



UNIVERSITÀ
DEGLI STUDI
FIRENZE

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

Tesi di Laurea

ANALISI DEL LINGUAGGIO RUST E DELLA
SUA ADOZIONE NEI SISTEMI OPERATIVI

ANALYSIS OF RUST AND ITS ADOPTION IN
OPERATING SYSTEMS

FRANCESCO BIRIBÒ

Relatore: *Rosario Pugliese*

Anno Accademico 2024-2025

INDICE

Elenco delle figure	3
Elenco delle tabelle	5
Elenco dei listati	7
1 Introduzione	11
1.1 Argomento	12
1.2 Motivazioni	12
1.3 Obiettivi	13
1.4 Struttura	14
1.5 Anteprima dei risultati	14
2 Rust	17
2.1 Contesto e motivazioni	17
2.2 Origine e primo sviluppo	19
2.3 Diffusione e adozione iniziali	19
2.4 Espansione e interesse crescenti	20
2.5 Supporto istituzionale e professionale	21
3 Gestione della memoria in Rust	23
3.1 Ownership	23
3.2 Borrowing	27
3.3 Lifetime	31
4 Sistemi Operativi	35
4.1 C: motivazioni e caratteristiche	35
4.2 Rust e C a confronto	38
4.2.1 Gestione delle risorse	40
4.2.2 Sicurezza della memoria	50
4.2.3 Prestazioni	64
4.2.4 Complessità del codice	68
5 Progetti e applicazioni reali	81
5.1 Kernel di Windows 11	81
5.2 Ubuntu: sudo-rs	82
5.3 Redox OS	84
5.4 Rust for Linux	86
5.4.1 Moduli in mainline	88
5.4.2 Moduli outside mainline	90
5.4.3 Impatto del progetto	91
5.5 Red Hat: Nova	93

2 INDICE

5.6	Considerazioni sui progetti analizzati	95
6	Conclusioni	97
	Bibliografia	101

ELENCO DELLE FIGURE

Figura 1	Tentativo di <i>borrow</i> mutabile di un valore non mutabile	29
Figura 2	Tentativo di due <i>borrow</i> mutabili allo stesso valore	30
Figura 3	Tentativo di <i>borrow</i> mutabile e non mutabile simultaneamente allo stesso valore	30
Figura 4	Generazione di un <i>dangling reference</i>	30
Figura 5	Esempio di violazione delle regole di <i>Lifetime</i> . . .	33
Figura 6	<i>Double free</i> in C	53
Figura 7	Tentativo di <i>double free</i> in Rust	53
Figura 8	<i>Access after free</i> in C	56
Figura 9	Tentativo di <i>access after free</i> in Rust	57
Figura 10	<i>Buffer overflow</i> in C	58
Figura 11	<i>Buffer overflow</i> in Rust	59
Figura 12	<i>Buffer overread</i> in C	60
Figura 13	<i>Buffer overread</i> in Rust	60
Figura 14	<i>Uninitialized memory access</i> in C	62
Figura 15	<i>Uninitialized memory access</i> in C con previa <i>free</i> . .	63
Figura 16	<i>Access after free</i> dovuto all'assenza di <i>lifetime</i> in C .	70
Figura 17	Tentativo di <i>access after free</i> in Rust	70
Figura 18	Traduzione di una <i>macro</i> in C	76
Figura 19	Limitazioni delle <i>macro</i> C	77
Figura 20	Compilazione di una <i>macro</i> in Rust	79
Figura 21	Errore durante la compilazione di una <i>macro</i> in Rust	79

ELENCO DELLE TABELLE

Tabella 1	Risultati n-body per Rust e C (implementazioni ideomatiche)	65
Tabella 2	Risultati del benchmark fasta per Rust e C (implementazioni ideomatiche)	66

ELENCO DEI LISTATI

1	Comportamento di Copy	25
2	Comportamento di Move	25
3	Comportamento di Clone	26
4	Trasferimento di ownership nelle chiamate a funzione . . .	26
5	Reference mutabile a variabile immutabile	28
6	Due reference mutabili alla stessa variabile	28
7	Coesistenza reference mutabile e immutabile	29
8	Reference non valida: <i>dangling reference</i>	29
9	Limitazione delle <i>lifetime</i> generiche	31
10	<i>Lifetime</i> esplicite	32
11	Limitazioni delle <i>lifetime</i>	32
12	Gestione ciclo di vita in memoria dinamica C	41
13	Gestione ciclo di vita in memoria dinamica Rust	41
14	I/O su file in C	44
15	I/O su file in Rust	44
16	Semplice server TCP in C	48
17	Semplice server TCP in Rust	49
18	Unfreed memory in C	51
19	Unfreed memory in Rust	51
20	Double free in C	52
21	Double free in Rust	52
22	Dangling pointer in C	54
23	Dangling reference in Rust	54
24	Access after free in C	55
25	Tentativo di <i>access after free</i> in Rust	56
26	Buffer overflow in C	57
27	Buffer overflow in Rust	58
28	Buffer overread in C	59
29	Buffer overread in Rust	59
30	Uninitialized memory access in C	61
31	Uninitialized memory access in C	62
32	Prevenzione di <i>access after free</i> tramite <i>lifetime annotations</i> .	69
33	<i>Access after free</i> dovuta alla mancanza di <i>lifetime</i> in C . . .	69
34	Programmazione generica in C	72

35	Programmazione generica in Rust	73
36	Definizione di <i>macro</i> in C	75
37	Utilizzo scorretto di <i>macro</i> in C	75
38	Definizione di <i>macro</i> dichiarativa in Rust	77

"Simplicity is prerequisite for reliability"
— Edsger Dijkstra

INTRODUZIONE

Negli ultimi anni il linguaggio di programmazione *Rust* ha suscitato un crescente interesse, attirando l'attenzione di sviluppatori e aziende. Una delle motivazioni principali risiede nel suo approccio alla gestione della memoria dinamica, da sempre un tema cruciale nello sviluppo di sistemi di basso livello.

Storicamente, si sono affermati due approcci alla gestione della memoria dinamica: quello *manuale*, adottato dai primi linguaggi di programmazione come l'assembly e, successivamente, dal C e quello *automatico*, basato su *Garbage Collection* (GC), introdotto nel 1959 da Lisp e oggi giorno impiegato da Java, Python, C# o anche GO, insieme a tanti altri linguaggi considerati di alto livello.

L'approccio manuale rende il programmatore interamente responsabile della gestione della memoria, introducendo potenzialmente una vasta gamma di errori dovuti al fattore umano; la GC, invece, solleva il programmatore da tale incarico, ma introduce un overhead non trascurabile e riduce la trasparenza e il controllo sulla memoria.

Nel contesto delle applicazioni di basso livello, soprattutto dei sistemi operativi, controllo diretto e prestazioni sono requisiti fondamentali; per questo, l'approccio manuale rappresenta nella pratica l'unica soluzione percorribile. È in questo contesto che si inserisce Rust, un linguaggio sviluppato per garantire una gestione sicura della memoria dinamica, senza impiegare GC, e, quindi, senza compromettere le prestazioni.

In questa tesi verrà analizzato Rust, con l'obiettivo di comprendere le motivazioni alla base della sua crescente diffusione e di valutare gli strumenti effettivamente offerti.

1.1 ARGOMENTO

La presente ricerca prende in esame la crescente popolarità di Rust nell'ambito della programmazione di basso livello, in particolare nello sviluppo di sistemi operativi.

Nel corso della trattazione verranno richiamati concetti tipici della programmazione di sistema, principalmente in ambiente Unix, come l'uso dei puntatori, le librerie POSIX e, in generale, elementi del linguaggio C, accompagnati da listati di codice. Per chiarire alcuni degli aspetti peculiari di Rust, specialmente quelli di alto livello o con un livello di astrazione maggiore, saranno proposte anche delle analogie con il linguaggio Java.

Nonostante una conoscenza di base sia di C che di Java possa agevolare la comprensione dei contenuti, non è strettamente necessaria: i concetti più tecnici saranno accompagnati da spiegazioni accessibili, ove possibile¹.

Rust ha una premessa chiara: prevenire intere classi di problemi legati alla gestione della memoria a livello di compilazione, fornendo allo stesso tempo prestazioni, in termini di velocità d'esecuzione, paragonabili a C e C++.

Vi sono, per tale motivo, varie iniziative che mirano a inserire il linguaggio come valida opzione nello sviluppo di sistemi operativi, ma la sua integrazione solleva questioni fondamentali, sia da un punto di vista di tradizione (in quanto storicamente C è un pilastro nello sviluppo di sistemi operativi), che di spesa, in termini di tempo, per imparare un nuovo linguaggio e per mantenere contemporaneamente due flussi di sviluppo paralleli (C e Rust).

1.2 MOTIVAZIONI

Ho scelto di approfondire il linguaggio Rust per un interesse personale nelle tematiche della programmazione di basso livello, specialmente in C, e della *Cybersecurity*.

¹ Rust offre diversi meccanismi riconducibili alla programmazione orientata agli oggetti (*Object-Oriented Programming, OOP*). Ho scelto di utilizzare delle analogie in Java per spiegare questi strumenti in quanto Java è un linguaggio frequentemente utilizzato per introdurre la *OOP* nei corsi universitari. Anche se non è indispensabile, ritengo che una familiarità, anche di base, con Java, può risultare utile: alcuni elementi caratteristici della *OOP* risultano difficili da spiegare, nella loro interezza, senza le basi del paradigma.

Durante il mio percorso di studi, ho avuto modo di confrontarmi più volte con il linguaggio C, sviluppando un certo interesse sia verso il linguaggio che verso le sue comuni applicazioni, come *sistemi embedded* e, in generale, software che interagisce con il sistema operativo.

Questo mi ha portato, più volte, a incorrere in comuni errori di memoria (principalmente *Segmentation Fault* e *Buffer Overflow*); inoltre, sempre durante il percorso di studi, ho avuto modo di acquisire conoscenze, sia teoriche che pratiche, di Cybersecurity (da un punto di vista offensivo, ovvero da *Penetration tester* o *White Hacker*) e ho avuto modo di osservare come la maggior parte degli errori o delle falle nelle applicazioni possano essere sfruttati per diversi scopi, anche malevoli (per eseguire *login bypass* o per accedere ad aree di memoria contenenti dati sensibili).

Non ho potuto quindi fare a meno di riflettere su come fosse possibile rendere codice scritto in C sicuro, con lo scopo di prevenire errori che potessero essere sfruttati. Tuttavia, proprio in questo contesto, sono arrivato a conoscenza di Rust, il quale ha attirato la mia attenzione, sia per la sua premessa del tutto innovativa, che per la sua crescente adozione, specialmente nella programmazione di basso livello.

1.3 OBIETTIVI

L'obiettivo principale di questa tesi è offrire una panoramica sul linguaggio di programmazione Rust, esponendo quelle che sono le motivazioni dietro il suo sviluppo e la sua popolarità; successivamente, verrà analizzato l'approccio alternativo di gestione della memoria implementato dal linguaggio.

Un altro obiettivo è quello di stabilire se, nella teoria, Rust rappresenti una valida opzione per la programmazione di basso livello. Verranno innanzitutto stabiliti i requisiti fondamentali di un linguaggio per lo sviluppo di sistemi operativi; successivamente Rust verrà confrontato con C, standard *de facto* per la programmazione di basso livello, sulla base di gestione delle risorse, della memoria e degli errori, complessità della sintassi e prestazioni, fornendo a tale scopo anche esempi di codice.

Come aspetto finale, verranno proposti progetti concreti che mostrano come la popolarità del linguaggio non sia solo teorica, ma rappresenti uno strumento pratico e in grado di portare risultati concreti.

L'analisi combina aspetti teorici (descrizione dei meccanismi ed esempi di codice) con lo studio di casi concreti, per evidenziare sia i principi alla base del linguaggio sia le sue applicazioni pratiche.

1.4 STRUTTURA

Questa tesi è strutturata in due parti principali, la prima adotterà un approccio di analisi teorica, mentre la seconda presenterà esempi concreti di applicazioni reali. La prima parte è suddivisa in tre capitoli: Capitolo 2, Capitolo 3 e Capitolo 4; mentre la seconda parte è composta interamente dal Capitolo 5.

Nel Capitolo 2, verrà fornita una panoramica sul linguaggio di programmazione Rust, con particolare attenzione sulle motivazioni che ne hanno determinato lo sviluppo e che sono alla base della sua crescente popolarità negli ultimi anni.

Nel Capitolo 3, verrà esaminato il modello di gestione della memoria implementato da Rust, con un'analisi approfondita dei meccanismi alla base; verranno inoltre forniti listati di codice che mostrano esempi del loro funzionamento.

Nel Capitolo 4, verranno esposti gli strumenti necessari che un linguaggio di programmazione dovrebbe avere per essere impiegato nella programmazione di sistema, con riferimento al linguaggio C. Successivamente, quest'ultimo verrà confrontato con Rust sotto gli aspetti di gestione della memoria, delle risorse e degli errori, complessità del codice e prestazioni, per mostrare come Rust metta a disposizione strumenti fondamentali per essere considerato una valida alternativa, teorica, nella programmazione di sistema.

Infine, nel Capitolo 5, verranno presentati progetti e applicazioni concreti che mostrano, nella pratica, come Rust non sia solo un linguaggio promettente dal punto di vista teorico, ma rappresenti un valido strumento, capace di ottenere risultati concreti, nella programmazione di basso livello.

1.5 ANTEPRIMA DEI RISULTATI

Nella prima parte della tesi (Capitolo 2, Capitolo 3 e Capitolo 4), viene evidenziato come l'approccio di gestione della memoria di Rust permetta di prevenire errori critici tipici dei linguaggi con approccio manuale, senza compromettere le prestazioni.

Il confronto con C evidenzia vantaggi in termini di sicurezza e manutenibilità, a fronte di una sintassi più complessa e una curva di apprendimento più ripida e lenta.

Nella seconda parte (Capitolo 5), verranno presi in considerazione progetti concreti (*Rust for Linux*, *Redox OS*, il kernel di Windows 11 e *sudo-rs*), evidenziando che Rust è già impiegato in contesti di sistemi operativi e che la sua adozione procede, seppur gradualmente, in maniera significativa.

RUST

In questo capitolo verrà offerta una panoramica sul linguaggio di programmazione Rust, con particolare attenzione riguardo le motivazioni che ne hanno determinato lo sviluppo e che sono alla base della sua crescente popolarità.

Verranno inoltre introdotti alcuni concetti fondamentali riguardanti il modello di gestione della memoria di Rust, che costituiranno il fulcro del Capitolo 3.

In chiusura, saranno presentati alcuni progetti attuali che evidenziano come Rust rappresenti una valida alternativa nel contesto della programmazione di sistemi.

Le informazioni contenute in questo capitolo sono tratte principalmente da [17].

2.1 CONTESTO E MOTIVAZIONI

Rust è un linguaggio di programmazione che, negli ultimi anni, ha ottenuto un'ottima reputazione tanto tra gli sviluppatori quanto tra le aziende. Il suo principale punto di forza risiede nella capacità, considerata a lungo un ideale solo teorico, di produrre codice che sia simultaneamente ad alte prestazioni e sicuro, in particolare sotto l'aspetto di gestione della memoria.

Che cosa rende Rust *speciale*, rispetto ad altri linguaggi di programmazione? Principalmente, il suo approccio innovativo e alternativo alla gestione della memoria, che evita sia la necessità di gestione manuale da parte dello sviluppatore che la dipendenza da meccanismi di automatici, come *garbage collection*.

Rust è spesso definito come un linguaggio di basso livello che offre astrazioni di alto livello a costo quasi nullo. La caratteristica principale che

lo distingue dalla maggior parte degli altri linguaggi è il meccanismo di gestione della memoria dinamica.

Tradizionalmente esistono due approcci principali alla gestione della memoria:

- **Gestione manuale:** tipica di linguaggi considerati di livello più basso, come C e C++, progettati per dare al programmatore il massimo controllo sulle risorse, compresa la memoria. Secondo questo approccio è il programmatore a decidere quando allocare e quando deallocare la memoria dinamica. Il codice risultante generalmente è molto performante in termini di velocità d'esecuzione, ma questo avviene a un costo: non si hanno garanzie sulla sicurezza e sull'integrità della memoria. L'accesso e la gestione diretti della memoria possono portare problemi come *unfreed memory*, *access after free*, *double free*, *dangling pointer* e non solo.
- **Gestione automatica:** tipica di linguaggi di livello più alto come Java, Python e C#, progettati per facilitare la gestione della memoria dinamica. È infatti gestita da un componente runtime, noto come *Garbage Collector* (GC): la memoria viene allocata implicitamente quando vengono creati oggetti, e liberata quando questi non sono più validi, in maniera trasparente agli occhi del programmatore. Il GC tiene traccia degli oggetti allocati e libera la loro memoria quando questi non sono più referenziati. Il codice prodotto generalmente è privo dalla maggior parte degli errori legati alla memoria, ma questo avviene a un costo: il codice relativo al GC viene incluso insieme al codice dell'applicazione, generando un file eseguibile di dimensioni maggiori; inoltre, il GC richiede risorse computazionali aggiuntive per compiere la sua funzione, eventualmente rallentando, se non bloccando momentaneamente, l'esecuzione dell'applicazione.

Non esiste un approccio 'oggettivamente corretto': durante lo sviluppo di un'applicazione la scelta sull'approccio utilizzato spesso ricade sulle specifiche esigenze. In applicazioni di basso livello, quali lo sviluppo di sistemi operativi, dove le prestazioni sono fondamentali, la scelta si riduce, nella pratica, solamente all'approccio manuale, rendendo linguaggi come C e i suoi derivati lo standard.

Il creatore di Rust si pose come obiettivo sviluppare un linguaggio che eliminasse il compromesso tra prestazioni e sicurezza della memoria, fornendo un'alternativa agli approcci tradizionali.

2.2 ORIGINE E PRIMO SVILUPPO

Rust nacque inizialmente come progetto personale di Graydon Hoare, sviluppatore presso l'azienda Mozilla, nel 2006. Il progetto nacque in seguito alla frustrazione di Hoare per i frequenti malfunzionamenti di software critici, spesso dovuti a problematiche di gestione della memoria¹.

L'obiettivo del progetto era molto ambizioso: sviluppare un linguaggio che consentisse di scrivere codice che fosse contemporaneamente *veloce e sicuro*, principalmente sotto l'aspetto di gestione della memoria.

Il nome '*Rust*' fu ispirato da un particolare tipo di fungo (chiamato appunto *Rust*, 'ruggine' in inglese), definito dal creatore stesso come '*sovra-ingegnerizzati per la sopravvivenza*', metafora per la resilienza e l'attenzione alla sicurezza che il linguaggio mirava a ottenere.

Le prime versioni di Rust adottavano un sistema *Garbage Collection* integrato nel linguaggio, dotato di una sintassi esplicita e dedicata. Tuttavia, dopo i primi esperimenti, il team di sviluppo decise di rinunciare completamente al GC, ritenendolo incompatibile con gli obiettivi di performance e controllo sulla memoria a cui mirava il linguaggio.

Fu così introdotto un nuovo modello, noto come '*modello di ownership*', il quale si fonda sui concetti di possesso (*ownership*) e prestito (*borrowing*). Questo modello definisce regole rigorose sul trasferimento e l'utilizzo dei valori, cercando di prevenire gli errori legati alla memoria già durante la fase di compilazione, senza necessità di un *runtime* dedicato (come il GC).

Questo modello divenne la base del sistema di memoria sicura senza GC che ancora oggi contraddistingue Rust.

2.3 DIFFUSIONE E ADOZIONE INIZIALI

Dopo il rilascio della prima versione stabile, nel 2015, Rust iniziò a essere adottato in contesti reali, seppure con una diffusione iniziale veramente limitata. Alcune aziende cominciarono a sperimentare con il linguaggio, per riscrivere componenti critici, mirando a ottenere un miglioramento in termini di sicurezza e prestazioni:

- **Mozilla**, promotrice del linguaggio, adottò Rust nella realizzazione del motore di rendering CSS *Servo*, dal quale derivano alcuni

¹ Secondo quanto riportato in [17], la frustrazione di Hoare era dovuta ai continui crash del software (presumibilmente scritto in C) dell'ascensore del palazzo in cui risiedeva, al ventunesimo piano.

componenti integrati nel motore di rendering di *Firefox*, *Gecko*;

- **Facebook** riscrisse in Rust parte degli strumenti interni dedicati alla gestione della propria codebase.

Questi primi esempi segnarono una fase di *‘adozione esplorativa’*, in cui Rust cominciava a essere percepito come una possibile alternativa ai linguaggi esistenti.

2.4 ESPANSIONE E INTERESSE CRESCENTI

Nel biennio 2020–2021 si verificò una prima vera e propria espansione dovuta a una maggiore stabilità del linguaggio. Le prime adozioni, da parte di Mozilla e Facebook, giocarono un ruolo cruciale: mostrarono le effettive capacità di Rust.

Come conseguenza, sempre più aziende cominciarono a introdurre codice Rust nella propria codebase, principalmente per i componenti critici sotto gli aspetti di prestazioni e sicurezza:

- **Dropbox** migrò il motore di sincronizzazione da Python a Rust, ottenendo un significativo miglioramento delle prestazioni;
- **Fusion Engineering**² utilizzò Rust per riscrivere i componenti relativi al sistema di controllo dei droni, precedentemente scritti in C++;
- **Discord** riscrisse la pipeline di encoding video e porzioni di backend relative alla gestione dell’autenticazione, dei messaggi e delle notifiche³;
- **Amazon** iniziò a introdurre Rust nei servizi della piattaforma AWS, rafforzando il legame tra il linguaggio e l’ambiente cloud.

Tuttavia, in questo periodo, la maggior parte degli sviluppatori Rust contribuivano al progetto senza un supporto economico diretto. I membri

² Si tratta di un’azienda che sviluppa sistemi software di controllo e gestione per droni. Fondata da Mara Bos, membro del *Rust Project* e co-leader del *Rust Library Team*.

³ Secondo dichiarazioni di Discord stessa, questa modifica ha generato un’incremento delle prestazioni fino a 10 volte. Tuttavia, è opportuno sottolineare che questo dato non va interpretato come un indicatore assoluto o una qualità intrinseca di Rust, in quanto un miglioramento di questa portata non è sempre replicabile nella pratica. Le prestazioni dipendono anche da altri fattori, quali il linguaggio di partenza e il livello di ottimizzazione del codice.

principali del team erano in gran parte impiegati in altre aziende (Mozilla, Amazon, Huawei e Microsoft) e si dedicavano a Rust come attività collaterale.

2.5 SUPPORTO ISTITUZIONALE E PROFESSIONALE

A partire dal 2021, Rust entrò in una nuova fase, contraddistinta da un crescente supporto industriale. Diverse aziende del settore tecnologico, in particolare alcune *Big Tech*, iniziarono a finanziare in modo diretto lo sviluppo del linguaggio, offrendo compensi economici ai membri della community che divennero, a tutti gli effetti, sviluppatori Rust a tempo pieno.

L'obiettivo di questi finanziamenti era quello di permettere agli sviluppatori di dedicarsi in modo continuativo al progetto, garantendo codice stabile, aggiornato e mantenibile.

Questo periodo segnò un passaggio cruciale: da un progetto guidato dalla passione della community a un ecosistema supportato anche da risorse industriali. Di conseguenza, Rust cominciò a consolidarsi come una scelta sempre più pragmatica per applicazioni ad alte prestazioni e con forti requisiti in termini di sicurezza della memoria.

Negli ultimi anni, Rust ha trovato crescente impiego anche a livello di sistema operativo, in particolare in ambito kernel. Tra gli esempi più rilevanti si trovano l'integrazione di Rust nel kernel di Windows 11 da parte di Microsoft, il progetto *Rust for Linux*, lo sviluppo di driver come *Nova* da parte di Red Hat o il tentativo di rendere *sudo* più sicuro da parte di Ubuntu, fino ad arrivare allo sviluppo di interi sistemi operativi in Rust, come *Redox OS*⁴.

⁴ Questi progetti verranno esplorati approfonditamente nel Capitolo 5.

GESTIONE DELLA MEMORIA IN RUST

In questo capitolo viene esaminato l'approccio implementato da Rust per gestire la memoria dinamica. In particolare, verranno approfonditi concetti chiave come *Ownership*, *Borrowing* e *Lifetime*, illustrandone il funzionamento. L'obiettivo è evidenziare come Rust possa garantire la sicurezza della memoria senza sacrificare le prestazioni.

Le informazioni presentate in questo capitolo sono tratte principalmente dalla documentazione ufficiale di Rust [9].

L'approccio implementato da Rust per gestire la memoria dinamica è noto come 'modello di ownership' (*Ownership Model*). Il modello si fonda sui concetti di *ownership*, *borrowing* e *lifetime*, i quali stabiliscono regole sulla gestione della memoria che devono essere rispettate affinché il codice sia compilabile. Il rispetto di tali regole costituisce un prerequisito per la compilazione. Il compilatore controlla che ogni vincolo sia soddisfatto e procede alla compilazione solo in caso positivo. Il modello garantisce un uso corretto e sicuro della memoria già a tempo di compilazione, evitando l'introduzione di overhead durante l'esecuzione¹.

3.1 OWNERSHIP

Il primo concetto alla base del modello di ownership di Rust è l'*ownership*² stessa, la quale definisce vincoli sui legami tra variabili e valori. Rust introduce il concetto di *proprietario*: ogni valore ha un proprietario, ovvero la variabile che lo possiede in un determinato istante.

¹ È opportuno precisare che un minimo overhead viene comunque introdotto, sotto forma di invocazione automatica di una funzione dedicata alla deallocazione.

² Sebbene innovativo, questo concetto non è unico di Rust, il quale ha infatti preso spunto da un pattern ampiamente utilizzato in C++: *RAII* (Resource Allocation Is Initialization), secondo il quale gli oggetti acquisiscono le risorse di cui hanno bisogno (ovvero vengono inizializzati) al momento della loro allocazione.

Si consideri, come esempio, la seguente assegnazione: $a = 5$. Nella maggior parte dei linguaggi di programmazione, l'istruzione verrebbe interpretata come *'a vale 5'*. In Rust il significato è diverso: *'a possiede il valore 5'* o *'a è il proprietario di 5'*.

Stabilire i proprietari è fondamentale per implementare un meccanismo di deallocazione della memoria deterministico e prevedibile: chi libera la memoria relativa a un determinato valore? Il suo proprietario.

Le regole di *ownership*, quindi, stabiliscono i vincoli sui legami tra valori e proprietari. Esse sono tre:

- Ogni valore ha un proprietario (owner);
- Può esserci un solo proprietario per volta (per ogni valore);
- Quando il proprietario di un valore esce dallo scope³, il relativo valore viene scartato.

In Rust la memoria viene liberata automaticamente quando il proprietario di un valore esce dallo scope. Per implementare questo, Rust invoca una funzione speciale, *drop*, in corrispondenza della fine di ogni scope. La funzione viene invocata su ogni variabile che esce dallo scope ed è il proprietario di un valore.

Grazie alle regole di *ownership*, Rust può garantire che la deallocazione della memoria avvenga sempre in modo corretto: poiché ogni valore ha un solo proprietario, non vi è ambiguità riguardo a chi sia responsabile della sua deallocazione. Valgono le seguenti due proprietà:

- l'esistenza di un proprietario garantisce che non vi siano valori allocati ma non più referenziati;
- l'unicità del proprietario impedisce che un'area di memoria venga deallocata più volte.

Le regole di *ownership* influenzano diversi aspetti del linguaggio, in particolare la condivisione dei valori tra variabili, l'assegnazione, il passaggio di parametri alle funzioni e operazioni analoghe.

Per comprendere a fondo il comportamento del modello di *ownership*, è utile esaminare come Rust gestisce i diversi tipi di dato. I tipi di dato si distinguono tra *semplici* e *complessi*:

³ Lo scope rappresenta un range nel programma all'interno del quale un oggetto è valido. Solitamente è delimitato da una coppia di parentesi graffe.

- I tipi semplici hanno dimensione fissa, nota a tempo di compilazione, e vengono quindi allocati nello stack;
- I tipi complessi, invece, hanno dimensione variabile o non determinabile a tempo di compilazione e vengono quindi allocati nell'heap.

In relazione alla gestione della memoria, Rust definisce due comportamenti distinti per l'assegnazione:

- *Copy*: l'intera memoria del valore viene duplicata tramite una copia bitwise; questo comportamento è adottato di default per i tipi semplici;
- *Move*: il valore viene trasferito da una variabile all'altra, invalidando quella di origine; questo comportamento è adottato di default per i tipi complessi.

I frammenti di codice nei Listati 1 e 2 illustrano il comportamento dell'assegnamento nei tipi semplici, che implementano Copy, e in quelli complessi, che implementano Move:

Listato 1: Comportamento di Copy

```
fn main() {  
    let x: i32 = 5;  
    let y: i32 = x;  
    println!("x: {}, y: {}", x, y);  
}
```

Listato 2: Comportamento di Move

```
fn main() {  
    let x: String = String::from("Sono in Heap!");  
    let y: String = x;  
    println!("x: {}, y: {}", x, y); // <-- Errore di compilazione  
}
```

Nel Listato 1 si osserva che `x` resta valido dopo l'assegnamento, poiché `i32` implementa `Copy`, trattandosi di un tipo semplice, allocato nello stack.

Nel Listato 2, invece, l'assegnamento provoca il trasferimento dell'ownership, che viene trasferita a `y`, invalidando `x`. `String` è infatti un tipo complesso, allocato nell'heap, che non implementa `Copy` ma utilizza il comportamento di `Move` per impostazione predefinita.

Come visto nel Listato 2, l'assegnamento di un tipo complesso trasferisce l'*ownership*, rendendo invalido il riferimento originale. Questo comportamento, tuttavia, può risultare limitante in alcuni scenari.

Per ovviare a tale limitazione, Rust fornisce un meccanismo per effettuare copie profonde anche con tipi che utilizzano Move: il *trait*⁴ *Clone*⁵.

Clone permette di definire esplicitamente come realizzare la copia profonda di un valore. L'invocazione del metodo *clone()* genera un nuovo valore, indipendente dal primo, ma con lo stesso contenuto.

Si consideri l'esempio nel Listato 3, che mostra il comportamento di *Clone*:

Listato 3: Comportamento di *Clone*

```
fn main() {
    let x: String = String::from("Sono in Heap!");
    let y: String = x.clone();
    println!("x: {}, y: {}", x, y); // Entrambi i riferimenti sono
                                   validi
}
```

Nel Listato 3 si osserva che sia *x* che *y* sono validi in seguito all'assegnamento, in quanto la chiamata a *clone()* ha generato una nuova stringa, con lo stesso contenuto, ma indipendente dalla prima.

Un'ulteriore limitazione del meccanismo di *ownership* si presenta durante l'invocazione di una funzione, in particolare, nel passaggio di un valore come parametro. Rust gestisce il passaggio di un parametro a una funzione in maniera analoga agli assegnamenti: le stesse regole di Copy e Move continuano ad applicarsi.

Si consideri l'esempio nel Listato 4 che illustra il comportamento di Rust nel passaggio a funzione di un valore di tipo semplice e di uno di tipo complesso:

Listato 4: Trasferimento di *ownership* nelle chiamate a funzione

```
fn my_func(first: i32, second: String) {
```

⁴ In Rust, un *trait* ('tratto') è un meccanismo analogo alle interfacce di Java, utilizzato per definire un comportamento comune per le *struct* e garantire un certo grado di polimorfismo. Nella pratica, un *trait* è un insieme di funzioni, o meglio, firme di funzioni, raggruppate e identificate con un nome. Le *struct* che intendono implementare un *trait* devono definire il corpo di ogni funzione presente in quest'ultimo.

⁵ La clonazione può introdurre costi in termini di prestazioni, in quanto richiede operazioni di lettura e scrittura nella memoria Heap.

```
println!("Parametro 'first' {}", first);
println!("Parametro 'second' {}", second);
}

fn main() {
    let x: i32 = 5;
    let y: String = String::from("move!");
    my_func(x, y);
    println!("Variabile originale x: {}", x); // <-- Ok
    println!("Variabile originale y: {}", y); // <-- Errore di
        compilazione: ownership trasferita
}
```

Nel Listato 4 si osserva che, passando un valore di tipo semplice come `x` (che implementa `Copy`), la variabile rimane valida anche dopo la chiamata alla funzione. Al contrario, passando `y`, di tipo `String` (che implementa `Move`), l'*ownership* viene trasferita al parametro `second`. Terminata l'esecuzione di `my_func`, `second` esce dallo scope e il valore viene deallocato, rendendo `y` una variabile non più valida.

Tale comportamento può risultare limitante in molte situazioni pratiche, dove si desidera utilizzare un valore senza trasferirne la proprietà. Per risolvere questo problema, Rust introduce un secondo meccanismo fondamentale: il *borrowing*.

3.2 BORROWING

Dopo l'*ownership*, il secondo concetto fondamentale su cui si fonda il modello di gestione della memoria di Rust è il *borrowing*. Questo meccanismo consente la condivisione sicura dei dati, permettendo di accedere temporaneamente a un valore senza acquisirne la proprietà.

Rust implementa concretamente il *borrowing* tramite il concetto di *reference*. Una *reference* può essere vista come un puntatore: rappresenta un indirizzo che può essere seguito per accedere al valore memorizzato a tale indirizzo; tale valore appartiene a un'altra variabile.

In generale, le variabili in Rust, sia di tipo primitivo che *reference*, sono immutabili per default: non possono essere modificate una volta assegnato un valore. Nel caso sia necessario modificare una variabile o una *reference*, è possibile utilizzare la parola chiave `mut` durante la sua dichiarazione, indicando che tale variabile (o *reference*) può essere modificata

durante l'esecuzione.

Il meccanismo che si occupa di controllare che le regole di *borrowing* vengano rispettate è noto come *Borrow Checker*. Le regole di *borrowing* sono le seguenti:

- in ogni istante può esistere una sola reference mutabile a un dato valore;
- se esiste almeno una reference immutabile, allora non possono esistere reference mutabili;
- tutte le reference devono essere sempre valide;
- una reference mutabile può essere creata solo da una variabile dichiarata come `mut`.⁶

Si considerino gli esempi nei Listati 5, 6, 7 e 8, che illustrano violazioni delle regole di *borrowing*. Essi hanno lo scopo di mostrare come tali comportamenti non siano ammessi dal Borrow Checker, generando errori in fase di compilazione.

Listato 5: Reference mutabile a variabile immutabile

```
fn main() {
    let x: u8 = 5;
    let mut y: &mut u8 = &mut x;
    *y += 1;
}
```

Listato 6: Due reference mutabili alla stessa variabile

```
fn main() {
    let mut x: u8 = 5;

    let first_ref = &mut x;
    let second_ref = &mut x;

    *first_ref += 1;
    *second_ref += 1;

    println!("Il valore di x ora \\'e {}", x);
}
```

⁶ Questo vincolo è spesso dato per implicito, ma è fondamentale nel comportamento del *Borrow Checker*.

Listato 7: Coesistenza reference mutabile e immutabile

```
fn main() {
    let mut x: u8 = 5;

    let mutable_ref = &mut x;
    let immutable_ref = &x;

    *mutable_ref += 1;

    println!("Il valore di x ora \\'e {}", immutable_ref);
}
```

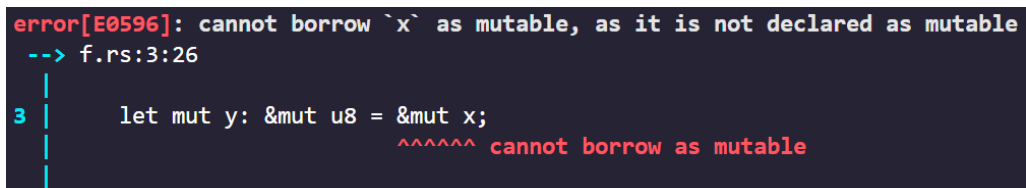
Listato 8: Reference non valida: *dangling reference*

```
fn main() {
    let dangling_reference: &u8;

    {
        let x: u8 = 5;
        dangling_reference = &x;
    }

    println!("Il valore che x aveva \\'e {}", *dangling_reference);
}
```

Nel Listato 5 si tenta di ottenere un riferimento mutabile a una variabile immutabile, violando un vincolo fondamentale del borrow checker. Questo genera un errore di compilazione, come riportato in Figura 1.



```
error[E0596]: cannot borrow `x` as mutable, as it is not declared as mutable
--> f.rs:3:26
3 |     let mut y: &mut u8 = &mut x;
  |                        ^^^^^^ cannot borrow as mutable
```

Figura 1: Tentativo di *borrow* mutabile di un valore non mutabile

Nel Listato 6 si tenta di ottenere due riferimenti mutabili simultaneamente, violando la prima regola di *borrowing*, come riportato nella Figura 2. Il Listato 7 mostra invece il tentativo di ottenere contemporaneamente una reference mutabile e una immutabile, violando la seconda regola di *borrowing*. L'errore di compilazione generato è riportato in Figura 3.

```

error[E0499]: cannot borrow `x` as mutable more than once at a time
--> f.rs:5:22
4 |     let first_ref = &mut x;
   |                      ----- first mutable borrow occurs here
5 |     let second_ref = &mut x;
   |                      ^^^^^^ second mutable borrow occurs here

```

Figura 2: Tentativo di due *borrow* mutabili allo stesso valore

Infine, nel Listato 8 si presenta una *dangling reference*, che viola la terza

```

error[E0502]: cannot borrow `x` as immutable because it is also borrowed as mutable
--> f.rs:5:25
4 |     let mutable_ref = &mut x;
   |                      ----- mutable borrow occurs here
5 |     let immutable_ref = &x;
   |                      ^^ immutable borrow occurs here

```

Figura 3: Tentativo di *borrow* mutabile e non mutabile simultaneamente allo stesso valore

regola di *borrowing*. Il risultato è un errore di compilazione, come mostrato in Figura 4. Come evidenziato dal Listato 8, alcune violazioni delle

```

error[E0597]: `x` does not live long enough
--> f.rs:6:30
5 |     let x: u8 = 5;
   |       - binding `x` declared here
6 |     dangling_reference = &x;
   |                        ^^ borrowed value does not live long enough
7 | }
   | - `x` dropped here while still borrowed

```

Figura 4: Generazione di un *dangling reference*

regole di *borrowing* non riguardano soltanto la mutabilità o il numero di reference presenti, ma coinvolgono anche la durata nel tempo di una reference rispetto al valore referenziato.

Rust gestisce queste relazioni temporali tramite il concetto di *lifetime*, che consente al compilatore di determinare se una reference sarà valida finché necessaria, evitando il rischio di *dangling reference*.

Il *Borrow Checker*, quindi, sfrutta anche il sistema delle *lifetime* per garantire che la condivisione dei valori avvenga sempre in modo sicuro e coerente. Questo meccanismo sarà approfondito nella prossima sezione.

3.3 LIFETIME

Il sistema delle *lifetime* rappresenta il terzo concetto fondamentale su cui si basa il modello di ownership di Rust. Questo meccanismo viene utilizzato dal *Borrow Checker* per determinare la validità temporale di un prestito (*borrow*).

Nella maggior parte dei casi, le *lifetime* sono implicite e dedotte automaticamente dal compilatore⁷, in questo caso, vengono definite *lifetime generiche*. In contesti particolari, tuttavia, è necessario esplicitarle, specialmente quando può crearsi ambiguità riguardo la durata di una reference, ad esempio quando si lavora simultaneamente con reference che appartengono a scope differenti.

Si consideri, a riguardo, l'esempio nel Listato 9, che mostra come le *lifetime* implicite non siano sufficienti in alcuni contesti:

Listato 9: Limitazione delle *lifetime* generiche

```
fn biggest(a: &u8, b: &u8) -> &u8 {  
    if *a > *b { a } else { b }  
}
```

Nel Listato 9 si tenta di restituire un riferimento all'intero più grande tra i due parametri forniti. Sebbene il codice possa sembrare corretto, il *Borrow Checker* lo rifiuta, in quanto non può garantire la validità della reference restituita nel tempo. Il problema nasce dal fatto che le reference *a* e *b* potrebbero riferirsi a variabili definite in scope differenti, con durate di vita (*lifetime*) diverse. Il compilatore, con le *lifetime generiche*, cerca di assegnare la stessa durata ad entrambe le reference, generando ambiguità.

Per esplicitare la durata di vita di una reference si utilizza un'annotazione di *lifetime* posta prima del tipo. Questa è rappresentata da un apostrofo seguito da un identificativo; per convenzione si usa una sola lettera, seguendo l'ordine alfabetico ('a', 'b' e così via). Riferendoci all'esempio del Listato 9, la versione corretta viene presentata nel Listato 10, con *lifetime esplicite*:

⁷ Per questo motivo, durante lettura di codice Rust, le annotazioni di *lifetime* sono spesso non visibili.

Listato 10: *Lifetime* esplicite

```
fn biggest<'a>(a: &'a u8, b: &'a u8) -> &'a u8 {
    if *a > *b { a } else { b }
}
```

Nel Listato 10 viene dichiarata una singola *lifetime*, *'a*, che indica al *Borrow Checker* che i parametri *a* e *b* devono avere la stessa durata di vita. Di conseguenza, se le *lifetime* dei parametri differiscono, la compilazione sarà rifiutata.

È importante fare una precisazione: esplicitare una *lifetime* **non modifica** l'effettiva durata di una variabile. Si tratta solamente di un'indicazione semantica che informa il compilatore dei vincoli temporali che devono essere rispettati tra le *reference* coinvolte. Il *Borrow Checker* utilizza queste annotazioni per verificare la validità dei prestiti nel tempo e può rifiutare il codice se i vincoli non sono coerenti.

A tale scopo, si consideri l'esempio nel Listato 11, che mostra come il *Borrow Checker* possa rifiutare una *reference* che non rispetti le *lifetime* specificate:

Listato 11: Limitazioni delle *lifetime*

```
fn biggest<'a>(a: &'a u8, b: &'a u8) -> &'a u8 {
    if *a > *b { a } else { b }
}

fn main (){
    let result;
    let smaller = 1;
    {
        let bigger = 2;
        result = biggest(&smaller, &bigger);
    }
    println!("Il piu grande e: {}", result);
}
```

Nel Listato 11, la funzione *biggest* viene invocata con due *reference* che hanno *lifetime* differenti. In particolare, *bigger* viene definita in uno scope interno, e scade prima che possa essere utilizzata la *reference* restituita da *biggest*. Il *Borrow Checker* si accorge di questa incoerenza e segnala un errore di compilazione, come mostrato in Figura 5.

In contesti dove le *lifetime* non vengono esplicitamente annotate, il com-

```

error[E0597]: `bigger` does not live long enough
--> f.rs:10:36
9 |         let bigger = 2;
  |         ----- binding `bigger` declared here
10 |         result = biggest(&smaller, &bigger);
    |                                ^^^^^^^ borrowed value does not live long enough
11 |     }
  |     - `bigger` dropped here while still borrowed
12 |     println!("Il piu grande e: {}", result);
    |                                ----- borrow later used here

```

Figura 5: Esempio di violazione delle regole di *Lifetime*

il compilatore applica automaticamente delle regole note come *lifetime elision rules*, che permettono di dedurre in modo implicito i vincoli temporali tra le *reference*. Queste regole sono tre:

- A ciascun parametro che è una *reference* viene assegnata una *lifetime* distinta;
- Se c'è una sola *reference* tra i parametri di ingresso, la sua *lifetime* viene assegnata automaticamente al valore di ritorno (se anch'esso è una *reference*);
- Se ci sono più *reference* in ingresso, ma una di esse è `&self`⁸ o `&mut self`, allora la *lifetime* di `self` viene propagata alle eventuali *reference* in uscita.

Infine, esiste una *lifetime* speciale, `'static`, la quale indica una *reference* valida per l'intera durata del programma. Questa *lifetime* tipicamente viene utilizzata per dati costanti o risorse allocate a tempo indeterminato.

Nel prossimo capitolo verrà mostrato come, grazie a questi concetti, Rust rappresenti un'alternativa valida per la programmazione di sistema, riuscendo a competere con linguaggi come C e C++.

⁸ In Rust i metodi delle *struct* ricevono un parametro implicito, `self`, che rappresenta l'istanza della struttura sulla quale il metodo è invocato; è un meccanismo analogo a `self` in Python e a `this` in Java. Anche `self` rispetta le regole di *ownership* e *borrowing*: il parametro `self` trasferisce l'*ownership* dell'istanza, il parametro `&self`, invece, indica un riferimento all'istanza (analogamente `&mut self` rappresenta un riferimento mutabile). `self` è un concetto che permette al programmatore di invocare una funzione su un valore, come se fosse un metodo (`x.function()` è equivalente a `function(x)`).

SISTEMI OPERATIVI

Questo capitolo esplorerà le capacità offerte da Rust nel contesto della programmazione di sistema. In apertura verrà fornita una panoramica sul linguaggio C, standard *de facto* nel contesto programmazione di sistema, con particolare attenzione alle caratteristiche che lo rendono popolare in questo ambito.

Successivamente Rust verrà confrontato con C sulla base degli aspetti analizzati, per valutare come Rust possa presentare un'alternativa valida sul piano teorico.

Nel Capitolo 5, l'analisi verrà spostata sul piano pratico, esplorando progetti concreti, che mostrano le potenzialità di Rust nella programmazione di basso livello.

Nel contesto dei sistemi operativi, quali lo sviluppo di kernel, driver o componenti embedded, i linguaggi tradizionalmente dominanti sono l'assembly e, con maggiore rilievo, C.

Nonostante l'esistenza di alternative come C++, Lisp, Forth o Bliss, la scelta spesso ricade esclusivamente su C, come testimoniano le codebase dei sistemi operativi più popolari oggi giorno: Windows, MacOS, Linux e Android sono scritti quasi interamente nel linguaggio C.

Ma cosa rende C così popolare in questo ambito? Quali caratteristiche lo distinguono dalle alternative?

4.1 C: MOTIVAZIONI E CARATTERISTICHE

C è un linguaggio di programmazione ampiamente utilizzato nella programmazione di sistema. Il linguaggio offre un livello di astrazione molto vicino al codice macchina (o meglio, all'assembly), garantendo un certo grado di portabilità tra architetture differenti.

Si consideri, per esempio, quanto detto dal creatore del linguaggio:

'[C has] the power of the assembly language and the convenience of ... assembly language.'

— *Dennis Ritchie, creatore del linguaggio C, 'Dennis Ritchie: The Shoulders Steve Jobs Stood On', Wired, 13 Oct 2011*

La popolarità di C in questo ambito è da attribuire a diverse sue caratteristiche, alcune delle quali ben documentate [15]. Queste sono principalmente: compilazione, assenza di dipendenze runtime, gestione diretta della memoria, manipolazione a basso livello di bit e corrispondenza quasi diretta al codice macchina.

COMPILAZIONE C è un linguaggio compilato: il file eseguibile generato risulta molto efficiente, sotto l'aspetto della velocità d'esecuzione, rispetto a linguaggi interpretati. Questo lo rende preferibile nell'ambito di sviluppo kernel, dove le prestazioni rappresentano un aspetto cruciale.

ASSENZA DI DIPENDENZE RUNTIME C può essere impiegato per realizzare codice *'bare metal'*, eseguibile direttamente sull'hardware, senza il supporto di un sistema operativo.

C non ha esigenze runtime significative: può funzionare anche senza un allocatore di memoria fornito dal sistema operativo¹. L'unica necessità è quella di un chiamante che invochi la funzione *main*².

ACCESSO DIRETTO ALLA MEMORIA I puntatori C consentono l'accesso diretto a indirizzi di memoria arbitrari, permettendo operazioni di lettura e scrittura dirette.

Tale controllo sulla memoria è cruciale per lo sviluppo di un sistema operativo, dove è richiesto di gestire le tabelle delle pagine, i dispositivi I/O mappati sulla memoria, i controllori DMA e altri meccanismi simili;

MANIPOLAZIONE DI BIT Molte interazioni con l'hardware avvengono tramite operazioni bitwise, come riportato da *Ada Computers* [1]. Esempi comuni sono:

- Operazioni di scrittura e lettura dei registri della CPU, come registri di flag di stato. Per esempio, in CPU x86 si trova il registro *EFLAGS*

1 Un programma che non usufruisce della memoria dinamica non richiede un allocatore. Inoltre, nello sviluppo di un sistema operativo, è il programmatore che deve realizzare il proprio sistema di allocazione.

2 In contesti a basso livello, questa chiamata può essere realizzata da un semplice bootstrap assembly.

che contiene una serie di bit di stato (*overflow*, *zero*, *carry*, *interrupt enable* e altri), il loro controllo di solito è fatto tramite l'applicazione di una maschera bitwise;

- Gestione delle periferiche mappate sulla memoria, tramite operazioni di abilitazione e disabilitazione dei singoli pin (ad esempio, registri *GPIO*);
- Il clock della CPU può essere gestito tramite operazioni bitwise.

Come riportato in [5], il linguaggio C supporta un ampio range di operazioni bitwise, offrendo i seguenti operatori: *AND* (&), *OR* (|), *XOR* (^), *SHL* (<<), *SHR* (>>), *NOT* (!) e il complemento (~).

SOMIGLIANZA AL CODICE MACCHINA C mantiene una corrispondenza quasi *1-to-1* con codice assembly. Questa trasparenza del linguaggio è fondamentale nello sviluppo di sistemi operativi, in quanto consente di comprendere l'effetto di ogni singola istruzione. C evita strutture dati complesse o astrazioni pesanti che potrebbero mascherare il comportamento a basso livello del programma.

DEBOLEZZA DI C Tuttavia, uno dei punti di forza di C rappresenta anche una sua criticità. L'accesso diretto e non controllato della memoria, unito all'assenza di protezioni a runtime, rende semplice commettere errori potenzialmente gravi come:

- **Accessi non autorizzati alla memoria:** un processo che legge da un indirizzo di memoria arbitrario potrebbe inavvertitamente compromettere la privacy di un'altro processo; una scrittura accidentale potrebbe causare la corruzione della memoria, condivisa o utilizzata da un altro processo;
- **Saturazione della memoria:** un processo che alloca memoria senza controllo o limiti potrebbe esaurire la memoria disponibile. Se lo *swapping* è disponibile, il sistema può degradare drasticamente le prestazioni; in caso non lo fosse, può verificarsi un crash.

4.2 RUST E C A CONFRONTO

Nonostante Rust offra numerose astrazioni di alto livello³, è progettato come un linguaggio di basso livello, adatto alla programmazione di sistema. Per comprendere le capacità di Rust nella programmazione di basso livello, verrà confrontato con il linguaggio C, sulla base delle caratteristiche che rendono quest'ultimo conveniente nella programmazione di sistema.

La maggior parte delle informazioni riportate nella sezione è stata ricavata dalla documentazione ufficiale di Rust [9].

- **Compilazione** (Come C): Anche Rust è un linguaggio compilato. Il compilatore genera, in base alla piattaforma, un file direttamente eseguibile;
- **Dipendenze runtime** (Come C): Rust, per configurazione predefinita, ha dipendenze runtime minime (principalmente un allocatore di memoria). Tuttavia, come riportato in *The Embedded Rust Book* [12], tramite la direttiva `#![no_std]` è possibile escludere la libreria standard: in questa configurazione, l'unico requisito è un bootstrap che invochi la funzione `_start` per fare iniziare l'esecuzione;
- **Accesso diretto alla memoria** (Come C): Tramite una parola chiave, *unsafe*, Rust mette a disposizione cinque operazioni denominate *superpoteri non sicuri*: tra queste si trovano anche la possibilità di dereferenziare un puntatore raw (come in C) e la possibilità di eseguire codice C o assembly;
- **Manipolazione di bit** (Come C): Rust offre lo stesso livello di manipolazione dei singoli bit di C, con un'unica differenza: le operazioni bitwise in Rust sono ben definite, evitando condizioni di *Undefined Behaviour* tramite controlli statici sulle dimensioni e tipi degli operandi;
- **Somiglianza al codice macchina** (Diverso da C): Rust generalmente offre astrazioni di alto livello, inoltre il compilatore può introdurre copie e spostamenti aggiuntivi per preservare le condizioni di

³ Rust offre delle astrazioni di alto livello definite *zero cost*: feature quali *iteratori*, *generics*, *smart pointers* e meccanismi di *async/await* che vengono compilate in codice dalle prestazioni equivalenti alle controparti *low-level* scritte a mano.

ownership o inserire controlli su accessi e indici per gli *slice*⁴. Tali comportamenti, pur aumentando la sicurezza, possono produrre un codice meno ‘trasparente’ rispetto alla controparte C;

Infine, a differenza di C, Rust garantisce l’integrità e la sicurezza della memoria già a tempo di compilazione, prevenendo errori comuni legati alla gestione della memoria grazie al *modello di ownership*. In quanto i controlli effettuati dal *Borrow Checker* avvengono a tempo di compilazione, non si hanno dipendenze o overhead introdotti a runtime.

UNSAFE RUST Come già accennato, Rust mette a disposizione, tramite la parola chiave *unsafe*, cinque operazioni principali. Gli sviluppatori del progetto Rust si riferiscono a queste operazioni con l’espressione *Unsafe Superpowers*:

- Dereference di un puntatore raw;
- Invocazione di una funzione o metodo non sicuri;
- Accesso e modifica di una variabile mutabile e statica;
- Implementazione di un trait non sicuro;
- Accesso ai campi di una *union*.

Le espressioni *Unsafe Rust* e *Unsafe Superpowers* possono risultare fuorvianti: il *Borrow Checker* esegue comunque controlli per garantire la validità delle reference.

La parola chiave permette solamente di eseguire operazioni che, per definizione, sono considerate non sicure. Il compilatore non può controllarne la sicurezza e l’integrità durante la compilazione; all’interno di un blocco *unsafe*, è il programmatore a diventare responsabile di garantire che gli accessi alla memoria siano validi.

La *best practice* riguardante i blocchi non sicuri prevede di ridurre il codice non sicuro il più possibile, utilizzandolo solo quando strettamente necessario. Inoltre, è preferibile incapsulare il codice non sicuro all’interno di astrazioni sicure, fornendo API sicure per il suo utilizzo.

⁴ *slice* è un tipo primitivo in Rust. Sono un riferimento a una porzione contigua (in memoria) di elementi di una collezione. Non memorizzano dati esplicitamente, si tratta di una vista su dati esistenti.

4.2.1 Gestione delle risorse

In questa sottosezione verranno analizzati e confrontati i meccanismi offerti da C e Rust per la gestione delle risorse. In particolare verrà analizzata la gestione di: memoria dinamica, file su disco, risorse di rete e *thread* con risorse condivise⁵.

Memoria dinamica

Oltre al *modello di ownership*, Rust incapsula la gestione della memoria dinamica tramite astrazioni note come *smart pointers*, con lo scopo di evitare problematiche comuni relative all'allocazione e deallocazione manuali della memoria. In questa sezione verranno trattati solamente i meccanismi di base per l'allocazione, l'accesso e la liberazione della memoria. L'analisi dettagliata degli errori comuni sarà il fulcro della sottosezione successiva: *Sicurezza della memoria* [4.2.2](#).

In C, la libreria standard (`stdlib.h`) fornisce primitive per la gestione manuale della memoria: `malloc` e `calloc` per l'allocazione, `realloc` per il ridimensionamento e `free` per la deallocazione. Entrambe `malloc` e `calloc` richiedono come argomento la quantità di memoria da allocare, espressa in byte.

Se l'allocazione ha successo, restituiscono un puntatore all'area di memoria allocata, altrimenti restituiscono `NULL`. Tuttavia, il controllo dell'esito è lasciato al programmatore, che deve verificare esplicitamente se il puntatore restituito è valido o meno. Nella pratica, spesso, questa verifica viene omessa, con potenziali gravi conseguenze.

Analogamente, `free` si limita a tentare la deallocazione dell'indirizzo fornito, ma non restituisce nessun esito; inoltre, se viene invocata più volte sullo stesso puntatore, può generare un errore noto come *double free*.

Rust adotta un approccio diverso, basato su *smart pointers* come `Box<T>` o `Vec<T>`, che incapsulano i valori memorizzati nell'heap e ne gestiscono automaticamente il ciclo di vita. Queste astrazioni si fondano sul pattern *RAII*, unendo le operazioni di allocazione e inizializzazione per evitare l'accesso a valori non inizializzati.

Si considerino gli esempi C e Rust, riportati rispettivamente nei Listati [12](#)

⁵ Gli esempi di codice riportati in questa sottosezione sono stati sviluppati e testati su ambiente Linux con `glibc`.

e 13, che mostrano la gestione del completo ciclo di vita (allocazione, inizializzazione, accesso e deallocazione) di un intero nella heap:

Listato 12: Gestione ciclo di vita in memoria dinamica C

```
#include <stdlib.h>
#include <stdio.h>
int main(void) {

    int* value = malloc(sizeof(int)); // Allocazione
    if(value == NULL) return -1;

    *value = 52; // Inizializzazione

    printf("%d\n", *value); // Accesso

    free(value); // Deallocazione

    return 0;
}
```

Listato 13: Gestione ciclo di vita in memoria dinamica Rust

```
fn main() {

    let value = Box::<i32>::new(52); // Allocazione e
    //      inizializzazione

    println!("{}", *value); // Accesso

} // <-- Fine scope: Deallocazione automatica
```

Entrambi i Listati 12 e 13 implementano la stessa logica: viene allocato spazio sufficiente per un intero nella memoria heap, successivamente viene inizializzato, stampato e infine deallocato.

Nel caso di C, le varie fasi sono distinte e scollegate: l'allocazione tramite `malloc`, l'inizializzazione e l'accesso tramite dereferenziazione e la deallocazione tramite `free`. Inoltre, è responsabilità del programmatore specificare la dimensione corretta della memoria da allocare, in questo caso tramite `sizeof(int)`. Tuttavia, non vi sono controlli per confermare che la dimensione inserita sia sufficiente o meno, rischiando di generare errori durante l'esecuzione.

Rust, al contrario, segue il paradigma *RAII*: la fase di allocazione

e di inizializzazione sono unificate nella creazione dello *smart pointer* (`Box::new`). La deallocazione avviene automaticamente al termine dello scope, prevenendo sia la mancata, che la doppia, liberazione. L'unico aspetto simile rispetto a C è l'accesso, gestito tramite *dereference*.

File

Sia Rust che C offrono strumenti per eseguire operazioni di input/output su file del filesystem. I due linguaggi si distinguono per il livello di astrazione offerto.

Nel linguaggio C, le principali operazioni di I/O vengono fornite dalle API POSIX, tramite funzioni definite nelle librerie `unistd.h` e `fcntl.h`. L'accesso è basato su *file descriptor*⁶, i quali vengono ottenuti tramite la funzione `open` e successivamente gestiti tramite `read`, `write` e `close`. L'interfaccia rispecchia la natura di basso livello del linguaggio, rappresentando un semplicissimo wrapper alle chiamate di sistema (*syscalls*). L'esito di un'operazione è determinabile attraverso il suo valore di ritorno, solitamente non negativo per indicare il successo e viceversa. La problematica principale deriva dal fatto che la gestione degli errori non è obbligatoria: sta alla discrezione del programmatore controllare la validità dei *file descriptor* o il numero di byte effettivamente letti o scritti, rispetto a quelli attesi. Spesso, purtroppo, queste verifiche vengono omesse e, come conseguenza, vi è il rischio di letture e scritture su *file descriptor* non validi, risultando spesso in *undefined behaviour*.

Rust d'altra parte, espone una API con un livello di astrazione più alto, tramite il modulo `std::fs`. Tale modulo mette a disposizione la struttura `File`, la quale incapsula interamente la gestione dei *file descriptor*. Le operazioni di lettura e scrittura avvengono tramite funzioni espone dal modulo `std::io`, sia con bufferizzazione che senza. In generale, per letture e scritture piccole, è sufficiente usare le funzioni `read` e `write`, mentre nella maggior parte dei casi, è consigliato l'utilizzo delle strutture con supporto alla bufferizzazione: `BufReader` e `BufWriter`. Le strutture fornite da Rust integrano un meccanismo per la gestione degli errori: le operazioni che possono fallire restituiscono il tipo `Result<T, E>`, forzando di conseguenza il programmatore a gestire esplicitamente l'errore, tramite costrutti `if-else` o `match` per definire una logica di gestione dell'errore, oppure propagando l'errore tramite l'operatore `?`.

⁶ Un *file descriptor* è un intero senza segno che identifica univocamente un file nel contesto di un determinato processo.

Il tipo `Result<T, E>` fornito da Rust permette di rappresentare il risultato di un'operazione che può avere successo o fallire; `T` ed `E` sono tipi generici:

- **`Ok(T)`**: rappresenta la variante di `Result` che indica successo, contenente il valore, di tipo `T`, restituito;
- **`Err(E)`**: è la variante di `Result` che indica un fallimento, contenente un valore descrittivo dell'errore, di tipo `E`, generato.

In Rust, il modulo `std::io` definisce `Result<T>`, in cui è sufficiente specificare solo il tipo di `Ok`, in quanto il tipo di `Err` è sempre `std::io::Error`. Insieme a `Result`, spesso viene utilizzato l'operatore `?` per propagare l'errore al chiamante, semplificando la gestione degli errori e migliorando la leggibilità del codice.

Nel caso di `'main'`, il chiamante è il sistema operativo, il quale gestirà l'errore in modo appropriato; però non vi è un valore effettivo da restituire insieme a `Ok` in caso di successo. Per questi scenari, Rust mette a disposizione il tipo `'vuoto'` `()`, il quale rappresenta l'assenza di un valore significativo, simile a `void` in C; di conseguenza in tali contesti viene restituito `Ok(())` (ovvero, *'successo senza un valore significativo'*).

Un altro aspetto fondamentale del linguaggio è quello dei panic, ovvero un ulteriore meccanismo impiegato nella gestione degli errori. Dal punto di vista esterno, sono molto simili alle eccezioni di Java, ovvero causano l'interruzione dell'esecuzione. Tuttavia, internamente al processo, i panic procedono in maniera leggermente diversa rispetto alle eccezioni, causando uno *stack unwinding*⁷: quando viene generato un panic, Rust percorre lo stack delle chiamate delle funzioni, invocando `drop` su ogni *owner* presente. Questo permette una gestione sicura perfino degli errori improvvisi, liberando tutte le risorse allocate al momento dell'errore⁸.

⁷ Conviene sottolineare che, anche in Java, si verifica uno *stack unwinding*, in cui viene percorso lo stack delle chiamate e, contemporaneamente, vengono marcati tutti gli oggetti come non referenziati. Anche questo provoca la deallocazione della memoria; tuttavia, non vi è garanzia riguardo le altre possibili risorse allocate (per esempio, un file aperto potrebbe rimanere tale). In Rust, la maggior parte delle strutture che allocano risorse, definiscono `drop` in maniera tale da liberarle; di conseguenza, anche in presenza di un panic, verrebbero deallocate correttamente.

⁸ Questo, tuttavia, non è sempre vero in pratica. Oltre allo *stack unwinding*, Rust permette di selezionare *abort* come gestore di panic. In questa configurazione, il processo viene terminato senza liberare le risorse, con un comportamento analogo a un `exit(FAILURE)` improvviso in C. La modifica del gestore di panic deve avvenire in maniera esplicita, in quanto Rust utilizza, di default, *stack unwinding*.

Si considerino gli esempi nei Listati 14 e 15 che mostrano un esempio di operazioni di I/O su file:

Listato 14: I/O su file in C

```
#include <unistd.h>
#include <stdlib.h>
#include <fcntl.h>
int main(void) {

    int read_from = open("./read_from.txt", O_RDONLY , S_IRUSR); //
        Apertura file
    int write_to = open("./write_to.txt", O_CREAT | O_WRONLY |
        O_APPEND , S_IWUSR); // Apertura file

    char buf[2048];
    ssize_t read_byte;

    while ( (read_byte = read(read_from, buf, sizeof(buf))) > 0) {
        // Lettura
        write(write_to, buf, read_byte); // Scrittura
    }

    close(write_to); // Chiusura
    close(read_from); // Chiusura
    return 0;
}
```

Listato 15: I/O su file in Rust

```
use std::fs::File;
use std::io::{Read, Write};
fn main() -> std::io::Result<()> {

    let mut read_from = File::open("read_from.txt")?; // Apertura in
        lettura
    let mut write_to = File::create("write_to.txt")?; // Creazione e
        apertura in scrittura

    let mut buf = [0u8; 2048];

    loop {
        let read_byte = read_from.read(&mut buf)?; // Lettura
```

```

    if read_byte == 0 { break }
    write_to.write_all(&buf[..read_byte]); // Scrittura
}
Ok(())
} // Fine scope: Chiusura dei file

```

La logica implementata in entrambi i Listati, 14 e 15, è la stessa: dalla directory attuale vengono aperti due file, `read_from` in lettura e `write_to` in scrittura, successivamente viene letto l'intero contenuto di `read_from` e viene scritto su `write_to`, tramite blocchi di 2048 byte.

Nel Listato 14, i controlli sugli errori sono stati intenzionalmente omessi per evidenziare che in C tale comportamento è ammesso (l'unico controllo effettuato è su `read`, per determinare la fine della lettura).

Nel Listato 15, invece, gli errori vengono gestiti tramite l'operatore `?`, il quale propaga l'errore alla funzione chiamante, garantendo simultaneamente sicurezza e maggiore leggibilità del codice.

A differenza di C, in cui la gestione degli errori viene lasciata alla discrezione del programmatore, Rust promuove un approccio più sicuro, tramite il tipo `Result` e l'operatore `?`. Questi meccanismi garantiscono la gestione dell'errore a livello di compilazione: il compilatore controlla che entrambi gli scenari (successo e fallimento) vengano gestiti esplicitamente⁹ oppure che l'errore venga propagato al chiamante tramite `?`, prevenendo scenari di *undefined behaviour*.

Socket

Entrambi i linguaggi mettono a disposizione strumenti per lo sviluppo di applicazioni di rete, in particolare per la comunicazione orientata alla connessione (ad esempio, socket TCP).

Nel linguaggio C, l'interfaccia di programmazione di rete è fornita principalmente dalle librerie di sistema POSIX:

- `socket`: creazione di un socket;

⁹ In realtà, per `Result` è disponibile una funzione, `unwrap`, la quale tenta di restituire il contenuto di `Ok`, generando un `panic` se al suo posto era presente un `Err`. Questa funzione andrebbe utilizzata quando si è certi che il `Result` su cui la si invoca contenga un `Ok`. Va considerato che, il comportamento ideomatico di Rust, prevederebbe di controllare il contenuto, verificando che sia presente un `Ok` o un `Err`, prima di accedere al valore di `Result`. Alternativamente, `unwrap` potrebbe essere utilizzato come indicatore che un eventuale errore in quel punto sia irrecuperabile (seguendo il pensiero ideomatico di Rust, in cui `panic` è da usare solo per gli errori irrecuperabili).

- `bind`: associazione di un indirizzo IP e porta;
- `listen`: messa in ascolto in attesa di connessioni;
- `accept`: accettazione di una connessione in arrivo;
- `write`: invio di dati sul socket;
- `read`: ricezione di dati dal socket;
- `close`: chiusura del socket;

Questa interfaccia, sebbene molto flessibile, è soggetta a errori comuni, principalmente dovuti alla gestione manuale dei file descriptor¹⁰ e alla mancata verifica dell'esito delle operazioni.

Rust, invece, fornisce un'interfaccia di alto livello tramite il modulo `std::net`, che incapsula i socket TCP in due strutture: `TcpListener` per il lato server e `TcpStream` per il lato client. Le principali operazioni sono:

- `TcpListener::bind`: unisce la creazione del socket, il bind e la messa in ascolto;
- `TcpListener::accept`: accetta una connessione in ingresso e restituisce un `TcpStream`;
- `TcpStream::connect`: si connette a un `TcpListener`;
- `TcpStream::read`: riceve dati dal socket, implementando il `Trait Read`;
- `TcpStream::write`: invia dati sul socket, implementando il `Trait Write`.

Queste astrazioni integrano la gestione degli errori avvalendosi del tipo `Result<T, E>`, obbligando il programmatore a gestire esplicitamente sia il caso di successo che quello di fallimento di un'operazione.

I Listati 16 e 17 mostrano la realizzazione di un server TCP in entrambi i linguaggi. La logica implementata è la stessa per entrambi: il server si mette in ascolto sulla porta 50000, accetta una connessione in ingresso, legge fino a 1024 byte dal socket e infine scrive, sempre su quest'ultimo, i dati ricevuti.

¹⁰ In ambiente Unix le socket vengono gestite tramite file descriptor.

Nel Listato 16 i controlli sugli errori sono stati omessi per evitare un'eccessiva complessità strutturale e facilitare la lettura, al fine di focalizzare l'attenzione sul flusso logico di una connessione TCP.

Per lo stesso motivo, e per mantenere una somiglianza strutturale con l'esempio precedente, nel Listato 17 gli errori non vengono gestiti esplicitamente, ma vengono propagati al chiamante tramite l'operatore `?`.

Analogamente a quanto visto per la gestione dei file, Rust promuove un approccio più sicuro, tramite il tipo `Result` e l'operatore `?`, che garantiscono la gestione dell'errore a livello di compilazione.

Thread

La gestione dei *thread* è fondamentale per un linguaggio di programmazione moderno, in quanto consente di suddividere il lavoro su più unità esecutive e sfruttare il parallelismo in sistemi multicore o multiprocessore. Sia C che Rust offrono meccanismi per la creazione e la gestione dei *thread*, ma con approcci differenti.

La programmazione parallela rappresenta un contesto ampio e complesso, per cui una trattazione sufficientemente dettagliata richiederebbe un capitolo a sé stante e non rientrerebbe nell'ambito di questa trattazione, che mira a un confronto di base tra i linguaggi.

In C, la programmazione concorrente si basa tipicamente sulla libreria `pthread.h`, parte dello standard POSIX. Questa libreria offre primitive per la creazione (`pthread_create`) e sincronizzazione (`pthread_join`, `pthread_mutex` e altri) dei *thread*. Il completo controllo sui *thread*, fornito dalla libreria, comporta, però, che sia il programmatore stesso a doverne garantire una corretta gestione: sincronizzando i *thread* e garantendo un accesso corretto alle risorse condivise. L'approccio del linguaggio C, infatti, prevede semplicemente di fornire al programmatore gli strumenti necessari, ma è una sua responsabilità adoperarli correttamente: non vengono effettuati controlli sull'utilizzo corretto di *mutex* o risorse condivise durante la fase di compilazione, permettendo di incorrere in errori durante l'esecuzione.

Rust, d'altra parte, tramite il modulo `std::thread`, offre un'interfaccia di livello più alto. Nonostante internamente sia basata su `pthread`, `std::thread` integra strumenti per la corretta gestione di *mutex* e accesso alle risorse condivise, tramite le strutture `Arc<T>` e `Mutex<T>`, che garantiscono la condivisione sicura di dati mutabili tra *thread*.

Listato 16: Semplice server TCP in C

```
#include <unistd.h>
#include <sys/socket.h>
#include <netinet/in.h>
#include <stdio.h>
int main(void) {

    int fd = socket(AF_INET, SOCK_STREAM, 0); // Creazione

    struct sockaddr_in addr = {
        .sin_family = AF_INET,
        .sin_port = htons(50000),
        sin_addr.s_addr = INADDR_ANY
    };
    bind(fd, (struct sockaddr*)&addr, sizeof(addr)); // Binding

    listen(fd, 1); // Messa in ascolto

    int client_fd = accept(fd, NULL, NULL); // Accettazione

    char buf[1024];
    ssize_t read_byte = read(client_fd, buf, sizeof(buf)); // Lettura

    write(client_fd, buf, read_byte); // Scrittura

    close(client_fd); // Chiusura
    close(fd);      // Chiusura
    return 0;
}
```

Listato 17: Semplice server TCP in Rust

```
use std::net::{TcpListener, TcpStream};
use std::io::{Read, Write};
fn main() -> std::io::Result<()> {

    let server = TcpListener::bind("0.0.0.0:50000"?); // Creazione,
                                                binding e messa in ascolto

    let (mut client, _) = server.accept()?; // Accettazione

    let mut buffer = [0; 1024];
    let n = client.read(&mut buffer)?; // Lettura

    client.write(&buffer[..n])?; // Scrittura

    Ok(())
} // Fine scope: Chiusura dei socket
```

`Arc<T>` è uno smart pointer a conteggio di riferimento (ha un contatore interno per tenere traccia dei riferimenti) che permette la condivisione non mutabile di dati tra *thread*. `Mutex<T>` è una struttura che fornisce mutua esclusione su una risorsa per l'accesso concorrente. La combinazione `Arc<Mutex<T>>` è la forma idiomatica di Rust per la condivisione sicura di dati mutabili tra *thread*.

Vengono inoltre messi a disposizione due Trait fondamentali: `Send`, per indicare che è possibile trasferire la ownership dei valori di un tipo tra più *thread* e `Sync`, per indicare che per un tipo è sicuro avere reference divise su più *thread*.

Il compilatore controlla l'implementazione dei Trait per determinare la validità degli accessi alle risorse, riducendo la probabilità di errori di concorrenza.

In sintesi, nonostante il modello di Rust possa sembrare più complesso o verboso rispetto alla flessibilità offerta dalle librerie C, esso garantisce un accesso sicuro alle risorse condivise e, in generale, riesce a prevenire a livello di compilazione una classe di errori comuni: *data race* (accesso concorrente ai dati di cui almeno uno in scrittura).

Conclusioni

In conclusione, sotto l'aspetto della gestione delle risorse, i due linguaggi offrono strumenti analoghi, ma con differenze non trascurabili.

C permette un controllo maggiore e diretto delle risorse, ma questo avviene a un costo, come la gestione manuale di puntatori, *file descriptor* e buffer; inoltre, la maggior parte delle funzioni utilizzate per interagire con file, *socket* o *thread* si basano su API POSIX (su sistemi *POSIX-compliant*), non garantendo la compatibilità tra più piattaforme, come Windows.

D'altra parte, Rust offre un controllo più restrittivo ma semplificato, spesso incapsulando le risorse in strutture astratte, permettendo una gestione sicura delle risorse e dei relativi errori. Inoltre, a differenza di C, Rust garantisce che il codice sviluppato sia *cross-platform* in quanto le strutture `File`, `TcpStream`, `TcpListener` e `Thread` sono indipendenti dalla piattaforma: l'interfaccia rimane la stessa, ma in base alla piattaforma specifica vengono utilizzati strumenti differenti di basso livello (ad esempio, i `pthread` POSIX, definiti in `pthread.h` e i `thread` Win32, definiti in `windows.h`).

4.2.2 Sicurezza della memoria

In questa sottosezione verranno analizzati alcuni tra gli errori più comuni legati alla gestione della memoria, mostrando esempi di codice scritti in C e confrontandoli, quando possibile, con le controparti in Rust.

Un aspetto interessante del modello di memoria di Rust è che alcuni errori non sono semplicemente rilevati durante l'esecuzione, o a tempo di compilazione, ma sono strutturalmente impossibili da generare: non vi è modo di scrivere codice Rust che generi tali errori. Come già accennato, il *modello di ownership* guida la struttura di un programma in Rust: spinge il programmatore a ragionare in termini di *ownership* e *lifetime*. Inoltre, l'allocazione della memoria dinamica è gestita tramite l'utilizzo di *smart pointers*, che forniscono un'astrazione sicura per la gestione della memoria dinamica.

Unfreed memory

La *unfreed memory* (memoria non liberata) rappresenta un errore comune durante la gestione della memoria dinamica tramite approccio manuale. Si presenta in quei contesti in cui aree di memoria vengono allocate, senza successiva deallocazione. Le aree di memoria interessate non possono es-

sere riutilizzate per allocazioni successive in quanto vengono considerate ancora utilizzate.

Si consideri l'esempio minimale C, riportato nel Listato 18, che genera un'errore di *unfreed memory*:

Listato 18: Unfreed memory in C

```
#include <stdlib.h>
int main(void) {
    void* mem = malloc(sizeof(int));
    return 0;
}
```

In Rust questo comportamento non si potrà mai realizzare, in quanto la deallocazione viene gestita implicitamente dal compilatore¹¹.

Nel Listato 19 è riportato un programma in Rust che, per quanto possibile, replica il comportamento del Listato 18:

Listato 19: Unfreed memory in Rust

```
fn main() {
    let mem = Box::<u32>::new(0);
}
```

Nel Listato 18 viene allocata memoria nella heap sufficiente a contenere un intero, successivamente il programma termina senza liberarla¹². Il programma viene compilato ed eseguito correttamente, nonostante la presenza di questo errore. Tuttavia, tale memoria risulta inaccessibile per ulteriori allocazioni da parte del processo durante la sua esecuzione.

Il Listato 19 rappresenta il codice Rust più vicino possibile a quello riportato nel Listato 18. La differenza principale è che la memoria viene deallocata automaticamente: al termine dello scope della funzione `main`, la variabile `mem` viene eliminata e la relativa memoria deallocata.

Di conseguenza, l'errore di *unfreed memory* non può manifestarsi, in quanto la deallocazione avviene in maniera trasparente agli occhi del programmatore.

11 Il compilatore inserisce chiamate alla funzione *drop* in corrispondenza della fine di uno scope. Questo è realizzabile in quanto il *Borrow Checker* controlla le relazioni di *ownership* e *borrowing* per determinare chi sia responsabile della deallocazione, ovvero su quali variabili invocare *drop*.

12 In realtà l'allocazione avviene nello spazio di memoria virtuale del processo; a termine dell'esecuzione, il sistema operativo recupera tutte le risorse allocate dal processo.

Double free

La *double free* (doppia liberazione) è un'errore di gestione della memoria che si verifica quando la stessa area di memoria viene deallocata più di una volta. Ciò può causare la corruzione della memoria heap, con conseguente comportamento indefinito o crash del programma.

Nel linguaggio C, questo si verifica invocando la funzione `free` sullo stesso puntatore più volte, come mostrato nel Listato 20.

Listato 20: Double free in C

```
#include <stdlib.h>
int main(void) {
    void* ptr = malloc(sizeof(int));
    free(ptr);
    free(ptr);
    return 0;
}
```

Nel Listato 21 è riportato un programma in Rust che, per quanto possibile, tenta di replicare il comportamento del Listato 20.

Listato 21: Double free in Rust

```
fn main() {
    let mem = Box::<u32>::new(0);
    std::mem::drop(mem);
    std::mem::drop(mem);
}
```

Nel Listato 20 la memoria puntata da `ptr` viene deallocata due volte. In fase di esecuzione questo può portare a un crash del programma, come mostrato in Figura 6.

Tuttavia, lo standard C non definisce un comportamento specifico da adottare in questo contesto, causando *undefined behaviour* nella pratica¹³.

Per quanto riguarda Rust, è possibile considerare due scenari:

- Il Listato 19 rappresenta il comportamento idiomatico di Rust: la memoria viene deallocata automaticamente alla fine di uno *scope*, senza necessità di interventi manuali;

¹³ In ambiente Linux con `glibc`, viene rilevato l'errore causando la stampa di un messaggio e l'interruzione dell'esecuzione. In ambienti Windows, l'errore può passare inosservato o causare un crash silenzioso.

- Nel Listato 21, invece, si tenta di deallocare esplicitamente la memoria con due chiamate consecutive a `std::mem::drop`¹⁴. Tuttavia, la funzione `drop` prende possesso del valore, rendendo la reference originale, e di conseguenza la seconda chiamata, invalida. Il compilatore si accorge della violazione della terza regola di *borrowing* (Tutte le reference devono essere valide), impedendo la compilazione e generando l'errore riportato in Figura 7.

```
frank@francyy:~/Documents/thesis$ gcc double-free.c -o double-free-c
frank@francyy:~/Documents/thesis$ ./double-free-c
free(): double free detected in tcache 2
Aborted (core dumped)
```

Figura 6: Double free in C

```
frank@francyy:~/Documents/thesis$ rustc double-free.rs
error[E0382]: use of moved value: `mem`
  --> double-free.rs:4:20
   |
2 |     let mem = Box::<u32>::new(0);
   |     --- move occurs because `mem` has type `Box<u32>`, which does not implement the `Copy` trait
3 |     std::mem::drop(mem);
   |                       --- value moved here
4 |     std::mem::drop(mem);
   |                       ^^^ value used here after move
```

Figura 7: Tentativo di double free in Rust

Dangling pointers

Un *dangling pointer* (puntatore pendente) è un puntatore che riferisce a una locazione di memoria non più valida o che è stata liberata. Generalmente si presenta quando un'area di memoria viene deallocata, ma gli eventuali puntatori che la referenziavano non vengono aggiornati.

Si consideri l'esempio C riportato nel Listato 22 che genera un *dangling pointer*¹⁵:

¹⁴ Questa funzione assume *ownership* del valore fornito, causandone la deallocazione una volta raggiunta la fine della funzione.

¹⁵ La presenza di uno scope interno non è funzionale all'esempio, ma non ne modifica nemmeno il risultato. È presente per garantire somiglianza strutturale con l'esempio Rust successivo.

Listato 22: Dangling pointer in C

```

int main(void) {
    int* ptr;
    {
        int* val = malloc(sizeof(int));
        *val = 30;
        ptr = val;
        free(val);
    }
}

```

Nel Listato 22, viene allocata memoria sufficiente per un intero e il suo indirizzo è memorizzato sia in `val` che in `ptr`. La memoria viene successivamente deallocata invocando la funzione `free` sul puntatore `val`. Tuttavia, il puntatore `ptr` non viene aggiornato, continuando a riferire alla stessa area di memoria, ormai non più valida.

È importante osservare che la sola esistenza di un *dangling pointer* non causa un errore immediato. Il comportamento indefinito si manifesta solo quando si tenta di dereferenziare tale puntatore, cercando di leggere o scrivere nella memoria a cui fa riferimento.

A dimostrazione di ciò, si consideri che Rust permette la creazione di una *dangling reference*, a patto che non venga utilizzata. Il Listato 23 mostra questa possibilità.

Listato 23: Dangling reference in Rust

```

fn main() {
    let _ref: &u32;
    {
        let val = Box::<u32>::new(30);
        _ref = &*val;
    }
}

```

Il Listato 23 tenta di replicare, per quanto possibile, il comportamento descritto nel Listato 22: alla variabile `_ref` viene assegnato un riferimento a `val`, allocata dinamicamente in uno scope interno. All'uscita da tale scope, `val` viene deallocato, lasciando `_ref` con un riferimento non valido.

Nonostante ciò, il compilatore non genera alcun errore. Questo comportamento è giustificato dal fatto che la reference non è mai utilizzata¹⁶

¹⁶ Il *Borrow Checker* lavora in maniera *lazy*: non verifica le condizioni di *borrowing* e *lifetime*

e, di conseguenza, il *Borrow Checker* non rileva alcuna violazione delle regole di *borrowing*.

Access after free

Sebbene la sola presenza di un *dangling pointer* non causi immediatamente problemi, essa rappresenta una condizione necessaria per un errore più grave: la *access after free* (accesso post-liberazione). Questo errore si verifica quando si tenta di accedere a memoria precedentemente deallocata, dereferenziando un *dangling pointer*.

A seconda del tipo di accesso, si possono avere conseguenze differenti:

- **Lettura:** può causare il recupero di dati non validi o incoerenti, eventualmente sovrascritti da allocazioni successive;
- **Scrittura:** può compromettere l'integrità della memoria, causando comportamenti indefiniti o crash del programma.

Nel Listato 24 è riportato un esempio C che, estendendo il Listato 22, genera un errore di *access after free*:

Listato 24: Access after free in C

```
#include <stdlib.h>
#include <stdio.h>
int main(void) {
    int* ptr;
    {
        int* val = malloc(sizeof(int));
        *val = 30;
        ptr = val;
        free(val);
    }
    printf("%d\n", *ptr);
}
```

Nel Listato 24 si tenta di dereferenziare il puntatore `ptr` dopo che la memoria a cui faceva riferimento è stata deallocata. In questo caso, l'accesso avviene per effettuare un'operazione di lettura, che rappresenta una forma meno distruttiva rispetto a una scrittura, ma che resta comunque pericolosa: i dati letti potrebbero essere non validi o casuali, in quanto la memoria potrebbe essere stata sovrascritta da allocazioni successive.

di una reference fino a che non viene utilizzata.

Questo comportamento rappresenta un esempio di *undefined behaviour*, come illustrato in Figura 8.

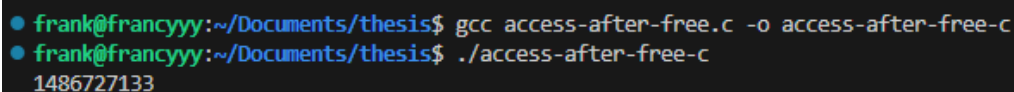
Nel Listato 25 è riportato un programma che mostra come Rust gestisce questa problematica. Per mantenere una somiglianza strutturale con la controparte C, il Listato estende il precedente esempio riportato nel Listato 23.

Listato 25: Tentativo di *access after free* in Rust

```
fn main() {
    let _ref: &u32;
    {
        let val = Box::<u32>::new(30);
        _ref = &*val;
    }
    println!("{}", _ref);
}
```

Il Listato 25 tenta di replicare il comportamento del Listato 24, cercando di accedere a una reference pendente per un'operazione di lettura.

Tuttavia, in questo caso viene generato un errore di compilazione, come mostrato in Figura 9. Il *Borrow Checker* rileva che `_ref` viene utilizzata successivamente alla deallocazione di `val`, rappresentando una violazione delle regole di *borrowing* e, di conseguenza, impedisce la compilazione.



```
frank@francyy:~/Documents/thesis$ gcc access-after-free.c -o access-after-free-c
frank@francyy:~/Documents/thesis$ ./access-after-free-c
1486727133
```

Figura 8: *Access after free* in C

Access out of bounds: buffer overflow e overread

L' *access out of bounds* è un errore comune associato alla gestione manuale della memoria dinamica. Si verifica quando si tenta di accedere a una porzione di memoria al di fuori dei limiti dell'area effettivamente allocata.

In base al tipo di accesso si distingue in due varianti: nel caso di una lettura si parla di *buffer overread* (lettura oltre i limiti) mentre, nel caso di una scrittura, di *buffer overflow* (sovraccarico o sovrascrittura del buffer).

```

frank@francyy:~/Documents/thesis$ rustc access-after-free.rs
error[E0597]: `*val` does not live long enough
  --> access-after-free.rs:5:16
   |
4  |         let val = Box::<u32>::new(30);
   |         --- binding `val` declared here
5  |         _ref = &*val;
   |                ^^^^^ borrowed value does not live long enough
6  |     }
   |     - `*val` dropped here while still borrowed
7  |     println!("{}", _ref);
   |                   ---- borrow later used here

```

Figura 9: Tentativo di *access after free* in Rust

BUFFER OVERFLOW L'esempio riportato nel Listato 26 mostra un programma in C che genera un errore di *buffer overflow*.

Listato 26: Buffer overflow in C

```

#include <stdlib.h>
#include <stdio.h>
int main(void) {
    int* vec = malloc(sizeof(int) * 3);
    printf("Allocato un vettore di 3 interi, indici validi: 0, 1,
          2\n");
    for(int i = 0; i <= 3; i++) {
        printf("Memorizzando %d all'indice %d\n", i * 10, i);
        vec[i] = i * 10;
    }
    free(vec);
    printf("Terminata la memorizzazione!\n");
    return 0;
}

```

Nel Listato 26 viene allocata memoria sufficiente per un vettore di tre interi, ma successivamente vengono inizializzati quattro elementi, dall'indice 0 fino a 3. In questo caso il programma termina senza errori, viene anche eseguita la stampa finale, come mostrato in Figura 10.

Nonostante ciò, questo comportamento non può essere considerato una garanzia in quanto, generalmente, la conseguenza di un *buffer overflow* è un *undefined behaviour*: non è possibile determinare a priori cosa contenga la memoria oltre il limite dell'allocazione o per cosa venga utilizzata¹⁷.

¹⁷ In generale un *buffer overflow* avviene all'interno dello spazio di indirizzamento virtuale

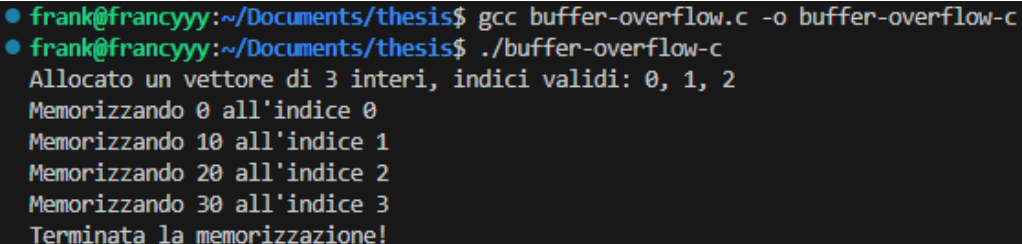
Nel Listato 27 è riportato un programma in Rust che mostra come viene gestito questo errore.

Listato 27: Buffer overflow in Rust

```
fn main() {
    let mut vec: Vec<u32> = vec![0; 3];
    println!("Allocato un vettore di 3 interi, indici validi: 0, 1,
        2");
    for i in 0u32..=3u32 {
        println!("Memorizzando {} all'indice {}", i * 10, i);
        vec[i as usize] = i * 10;
    }
    println!("Terminata la memorizzazione!");
}
```

Il Listato 27 tenta di replicare il comportamento del Listato 26, allocando un vettore di tre interi e successivamente cercando di inizializzarne quattro. In questo caso la compilazione va a buon fine, ma la stampa finale non avviene: il programma genera un *panic* durante l'esecuzione. Questo avviene durante l'accesso all'indice 3, come mostrato in Figura 11.

Questo comportamento è dovuto al fatto che il compilatore inserisce controlli di validità sugli indici, i quali vengono eseguiti a runtime.¹⁸



```
frank@francyy:~/Documents/thesis$ gcc buffer-overflow.c -o buffer-overflow-c
frank@francyy:~/Documents/thesis$ ./buffer-overflow-c
Allocato un vettore di 3 interi, indici validi: 0, 1, 2
Memorizzando 0 all'indice 0
Memorizzando 10 all'indice 1
Memorizzando 20 all'indice 2
Memorizzando 30 all'indice 3
Terminata la memorizzazione!
```

Figura 10: Buffer overflow in C

BUFFER OVERREAD L'esempio riportato nel Listato 28 mostra un programma in C che genera un errore di *buffer overread*.

di un processo. In questo caso, la memoria potenzialmente corrotta sarebbe del processo stesso. Tuttavia, nel caso in cui l'indirizzo di riferimento fosse al di fuori dello spazio di indirizzamento virtuale del processo verrebbe generato un'errore runtime di tipo *segmentation fault*, causando l'immediato crash del programma.

- 18 Infatti, il compilatore non può sapere a tempo di compilazione quali saranno gli indici che verranno utilizzati per accedere a un vettore, quindi si limita a inserire controlli sulla loro validità prima degli accessi effettivi.

```

frank@francyy:~/Documents/thesis$ rustc buffer-overflow.rs -o buffer-overflow-rust
frank@francyy:~/Documents/thesis$ ./buffer-overflow-rust
Allocato un vettore di 3 interi, indici validi: 0, 1, 2
Memorizzando 0 all'indice 0
Memorizzando 10 all'indice 1
Memorizzando 20 all'indice 2
Memorizzando 30 all'indice 3

thread 'main' panicked at buffer-overflow.rs:6:12:
index out of bounds: the len is 3 but the index is 3
note: run with `RUST_BACKTRACE=1` environment variable to display a backtrace

```

Figura 11: *Buffer overflow* in Rust

Listato 28: Buffer overread in C

```

#include <stdlib.h>
#include <stdio.h>
int main(void) {
    int* vec = malloc(sizeof(int) * 3);
    for(int i = 0; i < 3; i++) vec[i] = i * 10;
    printf("Contenuto del vettore:\n");
    for(int i = 0; i < 10; i++)
        printf("Indice: %d -> Valore: %d\n", i, vec[i]);
    free(vec);
    return 0;
}

```

Nel Listato 28 viene allocata memoria per un vettore di tre interi, i quali vengono correttamente inizializzati e, successivamente, si tenta di leggere dieci elementi dal vettore. In questo caso il programma termina senza errori, stampando effettivamente dieci elementi, come mostrato in Figura 12.

Come quanto detto per l'errore di *buffer overflow*, generalmente la conseguenza è un *undefined behaviour*: non è possibile determinare a priori cosa contenga la memoria (oltre il limite dell'allocazione) che viene letta¹⁹.

Rust gestisce questo errore in maniera analoga al *buffer overflow*. Nel Listato 27 è riportato un programma in Rust che mostra tale somiglianza.

Listato 29: Buffer overread in Rust

¹⁹ Vangono le stesse considerazioni fatte per il *buffer overflow*: un'accesso a un indirizzo che eccede lo spazio di indirizzi virtuali del processo genera un errore runtime di tipo *segmentation fault*, causando l'immediato crash del programma

```
fn main() {
    fn main() {
        let vec: Vec<u32> = vec![0, 10, 20];
        println!("Contenuto del vettore:");
        for i in 0..10 {
            println!("Indice: {} -> Valore: {}", i, vec[i]);
        }
    }
}
```

Il Listato 29 tenta di replicare il comportamento del Listato 28, allocando un vettore di tre interi e successivamente cercando di leggerne dieci.

Anche in questo caso la compilazione termina correttamente ma, durante l'esecuzione, in particolare durante l'accesso al quarto elemento, viene generato un *panic*. Questo comportamento può essere osservato in Figura 13.

```
frank@francyy:~/Documents/thesis$ gcc buffer-overread.c -o buffer-overread-c
frank@francyy:~/Documents/thesis$ ./buffer-overread-c
Contenuto del vettore:
Indice: 0 -> Valore: 0
Indice: 1 -> Valore: 10
Indice: 2 -> Valore: 20
Indice: 3 -> Valore: 0
Indice: 4 -> Valore: 0
Indice: 5 -> Valore: 0
Indice: 6 -> Valore: 1041
Indice: 7 -> Valore: 0
Indice: 8 -> Valore: 1768189513
Indice: 9 -> Valore: 540697955
```

Figura 12: *Buffer overread* in C

```
frank@francyy:~/Documents/thesis$ rustc buffer-overread.rs -o buffer-overread-rust
frank@francyy:~/Documents/thesis$ ./buffer-overread-rust
Contenuto del vettore:
Indice: 0 -> Valore: 0
Indice: 1 -> Valore: 10
Indice: 2 -> Valore: 20

thread 'main' panicked at buffer-overread.rs:5:52:
index out of bounds: the len is 3 but the index is 3
```

Figura 13: *Buffer overread* in Rust

Uninitialized memory access

La *uninitialized memory access* (accesso a memoria non inizializzata) è un errore che si presenta quando si tenta di leggere da un'area di memoria che non è stata inizializzata. Questo può portare a comportamenti imprevedibili, in quanto i dati letti potrebbero essere casuali e potenzialmente residui da allocazioni precedenti.

Nel linguaggio C, questo errore si presenta tipicamente quando viene allocata memoria tramite la funzione `malloc` e successivamente si tenta di accedervi senza previa inizializzazione.

Nel Listato 30 viene riportato un esempio minimale C che genera un errore di questo tipo.

Listato 30: Uninitialized memory access in C

```
#include <stdlib.h>
#include <stdio.h>
int main(void) {
    int* ptr = malloc(sizeof(int));
    printf("%d\n", *ptr);
    free(ptr);
    return 0;
}
```

Nel Listato 30 viene allocata memoria sufficiente a contenere un intero e, senza inizializzarne il contenuto, viene tentato un accesso in lettura. Il risultato è un comportamento indefinito: il valore letto potrebbe variare in base all'ambiente di esecuzione e tra esecuzioni, senza garanzia di coerenza.

È possibile osservare un caso particolare, in cui l'accesso può sembrare produrre un risultato coerente, come mostrato in Figura 14²⁰. Tuttavia, questa apparente stabilità non può essere considerata una garanzia, in quanto vi sono più fattori che possono influenzare il valore letto:

- Implementazioni specifiche dell'allocatore di memoria, che possono inizializzare la memoria allocata a zero o a un valore casuale;
- La posizione in memoria scelta dall'allocatore per il puntatore;
- Eventuali valori residui da allocazioni precedenti;

²⁰ L'esecuzione è avvenuta su Ubuntu LTS 24.04 con glibc. In questo contesto è possibile che `malloc` azzeri la memoria allocata, prima di restituirne l'indirizzo, ma non è un comportamento garantito dallo standard C.

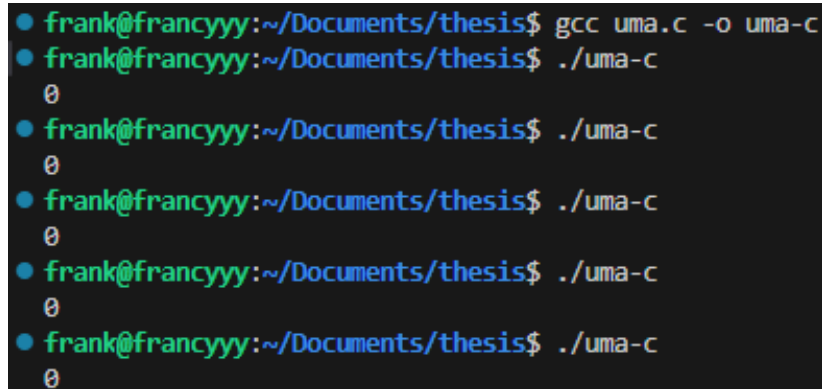
- La mappatura tra pagine virtuali e fisiche da parte del sistema operativo.

A conferma dell'inaffidabilità, si consideri l'esempio nel Listato 31, che estende il Listato 30.

Listato 31: Uninitialized memory access in C

```
#include <stdlib.h>
#include <stdio.h>
int main(void) {
    int* ptr = malloc(sizeof(int));
    free(ptr);
    ptr = malloc(sizeof(int));
    printf("%d\n", *ptr);
    free(ptr);
    return 0;
}
```

Il Listato 31 mostra come eseguire una `free` prima della `malloc` successiva può lasciare residui nella memoria. Di conseguenza il valore letto, questa volta, è diverso da zero, come si può osservare in Figura 15²¹.



```
frank@francyy:~/Documents/thesis$ gcc uma.c -o uma-c
frank@francyy:~/Documents/thesis$ ./uma-c
0
frank@francyy:~/Documents/thesis$ ./uma-c
0
frank@francyy:~/Documents/thesis$ ./uma-c
0
frank@francyy:~/Documents/thesis$ ./uma-c
0
frank@francyy:~/Documents/thesis$ ./uma-c
0
```

Figura 14: *Uninitialized memory access* in C

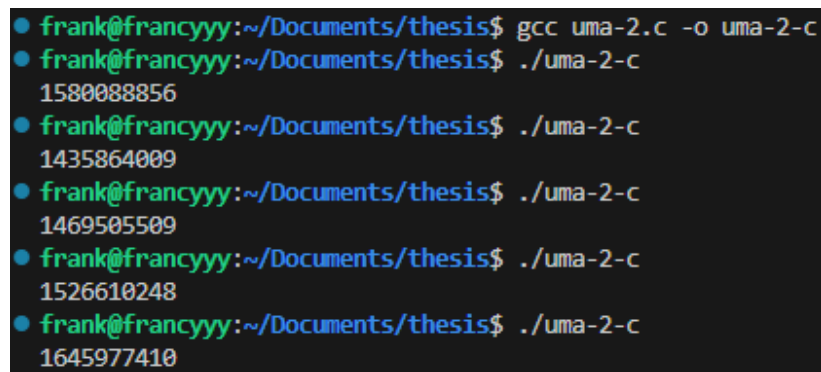
Rust previene questa problematica in maniera strutturale, non permettendo allocazioni dinamiche senza inizializzazione grazie all'utilizzo di smart pointers, i quali prevedono un'inizializzazione obbligatoria.

In Rust infatti, le principali primitive per l'allocazione dinamica richiedono sempre l'inizializzazione dei valori:

21 Anche in questo caso, si tratta di un comportamento specifico del sistema e non garantito dallo standard.

- **Box**: *smart pointer* impiegato per allocare un singolo valore nella heap. L'unico modo per crearne uno è tramite `Box::new`, che richiede un valore per l'inizializzazione. In *safe* Rust, non esiste un costrutto equivalente alla `malloc` in C;
- **Vec**: *smart pointer* utilizzato per una collezione dinamica di valori in heap. Anche se può essere creato vuoto (`Vec::new`), ogni accesso è verificato a runtime per evitare accessi oltre i limiti. In caso di violazioni viene generato un *panic*, terminando l'esecuzione, come è osservabile in Figura 13;

Di conseguenza, non è possibile sviluppare codice Rust che acceda a memoria non inizializzata senza ricorrere alla parola chiave `unsafe`²².



```
frank@francyy:~/Documents/thesis$ gcc uma-2.c -o uma-2-c
frank@francyy:~/Documents/thesis$ ./uma-2-c
1580088856
frank@francyy:~/Documents/thesis$ ./uma-2-c
1435864009
frank@francyy:~/Documents/thesis$ ./uma-2-c
1469505509
frank@francyy:~/Documents/thesis$ ./uma-2-c
1526610248
frank@francyy:~/Documents/thesis$ ./uma-2-c
1645977410
```

Figura 15: *Uninitialized memory access* in C con previa `free`

Conclusioni

Riassumendo, sotto l'aspetto della sicurezza della memoria, Rust si distingue da C garantendo un comportamento sicuro e deterministico, prevenendo alcuni errori durante la compilazione e rilevandone altri a runtime. In C, d'altra parte, tali errori possono passare inosservati, causando *undefined behaviour* o crash del programma.

²² Rust consente l'accesso a memoria potenzialmente non inizializzata, ma soltanto all'interno di blocchi *unsafe*. Un esempio è la funzione `std::mem::MaybeUninit`, ma meccanismi come questo sono riservati a casi particolari, in cui le garanzie di sicurezza devono essere gestite manualmente dal programmatore.

4.2.3 Prestazioni

In questa sottosezione verranno confrontate le prestazioni tipiche dei programmi scritti in C e Rust, basandosi su benchmark esistenti e pubblicamente disponibili.

Non verranno presentati né codice dettagliato né un'analisi sperimentale diretta, per i seguenti motivi: una valutazione accurata richiederebbe sviluppo di codice ottimizzato per entrambi i linguaggi, un set ampio di programmi da testare, differenti ambienti di esecuzione (diversi sistemi operativi, compilatori, architetture hardware) e una metodologia di confronto rigorosa, che andrebbero oltre gli obiettivi di questa trattazione.

Come accennato a inizio del capitolo, il linguaggio C è storicamente noto per le prestazioni, principalmente in termini di velocità d'esecuzione, dovute all'assenza di astrazioni complesse e al runtime minimo richiesto.

Rust, d'altra parte, mira a garantire la sicurezza della memoria senza compromettere le prestazioni, puntando a fornire velocità d'esecuzione paragonabili a C e C++.

Per offrire un confronto concreto, vengono riportati i risultati tratti da *The Computer Language Benchmarks Game*²³ [16], un progetto che confronta le prestazioni di diversi linguaggi di programmazione nell'esecuzione di algoritmi comuni. Il confronto si basa su quattro parametri:

- **secs**: tempo di esecuzione totale (*wall-clock time*);
- **cpu secs**: tempo effettivo utilizzato dalla CPU, trascurando attese dovute a I/O e context switch;
- **mem**: picco di memoria RAM utilizzata dal processo;
- **gz**: dimensione del file sorgente compresso con gzip, interpretato come indice di verbosità e complessità dell'implementazione.

Verranno analizzati due benchmark distinti: uno principalmente *CPU-bound* e uno principalmente *I/O-bound*. Non viene incluso un esempio della categoria '*contentious*', in quanto tali benchmark tendono a essere influenzati da librerie esterne e dalla specifica implementazione degli algoritmi, non tanto dalle caratteristiche intrinseche del linguaggio e, per questo, non rientrano negli obiettivi della trattazione.

²³ Nel sito vengono raccolti i risultati dell'esecuzione di vari algoritmi sviluppati in diversi linguaggi di programmazione. Il sito potrebbe variare nel tempo e i risultati riportati nella trattazione potrebbero non essere più i migliori e peggiori (data di riferimento 2025-07-18).

Confronto CPU-bound

Il benchmark *CPU-bound* considerato è *n-body*. Il problema consiste nel simulare il moto di *n* corpi che interagiscono attraverso la forza gravitazionale: a ogni passo viene calcolata la forza tra tutte le coppie di corpi, si aggiornano le velocità in base alle accelerazioni risultanti e si aggiornano le posizioni. L'algoritmo ha complessità quadratica ($O(n^2)$) ed è privo di I/O significativo: è un carico interamente *CPU-bound*.

Nella Tabella 1 sono riportati i risultati del benchmark *n-body* relativi a Rust e C. Per ciascun linguaggio è stata selezionata l'esecuzione migliore e quella peggiore in termini di **cpu secs**. Sono state considerate solo le implementazioni ideomatiche, escludendo quelle che ricorrono a *SIMD* scritti a mano²⁴ o a codice *unsafe* in Rust.

Linguaggio	secs	cpu secs	mem (KB)	gz (B)
Esecuzione migliore				
C	4.98	4.98	2482	1186
Rust	3.46	3.46	2937	1774
Esecuzione peggiore				
C	6.89	6.89	2490	1250
Rust	5.52	5.51	3027	1483

Tabella 1: Risultati *n-body* per Rust e C (implementazioni ideomatiche)

Dagli estratti riportati nella Tabella 1 è possibile osservare che, da un punto di vista temporale, anche in un benchmark puramente *CPU-bound*, Rust riesca a raggiungere C in termini di velocità d'esecuzione, senza ricorrere a codice *unsafe*. Nei benchmark, Rust ha addirittura superato C in velocità d'esecuzione: nel caso migliore, Rust ha impiegato circa il 30% di tempo in meno; nel caso peggiore circa il 20%.

Questo risultato supporta l'idea che Rust possa rappresentare un'alternativa a C anche in contesti ad alta intensità computazionale, pur garantendo la sicurezza della memoria.

Dal punto di vista dell'occupazione di risorse, invece, C si presenta vantaggioso, richiedendo una quantità inferiore di memoria durante

²⁴ Si tratta di un'ottimizzazione manuale del codice per sfruttare le istruzioni vettoriali della CPU: invece di scrivere codice 'standard', il programmatore scrive esplicitamente codice che utilizza i registri *Single Instruction, Multiple Data*.

l'esecuzione e producendo sorgenti più compatti (successivamente alla compressione).

Confronto I/O-bound

Il benchmark *I/O-bound* selezionato è *fasta*. Il problema consiste nella generazione di DNA (utilizzando i caratteri A, C, G e T) secondo pattern sia deterministici che pseudocasuali, seguita dalla loro scrittura sullo standard output. L'algoritmo presenta un carico di elaborazione minimo: rappresenta un esempio di algoritmo principalmente *I/O-bound*, in quanto la maggior parte del tempo di esecuzione è trascorso in scrittura sullo standard output. Inoltre, la presenza di generazione pseudocasuale e ripetitiva comporta un utilizzo intensivo di stringhe e buffer, rendendo il benchmark utile per verificare l'efficienza di gestione della memoria dinamica.

Nella Tabella 2 sono riportati i risultati del benchmark *fasta* relativi a Rust e C. Come specificato nell'analisi CPU-bound 4.2.3, per ciascun linguaggio viene riportata l'esecuzione migliore e quella peggiore in termini di **cpu secs**. Sono state considerate solamente le implementazioni ideomatiche, escludendo quelle che ricorrono a *SIMD scritti a mano* o codice *unsafe* in Rust.

Linguaggio	secs	cpu secs	mem (KB)	gz (B)
Esecuzione migliore				
C	0.79	0.78	2236	1469
Rust	2.03	2.03	3199	1235
Esecuzione peggiore				
C	8.27	8.27	2195	839
Rust	4.45	4.45	3113	1240

Tabella 2: Risultati del benchmark *fasta* per Rust e C (implementazioni ideomatiche)

Dai risultati mostrati nella Tabella 2 si osserva che non emerge un chiaro vantaggio costante tra i due linguaggi: nel caso migliore, C è circa 2.6 volte più veloce di Rust, mentre nel caso peggiore è Rust a richiedere circa la metà del tempo rispetto a C.

Dal punto di vista dell'occupazione di risorse, valgono ancora le considerazioni fatte durante l'analisi *CPU-bound*: C, generalmente, richiede una quantità minore di memoria RAM durante l'esecuzione e produce sorgenti più contenuti (una volta compressi), indicando una complessità minore dell'implementazione.

Conclusione

Dai risultati analizzati emerge che entrambi i linguaggi offrono prestazioni paragonabili, con differenze che variano in base al tipo di carico e all'implementazione specifica.

Con le opportune ottimizzazioni, sia a livello di scrittura del codice che di compilazione, Rust è in grado di raggiungere, e in alcuni casi superare, le prestazioni di C in termini di velocità d'esecuzione, pur garantendo sicurezza della memoria.

Tuttavia, un evidente svantaggio di Rust è rappresentato dagli elevati tempi di compilazione, specialmente rispetto a C, dovuti principalmente ai controlli effettuati dal *Borrow Checker* per determinare, e verificare, le relazioni di *ownership* e *borrowing*. Questo overhead in fase di compilazione è però compensato dalla sicurezza della memoria, garantita a tempo di compilazione, non richiedendo overhead aggiuntivi durante l'esecuzione.

Dal punto di vista della dimensione dei file eseguibili, Rust tende a generare file di dimensioni maggiori, principalmente per l'inclusione della libreria standard e dei simboli di debug²⁵.

Infine, per quanto riguarda l'utilizzo della memoria durante l'esecuzione, entrambi i linguaggi mostrano requisiti simili, con Rust che, tendenzialmente, può richiedere una quantità leggermente superiore di RAM, dovuta principalmente all'utilizzo di astrazioni. Infatti, nonostante quest'ultime siano definite *zero-cost* in teoria, nella pratica, spesso, il compilatore non riesce a ottimizzarle pienamente e possono introdurre un minimo overhead²⁶.

25 Tramite le opzioni del compilatore è possibile indicare l'ottimizzazione per lo spazio, che cerca di ridurre le dimensioni dell'eseguibile. Inoltre, in contesti di programmazione *bare-metal*, è possibile escludere la libreria standard (`#![no_std]`), in quanto non è presente un sistema operativo sottostante.

26 Per esempio, si consideri `Option<T>`, enumerazione che permette di esprimere la possibile assenza di un valore, `None` e `Some<T>`. In quanto Rust non permette che un riferimento sia nullo, tipi come `Box` o `String` non potranno mai valere 0. In questo scenario, `Option<Box>` e `Option<String>` occuperebbero lo stesso spazio di `Box` e `String`.

4.2.4 Complessità del codice

Una critica comune nei confronti di Rust, specialmente da coloro che provengono da C, è la maggiore complessità e verbosità del linguaggio. Generalmente, la sintassi di C è considerata più semplice e minimale, principalmente per la sua natura a basso livello, la mancanza di astrazioni complesse e la presenza di un numero limitato di costrutti.

Al contrario in Rust, dovendo gestire attentamente le relazioni di *ownership*, *borrowing* e *lifetime*, il codice può risultare più complesso; questi non sono gli unici aspetti che contribuiscono a questa percezione: gestione degli errori, controllo del flusso, supporto a *generics* e *macro* sono tutti elementi che possono rendere il linguaggio più verboso e complesso rispetto a C.

Annotazioni di *lifetime*

Una delle principali differenze tra Rust e C è dovuta alla presenza del *modello di ownership*, che impone annotazioni di *lifetime* nei casi in cui non siano deducibili implicitamente. Come già accennato nel capitolo tre: ‘Sistemi Operativi’ 3, queste permettono al *Borrow Checker* di determinare la validità di un prestito nel tempo, rifiutando codice che viola le annotazioni.

Queste annotazioni introducono un livello ulteriore di complessità, sia sintattica che concettuale: il programmatore deve specificare quanto a lungo un riferimento sarà valido, dovendo ragionare in termini di durata dei dati nel tempo.

Tuttavia, la complessità viene compensata da una maggiore sicurezza del codice: come verrà mostrato successivamente, permette di prevenire problemi quali *access after free*.

In C questo meccanismo non è presente, non esiste proprio il concetto di *lifetime* come elemento del linguaggio. In questi casi è il programmatore che deve garantire che i puntatori siano validi, senza supporto da parte del linguaggio.

Nei Listati 32 e 33 sono riportati due esempi che mostrano come le annotazioni di *lifetime* possano prevenire *access after free*.

rispettivamente: il compilatore ottimizza l’astrazione, utilizzando il valore 0 per None e considerando ogni valore diverso da 0 come Some. Tuttavia, per tipi che possono essere nulli, come `u8`, `Option<u8>` richiede un byte aggiuntivo per rappresentare l’assenza di valore, non permettendo al compilatore di ottimizzare l’uso dell’astrazione.

Listato 32: Prevenzione di *access after free* tramite *lifetime annotations*

```
fn biggest<'a>(a: &'a i32, b: &'a i32) -> &'a i32 {
    if *a > *b { a } else { b }
}
fn main (){
    let result;
    let smaller = Box::<i32>::new(1);
    {
        let bigger = Box::<i32>::new(2);
        result = biggest(&*smaller, &*bigger);
    }
    println!("Il piu grande e: {}", *result);
}
```

Listato 33: *Access after free* dovuta alla mancanza di *lifetime* in C

```
#include <stdlib.h>
#include <stdio.h>
int* biggest(int* a, int* b) {
    return (*a > *b) ? a : b;
}
int main(void) {
    int* result;
    int* smaller = malloc(sizeof(int));
    *smaller = 1;
    {
        int* bigger = malloc(sizeof(int));
        *bigger = 2;
        result = biggest(smaller, bigger);
        free(bigger);
    }
    printf("Il piu grande e: %d\n", *result);
    free(smaller);
}
```

Entrambi i Listati, [32](#) e [33](#), implementano la stessa logica: definiscono una funzione `biggest` che riceve in ingresso due riferimenti (o puntatori, nel caso di C) e restituisce quello che punta al valore maggiore. La funzione `main` alloca due interi, `smaller` e `bigger`, nella heap e passa un riferimento a ciascuno a `biggest`. Per come sono inizializzati, la funzione restituisce un riferimento a `bigger`. Tuttavia, `bigger` viene deallocato prima della stampa, lasciando un *dangling pointer*.

Qua è possibile osservare la differenza tra i due linguaggi:

- In C, il compilatore produce un eseguibile senza generare alcun avviso o errore. L'esecuzione sembra funzionare, ma in realtà viene generato un *access after free*, leggendo un valore corrotto, come illustrato in Figura 16;
- In Rust, invece, *biggest* specifica una sola *lifetime*, 'a, indicando che tutte le reference (le due in ingresso e quella in uscita) devono essere valide per la durata di tale *lifetime*. Il *Borrow Checker* rileva che *smaller* e *bigger*, essendo definiti in scope differenti, hanno *lifetime* diverse e impedisce la compilazione. È possibile osservare questo comportamento in Figura 17.

```
frank@francyy:~/Documents/thesis$ gcc no-lifetimes.c -o no-lifetimes-c
frank@francyy:~/Documents/thesis$ ./no-lifetimes-c
Il piu grande e: 1453304247
```

Figura 16: *Access after free* dovuto all'assenza di *lifetime* in C

```
frank@francyy:~/Documents/thesis$ rustc lifetimes.rs -o lifetimes-rs
error[E0597]: `*bigger` does not live long enough
--> lifetimes.rs:9:37
   |
 8 |         let bigger = Box::<i32>::new(2);
   |         ----- binding `bigger` declared here
 9 |         result = biggest(&*smaller, &*bigger);
   |                               ^^^^^^^^^ borrowed value does not live long enough
10 |     }
```

Figura 17: Tentativo di *access after free* in Rust

Questo è un esempio che mostra come Rust, tramite le annotazioni di *lifetime*, presenti una sintassi più complessa rispetto a C, ma allo stesso tempo più sicura.

Generics

Le espressioni *generics* o *generic programming* indicano un meccanismo di astrazione che consente di sviluppare codice indipendente da un tipo di dato specifico. Il tipo viene fornito come argomento, generalmente in fase di compilazione, consentendo il riuso del codice con diversi tipi di dato.

Il linguaggio C non supporta nativamente lo sviluppo di codice parametrico, in quanto comporterebbe un livello di astrazione superiore rispetto alla semplicità a cui il linguaggio aspira. Nonostante ciò, è possibile emulare un comportamento simile, principalmente tramite l'uso di puntatori `void*`, i quali possono riferirsi a valori di qualunque tipo. Questo meccanismo, per quanto flessibile, richiede controlli rigorosi da parte del programmatore:

- un puntatore `void*` non memorizza informazioni sulla dimensione del valore puntato, rendendo necessaria la conoscenza esplicita del tipo (solitamente memorizzandone la dimensione);
- le operazioni di accesso, sia in lettura che scrittura, devono essere effettuate tramite funzioni delicate come `memcpy`, in quanto non è possibile dereferenziare direttamente un puntatore `void*`.

Il Listato 34 riporta un'estratto da un'implementazione in C di un nodo di una lista generica. L'implementazione completa è disponibile sulla piattaforma Github²⁷ [4]. Nell'implementazione di riferimento sono state adottate alcune strategie per cercare di prevenire errori legati alla gestione manuale dei tipi e della memoria:

- **Tipo opaco:** la struttura dati è definita solo nel file sorgente, mentre l'header contiene solo una dichiarazione incompleta (*opaque type*). Questo impedisce l'accesso diretto ai campi, tra cui `element_size` di `clinkedlist`, la cui integrità è fondamentale per il corretto funzionamento della struttura;
- **Incapsulamento parziale:** la manipolazione della struttura è possibile solo tramite le funzioni fornite, obbligando l'utente a seguire un percorso logico e controllato per l'inserimento, la rimozione e la lettura dei dati. Questo permette l'introduzione di controlli interni riguardanti la validità e la coerenza dei dati.

Nonostante queste precauzioni, il modello rimane fragile, per via della natura del linguaggio: l'utilizzo di `void*` permette la memorizzazione di qualsiasi valore, tra cui anche, altri puntatori; tuttavia, una eccessiva stratificazione di indirezione (in altre parole, puntatori a puntatori) può compromettere la corretta deallocazione della memoria.

²⁷ Disponibile nel repository, [C-data-structures](#), personale del candidato, seguendo il percorso: `src/linear/`. Alcuni commenti sono stati rimossi o spostati per ragioni di spazio.

Listato 34: Programmazione generica in C

```

typedef struct clinkedlist {
    node* tail;
    size_t element_count; /* Number of elements present in the list */
    size_t element_size; /* Size of the elements stored in the list */
} clinkedlist;

typedef struct node {
    void* ptr; /* Pointer to the memory location that stores the
        node's value */
    node* next; /* Pointer to the next node */
} node;

node* node_create(void* value, size_t value_size) {

    node* n = NULL;
    if (value_size <= SIZE_MAX) {

        n = (node*)malloc(sizeof(node));
        if (n) {

            n->ptr = malloc(value_size);
            if (n->ptr) {

                memcpy(n->ptr, value, value_size);
                n->next = NULL;
            }
            else {
                free(n);
                n = NULL;
            }
        }
    }
    return n;
}

```

Listato 35: Programmazione generica in Rust

```

struct Clinkedlist<T> {
    tail: Option<Box<Node<T>>>,
    element_count: usize
}

struct Node<T> {
    value: T,
    next: Option<Box<Node<T>>>
}

impl<T> Node<T> {
    fn new(value: T) -> Node<T> {
        Node {
            value,
            next: None
        }
    }
}

```

Rust adotta un approccio diverso, supportando la programmazione generica tramite le cosiddette *zero-cost abstractions*²⁸. Durante la fase di compilazione, il compilatore applica un processo detto *monomorfizzazione*, trasformando ogni utilizzo di una funzione o una struttura generica in una versione specifica per il tipo utilizzato. Rust consente inoltre di specificare vincoli sui tipi generici (*trait bounds*), tramite `impl Trait` e `where`. Questi meccanismi sono equivalenti, rispettivamente, a `implements Interface` e `<T extends ...>` di Java.

Per ottenere un confronto sulla complessità del codice, viene riportato, nel Listato 35, l'implementazione in Rust del Listato 34: il supporto alla programmazione generica di Rust permette di definire strutture dati e funzioni generiche con codice sicuro e conciso.

L'implementazione risulta più semplice rispetto alla controparte C: non è necessario gestire manualmente la dimensione dei valori o gli accessi diretti alla memoria. A tal proposito, si consideri che `Node::<T>::new()` del Listato 35 è l'equivalente di `node_create` del Listato 34.

²⁸ Si tratta di meccanismi e implementazioni astratti che non introducono overhead in fase di esecuzione rispetto al codice esplicito (concreto) equivalente.

Gestione degli errori

I due linguaggi si distinguono per i meccanismi impiegati per la gestione degli errori. In C, tipicamente, tale gestione è opzionale: il linguaggio non impone alcun controllo esplicito, lasciando la responsabilità al programmatore. Il meccanismo maggiormente diffuso per segnalare errori è basato sul valore di ritorno, tipicamente interi, da parte delle funzioni, secondo convenzioni informali e non uniformi²⁹.

L'assenza di vincoli espliciti sul controllo dell'esito rende facile, e comune, la mancata verifica di errori. Questo tuttavia può portare a comportamenti indefiniti, principalmente dovuti all'elaborazione di dati non validi, derivanti da chiamate a funzioni fallite, il cui risultato viene interpretato come se fosse corretto.

Come riportato dalla documentazione ufficiale [9], Rust distingue tra due classi di errori, *recuperabili* e *irrecuperabili*, fornendo meccanismi distinti per la loro gestione.

Gli errori *irrecuperabili* rappresentano situazioni dalle quali non è sicuro proseguire con l'esecuzione del programma, come l'accesso oltre i limiti di un array o il tentativo di dereferenziare un valore `None`³⁰. Rust gestisce questo tipo di errore tramite *panic*, i quali causano l'immediata interruzione del programma e rilasciano correttamente tutte le risorse allocate dal processo.

Al contrario, gli errori *recuperabili* rappresentano situazioni meno gravi, per le quali è possibile definire una logica di gestione, senza necessità di interruzione: per esempio, il tentativo di apertura di un file inesistente potrebbe essere gestito creando il file, invece di terminare l'esecuzione. Per la gestione di questo tipo di errore Rust adotta l'enumerazione `Result<T, E>`.

A differenza del C, la gestione esplicita degli errori è obbligatoria in Rust: gli errori *irrecuperabili* vengono gestiti tramite *panic*; per quelli *recuperabili*, il compilatore impone al programmatore di gestire esplicitamente entrambe le varianti (`Ok` ed `Err`) tramite costrutti come `match` e `if`

29 Lo standard C non definisce una convenzione uniforme sugli interi da utilizzare come codice di errore. Alcune funzioni utilizzano 0 per indicare successo e ogni altro valore viene interpretato come fallimento. Per questo motivo, è sempre opportuno controllare la documentazione di una funzione, per determinare come interpretare l'esito: alcune funzioni potrebbero interpretare ogni valore non negativo con successo, come altre potrebbero utilizzare 0 per fallimento.

30 Rust utilizza l'enumerazione `Option<T>` per rappresentare un valore opzionale: `Some(T)` rappresenta la presenza di un valore, di tipo `T`, mentre `None` rappresenta l'assenza di un valore valido.

let oppure operatori come `?`, il quale propaga l'errore. Questo obbligo riduce la probabilità di errori non gestiti, rendendo il codice più sicuro e robusto rispetto all'approccio tradizionale di C.

Macro

I due linguaggi gestiscono le *macro* in maniera completamente differente. Come riportato dalla documentazione GCC [8], le *macro* in C sono frammenti di codice a cui viene associato un nome, dichiarate con la direttiva `#define`.

Durante la fase di pre-processamento (prima della compilazione), quando viene incontrato il nome di una *macro*, esso viene sostituito con il codice associato. Si tratta di una semplice sostituzione testuale, come mostrato nel Listato 36.

Listato 36: Definizione di *macro* in C

```
#define printline(x) printf("%s\n", x)
int main(void) {
    printline("Stampa di prova");
    return 0;
}
```

Nel Listato 36, viene definita la *macro* `printline`, che stampa la stringa `x`, andando a capo successivamente. All'interno della funzione `main` viene invocata `'printline("Stampa di prova")'`: questa sarà espansa in `'printf("%s\n", "Stampa di prova")'` dal pre-processore.

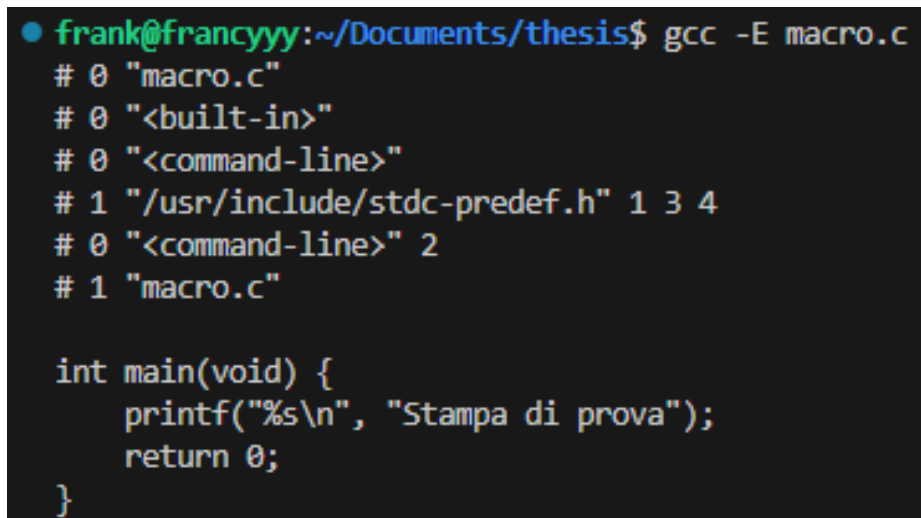
In quanto la sostituzione avviene durante la fase di pre-processamento, per osservare il codice generato è necessario utilizzare il flag `-E` durante la compilazione con GCC³¹, come mostrato in Figura 18.

La natura semplice delle *macro* C comporta delle limitazioni e falle di sicurezza, molte delle quali sono riportate sulla documentazione GCC [8]. A dimostrazione di ciò, si consideri l'esempio riportato nel Listato 37, modifica del Listato 36.

Listato 37: Utilizzo scorretto di *macro* in C

```
#include <stdio.h>
#define printline(x) printf("%s\n", x)
int main(void) {
```

31 Tramite la opzione `-E` viene indicato a GCC di fermarsi dopo la fase di pre-processamento, senza effettuare la compilazione. L'output della fase di pre-processamento verrà poi indirizzato allo standard output.



```

● frank@francyy:~/Documents/thesis$ gcc -E macro.c
# 0 "macro.c"
# 0 "<built-in>"
# 0 "<command-line>"
# 1 "/usr/include/stdc-predef.h" 1 3 4
# 0 "<command-line>" 2
# 1 "macro.c"

int main(void) {
    printf("%s\n", "Stampa di prova");
    return 0;
}

```

Figura 18: Traduzione di una *macro* in C

```

int my_val = 150;
println(my_val);
return 0;
}

```

Nel Listato 37 si tenta di sfruttare la stessa macro, `println`, per stampare il contenuto di un intero. È interessante osservare che, nonostante il mismatch dei tipi (`%s` indica di interpretare il parametro come stringa), il compilatore genera soltanto un warning, producendo lo stesso un file eseguibile. Tuttavia, tentando l'esecuzione si verifica un crash immediato dovuto a *segmentation fault*, come è osservabile in Figura 19.

In Rust le *macro* si distinguono in due classi: dichiarative e procedurali. Le *macro* procedurali sono molto complesse e richiederebbero un capitolo a parte, ma andrebbe oltre l'ambito di questa trattazione.

Le *macro* dichiarative, pur essendo meno complesse rispetto a quelle procedurali, sono caratterizzate comunque da una sintassi particolare, basata su *match di pattern*, permettendo perfino l'*overloading* in base al *pattern* come riportato dalla documentazione ufficiale [9]. Anche queste richiederebbero un capitolo dedicato per una spiegazione sufficientemente dettagliata.

Per mostrare la maggiore complessità, e differente gestione, delle *macro*

```

frank@francyyy:~/Documents/thesis$ gcc macro.c -o macro-c
macro.c: In function 'main':
macro.c:2:29: warning: format '%s' expects argument of type 'char *',
but argument 2 has type 'int' [-Wformat=]
  2 | #define printline(x) printf("%s\n", x)
    |                               ^~~~~~
macro.c:5:5: note: in expansion of macro 'printline'
  5 |     printline(t);
    |     ^~~~~~
macro.c:2:31: note: format string is defined here
  2 | #define printline(x) printf("%s\n", x)
    |                               ^
    |                               |
    |                           char *
    |                           %d
frank@francyyy:~/Documents/thesis$ ./macro-c
Segmentation fault (core dumped)

```

Figura 19: Limitazioni delle *macro* C

Rust rispetto alla controparte C, si consideri l'esempio rappresentativo³² nel Listato 38.

Listato 38: Definizione di *macro* dichiarativa in Rust

```

macro_rules! comp_eval {
    ($sx:expr; and $dx:expr) => {
        println!(
            "{:?} and {:?} is {:?}",
            stringify!($sx), stringify!($dx), $sx && $dx
        )
    };
    ($sx:expr; or $dx:expr) => {
        println!(
            "{:?} or {:?} is {:?}",
            stringify!($sx), stringify!($dx), $sx || $dx
        )
    };
    ($not:expr) => {
        println!(
            "not {:?} is {:?}",
            stringify!($not), !($not)
        )
    }
}

```

³² L'esempio fornito è basato su quello riportato nella documentazione ufficiale [9], nella sezione relativa all'overloading di macro.

```

    )
};
}
fn main() {
    comp_eval!(1 + 2 == 3; and true);
    comp_eval!(true; or false);
    comp_eval!(true);
}

```

Nel Listato 38 è definita la *macro* dichiarativa `comp_eval` che, in base al pattern, espande un codice diverso:

- Se viene fornita una coppia di espressioni, `$sx` e `$dx`, separate dalla parola chiave `and`, la *macro* espande una stampa della loro congiunzione logica;
- Se le espressioni sono separate dalla parola chiave `or`, espande una stampa della loro disgiunzione logica;
- Se invece viene fornita una singola espressione, `$not`, la *macro* espande una stampa della sua negazione logica.

Il risultato dell'esecuzione di questo codice è osservabile in Figura 20.

Le *macro* Rust rappresentano uno strumento molto potente rispetto alle direttive pre-processor di C: permettono di definire comportamenti differenti in base alla struttura dell'invocazione, cosa non possibile con la semplice sostituzione testuale.

Un'altra differenza rispetto alle *macro* C riguarda la loro espansione: le *macro* in Rust vengono espanse a livello di compilazione e in caso di mancata corrispondenza con qualsiasi *pattern* previsto viene generato un errore di compilazione.

Ad esempio, aggiungendo l'invocazione `'comp_eval!("test")'` all'interno della funzione `main` nel Listato 38, si otterrà un errore durante la compilazione in quanto il pattern fornito non corrisponde a nessuna delle regole definite nella *macro*³³. Il messaggio di errore è osservabile in Figura 21.

³³ In questo esempio il pattern fornito è una singola stringa, quindi la regola adeguata sarebbe la terza, che accetta un'unica espressione. Tuttavia, l'operatore di negazione logica non può essere applicato a una stringa, di conseguenza il compilatore rifiuta il codice, generando un errore.

```
frank@francyy:~/Documents/thesis$ rustc macro.rs -o macro-rs
frank@francyy:~/Documents/thesis$ ./macro-rs
"1 + 2 == 3" and "true" is true
"true" or "false" is true
not "true" is false
```

Figura 20: Compilazione di una *macro* in Rust

```
frank@francyy:~/Documents/thesis$ rustc macro.rs -o macro-rs
error[E0600]: cannot apply unary operator `!` to type `&'static str`
  --> macro.rs:17:31
   |
17 |         stringify!($not), !($not)
   |                             ^^^^^^ cannot apply unary operator `!`
...
25 |     comp_eval!("test");
   |     ----- in this macro invocation
```

Figura 21: Errore durante la compilazione di una *macro* in Rust

Conclusioni

Nonostante siano limitati, gli aspetti esaminati mettono in luce un tratto distintivo di Rust, specialmente rispetto a C: il linguaggio richiede uno sforzo iniziale maggiore, dovuto alla maggiore complessità della sintassi, ma questo viene ripagato garantendo meccanismi di sicurezza della memoria affidabili e integrati nel linguaggio stesso.

PROGETTI E APPLICAZIONI REALI

In questo capitolo vengono presentati progetti concreti che dimostrano come Rust possa costituire una scelta valida non solo in termini teorici, ma anche nella pratica per lo sviluppo di basso livello quali driver, kernel e addirittura sistemi operativi completi.

I progetti analizzati sono in costante sviluppo ed evoluzione; le informazioni riportate in questa trattazione sono aggiornate ad Agosto 2025, ma potrebbero subire variazioni nel tempo.

Tra le applicazioni più rilevanti sono presenti: l'integrazione di Rust nel kernel di Windows 11, il progetto *Rust for Linux* (RfL) con i relativi sottoprogetti, l'implementazione di *sudo-rs* promossa da Ubuntu e il sistema operativo *Redox OS*.

5.1 KERNEL DI WINDOWS 11

L'adozione di Rust da parte di Microsoft risale al 2023, quando l'azienda ha iniziato a riprogettare sezioni critiche del kernel del sistema operativo Windows 11, come riportato in [10].

Prima di analizzare quest'adozione, è utile un breve excursus storico sull'evoluzione del kernel Windows, principalmente per comprendere l'importanza di questo cambiamento.

Lo sviluppo dei sistemi Microsoft iniziò con un kernel interamente scritto in assembly (ASM8086) nelle prime versioni di *MS-DOS*. Con *MS-DOS 3.0* venne introdotto il linguaggio C, ma il primo vero kernel scritto interamente in C si ebbe con Windows NT 3.1 (**KERNEL.EXE**, 1993). Successivamente, C è rimasto il linguaggio principale per il kernel di Windows, con integrazioni C++ nelle versioni successive.

Il passaggio a Rust in Windows 11 è stato motivato da una combinazione di esigenze aziendali e del crescente interesse per il linguaggio Rust, che iniziò a essere popolare proprio in quel periodo, grazie al modello offerto: garantire la sicurezza della memoria senza sacrificare le prestazioni.

In particolare, Rust ha permesso a Microsoft di affrontare due aspetti fondamentali:

- **Sicurezza della memoria:** Windows ha una storia documentata alle spalle di errori legati alla memoria, quali *null reference*, *buffer overflow*, scritture illegali e altri. Il *modello di ownership* offerto da Rust consente di prevenire queste problematiche già a tempo di compilazione;
- **Gestione della concorrenza:** Data la complessità Windows, anche gli errori di concorrenza sono frequenti. Il *Borrow Checker* di Rust, insieme all'uso di *smart pointers* per la condivisione dei dati, garantisce integrità e un uso corretto delle risorse condivise, eliminando errori quali *data race*.

Va sottolineato che, ad oggi, si tratta di un'adozione parziale: il kernel di Windows 11 è tuttora scritto prevalentemente in C, con alcune porzioni in C++. L'obiettivo dichiarato di Microsoft è intervenire sulle sezioni chiave del kernel e sulle nuove funzionalità, valutando caso per caso l'impiego di Rust.

Nonostante si tratti di un'adozione parziale, questa scelta rappresenta un traguardo significativo per Rust: guadagnare la fiducia di un colosso tecnologico come Microsoft ha contribuito alla crescita della popolarità del linguaggio, specialmente nel panorama della programmazione di sistema.

5.2 UBUNTU: SUDO-RS

Sebbene non faccia parte del kernel, il comando `sudo` è uno degli strumenti più importanti e diffusi nel mondo Linux. La sua funzione è fondamentale: *Substitute User DO* consente l'esecuzione di comandi con i privilegi di un'altro utente, solitamente più elevati, come quelli di un utente privilegiato o di super-utente. Proprio per questo, è uno degli strumenti più delicati dal punto di vista della sicurezza di un sistema.

Dalla sua prima introduzione fino a oggi, `sudo` è stato sviluppato nel linguaggio C. Come riportato in [14], tuttavia, Canonical ha annunciato

che, a partire dalla versione 25.10 di Ubuntu, verrà introdotta una nuova implementazione del comando, scritta interamente in Rust, `sudo-rs`.

La scelta di Rust deriva dalla necessità di garantire una maggiore sicurezza della memoria e ridurre le vulnerabilità tipiche delle implementazioni in C, come accessi a memoria non valida. Il *modello di ownership* di Rust rappresenta una scelta perfetta per questo scopo, eliminando intere classi di errori pur mantenendo prestazioni comparabili a quelle di C.

Il progetto, guidato dalla *Trifecta Tech Foundation*, punta a migliorare la sicurezza di uno degli strumenti più sensibili dei sistemi Linux. Alla sua realizzazione contribuisce anche Todd Miller, storico maintainere del comando negli ultimi trent'anni, fornendo supporto tecnico e linee guida al team di sviluppo del progetto.

Le principali vulnerabilità della versione attuale riguardano semplici errori di gestione della memoria, in particolare *use-after-free* e accessi oltre i limiti. Queste problematiche possono essere sfruttate per eseguire attacchi di tipo *privilege escalation*¹.

Gli obiettivi del progetto `sudo-rs` includono:

- Evitare errori di tipo *use-after-free* e *access out of bounds*, per evitare, come in passato, *privilege escalation*;
- Migliorare la gestione dell'escaping dei caratteri speciali nella shell, evitando l'esecuzione di comandi indesiderati e potenzialmente malevoli²;
- Integrare il controllo sui profili di AppArmor;
- Supportare l'utilizzo di `sudoedit`;
- Garantire retrocompatibilità con kernel precedenti alla versione 5.9, come Ubuntu 20.04 LTS.

La filosofia del gruppo di sviluppo segue il principio '*less is more*': evitare un eccesso di funzionalità iniziali per ridurre la complessità e il rischio di introdurre errori logici da risolvere in seguito.

¹ La *privilege escalation* rappresenta uno scenario in cui un utente con permessi limitati (ad esempio, *guest*) ottiene, sfruttando le vulnerabilità di un sistema, privilegi più elevati (ad esempio, *root*).

² A parte il primo obiettivo, che rappresenta un miglioramento rispetto alla versione attuale di `sudo`, gli altri quattro esprimono dei miglioramenti rispetto alle prime versioni di `sudo-rs`. Essendo una nuova implementazione, nelle prime versioni erano presenti diversi errori logici, tra cui uno che permetteva *path traversal* (come riportato in [2]), ovvero la manipolazione dei percorsi, tramite username con caratteri speciali (`/` e `.`).

Per questo motivo, `sudo-rs` non mira a essere un rimpiazzo totale di `sudo`: molte feature, specialmente quelle meno utilizzate o considerate meno importanti potrebbero venire non implementate inizialmente.

Il nuovo comando è già disponibile per gli utenti che desiderano testarlo³ e fornire feedback alla comunità.

Come nota finale, Canonical garantisce che la versione originale in C del comando continuerà a essere mantenuta e distribuita, così da lasciare all'utente la libertà di scegliere quale versione adottare, in base alle proprie preferenze o esigenze.

5.3 REDOX OS

Redox OS rappresenta una pietra miliare per il linguaggio Rust, in particolare modo nell'ambito della programmazione di sistema. Redox, come riportato dal sito ufficiale [11], è un sistema operativo general purpose Unix-like basato su microkernel; molti sistemi operativi rientrano in questa categoria (come Minix o BlackBerry QNX), ma ciò che distingue Redox risiede nell'implementazione, interamente in Rust.

La scelta di usare solamente Rust ha l'obiettivo di mettere in luce le capacità pratiche del linguaggio nello sviluppo sia di (micro)kernel che di programmi general purpose, fornendo un'alternativa completa a Linux o BSD.

Il progetto Redox ha riscontrato successo grazie a una combinazione di design basato su microkernel, impiego di Rust, compatibilità con POSIX e con la maggior parte dei programmi Linux/BSD grazie alla propria libreria C (anch'essa sviluppata in Rust).

DESIGN BASATO SU MICROKERNEL Un microkernel è un insieme di software minimale che offre i meccanismi necessari per implementare un sistema operativo: il microkernel da solo non è sufficiente, è necessario integrare moduli aggiuntivi per ricoprire tutte le funzionalità di un sistema operativo.

Questo permette il caricamento, la modifica e la rimozione di moduli a runtime, senza la necessità di riavviare il sistema come conseguenza di ogni cambiamento. Inoltre, in quanto i moduli sono esterni al microkernel, essi vivono nello spazio utente, garantendo isolamento dei bug: anche se

³ È possibile installare il comando tramite il package manager apt, cercando `rust-sudo-rs`; successivamente, sarà disponibile come `'sudo-rs'`.

un intero modulo andasse in crash, non avrebbe conseguenze sul resto del sistema⁴ (il kernel non risentirebbe del crash).

SCRITTURA IN RUST Redox trae beneficio dai vari aspetti caratteristici del linguaggio, nonché da molte delle garanzie offerte:

- Il *modello di ownership* impone regole che prevengono la maggior parte di errori legati alla memoria, come *null reference*, *use-after-free*, *unfreed memory* e *double free*, nonché *data race*;
- Il *Borrow Checker* vieta la condivisione non sicura di risorse, imponendo vincoli rigidi per garantire l'integrità dei dati condivisi da più thread o processi;
- L'utilizzo di astrazioni sicure come `Option<T>` e `Result<T,E>` obbliga il programmatore a gestire esplicitamente sia il successo sia l'eventuale fallimento di un'operazione, evitando accessi a dati non validi;
- Tutti questi meccanismi, uniti alla sintassi restrittiva del linguaggio, eliminano classi intere di bug, lasciando solo quelli legati alla logica applicativa⁵.

Sia il microkernel che i vari driver sono implementati interamente in Rust, sfruttando le caratteristiche sopra citate e senza richiedere un ulteriore sforzo per garantire la *thread-safety* o la *memory-safety*⁶.

COMPATIBILITÀ CON POSIX Redox mira a essere compatibile con la maggior parte delle applicazioni Linux e, più in generale, a rispettare lo standard POSIX. Tale compatibilità non è di tipo binario, ma è ottenuta a livello di codice sorgente: in molti casi è sufficiente ricompilare l'applicazione per renderla eseguibile su Redox.

⁴ Questo rappresenta uno dei maggiori vantaggi dell'approccio microkernel. Non vi è il rischio che l'intero sistema si corrompa in seguito al crash di una singola applicazione o modulo.

⁵ Queste considerazioni valgono solamente se si sviluppa codice Rust ideomatico, utilizzando correttamente gli strumenti forniti e rispettando i vincoli del *Borrow Checker*. Per esempio, utilizzare `unsafe` con l'unico scopo di aggirare alcuni vincoli del *Borrow Checker*, oltre a rappresentare un utilizzo non ideomatico di Rust, introduce potenzialmente errori (come inconsistenze tra letture e scritture),

⁶ È richiesto uno sforzo maggiore solo durante la fase iniziale dello sviluppo, ma una volta raggiunta la compilazione, si ha la garanzia che la memoria e la concorrenza sono gestite senza incoerenze.

Questo risultato è reso possibile da `relibc`, la libreria C di Redox, scritta interamente in Rust per mantenere coerenza con la filosofia del progetto. Grazie a `relibc`, Redox supporta già numerose applicazioni fondamentali, tra cui GNU `bash`, `Git`, `Ffmpeg`, `GCC` e `LLVM`, rendendolo potenzialmente utilizzabile anche da un utente finale.

CONSIDERAZIONI SU REDOX È necessario osservare che Redox, nonostante sia nato come sistema operativo general purpose, è ancora in fase di sviluppo, non ancora maturo per un utilizzo quotidiano, specialmente se paragonato a colossi del calibro di Windows, MacOS o Ubuntu, i quali forniscono una buona esperienza utente già *'out of the box'*.

Ciononostante, Redox costituisce una dimostrazione concreta della maturità e dell'usabilità di Rust per lo sviluppo di sistemi operativi completi, confermandone l'applicabilità ben oltre la teoria.

5.4 RUST FOR LINUX

Il progetto *Rust for Linux* nasce con l'obiettivo di integrare il supporto a Rust come linguaggio di programmazione utilizzabile all'interno del kernel Linux, dimostrandone l'idoneità nello sviluppo di componenti di basso livello, principalmente driver, tradizionalmente implementati in C.

L'interesse verso Rust deriva principalmente dalle garanzie offerte dal linguaggio sotto gli aspetti di *memory-safety* e *thread-safety* senza costi di esecuzione aggiuntivi. Grazie al *modello di ownership* intere classi di errori vengono completamente prevenute a livello di compilazione, come accessi a memoria non inizializzata, memoria non liberata, *double free* a altri ancora, comuni nei sistemi sviluppati in C⁷.

Il progetto raccoglie un insieme di contributi al kernel, in gran parte rappresentati da driver per dispositivi fisici (schede di rete, NVMe) e virtuali (GPU virtuali), sviluppati interamente in Rust.

La documentazione ufficiale è consultabile dal sito di Rust for Linux [13]. Inoltre, è opportuno osservare che il progetto è in continua evoluzione, le informazioni riportate in questa trattazione sono aggiornate ad Agosto 2025, ma potrebbero variare nel futuro.

Per comprendere l'importanza di questo progetto, è utile un breve excursus sull'evoluzione dei linguaggi supportati dal kernel linux.

⁷ Il *modello di ownership* viene esposto nel Capitolo 3, mentre nel Capitolo 4.2.2 è possibile osservare quali errori vengono effettivamente eliminati grazie ad esso.

Il kernel Linux venne rilasciato inizialmente nel 1991 da Linus Torvalds ed era quasi interamente scritto in C, con alcune sezioni in assembly. Questa situazione rimase invariata fino al 2007, quando venne proposta l'integrazione di C++ nel kernel.

Tuttavia, l'idea venne rifiutata da Torvalds stesso, contrario all'utilizzo di un linguaggio considerato meno trasparente e più complesso rispetto a C, come riportato in [6], dove vengono esposte alcune motivazioni specifiche a riguardo.

Solo nel 2022, con la versione 6.1, venne aggiunto un primo supporto sperimentale a Rust, sufficiente per permettere agli sviluppatori di iniziare a scrivere codice Rust nel kernel e testarne l'integrazione.

Al giorno d'oggi il kernel Linux consente lo sviluppo di moduli e driver, sia in C che in Rust, definiti *out-of-tree*: moduli sviluppati e mantenuti separatamente dal codice sorgente del kernel, spesso proprietari o specifici per hardware particolare.

Il progetto *Rust for Linux* si concentra soprattutto sullo sviluppo di moduli *in-tree*, inclusi direttamente nella *mainline* del kernel (in altre parole, incluse nel sorgente del kernel). Tuttavia, il progetto raccoglie anche moduli *out-of-tree*, permettendo agli sviluppatori un ambiente sicuro per sperimentare e validare soluzioni prima di proporle per un'inclusione ufficiale nella *mainline*.

Un concetto chiave introdotto dal progetto è quello di *reference driver* (driver di riferimento), ovvero implementazioni in Rust che possono essere integrate nei sottosistemi senza sostituire i driver C esistenti.

Questi driver hanno diverse funzioni:

- Definire astrazioni sicure per i nuovi driver, evitando la riscrittura o duplicazione del codice esistente;
- Fornire un modello di riferimento per gli sviluppatori C, mostrando come un driver equivalente possa essere realizzato in Rust;
- Sfruttare le infrastrutture già presenti *in-tree* per preparare i sottosistemi a un'integrazione graduale e progressiva di Rust;
- Facilitare l'apprendimento graduale del linguaggio da parte degli sviluppatori del kernel;

- Valutare la convenienza dell'adozione: verificare quanta parte del codice possa essere scritta in modalità *safe*, quanti bug del driver originale C verrebbero effettivamente eliminati e quale sia l'impatto sulla manutenzione.

In molti casi, un *reference driver* può essere semplicemente un prototipo o un banco di prova, più che un driver destinato all'uso in produzione.

5.4.1 Moduli in mainline

Come precedentemente accennato, i moduli e driver *in mainline* fanno ufficialmente parte del kernel: sono inclusi nel sorgente e compilati insieme ad esso, risultando disponibili senza necessità di installazioni aggiuntive. La categoria *in mainline* di *Rust for Linux* comprende principalmente driver per schede di rete (*Network Interface Cards*) e per GPU, oltre ad alcuni moduli di utilità.

DRIVER PER NIC Due dei principali moduli *in-tree* del progetto sono driver per dispositivi di rete, **AMCC QT2025 PHY** e **ASIX PHY**.

Il primo è un driver per l'omonimo dispositivo (transceiver **AMCC QT2025**) e fornisce un'interfaccia verso lo stack di rete del kernel, facilitando l'interazione tra il dispositivo e il sistema operativo e la gestione delle funzionalità di rete. È stato integrato nel kernel nella versione 6.12.

Il secondo è destinato ai dispositivi Ethernet del produttore **ASIX**, ma ha finalità principalmente dimostrative: è sviluppato come *reference driver*, per fornire un esempio di implementazione di un driver PHY (layer fisico dello stack ISO/OSI) in Rust. È stato introdotto nella versione 6.8.

DRIVER PER GPU Tra i moduli *in-tree* rientrano anche due driver per schede grafiche: **Nova GPU** e **Tyr GPU**.

Nova GPU è un driver per le schede grafiche NVIDIA a partire dalla serie RTX 2000, concepito come successore dell'attuale *Nouveau*. La descrizione dettagliata del driver è riportata nella Sezione 5.5.

Tyr GPU è un driver *Direct Rendering Manager (DRM)* per le GPU *Arm Mali* basate su CSF (Command Stream Frontend), sviluppato come porting in Rust dell'attuale driver *Panthor* (in C). Il progetto ha un team di sviluppo che comprende ingegneri provenienti da *Collabora*, *Arm* e *Google*, e mira a fornire la stessa API per spazio utente attualmente offerta da *Panthor*, così da poter sostituire direttamente il driver nel contesto di *PanVK* (driver Vulkan).

Lo sviluppo di *Tyr* procede su due rami distinti:

- **Upstream:** Attualmente in grado di rilevare GPU su SoC (*System-on-Chip*) RK3588, leggere alcune sezioni della ROM della GPU e trasferirle allo spazio utente tramite chiamate API;
- **Downstream:** Utilizzato come *reference driver* per testare le nuove astrazioni proposte, prima della loro integrazione nella *upstream*. Attualmente è in grado di inviare piccoli pacchetti di lavoro alla GPU.

Il progetto *Tyr* è iniziato a giugno 2025 e si trova ancora in fase iniziale, instabile e fortemente sperimentale. Secondo la documentazione ufficiale (Sezione *Users – in mainline/Tyr GPU Driver*) [13], non gestisce ancora il controllo della corrente elettrica e implementa funzionalità limitate di recupero dagli errori, ma il team di sviluppo prevede di estenderle nei prossimi mesi.

DRIVER PER NULL BLOCK Il *Null Block device* (`/dev/null`) è un dispositivo a blocchi virtuale utilizzato principalmente per test e benchmarking: scarta tutti i dati scritti e restituisce EOF se si legge da esso, senza usufruire di memoria o spazio di archiviazione fisico.

Il driver attuale, `null_blk`, è scritto interamente in C e ha una nota storia di vulnerabilità legate alla gestione della memoria: un'analisi dei commit relativi al driver, consultabile presso [7], mostra che circa il 41% dei fix sono dovuti a errori di sicurezza della memoria. Ciò lo ha reso un ottimo candidato per un'implementazione in Rust.

A questo scopo, è stato sviluppato `rnull`, un driver scritto interamente in safe Rust, con porzioni minime di codice unsafe, incapsulate in astrazioni sicure per interagire con le API C del kernel.

Attualmente `rnull` replica gran parte delle funzionalità di `null_blk`, ma non è ancora completo e manca di alcune feature presenti nell'implementazione originale.

GENERATORE DI CODICI QR PER DRM PANIC Questo modulo è destinato ai sottosistemi *Direct Rendering Manager* (**DRM**) e ha lo scopo di semplificare l'analisi degli *stack trace* generati in seguito a un *kernel panic*.

Il problema di base è che i messaggi di errore del DRM possono essere molto lunghi e poco pratici da copiare manualmente. Il modulo genera un codice QR, scansionabile con uno smartphone, contenente le informazioni sull'errore, così da ottenere rapidamente i dettagli necessari all'analisi.

L'implementazione non richiede memoria o spazio di archiviazione aggiuntivi: sfrutta lo spazio libero nel buffer già riservato al processo per memorizzare il codice QR. È stato integrato nel kernel nella versione 6.12.

5.4.2 *Moduli outside mainline*

Oltre a moduli inclusi nella *mainline*, il progetto raccoglie anche sviluppi *outside mainline*, ovvero non integrati direttamente nel kernel: pur essendo compatibili con una determinata versione del kernel, vengono mantenuti e distribuiti separatamente, richiedendo quindi installazioni aggiuntive prima di essere utilizzabili.

In questa categoria rientrano driver per dispositivi di storage, GPU proprietarie, filesystem e moduli per Android.

Il progetto *Rust for Linux* privilegia lo sviluppo di moduli *in-tree*, ma ciò non rende meno rilevanti quelli *out-of-tree*: spesso si tratta di soluzioni pensate per contesti o hardware specifici; per questo, solitamente, la documentazione disponibile può risultare limitata.

ANDROID ASHMEM Ashmem (*Anonymous Shared Memory Subsystem for Android*) è un allocatore di memoria condivisa per Android, concettualmente simile a POSIX SHM ma con un'API più semplice e basata su file.

È progettato per liberare automaticamente le regioni di memoria condivisa quando il sistema è sotto pressione (ovvero quando la memoria fisica si sta saturando), caratteristica che lo rende particolarmente adatto a dispositivi con risorse limitate.

DRIVER PER ANDROID BINDER Questo progetto mira a riscrivere in Rust il driver kernel per il **Binder** di Android. Il *Binder* è un componente fondamentale per la sicurezza e le prestazioni dei sistemi Android: gestisce la IPC (*Inter-Process Communication*) all'interno del *sandbox*⁸ di Android, permettendo la comunicazione tra applicazioni isolate.

La sua natura critica lo rende particolarmente vulnerabile a errori legati alla gestione della memoria, per cui trarrebbe significativi vantaggi dalle garanzie offerte dal *modello di ownership* di Rust, sia intermini di sicurezza che di prestazioni.

⁸ Android isola l'esecuzione delle applicazioni in ambienti detti Sandbox, per cui ogni applicazione ha il proprio ambiente privato per la sua esecuzione, separato dalle altre.

DRIVER PER APPLE AGX Questo driver è destinato alla GPU **AGX** di Apple ed è accompagnato da binding *DRM (Direct Rendering Manager)* per lo spazio utente. Oltre a far parte di *Rust for Linux*, il driver rientra anche nel progetto *Asahi Linux*, il quale mira a portare supporto Linux sulle CPU Apple Silicon.

L'interesse principale verso *AGX* deriva proprio dal contesto di *Asahi*: la documentazione tecnica più dettagliata è presente nelle pagine ufficiali di tale progetto [3].

Attualmente lo sviluppo è focalizzato sull'implementazione di driver per *OpenGL* e *Vulkan* e sul *reverse engineering* del set di istruzioni supportato dalla GPU.

DRIVER PER NVME Questo driver rappresenta il tentativo di sviluppare un driver per dispositivi di *storage* interamente in safe Rust, concepito principalmente come *reference driver*. L'obiettivo è dimostrare la fattibilità di astrazioni sicure per dispositivi ad alte prestazioni, oltre che fornire un esempio concreto per sviluppi futuri.

Allo stato attuale si trova in fase sperimentale e instabile, non adatta a un uso in produzione.

DRIVER PER FILESYSTEM PUZZLEFS **PuzzleFS** è un filesystem per container progettato per superare alcune limitazioni dell'attuale stack *OCI (Open Container Initiative)*. Il progetto mira a ridurre la duplicazione dei dati, garantire build riproducibili, supportare il *mounting* diretto e garantire la sicurezza della memoria.

Il driver, scritto in Rust, implementa queste caratteristiche attraverso:

- **Riduzione della duplicazione:** utilizzo dell'algoritmo *FastCDC* per condividere segmenti di memoria tra i vari layer;
- **Build riproducibili:** definizione di una rappresentazione canonica del formato delle immagini;
- **Sicurezza della memoria:** ottenuta implicitamente grazie al *modello di ownership* di Rust.

5.4.3 Impatto del progetto

Tra i recenti esempi di integrazione di Rust in progetti esistenti, *Rust for Linux* è senza dubbio quello che ha avuto il maggiore impatto tec-

nico e mediatico, generando anche un ampio dibattito all'interno della community *open source*.

Il motivo è semplice ma fondamentale: il progetto coinvolge il kernel Linux, cuore di vari sistemi operativi alla base di gran parte dell'infrastruttura digitale odierna, sviluppato e mantenuto da una comunità vasta e con una cultura tecnica consolidata.

A confronto con:

- **Kernel di Windows 11:** Windows è un sistema proprietario sviluppato internamente da Microsoft, quindi l'integrazione di Rust non genera lo stesso dibattito pubblico né ha lo stesso impatto sulla community *open source*;
- **sudo-rs:** riscrittura in Rust del comando `sudo`, importante per la sicurezza, ma limitato a un singolo strumento, non all'intero sistema;
- **Redox OS:** sistema operativo scritto interamente in Rust, ma con adozione limitata (è considerato principalmente un progetto '*di nicchia*') rispetto a Linux, il quale è alla base di Android e della maggior parte dei server.

MOTIVAZIONI DELL'IMPORTANZA L'integrazione di Rust nel kernel Linux rappresenta il primo tentativo concreto di introdurre un linguaggio con garanzie di sicurezza della memoria nella *mainline* del kernel. L'obiettivo è ridurre la classe di vulnerabilità legate a errori di gestione della memoria, mantenendo le prestazioni richieste da un sistema operativo.

L'importanza è amplificata dal fatto che Linux è alla base di una porzione enorme dell'ecosistema tecnologico globale: dai dispositivi mobili Android ai vari server che alimentano servizi cloud e web. Qualsiasi cambiamento strutturale al kernel ha quindi effetti su larga scala.

Infine, si tratta di un cambiamento che rompe una tradizione radicata: dal 1991 il kernel Linux è scritto quasi interamente in C, e in passato altri linguaggi (incluso C++) non sono stati accettati. L'introduzione di Rust non è solo un aggiornamento tecnico, ma anche una modifica della filosofia di sviluppo del progetto.

CONSEGUENZE SOCIO-CULTURALI L'ultima proposta per introdurre un nuovo linguaggio nel kernel risale nel 2007, quando Torvalds rifiutò l'adozione di C++. Questa storia rende l'accettazione, seppure parziale, di Rust un segnale di apertura.

In una prima occasione, Linus Torvalds ha espresso un approccio pragmatico verso Rust, riconoscendo sia le resistenze culturali che i possibili benefici tecnici:

‘Rust is a very different thing, and there are a lot of people who are used to the C model. They don’t like the differences, but that’s OK [...] Clearly, some people just don’t like the notion of Rust and having Rust encroach on their area. But we’ve only been doing Rust for a couple of years, so it’s way too early to say Rust is a failure’.

– *Linus Torvalds, discussione alla ‘Linux Kernel Mailing List’, 2022*

In un’altra occasione, ha chiarito di non considerare Rust un rimpiazzo totale di C, ma uno strumento aggiuntivo, utile in casi specifici:

‘I do not think Rust will take over the kernel, and I don’t think anybody is even suggesting that. But I do think Rust can be a good tool for some things, and we should use the best tool for the job’.

– *Linus Torvalds, conferenza ‘Linux in the Multiverse’, 2024*

Queste dichiarazioni evidenziano due aspetti fondamentali: da un lato, l’esistenza di una divisione culturale tra sviluppatori più legati al modello C e sostenitori di Rust; dall’altro, la volontà di valutare Rust su basi pratiche e a lungo termine, senza pregiudizi definitivi e senza trasformarlo in una questione ideologica.

5.5 RED HAT: NOVA

Nova è un progetto sviluppato dall’azienda Red Hat, parte integrante dell’iniziativa *Rust for Linux*. Si tratta di un driver per GSP, un componente presente nelle schede grafiche NVIDIA di nuova generazione, dalla serie RTX 2000 in poi.

Il GSP è un componente hardware e firmware integrato nella GPU che consente di gestire quest’ultima come se fosse un sistema embedded. Tra le sue funzioni principali si trovano: la gestione dell’alimentazione, la gestione del clock, l’inizializzazione dell’hardware, lo scheduling delle code e il controllo termico. Permette di comunicare con la GPU come se fosse un’entità autonoma interna al sistema, sollevando la CPU da tutte le funzionalità precedentemente elencate.

Prima dell'introduzione del *GSP*, il driver open source *Nouveau*, sviluppato principalmente tramite *reverse engineering*⁹, soffriva di limitazioni in termini di prestazioni e stabilità rispetto ai driver proprietari forniti da NVIDIA.

Con il *GSP*, molte delle funzionalità ricostruite tramite *reverse engineering* sono ora gestite direttamente da questo processore dedicato, il quale funge da strato di astrazione che permette la comunicazione tra GPU e kernel tramite *IPC*. Questo cambiamento strutturale ha reso necessario lo sviluppo di un nuovo driver specifico.

I motivi che hanno portato Red Hat a scegliere Rust per lo sviluppo di *Nova* sono fondamentalmente tre:

- **memory-safety**: il motivo principale risiede nella garanzia di sicurezza della memoria offerta da Rust, così da prevenire in partenza la maggior parte di bug;
- **thread-safety**: le GPU sono composte da un numero elevato di *thread*, per cui il driver trae vantaggio dalla gestione sicura della concorrenza di Rust;
- **sperimentazione**: trattandosi di un nuovo driver e non di una riscrittura, vi era l'occasione per sperimentare e testare il linguaggio.

Nova è strutturato in due componenti principali:

- **nova-core**: esegue le operazioni a basso livello, come l'avvio del *GSP* e l'interazione, tramite quest'ultimo, con l'hardware;
- **nova-DRM**: fornisce un'interfaccia astratta conforme al *DRM* (*Direct Rendering Manager*), fondamentale per la comunicazione con lo spazio utente.

Questa architettura modulare permette di combinare *nova-core* con differenti driver grafici, oltre a *nova-DRM*. Ad esempio, è possibile impiegare *VFIO* per assegnare la GPU a macchine virtuali, sfruttando la possibilità, supportata a livello firmware, di creare più *vGPU* (GPU virtuali).

Attualmente, *nova-DRM* è sviluppato principalmente come driver grafico per ambienti virtualizzati, ma può essere utilizzato anche su sistemi fisici, sebbene non sia ottimizzato per questo ambito.

⁹ NVIDIA è nota per non rilasciare pubblicamente la documentazione delle proprie GPU, principalmente per motivi di marketing, per poter offrire il proprio driver proprietario, maggiormente ottimizzato rispetto alle alternative open source.

5.6 CONSIDERAZIONI SUI PROGETTI ANALIZZATI

In questo capitolo, sono stati esaminati diversi progetti concreti che mostrano come Rust possa essere integrato nei sistemi operativi, sia tramite l'aggiunta di moduli specifici che tramite la riscrittura completa di componenti esistenti.

Un aspetto comune di tutti gli esempi presentati è la volontà di affrontare uno dei principali punti deboli dell'uso del C: la gestione della memoria. Il *modello di ownership* di Rust, come discusso nel Capitolo 3 e mostrato nel Capitolo 4, riesce a eliminare già a tempo di compilazione molti degli errori legati alla gestione della memoria, i quali in C emergerebbero solo a tempo di esecuzione.

Le strategie adottate, tuttavia, sono diverse: alcuni progetti optano per un'interazione graduale (come il kernel di Windows 11 e *Rust for Linux*), mentre altri per una sostituzione totale (come *sudo-rs*). Allo stesso modo, in alcuni casi l'attenzione è posta interamente sulla sicurezza del sistema (Windows 11 e *Rust for Linux*), mentre in altri viene data importanza all'esperienza dell'utente finale (*Redox OS*).

Nel complesso, queste iniziative mostrano come Rust non sia più solamente un linguaggio *promettente* e sperimentale, ma uno strumento concreto, impiegabile in contesti reali e complessi, con risultati effettivi nella direzione di maggiore sicurezza e affidabilità dei sistemi.

CONCLUSIONI

Il Capitolo 2, ha ripercorso le origini del linguaggio e le motivazioni alla base del suo sviluppo, evidenziando come si sia evoluto in breve tempo, da progetto personale e sperimentale, in uno strumento adottato da un ampio numero di sviluppatori e aziende.

Nel Capitolo 3, sono stati analizzati a fondo i meccanismi di *ownership*, *borrowing* e *lifetime*, che costituiscono il fulcro del modello di gestione della memoria del linguaggio. Questi strumenti hanno mostrato come Rust riesce a garantire simultaneamente sicurezza e controllo, aspetti fondamentali per lo sviluppo di basso livello.

Il Capitolo 4, ha preso in analisi il linguaggio C come riferimento storico e pratico, ricostruendo le caratteristiche che lo hanno reso centrale nella programmazione di sistema: *compilazione*, *assenza di runtime*, *manipolazione della memoria* e dei *bit*. Rust è stato successivamente confrontato con C sia negli aspetti precedentemente elencati, sia in aspetti trasversali, quali *gestione delle risorse*, *sicurezza della memoria*, *complessità della sintassi* e *prestazioni*.

Ne è emerso un quadro in cui Rust rappresenta una valida alternativa, nella teoria, a C, a costo di una curva di apprendimento più ripida rispetto a quest'ultimo.

Infine, il Capitolo 5, ha illustrato progetti concreti già in corso che sperimentano l'uso del linguaggio nello sviluppo di sistemi operativi e componenti critici. Queste iniziative mostrano come il linguaggio non sia più una premessa puramente teorica, ma uno strumento concreto già in grado di produrre risultati significativi.

Nel complesso, l'analisi ha messo in luce le potenzialità e i limiti di Rust nello sviluppo di basso livello. Rust appare in grado di portare benefici significativi nello sviluppo di applicazioni di basso livello e, in particolare,

di sistemi operativi. Grazie al *modello di ownership*, gli errori di gestione della memoria dinamica vengono di fatto eliminati a tempo di compilazione; inoltre le astrazioni *zero-cost* e l'impiego di *smart pointer* consentono una gestione sicura della concorrenza, prevenendo interamente le *data race*.

Tuttavia, Rust non rappresenta una soluzione magica né priva di limitazioni. La sintassi e il *Borrow Checker* richiedono tempo e dedizione per essere padroneggiati, definendo una curva di apprendimento molto più ripida rispetto a linguaggi più permissivi come Python o anche C. A questo va aggiunto il fatto che Rust si limita a prevenire gli errori di gestione della memoria, ma non quelli logici: la correttezza di un algoritmo o di un'implementazione rimane ancora pienamente responsabilità del programmatore.

La mia introduzione al linguaggio non è stata troppo ardua inizialmente. La mia esperienza con il C e la programmazione di sistema in Unix, seppure da livello universitario, mi ha permesso, in una fase iniziale, di comprendere facilmente i concetti di *Ownership* e *Borrowing*. A questo si può aggiungere la conoscenza del linguaggio Java, il quale mi aiutato a comprendere *Generics* e gli altri aspetti riconducibili alla OOP (come i *trait* e l'utilizzo di *self* per simulare oggetti e metodi). Sotto questo aspetto, a primo impatto il linguaggio può sembrare molto astratto e di alto livello, per questo è stato sorprendente osservare, sia di prima persona che da estratti online, velocità di esecuzione così elevate; il merito risiede nel compilatore, *rustc*, e nella sua *ottimizzazione nell'ottimizzare il codice*, elemento che, oltre a giustificare, spiega il motivo dietro ai tempi di compilazione relativamente lunghi.

Le prime frustrazioni non hanno, però, tardato ad arrivare: è stato sufficiente tentare di sviluppare una semplice lista concatenata per scontrarmi con il *Borrow Checker* e con le *lifetime*, probabilmente il concetto più complicato del *modello di ownership*. Anche programmando in C si ha a che fare questo concetto, ma *fuori* dal linguaggio: lo sviluppatore deve tenere traccia dei puntatori, delle allocazioni e delle deallocazioni, per garantire un funzionamento corretto dell'applicazione. In Rust, invece, il concetto è esplicito: si deve spiegare, e giustificare, al *Borrow Checker* le validità temporali. Tuttavia, quest'ultimo analizza perfino casi limite, i quali potrebbero sfuggire al programmatore, rifiutando codice che a prima vista sembra corretto, ma non copre ogni singolo *edge-case*.

Per la mia esperienza, considero il linguaggio affascinante, ma in un *suo mondo*: la maggior parte della complessità deriva proprio dal *Borrow*

Checker e dalle *lifetime*, i quali, seppur non immediatamente intuitivi, permettono di evitare potenziali problemi. Proprio per la complessità, il linguaggio rappresenta un *trade-off*, in cui si scambia tempo (per lo sviluppo e l'acquisizione del linguaggio), per un codice sicuro. Non è sensato utilizzare Rust solo per il gusto di farlo (eccezione potrebbe essere fatta per scopi didattici), in quanto si rischia di nullificarne i benefici.

È quindi opportuno, o comunque consigliato, evitare un impiego indiscriminato del linguaggio. Da un lato, utilizzare Rust in contesti dove i suoi vantaggi non si traducono in benefici concreti (ad esempio, un'applicazione che non lavora con una quantità significativa di memoria dinamica) rischia di introdurre soltanto complessità; dall'altro, utilizzare Rust come se fosse un altro linguaggio, aggirando o utilizzando in maniera scorretta i costrutti disponibili (per esempio scrivendo ampie porzioni di codice *unsafe*) vanifica gran parte dei principi alla base del linguaggio stesso.

La prospettiva più equilibrata è dunque quella di impiegare Rust nei contesti critici, in cui sia le prestazioni che la sicurezza della memoria sono fondamentali e in cui i vantaggi offerti si traducono in maniera concreta, evitando un uso eccessivo o improprio che ne comprometterebbe i punti di forza.

BIBLIOGRAFIA

- [1] Ada Computer Science Team. Bitwise Manipulation. *University of Cambridge, Raspberry Pi Foundation*, n.d. Esempi sull'importanza delle operazioni bitwise.
- [2] Andrea Jegher, Radically Open Security. CVE-2023-42456. sudo-rs Session File Relative Path Trasversal vulnerability. *National Vulnerability Database (NVD)*, November 2023.
- [3] Asahi Linux Developers. Asahi Linux Project. <https://asahilinux.org/>, 2020. Usato semplicemente come riferimento per i lettori interessati ai dettagli del driver per GPU AGX.
- [4] Francesco Biribò. C-data-structures. <https://github.com/whocaresleft/C-data-structures>, February 2025.
- [5] Mohamed Amin Bouali. Bit Manipulation in C. *Medium*, May 2023.
- [6] Dario Meoli. Linus Torvalds odia il C++: "E' un linguaggio orribile, parola di Linus". *Zeus News*, September 2007. L'articolo esplora a fondo alcuni motivi per cui Torvalds non apprezza C++, nella trattazione è stato citato in maniera molto riassuntiva.
- [7] Git Summary: Linus Torvalds and Linux Kernel Contributors. History log dei commit su null_blk. [Git Kernel Commit Archive](#), 2022. Usato come riferimento per mostrare la mole di errori relativi al blocco nullo.
- [8] GNU foundation. Macros (The C Preprocessor). [GCC, the GNU Compiler Collection - Online Documentation](#), 2025.
- [9] Steve Klabnik and Caron Nichols. *The Rust Programming Language*. No Starch Press, San Francisco, 2nd edition, 2023.
- [10] Gary Olsen. Does Windows kernel use Rust and what does it mean for IT? *SearchEnterpriseDesktop (TechTarget)*, March 2025. Integrazione di Rust nel kernel di Windows 11 e l'evoluzione del kernel Windows.
- [11] Redox Developers. Redox OS: A Unix-like operating system written in Rust. <https://www.redox-os.org>, 2025. Project homepage.

- [12] Rust Embedded Working Group. *The Embedded Rust Book*. Rust Embedded Working Group, Online Documentation, 1st edition, 2019.
- [13] Rust for Linux Developers. Rust for linux project. <https://rust-for-linux.com>, 2025. Raccolte informazioni sui progetti sia dentro che fuori la mainline.
- [14] Joey Sneddon. Ubuntu 25.10 Switches to Rust-based Sudo. *OMG! Ubuntu*, May 2025.
- [15] Dark Bears Team. Why C Continues to be the Preferred Systems Programming Language. *Dark Bears Blog*, October 2018.
- [16] The Computer Language Benchmarks Game Team. The Computer Language Benchmarks Game. *The Computer Language Benchmarks Game (Debian)*, 2025. Raccolti estratti di benchmark: (n-body e fasta). Data di accesso: 2025-07-18.
- [17] Clive Thompson. How Rust went from a side project to the world's most-loved programming language. *MIT Technology Review*, February 2023.