel consentirebbe un'operazione se e solo se chi richiede di eseguire l'operazione ha la "giusta capability" per eseguire l'operazione su quel file.

2.1.1 *Proprietà delle capability*

Interposition → le capability hanno la proprietà di mettersi in mezzo (interpose) tra chi crea una capability e l'effettivo accesso ad una risorsa: se un utente dà una capability ad un oggetto esso non è in grado di sapere cosa effettivamente sia quell'oggetto, può chiaramente utilizzarlo senza però sapere che tipo di oggetto sia.

Delegation → le capability supportano la delegazione dei privilegi tra gli utenti: l'utente X ha un oggetto e vuole dare accesso ad esso anche all'utente Y; X può creare una nuova capability e darla ad Y senza conservare nessun riferimento all'utente X che l'ha creata, la nuova capability può anche avere meno diritti di accesso (esempio solo lettura invece di lettura e scrittura) e inoltre X in qualsiasi momento può revocare l'accesso ad Y distruggendo la capability.

2.2 HARD REAL-TIME SYSTEMS

Un *Hard Real-Time System* è un sistema in cui il mancato rispetto di una scadenza può portare al fallimento dell'intero sistema. Un esempio molto alla mano di tutti e semplificato può essere l'autopilot di un'automobile; un veicolo dotato di un software di guida autonoma richiede la presenza di un numero estremamente elevato di sensori esterni ed interni al veicolo e il computer di bordo deve leggere, elaborare e dare una risposta immediata ad ogni minimo cambiamento di un valore proveniente da questi sensori, se ad un certo punto l'elaborazione di un dato richiede più del tempo dovuto, anche solo di qualche millisecondo, c'è il rischio che questo comporti una serie di ritardi a catena che ad esempio portano al non rilevamento di un oggetto che si sta avvicinando al veicolo, oppure alla mancata correzione della traiettoria e quindi l'abbandono della carreggiata, con conseguenze anche catastrofiche. → Lo scheduling dei processi in seL4 è basato sulla priorità, il kernel di sua iniziativa non

cambierà mai la priorità di un processo, è sempre decisa dall'utente.

→ SeL4 quando esegue delle operazioni in modalità kernel queste sono esenti dagli interrupt, all'apparenza questo può sembrare catastrofico se non fosse per il fatto che le chiamate di sistema sono tutte brevi, solo la revocazione di una capability può richiedere tempi più lunghi ma in presenza di queste operazioni seL4 adotta una politica di divisione dell'esecuzione in sotto operazioni più brevi, ed inoltre ognuna di esse può essere annullata e poi ripresa da quel punto in poi, così da poter gestire degli eventuali interrupt in attesa.

Questi due punti appena elencati sono requisiti fondamentali per gli Hard Real-Time system, scheduling dei processi basato sulla priorità, che sia quindi facilmente analizzabile, e latenza degli interrupt limitata, essendo gli interrupt disabilitati non ci sarà nessuna latenza dovuta al cambio di contesto per gestire subito l'interrupt e dato che le operazioni sono tutte brevi questo non risulta un problema. **Per seL4 è stata eseguita una worst-case execution time (WCET), questo vuol dire che è stato determinato un limite superiore di latenza di ogni system call nel caso peggiore e ciò implica anche il caso peggiore di latenza di un interrupt**

2.2.1 *Mixed-criticality systems*

Un mixed-criticality system (MCS) è un sistema fatto da più componenti che interagiscono tra di loro e che hanno differenti livelli di criticità, in questi sistemi è imperativo che il fallimento di un componente non influenzi gli altri componenti critici, che questi siano quindi isolati e protetti da componenti meno critici.

Un approccio classico per questo tipo di sistemi è isolare le criticità sia per quanto riguarda il tempo che lo spazio, chiamato anche *strict time and space partitioning* (TSP), ma questo implica dover assegnare staticamente l'area di memoria, il tempo di esecuzione e quindi lo scheduling, e per farlo si utilizzano dati misurati precedentemente nel caso pessimo. Essendo sistemi real-time, come già detto prima, ogni operazione deve avere dei limiti di tempo quindi un'operazione a cui è stato misurato, nel caso peggiore, 5 millisecondi deve avere questa durata, non 4ms nè tantomeno 6ms. Chiaramente determinando staticamente i tempi e gli spazi nel caso peggiore siamo sicuri che questi vengano rispettati ma fortunatamente non sempre si presentano dei casi pessimi e quindi porta ad uno scarso utilizzo delle risorse, inoltre la latenza di un interrupt nel caso pessimo può essere molto costosa.

SeL4 supporta il mixed-criticality system: per quanto riguarda l'isolamen-

to abbiamo già visto che le capability, in termini di spazio, intrinsecamente lo garantiscono. Resta quindi da esaminare il comportamento da un punto di vista temporale.

SeL4 normalmente utilizza due parametri per gestire lo scheduling dei processi: la priorità e la quantità di tempo; la *priorità* determina l'ordine di esecuzione dei processi mentre la *quantità di tempo* (time slice) determina quanto tempo il kernel lascerà in esecuzione un thread prima di stopparlo per selezionare un altro processo, quest'ultimo verrà scelto tra i processi pronti in base alla priorità, con una politica round-robin tra i pari livelli di priorità.

La versione MCS di seL4 si comporta diversamente, l'accesso al processore viene controllato dalle capability, un componente può ottenere la CPU solo se ha una capability che glielo permette e il tempo di esecuzione è codificato in essa, questa politica si chiama *scheduling-context capabilities*. Lo scheduling-context contiene due attributi principali:

1. *time budget* che sostituisce il vecchio time slice
2. *time period* che determina invece quante volte un budget può essere usato per periodo, in questo modo viene evitato che un processo monopolizzi la CPU indipendentemente dalla sua priorità

2.3 SICUREZZA E PERFORMANCE

Come già detto nelle prime righe di questo capitolo la famiglia dei microkernel L4 nasce per sopperire alle scarse performance dei suoi predecessori, finora è stata fatta una descrizione del funzionamento generale di seL4 con particolare riguardo sulla sicurezza di questo sistema; per chi è dell'ambito sa già che spesso sicurezza e buone performance non vanno molto d'accordo, garantire la sicurezza vuol dire attenersi a regole ben precise e controlli che spesso poi portano a rallentamenti e quindi vanno a influire sulle performance di un sistema, quindi è lecito domandarsi se questo microkernel sia performante oppure no.

Nonostante non fosse nelle prerogative dello sviluppo di seL4 questo, alla fine, si è rivelato il più performante dei microkernel della famiglia L4, inoltre sono state fatte altre pubblicazioni indipendenti che mettono a confronto seL4 con altri microkernel per studiarne le performance, in particolare Fiasco.OC, Zicron [20] e CertiKOS [3], confrontando i costi dell'IPC si può vedere che seL4 ha un bel vantaggio anche di oltre un fattore due rispetto agli altri microkernel.

IMPOSTAZIONE DI SEL4

Come primo approccio per arrivare alla scrittura di questa tesi ho innanzitutto fatto una ricerca sulla letteratura che si trova disponibile riguardo a seL4, nonostante sia poca e principalmente fornita da Trustworthy (TS) stesso comunque sufficiente per avere una conoscenza abbastanza approfondita del microkernel.

SeL4 è un sistema open-source dunque lo step successivo è stato quello di scaricare seL4 e sperimentare con mano le funzionalità, ovviamente questo ha richiesto un approfondimento più tecnico e specifico, rispetto a quanto fatto finora, di alcuni aspetti come la gestione della memoria fisica e virtuale, l'IPC ecc. che verranno trattati in questo capitolo.

3.1 PREREQUISITI

Come prima cosa ho installato sul mio portatile VirtualBox in quanto come consigliato dalle guide fornite da TS sarebbe ottimale lavorare in ambiente Linux, non avendo una partizione del portatile con Linux ho inizialmente pensato di utilizzare una macchina virtuale così da lasciare inalterato il mio computer e comunque avere a disposizione un sistema operativo Linux su cui lavorare. Andando avanti con il set-up del sistema per iniziare a lavorare su seL4 però ho incontrato una prima difficoltà che è stata lo spazio: purtroppo lo spazio nel portatile non era tantissimo, la macchina virtuale, considerando il sistema operativo e l'installazione dei vari prerequisiti per poter far girare il microkernel, cominciava ad occupare molto spazio, dunque ho dovuto cercare un'alternativa; Per sopperire al problema mi sono procurato un SSD su cui sono andato a copiare la partizione creata in VirtualBox continuando la sperimentazione sul microkernel lavorando sull'SSD esterno collegato via USB.

Per lavorare su seL4 è necessario avere installato sul sistema dei programmi che simulino un'architettura su cui farlo girare, per fare ciò è necessario installare delle dipendenze (prerequisiti) cioè compilatori,

emulatori software vari e librerie che devono essere installate affinché sia possibile utilizzare seL4.

Prima di tutto ho installato Google repo, così da poter clonare i repository git:

```
sudo apt-get install repo
```

build-essential, cmake, ninja, curl, python e QEMU abbreviazione di Quick EMUlator, un emulatore open-source che permette di emulare un'architettura informatica e che permette di simulare diversi sistemi operativi, in questo caso è fondamentale perchè permette l'esecuzione di seL4:

```
sudo apt-get install build-essential
sudo apt-get install cmake ccache ninja-build cmake-curses-gui
sudo apt-get install libxml2-utils ncurses-dev
sudo apt-get install curl git doxygen device-tree-compiler
sudo apt-get install u-boot-tools
sudo apt-get install python3-dev python3-pip python-is-python3
sudo apt-get install protobuf-compiler python3-protobuf
sudo apt-get install qemu-system-arm qemu-system-x86 qemu-system-misc
pip3 install --user setuptools
pip3 install --user sel4-deps
```

Altro componente fondamentale è CAMkES (component architecture for microkernel-based embedded systems), un framework per realizzare velocemente sistemi multiserver affidabili basati su microkernel:

```
pip3 install --user camkes-deps
curl -sSL https://get.haskellstack.org/ | sh
sudo apt-get install haskell-stack
sudo apt-get install clang gdb
sudo apt-get install libssl-dev libclang-dev libcunit1-dev libsqlite3-dev
sudo apt-get install qemu-kvm
```

Dopodiché sono passato alle dipendenze per l'installazione di Isabelle (theorem prover) che serve per la verifica automatica di sistemi software e hardware:

```
sudo apt-get install \
  python3 python3-pip python3-dev \
  gcc-arm-none-eabi build-essential libxml2-utils ccache \
  ncurses-dev librsvg2-bin device-tree-compiler cmake \
  ninja-build curl zlib1g-dev texlive-fonts-recommended \
  texlive-latex-extra texlive-metapost texlive-bibtex-extra \
  mltton-compiler haskell-stack repo
```

Ancora dipendenze Python e Haskell

```
pip3 install --user --upgrade pip
pip3 install --user sel4-deps

stack upgrade --binary-only
which stack # should be $HOME/.local/bin/stack
stack install cabal-install
```

Con questa serie di comandi bash il sistema operativo Linux, per la precisione Ubuntu **VERSIONE**, ha tutti i prerequisiti necessari per procedere alla configurazione.

3.2 CONFIGURAZIONE

Lo step successivo è stato quello di recuperare, attraverso repo, la collezione di repository necessari per la verifica di seL4; in particolare contiene il sorgente del kernel, il theorem prover Isabelle/HOL e HOL4 e lo strumento di verifica binaria.

```
mkdir verification
cd verification
repo init -u https://git@github.com:seL4/verification-manifest.git
repo sync
```

A questo punto si avrà quindi una cartella con questa struttura:

```
verification
├── HOL4/
├── graph-refine/
├── isabelle/
├── l4v/
└── seL4/
```

Il che indica che l'importazione delle repository è andata a buon fine, quindi possiamo procedere alla configurazione di Isabelle posizionandoci nella cartella l4v:

```
mkdir -p ~/.isabelle/etc
cp -i misc/etc/settings ~/.isabelle/etc/settings
./isabelle/bin/isabelle components -a
./isabelle/bin/isabelle jedit -bf
./isabelle/bin/isabelle build -bv HOL
```

Questa serie di comandi bash daranno come risultato:

- la creazione di una cartella per le impostazioni utente di Isabelle.

- installazione delle impostazione Isabelle per L4.verified [2] il rt-bj6nmgquale è una repository che contiene formalismi per la verifica di seL4.
- download di Scala, Java JDK, PolyML ed altri dimostratori (prover) esterni.
- compilazione del Prover IDE (PIDE) jEdit di Isabelle

3.3 AVVIO DI SEL4

Terminata la prima fase di installazione dei prerequisiti e di configurazione mi sono procurato ciò che servirà poi per eseguire i test delle varie funzionalità di seL4:

```
mkdir sel4test
cd sel4test
repo init -u https://github.com/seL4/sel4test-manifest.git
repo sync
```

Con questi comandi si va a creare una directory `sel4test` al cui interno, attraverso il comando `repo`, che conterrà tutte le direttive e le librerie necessarie per eseguire i vari test (quindi anche il kernel stesso). Dopodichè è stato necessario creare una cartella `build-x86` di configurazione per QEMU in modo da indicargli il target su cui eseguire le simulazioni:

```
mkdir build-x86
cd build-x86
../init-build.sh -DPLATFORM=x86_64 -DSIMULATION=TRUE
ninja
```

Il comando `ninja`, che si vedrà spesso a seguire, è un assembler che permette di fare il build di sistemi anche complessi molto velocemente. A questo punto è possibile eseguire il comando `./simulate` che farà partire la simulazione e dopo una lunga serie di test (IPC, chiamate di sistema, thread...) che appariranno nel terminale, concluderà, se tutto è andato a buon fine, con:

All is well in the universe

Il che indica che seL4 può essere utilizzato in questo ambiente simulato.

```
Running test VSPACE0006 (Test touching all available ASID pools)
Test VSPACE0006 passed
Starting test 121: Test all tests ran
Test suite passed. 121 tests passed. 57 tests disabled.
All is well in the universe
```

3.4 PROGRAMMAZIONE CON LE API LIVELLO KERNEL DI SEL4

Una volta che mi sono procurato tutti i prerequisiti necessari e appurato che seL4 può girare senza problemi ho iniziato a prendere familiarità con il sistema seguendo dei tutorial forniti dalla seL4 Foundation che sono programmi semicompleti creati appositamente per sperimentare e far comprendere le funzionalità del sistema, in particolare con le API di seL4. [1]

Come ormai già visto più volte sopra ho recuperato l'ambiente per eseguire i tutorial attraverso l'uso di repo:

```
mkdir sel4-tutorials-manifest
cd sel4-tutorials-manifest
repo init -u https://github.com/sel4/sel4-tutorials-manifest
repo sync
```

Ogni tutorial ha una sua repository da importare nell'ambiente di lavoro nel quale, tra gli altri file e cartelle, c'è (solitamente) un main.c che sarà quello su cui andare a fare le modifiche per completare il tutorial.

3.4.1 *Capability*

Prima di tutti ho fatto un approfondimento sulle capability:

Come già detto nel capitolo precedente una capability è un token unico che dà accesso ad un'entità del sistema, un puntatore con dei diritti di accesso. In seL4 ci sono 3 tipi di capability:

1. capability che controllano l'accesso ad entità del kernel come i thread control block (TCB)
2. capability che controllano l'accesso a risorse astratte tipo gli interrupt

3. untyped capability che sono responsabili per la gestione della memoria

Tutte le capability delle risorse del kernel sono date dal processo root all'inizializzazione del sistema, un po' come il processo `init` nei sistemi unix che è padre di tutti i processi. Quando parliamo di capability ci sono 3 termini fondamentali CNode, CSlot e CSpace: il primo di questi è l'abbreviazione di capability-node ed è un oggetto che contiene delle capability, possiamo pensarlo come un vettore (array) di capability, questi elementi dell'array sono chiamati CSlot (capability-slot) ed ogni CSlot può avere due stati: `empty` o `full`, ciò equivale che il CNode ha, rispettivamente, una capability nulla oppure una capability ad una risorsa del kernel, per convenzione il primo CSlot quindi quello situato alla posizione 0 del vettore è nullo. Invece un CSpace (capability-space) è il range completo di capability accessibili da un thread che può essere composto da uno o più CNode. Per fare riferimento ad una capability ed eseguire operazioni su di essa è necessario fare un adress (indirizzamento) della capability, per farlo ci sono due modi per farlo in sel4: tramite invocazione o con l'indirizzamento diretto. Invocazione: ogni thread ha uno speciale CNode installato nel suo TCB come CSpace `root`, questo può essere nullo, ad esempio quando il thread non è autorizzato a invocare nessuna capability, o può avere una capability ad un certo CNode. Quando si vuole fare un addressing di una capability attraverso invocazione un CSlot viene indirizzato implicitamente invocando il CSpace `root` del thread che sta facendo l'invocazione. Indirizzamento diretto: questo metodo permette di specificare il CNode piuttosto che utilizzare implicitamente il CSpace `root`, questo tipo di addressing è usato principalmente per costruire e manipolare la forma dei CSpace, potenzialmente il CSpace di un altro thread. L'esercizio proposto in questa sezione è un programma in linguaggio C con una serie di errori da risolvere, il primo tra questi è un errore nel settaggio del numero di byte del CNode:

```
int main(int argc, char *argv[]) {

    /* parse the location of the sel4_BootInfo data structure from
    the environment variables set up by the default crt0.S */
    sel4_BootInfo *info = platsupport_get_bootinfo();

    size_t initial_cnode_object_size = BIT(info->initThreadCNodeSizeBits);
    printf("Initial CNode is %zu slots in size\n",
    initial_cnode_object_size);
    size_t initial_cnode_object_size_bytes = 0; // TODO
```

```
printf("The CNode is %zu bytes in size\n",
initial_cnode_object_size_bytes);
```

Chiaramente `initial_cnode_object_size_bytes` non può essere 0, il suo valore invece sarà dato dal numero degli slot del CNode moltiplicato per le dimensione in bit di ognuno di essi \rightarrow `initial_cnode_object_size * (1u « sel4_SlotBits)`.

Eseguendo nuovamente il codice questo darà un errore: "Attempted to invoke a null cap" questo accade perchè il codice cerca di impostare la priorità del TCB del thread invocando l'ultimo CSlot del CSpace che è vuoto

```
sel4_CPtr first_free_slot = info->empty.start;
sel4_Error error = sel4_CNode_Copy(sel4_CapInitThreadCNode, first_free_slot
, sel4_WordBits, sel4_CapInitThreadCNode, sel4_CapInitThreadTCB,
sel4_WordBits, sel4_AllRights);
ZF_LOGF_IF(error, "Failed to copy cap!");
%sel4_CPtr last_slot = info->empty.end - 1;
// TODO

/* set the priority of the root task */
error = sel4_TCB_SetPriority(last_slot, last_slot, 10);
ZF_LOGF_IF(error, "Failed to set priority");
```

Dunque per risolvere il problema è necessario fare un'altra copia della capability del TCB dentro l'ultimo slot del CNode: per fare ciò utilizziamo `sel4_CNode_Copy` che prende come parametri (destination root, slot, depth, source root, slot, depth, rights) dove depth indica quanto bisogna attraversare il CNode per arrivare al CSlot e rights sono i diritti ereditati dalla nuova capability:

```
sel4_CNode_Copy(sel4_CapInitThreadCNode, last_slot, sel4_WordBits,
sel4_CapInitThreadCNode, first_free_slot, sel4_WordBits, sel4_AllRights
);
```

Dove `first_free_slot` è lo slot in cui è stata fatta una copia della capability del TCB del thread iniziale qualche riga di codice sopra.

Rieseguendo il programma non viene più mostrato l'errore precedente ma c'è comunque un altro errore: "first_free_slot is not empty" questo avviene perchè il codice cerca di spostare `first_free_slot` e `last_slot` in se stessi, questo non è possibile (perchè è già presente una capability, cioè se stessa) ed è in realtà un escamotage per controllare se un CSlot è vuoto.

```
// TODO

// check first_free_slot is empty
error = sel4_CNode_Move(sel4_CapInitThreadCNode, first_free_slot,
    sel4_WordBits, sel4_CapInitThreadCNode, first_free_slot, sel4_WordBits)
;
ZF_LOGF_IF(error != sel4_FailedLookup, "first_free_slot is not empty");

// check last_slot is empty
error = sel4_CNode_Move(sel4_CapInitThreadCNode, last_slot, sel4_WordBits,
    sel4_CapInitThreadCNode, last_slot, sel4_WordBits);
ZF_LOGF_IF(error != sel4_FailedLookup, "last_slot is not empty");
```

Quindi per risolvere il problema è necessario eliminare le due capability, questo può essere fatto in due modi: eliminando le due copie delle capability usando `sel4_CNode_Delete` oppure con `sel4_CNode_Revoke` sulla capability originale da cui sono state fatte le copie, quest'ultima API elimina tutte le capability figlie di essa. Per fare più velocemente ho utilizzato il secondo metodo che richiede come parametri il CNode e la posizione dentro di esso in cui andare a recuperare la capability (CNode, index, depth):

```
sel4_CNode_Revoke(sel4_CapInitThreadCNode, sel4_CapInitThreadTCB,
    sel4_WordBits);
```

L'esercitazione conclude con sospensione del thread corrente: ¹

```
sel4_TCB_Suspend(sel4_CapInitThreadTCB);
```

3.4.2 Gestione delle memoria

Nella sezione subito prima di questa quando ho elencato i tipi di capability presenti in sel4 come terza ho elencato le untyped capability, queste sono il modo con il quale è possibile gestire la memoria fisica nel microkernel sel4. Ad accezione di una piccola parte di memoria del kernel tutta la restante è gestita a livello utente, le capability a tutte la memoria fisica disponibile sono passate al processo root come capability all'untyped memory, non altro è che un blocco contiguo di memoria fisica con una dimensione ben specifica, di conseguenza avremo le untyped capability che sono capability all'untyped memory. Inoltre le untyped capability possono essere riscritte in oggetti del kernel insieme alla capa-

¹ codice completo [13]

bility oppure in ulteriori untyped capability più piccole.

Le untyped capability hanno anche un flag booleano `device` che indica se la memoria è scrivibile dal kernel oppure no: può non essere supportata dalla RAM ma da altri dispositivi o essere in un'area non accessibile dal kernel.

Invocazione: esiste un unico modo per invocare una untyped capability in sel4 cioè attraverso l'utilizzo dell'API `seL4_Untyped_Retype` che serve appunto per creare una nuova capability da una untyped capability, nello specifico, questo `retype` darà accesso a un sottoinsieme del range di memoria della capability di origine o come una untyped capability più piccola o puntando ad un nuovo oggetto con un tipo specifico.

```
seL4_Untyped_Retype(parent_untyped, // the untyped capability to retype
                    seL4_UntypedObject, // type
                    untyped_size_bits, //size
                    seL4_CapInitThreadCNode, // root
                    0, // node_index
                    0, // node_depth
                    child_untyped, // node_offset
                    1); // num_caps
```

Le untyped capability sono ritipate in maniera incrementale seguendo una politica greedy a partire dall'untyped invocato, ogni untyped capability mantiene un singolo watermark, con gli indirizzi prima di esso non disponibili e quelli dopo liberi, la memoria non può essere liberata fino a che entrambi i figli non vengono revocati, dove i figli non altro sono che le nuove capability che vengono create da una untyped capability.

Come per la sezione sopra anche qui c'è una repository da scaricare con all'interno un file `main.c` che una volta compilato e avviato stampa a video una lista di tutte le untyped capability fornite dal processo root all'avvio alla fine di questa segnala un errore "Untyped Retype: Requested UntypedItem size too small." e ciò succede perchè il programma sta tentando di creare un untyped di dimensione 0 come si può vedere nel codice:

```
int main(int argc, char *argv[]) {
    /* parse the location of the seL4_BootInfo data structure from
       the environment variables set up by the default crt0.S */
    seL4_BootInfo *info = platsupport_get_bootinfo();

    printf("    CSlot    \tPAddr        \tSize\tType\n");
    for (seL4_CPtr slot = info->untyped.start; slot != info->untyped.end; slot++) {
        seL4_UntypedDesc *desc = &info->untypedList[slot - info->untyped.start];
        printf("%8p\t%16p\t2^%d\t%s\n", (void *) slot, (void *) desc->paddr, desc->sizeBits, desc->isDevice ? "device untyped" : "untyped");
    }
    seL4_Error error;

    // list of general seL4 objects
```

```

seL4_Word objects[] = {seL4_TCBObject, seL4_EndpointObject, seL4_NotificationObject};
// list of general seL4 object size_bits
seL4_Word sizes[] = {seL4_TCBBits, seL4_EndpointBits, seL4_NotificationBits};

// TODO
seL4_Word untyped_size_bits = 0; //ERRORE GENERATO QUI
seL4_CPtr parent_untyped = 0;
seL4_CPtr child_untyped = info->empty.start;

// First, find an untyped big enough to fit all of our objects
for (int i = 0; i < (info->untyped.end - info->untyped.start); i++) {
    if (info->untypedList[i].sizeBits >= untyped_size_bits && !info->untypedList[i].isDevice) {
        parent_untyped = info->untyped.start + i;
        break;
    }
}

```

Per risolvere questo problema è necessario assegnare una dimensione consona alla variabile `untyped_size_bits`, dato che poi dobbiamo cercare uno spazio per tutti gli elementi di `objects[]` e considerato che la somma di `seL4_EndpointBits` e `seL4_NotificationBits` è inferiore a `seL4_TCBBits` possiamo attribuire al valore della variabile `seL4_TCBBits + 1`, in quanto essendo questo in bit quel +1 fa poi raddoppiare il numero di byte visto che lo spazio sarà $2^{\text{seL4_TCBBits}+1}$ che sono sufficienti per contenere tutti e tre gli elementi. Eseguendo di nuovo il programma questo procederà fino a che non segnerà un ulteriore errore “Failed to set priority”

```

// create an untyped big enough to retype all of the above objects from
error = seL4_Untyped_Retype(parent_untyped, seL4_UntypedObject,
    untyped_size_bits, seL4_CapInitThreadCNode, 0, 0, child_untyped, 1);
ZF_LOGF_IF(error != seL4_NoError, "Failed to retype");

// use the slot after child_untyped for the new TCB cap:
seL4_CPtr child_tcb = child_untyped + 1;
// TODO

// try to set the TCB priority
error = seL4_TCB_SetPriority(child_tcb, seL4_CapInitThreadTCB, 10);
ZF_LOGF_IF(error != seL4_NoError, "Failed to set priority");

```

L’errore viene generato perchè `child_tcb` è un CSlot vuoto, per risolvere è sufficiente assegnare al CSlot una capability creando un TCB object da `child_untyped`

```

seL4_Untyped_Retype(child_untyped, seL4_TCBObject, 0,
    seL4_CapInitThreadCNode, 0, 0, child_tcb, 1);

```

Con questa linea di codice il problema viene risolto ma l’esecuzione viene stoppata da un altro errore “Endpoint cap is null cap”

```

// use the slot after child_tcb for the new endpoint cap:

```

```

sel4_CPtr child_ep = child_tcb + 1;
// TODO

// identify the type of child_ep
uint32_t cap_id = sel4_DebugCapIdentify(child_ep);
ZF_LOGF_IF(cap_id == 0, "Endpoint cap is null cap");

```

Questo errore è molto simile al precedente si sta cercando di identificare un endpoint nullo, quindi per risolvere il problema è necessario creare un endpoint object sempre da `child_untyped` e mettere la capability nel CSlot `child_ep`

```

sel4_Untyped_Retype(child_untyped, sel4_EndpointObject, 0,
    sel4_CapInitThreadCNode, 0, 0, child_ep, 1);

```

Alla fine il programma tenta di allocare tutto il `child_untyped` come endpoint ma fallisce perchè tutto lo spazio è stato consumato dalla allocazioni fatte precedentemente, quindi la soluzione è fare una `sel4_CNode_Revoke` (vista sopra) su di esso in modo che tutto lo spazio venga liberato e così facendo il programma termina con successo.²

```

// revoke the child untyped
error = sel4_CNode_Revoke(sel4_CapInitThreadCNode, child_untyped,
    sel4_WordBits);

// allocate the whole child_untyped as endpoints
// Remember the sizes are exponents, so this computes 2^untyped_size_bits /
// 2^sel4_EndpointBits:
sel4_Word num_eps = BIT(untyped_size_bits - sel4_EndpointBits);
error = sel4_Untyped_Retype(child_untyped, sel4_EndpointObject, 0,
    sel4_CapInitThreadCNode, 0, 0, child_tcb, num_eps);
ZF_LOGF_IF(error != sel4_NoError, "Failed to create endpoints.");

printf("Success\n");

```

3.4.3 Virtual memory management

Incipit, cos'è la memoria virtuale: "la memoria virtuale è un'architettura di sistema capace di simulare uno spazio di memoria centrale (memoria primaria) maggiore di quello fisicamente presente o disponibile, dando l'illusione all'utente di un enorme quantitativo di memoria". [9] Questa tecnica porta con sé diversi vantaggi, uno tra questi la sicurezza dovuta

² codice completo [18]

all'isolamento della memoria, la possibilità di condivisione di alcune pagine di memoria tra più processi, tipo quelle contenenti le librerie che ovviamente non cambiano e quindi possono essere usate in contemporanea da più processi e infine l'ultimo ma allo stesso tempo il principale vantaggio avere a disposizione molta più memoria centrale di quella che in realtà disponibile. Giustamente viene da chiedersi com'è possibile tutto ciò e il meccanismo alla base è quello di utilizzare una memoria ausiliaria, solitamente la memoria di massa, per allocare una certa parte di memoria che non è stata utilizzata recentemente, nel momento in cui ne viene richiesto nuovamente la porzione di dati salvati nella memoria ausiliaria (oppure si libera spazio nella memoria centrale) questi vengono nuovamente spostati in memoria centrale, processo che prende il nome di *swapping*. In presenza di memoria virtuale quindi vengono mappati gli indirizzi di memoria in indirizzi fisici e logici, i programmi lavoreranno solo con indirizzi logici (quindi viene anche facilitata la programmazione) e poi a livello di CPU avverrà un processo di traduzione negli indirizzi fisici.

SeL4 non fornisce strumenti per la gestione della memoria virtuale al di là delle primitive per gestione dell'hardware, quindi il servizio di mappatura della memoria e lo swapping deve essere gestito a livello utente che ha tutta la libertà di gestirlo in base alle esigenze del sistema. SeL4 mette dunque a disposizione degli oggetti appositi chiamati *VSpace* (virtual address space), simili ai *Cspace*, e sono composti da oggetti forniti dal kernel che variano in base all'architettura hardware (x86_64, RISC-V, ARM). Per mappare le pagine sono necessari "intermediate hardware virtual memory objects", praticamente per mappare una pagina è necessario creare una struttura intermedia che varia in base all'architettura: ad esempio nei sistemi x86_64 per mappare una pagina sono necessari questi 3 oggetti `seL4_PDPT`, `seL4_PageDirectory`, `seL4_PageTable`. Le API di seL4 forniscono varie funzioni per la mappatura della memoria in base all'architettura in cui sta girando seL4, tutte le funzione di mapping prendono 3 argomenti:

- il *VSpace* in cui mappare l'oggetto
- l'indirizzo virtuale su cui mappare l'oggetto
- attributi della memoria virtuale che dipendono dall'architettura

Esempio di mappatura di un oggetto `seL4_PDPT` ad un certo indirizzo `TEST_VADDR`:

```
sel4_X86_PDPT_Map(pdpt, sel4_CapInitThreadVSpace, TEST_VADDR,
    sel4_X86_Default_VMAttributes);
```

Una volta che le strutture di paginazione intermedie sono state mappate in un certo range di indirizzi virtuali i frame fisici possono essere mappati in quel range attraverso l'invocazione del frame capability.

Esempio di mappatura di un frame:

```
sel4_X86_Page_Map(frame, sel4_CapInitThreadVSpace, TEST_VADDR, sel4_CanRead
    , sel4_X86_Default_VMAttributes);
```

Come si vede questo metodo prende un argomento in più perchè per mappare i frame vengono richiesti anche i diritti che determineranno il tipo di mappatura(nell'esempio sopra diritti di sola lettura). Il tutorial di questa sezione fornisce un programma che all'avvio termina lanciando l'errore "Missing intermediate paging structure at level 30"

```
int main(int argc, char *argv[]) {
    /* parse the location of the sel4_BootInfo data structure from
       the environment variables set up by the default crt0.S */
    sel4_BootInfo *info = platsupport_get_bootinfo();
    sel4_Error error;
    sel4_CPtr frame = alloc_object(info, sel4_X86_4K, 0);
    sel4_CPtr pdpt = alloc_object(info, sel4_X86_PDPTObject, 0);
    sel4_CPtr pd = alloc_object(info, sel4_X86_PageDirectoryObject, 0);
    sel4_CPtr pt = alloc_object(info, sel4_X86_PageTableObject, 0);

    // TODO

    // TODO

    /* map a PDPT at TEST_VADDR */
    error = sel4_X86_PDPT_Map(pdpt, sel4_CapInitThreadVSpace, TEST_VADDR,
        sel4_X86_Default_VMAttributes);

    /* map a read-only page at TEST_VADDR */
    error = sel4_X86_Page_Map(frame, sel4_CapInitThreadVSpace, TEST_VADDR,
        sel4_CanRead, sel4_X86_Default_VMAttributes);
    if (error == sel4_FailedLookup) {
        printf("Missing intermediate paging structure at level %lu\n",
            sel4_MappingFailedLookupLevel());
    }
    ZF_LOGF_IF(error != sel4_NoError, "Failed to map page");
}
```

Errore dovuto al fatto che per mappare una pagina tutte le strutture di paginazione intermedie devono essere mappate; il valore 30 equivale alla

costante `SEL4_MAPPING_LOOKUP_NO_PD` il che ci indica che è necessario mappare un oggetto page directory che può essere fatto con l'apposito metodo `seL4_X86_PageDirectory_Map`:

```
seL4_X86_PageDirectory_Map(pd, seL4_CapInitThreadVSpace, TEST_VADDR,
    seL4_X86_Default_VMAAttributes);
```

Ricompilando ed eseguendo il codice appare un errore simile al precedente "Missing intermediate paging structure at level 21" dove il valore 21 questa volta indica la costante `SEL4_MAPPING_LOOKUP_NO_PT` che suggerisce di mappare un oggetto page table:

```
seL4_X86_PageTable_Map(pt, seL4_CapInitThreadVSpace, TEST_VADDR,
    seL4_X86_Default_VMAAttributes);
```

Adesso il codice procede mappando la pagina però successivamente, come si può leggere nel codice sotto, avviene un tentativo di scrittura sulla pagina questo genera un errore perchè la pagina era stata mappata in sola lettura `seL4_CanRead`, quindi l'errore può essere aggiustato facendo una rimappatura della pagina questa volta in lettura e scrittura:³

```
seL4_X86_Page_Map(frame, seL4_CapInitThreadVSpace, TEST_VADDR,
    seL4_ReadWrite, seL4_X86_Default_VMAAttributes);
```

Ovviamente le pagine possono anche essere "unmappate" utilizzando `unmap` sulla pagina o su qualsiasi struttura intermedia di paginazione oppure eliminando la capability finale di qualsiasi struttura di paginazione.

3.4.4 Thread

SeL4 per rappresentare l'esecuzione di un processo e gestire i tempi di esecuzione fornisce i thread, essi sono realizzati attraverso *thread control block object* (TCBs) e ce ne sono uno per ogni thread del kernel. Un TCBs contiene:

- **VEDIAMO SE METTERLO OPPURE NO**

Come sappiamo in un SO è lo scheduler a decidere quale processo e per quanto tempo può utilizzare la CPU, in seL4, come avevamo già visto nel capitolo precedente, la politica di scheduling è un'integrazione di round-robin e scheduling sulla priorità: lo scheduler sceglie i thread con

³ codice completo [19]

maggiore priorità che sono pronti e se ci sono processi con la stessa priorità questi saranno scelti in ordine FIFO seconda la politica round-robin. La priorità è determinata da un range che va da 0 (`seL4_MinPrio`) a 255 (`seL4_MaxPrio`), oltre alla priorità un TCBs contiene anche un *maximum control priority* (MCP) che serve per controllare che un processo non modifichi la priorità di un altro processo (o di se stesso) impostandola più alta della sua, quindi un processo che vuole modificare una priorità deve fornire la sua capability (di thread) in modo da determinare se è autorizzato a impostare quella priorità. **domain scheduling** L'esercizio per questa sezione, se fatto partire senza nessuna modifica inizialmente mostrerà a video una tabella di tutti i TCBs, questo ottenuto tramite una chiamata di sistema di debug `seL4_DebugDumpScheduler()` e successivamente lancia un errore `Failed to retype thread: 2` come in Figura 5. Questo errore avviene perchè il codice ha un errore nell'invocazione del

```
Hello, World!
Dumping all tcbs!
Name                               State      IP                Prio   Core
-----
tcb_threads                        running 0x4012f2          254    0
idle_thread                        idle 0           0      0
rootserver                        inactive 0x4014bf          255    0
seL4(CPU 0) [Cond failed: result]
main@threads.c:47 [Cond failed: result]
Failed to retype thread: 2
```

Figura 3: Avvio esercizio Thread

metodo: `seL4_Untyped_Retype()`

```
// the root CNode of the current thread
extern seL4_CPtr root_cnode;
// VSpace of the current thread
extern seL4_CPtr root_vspace;
// TCB of the current thread
extern seL4_CPtr root_tcb;
// Untyped object large enough to create a new TCB object

extern seL4_CPtr tcb_untyped;
extern seL4_CPtr buf2_frame_cap;
extern const char buf2_frame[4096];

// Empty slot for the new TCB object
extern seL4_CPtr tcb_cap_slot;
// Symbol for the IPC buffer mapping in the VSpace, and capability to the
// mapping
extern seL4_CPtr tcb_ipc_frame;
extern const char thread_ipc_buff_sym[4096];
// Symbol for the top of a 16 * 4KiB stack mapping, and capability to the
// mapping
```

```
extern const char tcb_stack_base[65536];
static const uintptr_t tcb_stack_top = (const uintptr_t)&tcb_stack_base +
    sizeof(tcb_stack_base);

int new_thread(void *arg1, void *arg2, void *arg3) {
    printf("Hello2: arg1 %p, arg2 %p, arg3 %p\n", arg1, arg2, arg3);
    void (*func)(int) = arg1;
    func(*(int *)arg2);
    while(1);
}

int main(int c, char* argv[]) {

    printf("Hello, World!\n");

    sel4_DebugDumpScheduler();
    // TODO
    sel4_Error result = sel4_Untyped_Retype(sel4_CapNull, sel4_TCBObject,
        sel4_TCBBits, sel4_CapNull, 0, 0, sel4_CapNull, 1);
    ZF_LOGF_IF(result, "Failed to retype thread: %d", result);
    sel4_DebugDumpScheduler();
}
```

Come si può vedere al metodo viene passato un oggetto `sel4_CapNull` come oggetto da ritipare che ovviamente genera l'errore, dunque un modo corretto per sistemare questo errore è utilizzare gli oggetti creati nelle variabili globali del codice

```
sel4_Error result = sel4_Untyped_Retype(tcb_untyped, sel4_TCBObject,
    sel4_TCBBits, root_cnode, 0, 0, tcb_cap_slot, 1);
```

Rieseguendo il codice vedremo che l'errore è risolto e tra la lista dei TCB adesso è presente anche quello appena creato. Dopo aver risolto questo problema se ne presenta un altro
 textttFailed to configure thread: 2 in quanto la configurazione del TCB viene fatta tutta su valori nulli

```
result = sel4_TCB_Configure(sel4_CapNull, sel4_CapNull, 0, sel4_CapNull, 0,
    0, (sel4_Word) NULL, sel4_CapNull);
ZF_LOGF_IF(result, "Failed to configure thread: %d", result);
```

Il metodo `sel4_TCB_Configure` prende come parametri:

```
sel4_Untyped_Retype(tcb, // tcb su cui operare
    fault_ep, // chi riceverà l'IPC quando il thread fallisce
    cspace_root, // nuovo CSpace root
    cspace_root_data, // opzionale: setta il nuovo CNode
    vspace_root, // nuovo VSpace root
```



```

vspace_root_data, // non ha effetto su x86 e ARM
buffer, // locazione dell'IPC buffer
bufferFrame); // IPC buffer

```

Avendo adesso il TCB creato poco sopra è possibile configurarlo in modo da avere lo stesso CSpace e VSpace del thread corrente:

```

result = sel4_TCB_Configure(tcb_cap_slot, sel4_CapNull, root_cnode, 0,
    root_vspace, 0, (sel4_Word) thread_ipc_buff_sym, tcb_ipc_frame);

```

Adesso l'errore che si presenta sarà un altro Failed to set the priority for the new TCB object. questo perchè la priorità data al thread ha valore 0

```

result = sel4_TCB_SetPriority(tcb_cap_slot, sel4_CapNull, 0);
ZF_LOGF_IF(result, "Failed to set the priority for the new TCB object.\n");
sel4_DebugDumpScheduler();

```

Il thread corrente ha un MCP di 254 quindi è possibile assegnare questo valore come priorità, per poterlo fare è necessario anche cambiare il valore sel4_CapNull e sostituirlo con il tcb del thread corrente root_tcb. Dopodiché è necessario impostare in maniera adeguata i registri iniziali in particolare il *program counter* e lo *stack pointer* che è possibile farlo grazie alle utility contenute in libsel4utils

```

sel4_UserContext regs = {0};
int error = sel4_TCB_ReadRegisters(tcb_cap_slot, 0, 0, sizeof(regs)/sizeof(
    sel4_Word), &regs);
ZF_LOGF_IFERR(error, "Failed to read the new thread's register set.\n");

// TODO
sel4utils_set_instruction_pointer(&regs, (sel4_Word)new_thread);
// TODO
sel4utils_set_stack_pointer(&regs, tcb_stack_top);
// TODO
error = sel4_TCB_WriteRegisters(tcb_cap_slot, 0, 0, sizeof(regs)/sizeof(
    sel4_Word), &regs);
ZF_LOGF_IFERR(error, "Failed to write the new thread's register set.\n"
    "\tDid you write the correct number of registers? See
    arg4.\n");
sel4_DebugDumpScheduler();

```

A questo punto è possibile far partire il thread ma per farlo è necessario fare un piccolo aggiustamento nel codice

```

//resume the new thread

```

```
error = sel4_TCB_Resume(sel4_CapNull);
ZF_LOGF_IFERR(error, "Failed to start new thread.\n");
while(1);
return 0;
}
```

Chiaramente il resume va fatto sul nostro `tcb_cap_slot` non su `sel4_CapNull`, a questo punto il nuovo thread viene eseguito e mostra a video Hello2: `arg1 0, arg2 0, arg3 0`. Come si può vedere i valori passati al nuovo thread sono 0 se volessimo passare valori differenti potremmo utilizzare la funzione `sel4utils_arch_init_local_context` e facendo degli aggiustamenti al codice:⁴

```
UNUSED sel4_UserContext regs = {0};
int error = sel4_TCB_ReadRegisters(tcb_cap_slot, 0, 0, sizeof(regs)/sizeof(
    sel4_Word), &regs);
ZF_LOGF_IFERR(error, "Failed to read the new thread's register set.\n"
    "\tDid you write the correct number of registers? See arg4.\n"
    ");

sel4utils_arch_init_local_context((void*)new_thread,
    (void *)1, (void *)2, (void *)3,
    (void *)tcb_stack_top, &regs);
error = sel4_TCB_WriteRegisters(tcb_cap_slot, 0, 0, sizeof(regs)/sizeof(
    sel4_Word), &regs);
ZF_LOGF_IFERR(error, "Failed to write the new thread's register set.\n"
    "\tDid you write the correct number of registers? See arg4.\n"
    ");
```

3.4.5 IPC

InterProcess Communication è il meccanismo che utilizza il microkernel per sincronizzare lo scambio di piccole quantità di dati e capability tra i processi. In sel4 l'IPC è facilitato dal fatto che gli oggetti del kernel sono di piccole dimensioni noti come *endpoint* che fungono da porte per la comunicazione, per mandare e ricevere messaggi IPC lo si fa attraverso invocazioni sugli endpoint.

I thread possono mandare messaggi sugli endpoint con la system call `sel4_Send` che è bloccante, mentre possono usare `sel4_Recv` per ricevere messaggi; `sel4_Call` invece è una chiamata di sistema che combina le due precedenti con una differenza: nella fase di ricezione il thread che

⁴ codice completo [17]

usa questa funzione è bloccato su una "one-time capability" chiamata *reply capability* e non l'endpoint stesso. La *replay capability* è contenuta nel TCB del ricevente, la system call `sel4.Reply` invoca questa capability la quale manderà un IPC che farà risvegliare il processo bloccato. `sel4.ReplyRecv` fa lo stesso ma invia la risposta e blocca l'endpoint fornito in una chiamata di sistema combinata. Ogni thread ha un buffer che contiene il payload del messaggio IPC composto da dati e capability. Il mittente del messaggio specifica la lunghezza e il kernel copia questa quantità tra il mittente e il destinatario dell'IPC buffer. Il buffer IPC contiene un'area limitata di *message register* (MR) che sono utilizzati per trasmettere dati sull'IPC, ogni registro ha dimensione di una parola della macchina e la lunghezza massima di un messaggio è contenuta nella costante `sel4.MsgMaxLength`, per caricare un messaggio dentro il buffer è possibile utilizzare `sel4.SetMR` mentre per estrarlo `sel4.GetMR`, la quantità di parole che possono entrare in un registro è disponibile nelle costante `sel4.FastMessageRegisters`.

Insieme al messaggio il kernel consegna inoltre il *badge* dell'endpoint capability sul quale il mittente ha fatto l'invocazione per mandare il messaggio; gli endpoint possono eddere "bedgati" usando `sel4.CNode.Mint` oppure `sel4.CNode.Mutate`, una volta che è stato messo il badge sull'endpoint questo viene trasferito a tutti i destinatari che ricevono un messaggio su quell'endpoint.

SEL4 usa la struttura dati `sel4.MessageInfo_t` per codificare la descrizione di un messaggio IPC in una singola parola (word) ed è composta dai seguenti campi:

- `length` la quantità di dati nel messaggio
- `extraCaps` numero di capability nel messaggio
- `capsUnwrapped` marca le capability unwrapped dal kernel
- `label` dati che verranno trasferiti che non sono stati modificati dal kernel

Come già accennato, insieme ai dati, attraverso l'IPC, è possibile scambiare anche capability, in gergo questo viene chiamato *cap transfer*

```
//Invio di una capability via IPC
sel4.MessageInfo info = sel4.MessageInfo_new(0, 0, 1, 0);
sel4.SetCap(0, free_slot);
sel4.Call(endpoint, info);
```

```
//Ricezione di una capability
sel4_SetCapReceivePath(cnode, badged_endpoint, sel4_WordBits);
sel4_Recv(endpoint, &sender);
```

Il numero di capability trasferita è codificato nella struttura dati `sel4_MessageInfo_t` come `extraCaps`. Inoltre `sel4` può fare la cosiddetta *unwrap* (scartare) delle capability sull'IPC, se l'*n*-esima capability nel messaggio si riferisce all'endpoint attraverso il quale il messaggio viene inviato, la capability viene *unwrapped*: il suo badge viene inserito nell'*n*-esima posizione dell'IPC buffer del destinatario (`caps_or_badges`) e il kernel setta l'*n*-esimo bit nel campo `capsUnwrapped` del `sel4_MessageInfo_t`.

Fastpath: avere un IPC veloce è un elemento fondamentale per i sistemi basati su microkernel in quanto tutti i servizi a livello utente ed essendo isolati l'unico modo per comunicare è attraverso l'IPC, dunque è necessario avere quello che si chiama *fastpath*, cioè un cammino nel kernel altamente ottimizzato che garantisce velocità dell'IPC; per potersi definire tale deve soddisfare cinque condizioni:

- devono essere usate le system call `sel4_Call` o `sel4_ReplyRecv`
- i dati del messaggio devono entrare nel registro `sel4_FastMessageRegisters`
- i processi devono avere spazi di indirizzi validi
- non dovrebbero essere trasferite capability
- nessun altro thread nello scheduler con priorità superiore a quello sbloccato dall'IPC può essere in esecuzione

In questa sezione l'esercizio è un pò diverso, non c'è un unico file main in cui è contenuto tutto il codice ma c'è un `server.c` e due client: `client_1.c` e `client_2.c` i quali manderanno dei messaggi al server che farà da echo; tutti i processi hanno accesso ad un unica endpoint capability che fornisce accesso allo stesso endpoint object.

Al primo avvio si ha questo output: Gli errori sono dovuti al fatto che en-

```
Booting all finished, dropped to user space
Client 2: waiting for badged endpoint
Badged 2
Client 1: waiting for badged endpoint
Badged 1
Assertion failed: sel4_MessageInfo_get_extraCaps(info) == 1 (/home/elia/sel4-tutorials-manifest/ipc/client_2.c: main: 23)
Assertion failed: sel4_MessageInfo_get_extraCaps(info) == 1 (/home/elia/sel4-tutorials-manifest/ipc/client_1.c: main: 23)
```

Figura 4: Prima esecuzione

trambi i client si mettono in attesa sull'endpoint in attesa di un endpoint

con badge tramite cap transfer che però il server non invierà in quanto il server risponde solo ai messaggi dei client.

```
// cslot containing IPC endpoint capability
extern sel4_CPtr endpoint;
// cslot containing a capability to the cnode of the server
extern sel4_CPtr cnode;
// empty cslot
extern sel4_CPtr free_slot;

int main(int c, char *argv[]) {

    sel4_Word sender;
    sel4_MessageInfo_t info = sel4_Recv(endpoint, &sender);
    while (1) {
        sel4_Error error;
        if (sender == 0) {

            /* No badge! give this sender a badged copy of the endpoint */
            sel4_Word badge = sel4_GetMR(0);
            sel4_Error error = sel4_CNode_Mint(cnode, free_slot,
            sel4_WordBits,
                                                    cnode, endpoint,
            sel4_WordBits,
                                                    sel4_AllRights, badge);

            printf("Badged %lu\n", badge);

            // TODO

            /* reply to the sender and wait for the next message */
            sel4_Reply(info);

            /* now delete the transferred cap */
            error = sel4_CNode_Delete(cnode, free_slot, sel4_WordBits);
            assert(error == sel4_NoError);

            /* wait for the next message */
            info = sel4_Recv(endpoint, &sender);
        }
    }
}
```

Dunque per risolvere questo problema è necessario impostare il cap transfer in modo che i client ricevano l'endpoint badgato:

```
sel4_SetCap(0, free_slot);
info = sel4_MessageInfo_new(0, 0, 1, 0);
```

Compilando e riavviando il programma sembra vada tutto bene eccetto che il sistema si blocca. Questo succede perchè al server manca l'imple-

```

Booting all finished, dropped to user space
Client 2: waiting for badged endpoint
Badged 2
Client 1: waiting for badged endpoint
Badged 1
Client 2: received badged endpoint
Client 1: received badged endpoint

```

Figura 5: Esecuzione dopo invio badge

mentazione della sua funzione di echo dei messaggi che gli vengono inviati la quale può essere fatta scorrendo e stampando a video il contenuto dei message register.

I client mandano rispettivamente le stringhe "quick", "fox", "over", "lazy" il client_1 mentre il client_2 "the", "brown", "jumps", "the", "dog".

```

for (int i = 0; i < seL4_MessageInfo_get_length(info); i++) {
    printf("%c", (char) seL4_GetMR(i));
}
printf("\n");

```

A questo punto però vedremo stampata a video sempre la stessa parola "PAROLA" in loop perchè il server non manda un feedback di risposta al client e di conseguenza continua a stampare l'ultima parola ricevuta.

```

for (int i = 0; i < seL4_MessageInfo_get_length(info); i++) {
    printf("%c", (char) seL4_GetMR(i));
}
printf("\n");

// reply to the client and wait for the next message
info = seL4_ReplyRecv(endpoint, info, &sender);

```

A questo punto l'output sarà la stampa a video prima di tutte le parole inviate dal client 2 seguite da quelle del client 1, possiamo modificare il codice in modo da alternare le stampe dei due client utilizzando `seL4_CNode_SaveCaller` usando `free_slot` per salvare le risposte.

```

for (int i = 0; i < seL4_MessageInfo_get_length(info); i++) {
    printf("%c", (char) seL4_GetMR(i));
}
printf("\n");

```

```
error = sel4_CNode_SaveCaller(cnode, free_slot, sel4_WordBits);
assert(error == 0);
info = sel4_Rcv(endpoint, &sender);
for (int i = 0; i < sel4_MessageInfo_get_length(info); i++) {
    printf("%c", (char) sel4_GetMR(i));
}
printf("\n");
sel4_Send(free_slot, sel4_MessageInfo_new(0, 0, 0, 0));

// reply to the client and wait for the next message
info = sel4_ReplyRcv(endpoint, info, &sender);
```

A questo punto l'output finale sarà il seguente.⁵⁶⁷

```
Client 2: received badged endpoint
the
Client 1: received badged endpoint
quick
fox
brown
jumps
over
lazy
the
dog
```

5 codice completo server [16]

6 codice completo client_1 [14]

7 codice completo client_2[15]

BIBLIOGRAFIA

- [1] Sel4 foundation. Api reference, 2023. URL: <https://docs.sel4.systems/projects/sel4/api-doc.html>. (Cited on page 19.)
- [2] SeL4 Foundation. l4v, 2023. Ultima modifica 19 luglio 2023. URL: <https://github.com/sel4/l4v>. (Cited on page 18.)
- [3] Ronghui Gu, Zhong Shao, Hao Chen, Xiongnan (Newman) Wu, Jieung Kim, Vilhelm Sjöberg, , and David Costanzo. Certikos: An extensible architecture for building certified concurrent os kernels, 2016. USENIX Symposium on Operating Systems Design and Implementation. (Cited on page 13.)
- [4] Gernot Heiser. The sel4 microkernel - an introduction. *The sel4 Foundation*, Revision 1.2, 2020. (Cited on page 9.)
- [5] JavaTpoint. Hard and soft real-time operating system. URL: <https://www.javatpoint.com/hard-and-soft-real-time-operating-system>.
- [6] Wikipedia, l'enciclopedia libera. Hypervisor, 2023. Ultima modifica 25 luglio 2023. URL: <https://en.wikipedia.org/wiki/Hypervisor>.
- [7] Wikipedia, l'enciclopedia libera. Isabelle (proof assistant), 2023. Ultima modifica 1 marzo 2023. URL: [https://en.wikipedia.org/wiki/Isabelle_\(proof_assistant\)](https://en.wikipedia.org/wiki/Isabelle_(proof_assistant)).
- [8] Wikipedia, l'enciclopedia libera. Kernel, 2023. Ultima modifica 7 giu 2023. URL: <https://it.wikipedia.org/wiki/Kernel>.
- [9] Wikipedia, l'enciclopedia libera. Memoria virtuale, 2023. Ultima modifica 16 giugno 2023. URL: https://it.wikipedia.org/wiki/Memoria_virtuale. (Cited on page 25.)
- [10] Wikipedia, l'enciclopedia libera. Operating system, 2023. Ultima modifica 16 luglio 2023. URL: https://en.wikipedia.org/wiki/Operating_system.

- [11] Wikipedia, l'enciclopedia libera. Principle of least privilege, 2023. Ultima modifica 2 agosto 2023. URL: https://en.wikipedia.org/wiki/Principle_of_least_privilege.
- [12] Wikipedia, l'enciclopedia libera. Qemu, 2023. Ultima modifica 30 giugno 2023. URL: <https://en.wikipedia.org/wiki/QEMU>.
- [13] Elia Matteini. Tutorial capability codice completo, 2023. URL: <https://github.com/Elia-dev/Tesi/blob/main/sel4-tutorials-manifest/capabilities/src/main.c>. (Cited on page 22.)
- [14] Elia Matteini. Tutorial IPC codice completo client_1, 2023. URL: https://github.com/Elia-dev/Tesi/blob/main/sel4-tutorials-manifest/ipc/client_1.c. (Cited on page 37.)
- [15] Elia Matteini. Tutorial IPC codice completo client_2, 2023. URL: https://github.com/Elia-dev/Tesi/blob/main/sel4-tutorials-manifest/ipc/client_2.c. (Cited on page 37.)
- [16] Elia Matteini. Tutorial IPC codice completo server, 2023. URL: <https://github.com/Elia-dev/Tesi/blob/main/sel4-tutorials-manifest/ipc/server.c>. (Cited on page 37.)
- [17] Elia Matteini. Tutorial threads codice completo, 2023. URL: <https://github.com/Elia-dev/Tesi/blob/main/sel4-tutorials-manifest/threads/threads.c>. (Cited on page 32.)
- [18] Elia Matteini. Tutorial untyped capability codice completo, 2023. URL: <https://github.com/Elia-dev/Tesi/blob/main/sel4-tutorials-manifest/untyped/src/main.c>. (Cited on page 25.)
- [19] Elia Matteini. Tutorial virtual memory management codice completo, 2023. URL: <https://github.com/Elia-dev/Tesi/blob/main/sel4-tutorials-manifest/mapping/src/main.c>. (Cited on page 28.)
- [20] Zeyu Mi, Dingji Li, Zihan Yang, Xinran Wang, , and Haibo Chen. Skybridge: Fast and secure inter-process communication for microkernels, 2019. EuroSys Conference. (Cited on page 13.)
- [21] Ninja-build. The ninja build system, 2022. Ultima modifica 30 agosto 2022. URL: <https://ninja-build.org/manual.html>.

- [22] Rosario Pugliese. Chiedere al prof come citare le sue slide.