

Propuesta de Tesis de Grado de Ingeniería en Informática

Detección de Deadlocks en Rust en tiempo de compilación mediante Redes de Petri

Director: Ing. Pablo A. Deymonnaz

Alumno: Horacio Lisdero Scaffino, (*Padrón # 100.132*)
hlisdero@fi.uba.ar

Facultad de Ingeniería, Universidad de Buenos Aires

17 de febrero de 2023

Índice

1. Introducción	2
1.1. El problema de la correctitud en programación concurrente	2
1.2. Motivación	3
2. Estado del arte / Literatura relacionada	4
3. Objetivos	4
4. Metodologías propuestas	5
5. Conjuntos de datos (dependiendo del área de la tesis)	5
6. Recursos informáticos (opcional)	5
6.1. Hardware	5
6.2. Software	5

6.2.1. Otras herramientas	5
7. Cronograma de trabajo	5

1. Introducción

1.1. El problema de la correctitud en programación concurrente

En el área de computación concurrente, uno de los desafíos principales es probar la correctitud de un programa concurrente. A diferencia de un programa secuencial donde para cada entrada se obtiene siempre la misma salida, en un programa concurrente la salida puede depender de cómo se intercalaron las instrucciones de los diferentes procesos o threads durante la ejecución.

La correctitud de un programa concurrente se define entonces en términos de propiedades del cómputo realizado y no en términos del resultado obtenido. En la literatura se definen dos tipos de propiedades de correctitud[1][3][26]:

- Propiedades de *safety*: Propiedades que se deben cumplir *siempre*.
- Propiedades de *liveness*: Propiedades que se deben cumplir *eventualmente*.

Dos de las propiedades de tipo *safety* deseables en un programa concurrente son:

- **Exclusión mutua**: dos procesos no deben acceder a recursos compartidos al mismo tiempo.
- **Ausencia de *deadlock***: un sistema en ejecución debe poder continuar realizando su tarea, es decir, avanzar produciendo trabajo útil.

Usualmente se utilizan primitivas de sincronización tales como mutexes, semáforos, monitores y *condition variables* para implementar el acceso coordinado de los procesos o hilos a los recursos compartidos. No obstante, el uso correcto de estas primitivas es difícil de lograr en la práctica y se pueden introducir errores difíciles de detectar y corregir. Actualmente la mayoría de lenguajes de uso general, ya sean compilados o interpretados, no permiten detectar estos errores en todos los casos.

Dada la creciente importancia de la programación concurrente debida a la proliferación de sistemas de hardware multihilo y multiproceso, reducir el número de *bugs* ligados a la sincronización de los procesos o hilos es de vital importancia para la industria. Especialmente en programación de bajo nivel y sistemas de misión crítica, evitar los deadlocks es un requerimiento ineludible en el desarrollo.

1.2. Motivación

Uno de los lenguajes de programación modernos más prometedores para programación concurrente es Rust[23]. Su modelo de memoria basado en el concepto de *ownership* y su expresivo sistema de tipos permite eliminar una amplia variedad de errores relacionados al manejo de memoria (*double free*[22], *use-after-free*[22], referencia colgante ”*dangling pointers*”[21], entre otros) y errores relacionados a la programación concurrente (*data races*[21], pasaje de variables de tipo *non-thread-safe* entre hilos[20], entre otros) en tiempo de compilación.

La seguridad del código generado es una de los objetivos principales del lenguaje Rust. Por un lado la generación de código incluye una serie de mitigaciones a *exploits* de diversos tipos[25]. Por otro lado, si bien la librería estándar no está exenta de errores[4], los procesos de gobierno open-source y transparentes basados en el modelo RFC (*Requests for Comments*)[24] aseguran una mejora continua del lenguaje y su funcionalidad.

Diversas investigaciones han llegado a la conclusión que 70 % de las vulnerabilidades encontradas en proyectos grandes en C/C++ ocurren debido a errores en el manejo de la memoria. Este hecho preocupante se puede observar en proyectos tales como Android[15], los componentes de Bluetooth y media de Android[17], Chrome[12], el componente CSS de Firefox[8], iOS y macOS[9], productos de Microsoft[10][5] y Ubuntu[7].

En los últimos años, varios proyectos de gran importancia en el ambiente Open Source han decidido incorporar Rust a fin de reducir el número de bugs relacionados al manejo de la memoria. Entre ellos podemos nombrar al Android Open Source Project[16] y al kernel Linux que desde su versión 6.1 introduce soporte para programar componentes en Rust[14][2]. Asimismo Meta aprueba y fomenta el uso de Rust como lenguaje para desarrollo *server-side* desde el 2022[6]. No podemos olvidar que Rust ha sido elegido durante 7 años consecutivos como el lenguaje de programación más querido por los programadores en la encuesta anual de Stack Overflow[11].

En el presente trabajo nos proponemos estudiar la detección de *deadlocks* y *lost signals* en el lenguaje de programación Rust. Se utilizará un modelo teórico basado en Redes de Petri para encontrar los errores en el código fuente. Mediante una traducción del código fuente en tiempo de compilación, se obtendrá una red de Petri que luego podrá ser analizada métodos de verificación de modelos para garantizar la ausencia de *deadlocks*.

El trabajo contará con una primera implementación del traductor que podrá ser utilizado como *plugin* del gestor de paquetes estándar de Rust *cargo*[18]. El objetivo es contribuir a la comunidad de Rust aportando una primera versión de esta herramienta que podría luego ser extendida para soportar casos más complejos. Se busca que el uso de la herramienta sea lo más sencillo y accesible posible para fomentar su uso y aplicación a proyectos reales de software. A largo plazo, se podría incorporar la herramienta al compilador como un pase adicional opcional en el proceso de compilación. Este pase verificaría que no se pueden producir *deadlocks*, lo que haría de Rust un lenguaje de programación aún más seguro y confiable.

2. Estado del arte / Literatura relacionada

Existen varias herramientas de verificación automática en Rust. Una primera aproximación recomendable es el resumen producido por Alastair Reid, investigador en Intel. En ella se lista explícitamente que la mayoría de las herramientas formales de verificación no soportan concurrencia[13].

El intérprete *Miri* desarrollado por el *Rust project* en GitHub es un intérprete experimental para la representación intermedia del lenguaje Rust (*mid-level intermediate representation*, conocida comúnmente por la sigla "MIR") que permite ejecutar binarios de proyectos de *cargo* de forma granularizada, instrucción a instrucción, para verificar la ausencia de *Undefined Behaviour* (UB) y otros errores en el manejo de la memoria. Detecta *memory leaks*, accesos no alineados a memoria, *data races* y violaciones de precondiciones o invariantes en código marcado como *unsafe*. [19]

3. Objetivos

El objetivo general del trabajo es ...

Los objetivos particulares son:

1. 1
2. 2
3. 3
 - a) 3.1
 - b) 3.2

4. Metodologías propuestas

5. Conjuntos de datos (dependiendo del área de la tesis)

6. Recursos informáticos (opcional)

6.1. Hardware

6.2. Software

6.2.1. Otras herramientas

7. Cronograma de trabajo

Cantidad de horas: [768-1000]

Tareas	Meses											
	1	2	3	4	5	6	7	8	9	10	11	12
Tarea A												
Tarea B												
Tarea C												
Tarea D												
Tarea E												

- Tarea A [100 horas]: (descripción)
- Tarea B [180 horas]:
- ...

Referencias

- [1] BEN-ARI, M. *Principles of Concurrent and Distributed Programming*, 2nd ed. Pearson Education, 2006.
- [2] CORBET, J. The 6.1 kernel is out. <https://lwn.net/Articles/917504/>. Accedido: 2022-02-24.
- [3] COULOURIS, G., DOLLIMORE, J., KINDBERG, T., AND BLAIR, G. *Distributed Systems, Concepts and Design*, 5th ed. Pearson Education, 2012.
- [4] DAVIDOFF, S. S. How rust's standard library was vulnerable for years and nobody noticed. <https://shnatsel.medium.com/how-rusts-standard-library-was-vulnerable-for-years-and-nobody-noticed-aebf0503c3d6>. Accedido: 2022-02-20.
- [5] FERNANDEZ, S. A proactive approach to more secure code. <https://msrc.microsoft.com/blog/2019/07/a-proactive-approach-to-more-secure-code/>. Accedido: 2022-02-24.
- [6] GARCIA, E. Programming languages endorsed for server-side use at meta. <https://engineering.fb.com/2022/07/27/developer-tools/programming-languages-endorsed-for-server-side-use-at-meta/>. Accedido: 2022-02-24.
- [7] GAYNOR, A. What science can tell us about c and c++'s security. <https://alexgaynor.net/2020/may/27/science-on-memory-unsafety-and-security/>. Accedido: 2022-02-24.
- [8] HOSFELT, D. Implications of rewriting a browser component in rust. <https://hacks.mozilla.org/2019/02/rewriting-a-browser-component-in-rust/>. Accedido: 2022-02-24.
- [9] KEHRER, P. Memory unsafety in apple's operating systems. <https://langui.sh/2019/07/23/apple-memory-safety/>. Accedido: 2022-02-24.
- [10] MILLER, M. Trends, challenges, and strategic shifts in the software vulnerability mitigation landscape. <https://www.youtube.com/watch?v=PjbGojjnBZQ>. Accedido: 2022-02-24.
- [11] OVERFLOW, S. 2022 developer survey. <https://survey.stackoverflow.co/2022/#section-most-loved-dreaded-and-wanted-programming-scripting-and-markup-languages>. Accedido: 2022-02-22.
- [12] PROJECTS, T. C. Memory safety. <https://www.chromium.org/Home/chromium-security/memory-safety/>. Accedido: 2022-02-24.

- [13] REID, A. Automatic rust verification tools (2021). <https://alastairreid.github.io/automatic-rust-verification-tools-2021/>. Accedido: 2022-02-20.
- [14] SIMONE, S. D. Linux 6.1 officially adds support for rust in the kernel. <https://www.infoq.com/news/2022/12/linux-6-1-rust/>. Accedido: 2022-02-24.
- [15] STEPANOV, E. Detecting memory corruption bugs with hwasan. <https://android-developers.googleblog.com/2020/02/detecting-memory-corruption-bugs-with-hwasan.html>. Accedido: 2022-02-24.
- [16] STOEP, J. V., AND HINES, S. Rust in the android platform. <https://security.googleblog.com/2021/04/rust-in-android-platform.html>. Accedido: 2022-02-22.
- [17] STOEP, J. V., AND ZHANG, C. Queue the hardening enhancements. <https://android-developers.googleblog.com/2020/02/detecting-memory-corruption-bugs-with-hwasan.html>. Accedido: 2022-02-24.
- [18] THE RUST PROJECT. The cargo book. <https://doc.rust-lang.org/stable/cargo/>. Accedido: 2022-02-20.
- [19] THE RUST PROJECT. Miri. <https://github.com/rust-lang/miri>. Accedido: 2022-02-20.
- [20] THE RUST PROJECT. Rust book: Extensible concurrency with the sync and send traits. <https://doc.rust-lang.org/book/ch16-04-extensible-concurrency-sync-and-send.html>. Accedido: 2022-02-20.
- [21] THE RUST PROJECT. Rust book: References and borrowing. <https://doc.rust-lang.org/book/ch04-02-references-and-borrowing.html>. Accedido: 2022-02-20.
- [22] THE RUST PROJECT. Rust book: What is ownership. <https://doc.rust-lang.org/book/ch04-01-what-is-ownership.html>. Accedido: 2022-02-20.
- [23] THE RUST PROJECT. Rust programming language. <https://www.rust-lang.org/>. Accedido: 2022-02-20.
- [24] THE RUST PROJECT. The rust rfc book. <https://rust-lang.github.io/rfcs/>. Accedido: 2022-02-20.
- [25] THE RUST PROJECT. The rustc book: Exploit mitigations. <https://doc.rust-lang.org/rustc/exploit-mitigations.html>. Accedido: 2022-02-20.
- [26] VAN STEEN, M., AND TANENBAUM, A. S. *Distributed Systems*, 3rd ed. Pearson Education, 2017.