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

Horacio Lisdero Scaffino

Facultad de Ingeniería
Universidad de Buenos Aires

19 de febrero del 2023

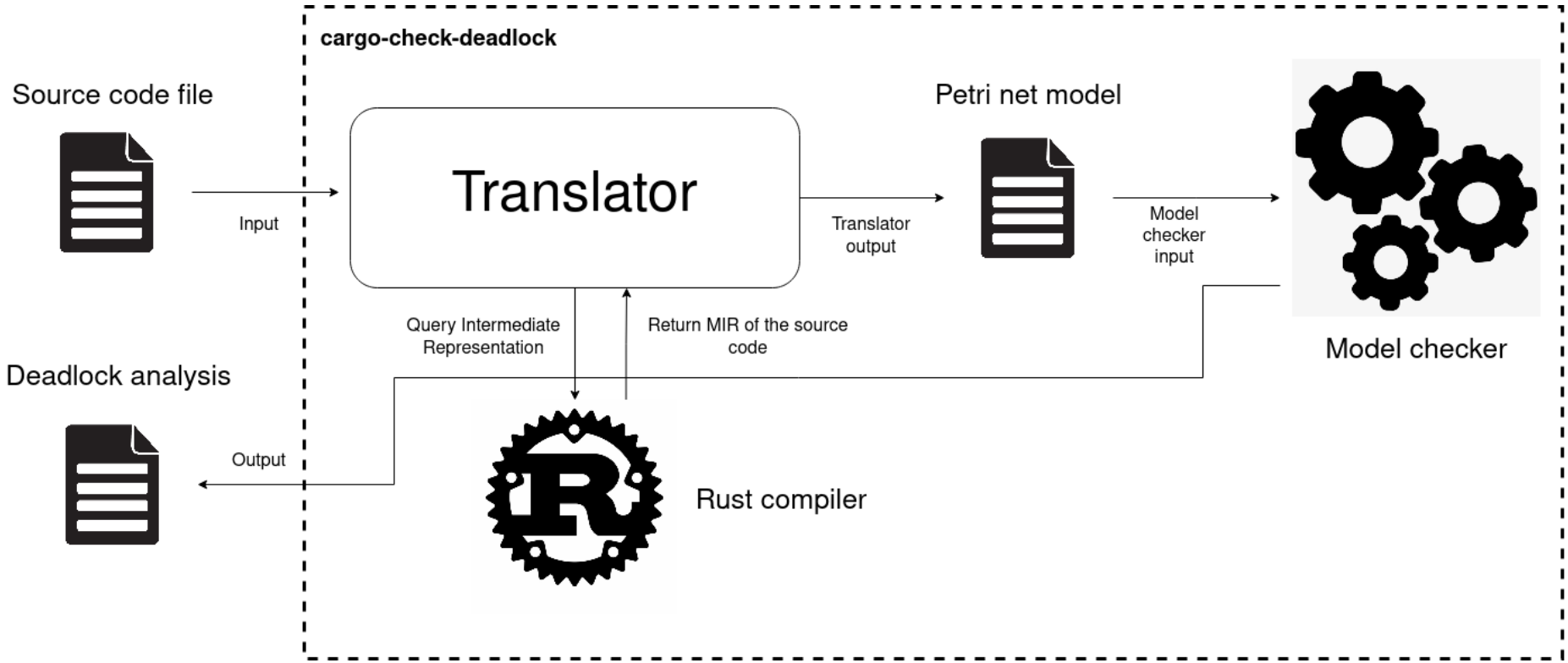
.UBAfiuba
FACULTAD DE INGENIERÍA

Agenda

Agenda

Una vista general de la herramienta

El traductor es el componente principal. El verificador de modelos y el compilador de Rust, *rustc*, son dependencias.



Agenda

Agenda

¿Qué es Rust?

Rust es un lenguaje de programación multiparadigma de uso general que busca ofrecer a los desarrolladores una forma segura y eficiente de escribir código de bajo nivel.

¿Qué es Rust?

Rust es un lenguaje de programación multiparadigma de uso general que busca ofrecer a los desarrolladores una forma segura y eficiente de escribir código de bajo nivel.

- Uso seguro de la memoria (*Memory safe*)
- Compilado a código máquina, con un *runtime* mínimo
- Expresividad de un lenguaje de alto nivel
- Performance de un lenguaje de bajo nivel (similar a C o C++)

Breve línea de tiempo de Rust

- 2007 Inicio como un proyecto personal de Graydon Hoare, un programador en Mozilla
- 2009 Mozilla comienza a patrocinar oficialmente el proyecto
- 2015 Primera versión estable 1.0
- 2016 Mozilla lanza Servo, un motor de navegador construido con Rust
- 2019 Se estabiliza el soporte de `async`/`await`
- 2021 AWS, Huawei, Google, Microsoft y Mozilla crean la Rust Foundation.
- 2021 El proyecto Android fomenta el uso de Rust para los componentes del SO por debajo del ART
- 2022 El kernel de Linux añade soporte para Rust junto con C
- 2023 8 años consecutivos el lenguaje de programación más querido en la Stack Overflow Developer Survey

Uso seguro de la memoria

Logra un uso seguro de la memoria sin recurrir a un recolector de basura o a un contador de referencias. En su lugar, utiliza el concepto de **ownership** (propiedad) y **borrowing** (préstamo).

Uso seguro de la memoria

Logra un uso seguro de la memoria sin recurrir a un recolector de basura o a un contador de referencias. En su lugar, utiliza el concepto de **ownership** (propiedad) y **borrowing** (préstamo).

Evita una gran variedad de clases de errores en tiempo de compilación:

- Double free
- Use after free
- Punteros colgantes (*Dangling pointers*)
- Condiciones de carrera (*Data races*)
- Pasaje de variables no seguras entre hilos

Si se encuentra una violación de las reglas del compilador, el programa simplemente no compilará.

Agenda

Memory safety is critical for reliability and security

Empirical investigations have concluded that around 70 % of the vulnerabilities found in large C/C++ codebases are due to memory handling errors. This high figure can be observed in projects such as:

- Android Open Source Project [1],
- the Bluetooth and media components of Android [2],
- the Chromium Projects behind the Chrome web browser [3],
- the CSS component of Firefox [4],
- iOS and macOS [5],
- Microsoft products [6, 7],
- Ubuntu [8]

Rust adoption is increasing fast

- The Android Open Source Project encourages the use of Rust for the SO components below the ART [9].
- The Linux kernel introduces in version 6.1 official tooling support for programming components in Rust [10, 11].
- At Mozilla, the Oxidation project was created in 2015 to increase the usage of Rust in Firefox and related projects. As of March 2023, the lines of code in Rust represent more than 10 % of the total in Firefox Nightly [12].
- At Meta, the use of Rust as a development language server-side is approved and encouraged since July 2022 [13].
- At Cloudflare, a new HTTP proxy in Rust was built from scratch to overcome the architectural limitations of NGINX, reducing CPU usage by 70 % and memory usage by 67 % [14].
- At Discord, reimplementing a crucial service in Rust provided great benefits in performance and solved a performance penalty due to the garbage collection in Go [15].
- At npm Inc., the company behind the npm registry, Rust allowed scaling CPU-bound services to more than 1.3 billion downloads per day [16].
- A study of Rust-based code found it runs so efficiently that it uses half as much electricity as a similar program written in Java, a language commonly used at AWS [17].

The tooling is great and works out of the box

- cargo, the official package manager: Format, build, test, lint, update, and publish packages.

The tooling is great and works out of the box

- cargo, the official package manager: Format, build, test, lint, update, and publish packages.
- Test harness for unit tests, integration tests, and tests in documentation comments (doctests). No third-party libraries needed.

The tooling is great and works out of the box

- cargo, the official package manager: Format, build, test, lint, update, and publish packages.
- Test harness for unit tests, integration tests, and tests in documentation comments (doctests). No third-party libraries needed.
- An official public registry for Rust packages (called “crates”): `https://crates.io/`.

The tooling is great and works out of the box

- cargo, the official package manager: Format, build, test, lint, update, and publish packages.
- Test harness for unit tests, integration tests, and tests in documentation comments (doctests). No third-party libraries needed.
- An official public registry for Rust packages (called “crates”): `https://crates.io/`.
- Automatic generation of a static website from the doc comments of the project. It is published automatically to `https://docs.rs/`.

The tooling is great and works out of the box

- cargo, the official package manager: Format, build, test, lint, update, and publish packages.
- Test harness for unit tests, integration tests, and tests in documentation comments (doctests). No third-party libraries needed.
- An official public registry for Rust packages (called “crates”): `https://crates.io/`.
- Automatic generation of a static website from the doc comments of the project. It is published automatically to `https://docs.rs/`.
- An official linter included with the default installation that catches even more errors and spots non-idiomatic code: clippy.

The tooling is great and works out of the box

- cargo, the official package manager: Format, build, test, lint, update, and publish packages.
- Test harness for unit tests, integration tests, and tests in documentation comments (doctests). No third-party libraries needed.
- An official public registry for Rust packages (called “crates”): `https://crates.io/`.
- Automatic generation of a static website from the doc comments of the project. It is published automatically to `https://docs.rs/`.
- An official linter included with the default installation that catches even more errors and spots non-idiomatic code: clippy.
- Integration with git, GitHub, VSCode, IntelliJ is great and easy to use.

The tooling is great and works out of the box

- cargo, the official package manager: Format, build, test, lint, update, and publish packages.
- Test harness for unit tests, integration tests, and tests in documentation comments (doctests). No third-party libraries needed.
- An official public registry for Rust packages (called “crates”): `https://crates.io/`.
- Automatic generation of a static website from the doc comments of the project. It is published automatically to `https://docs.rs/`.
- An official linter included with the default installation that catches even more errors and spots non-idiomatic code: clippy.
- Integration with git, GitHub, VSCode, IntelliJ is great and easy to use.
- A new stable compiler release every 6 weeks [18].

Agenda

Agenda

Definición informal

A Petri net is a mathematical modeling tool used to describe and analyze the behavior of concurrent systems. It provides a graphical representation of the system's state and its transitions, allowing for visual and formal analysis of complex processes.

- Places: Represent states in the system (*circles*)
- Transitions: Represent usually events or actions that occur in the system (*rectangles*)
- Tokens: Marks inside of places that are created and consumed by transitions (*points inside of places*)

Mathematical definition

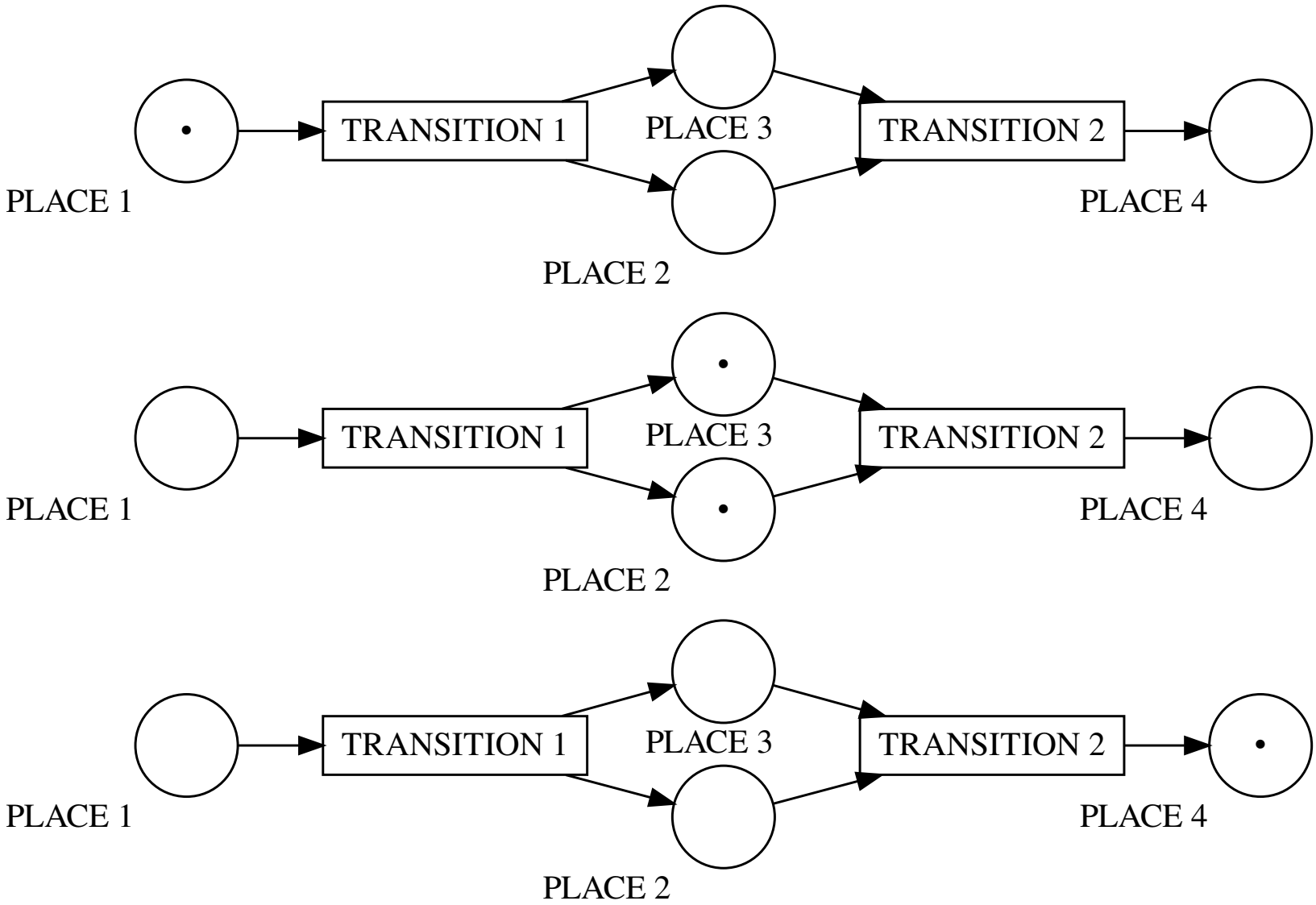
A Petri net is a 5-tuple, $PN = (P, T, F, W, M_0)$ where:

$P = \{p_1, p_2, \ldots, p_m\}$ *is a finite set of places,*
$T = \{t_1, t_2, \ldots, t_n\}$ *is a finite set of transitions,*
$F \subseteq (P \times T) \cup (T \times P)$ *is a set of arcs (flow relation),*
$W : F \to \{1, 2, 3, ...\}$ *is a weight function for the arcs,*
$M_0 : P \to \{0, 1, 2, 3, ...\}$ *is the initial marking,*
$P \cap T = \varnothing$ *and* $P \cup T \neq \varnothing$

The graph is by definition *bipartite*. There can only be edges:

- from places to transitions or
- from transitions to places

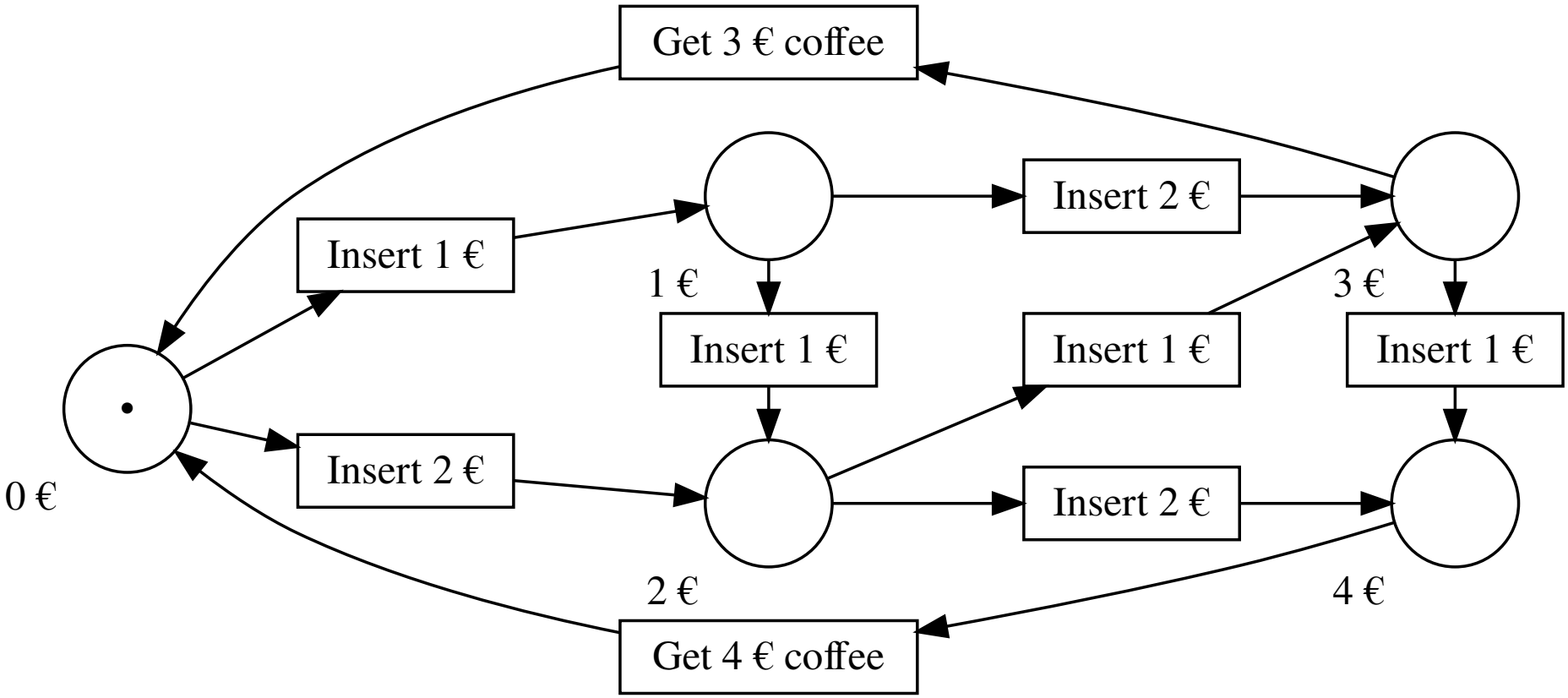
Transition firing rule



Agenda

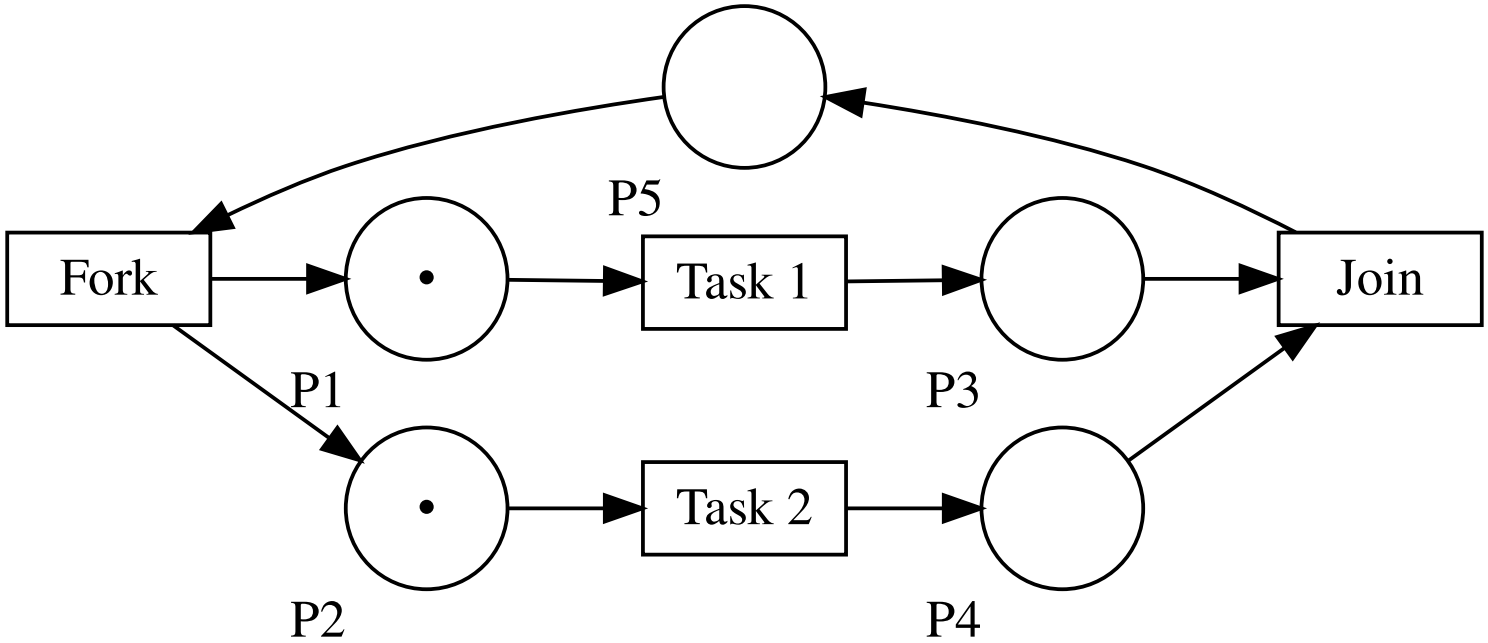
Vending machine

This is a finite-state machine (FSM), a subclass of Petri nets.



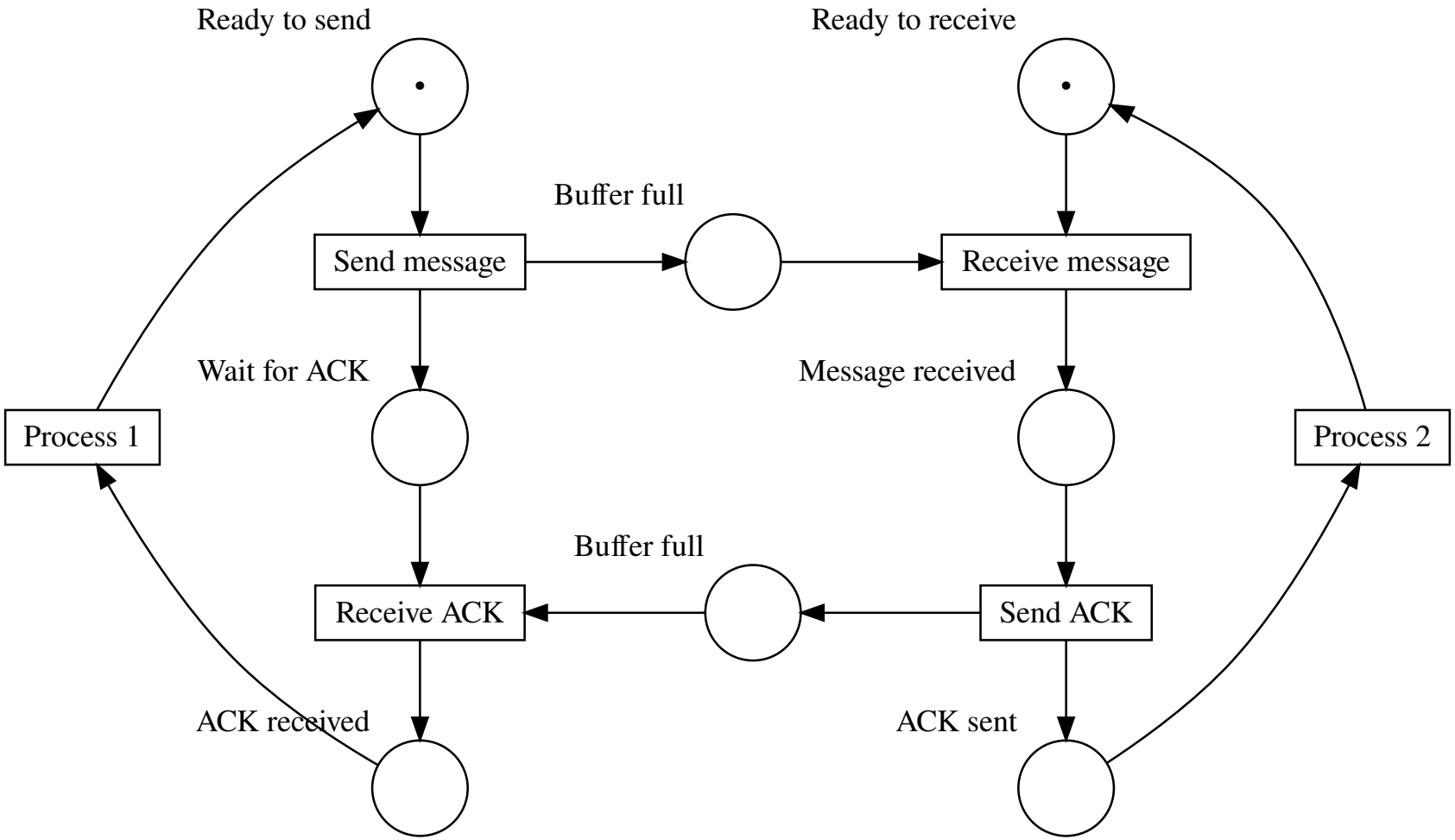
Parallel activities: Fork/Join

This is a marked graph (MG), a subclass of Petri nets. Observe the concurrency between Task 1 and Task 2. This cannot be modeled by a single finite-state machine.



Communication protocols: Send with ACK

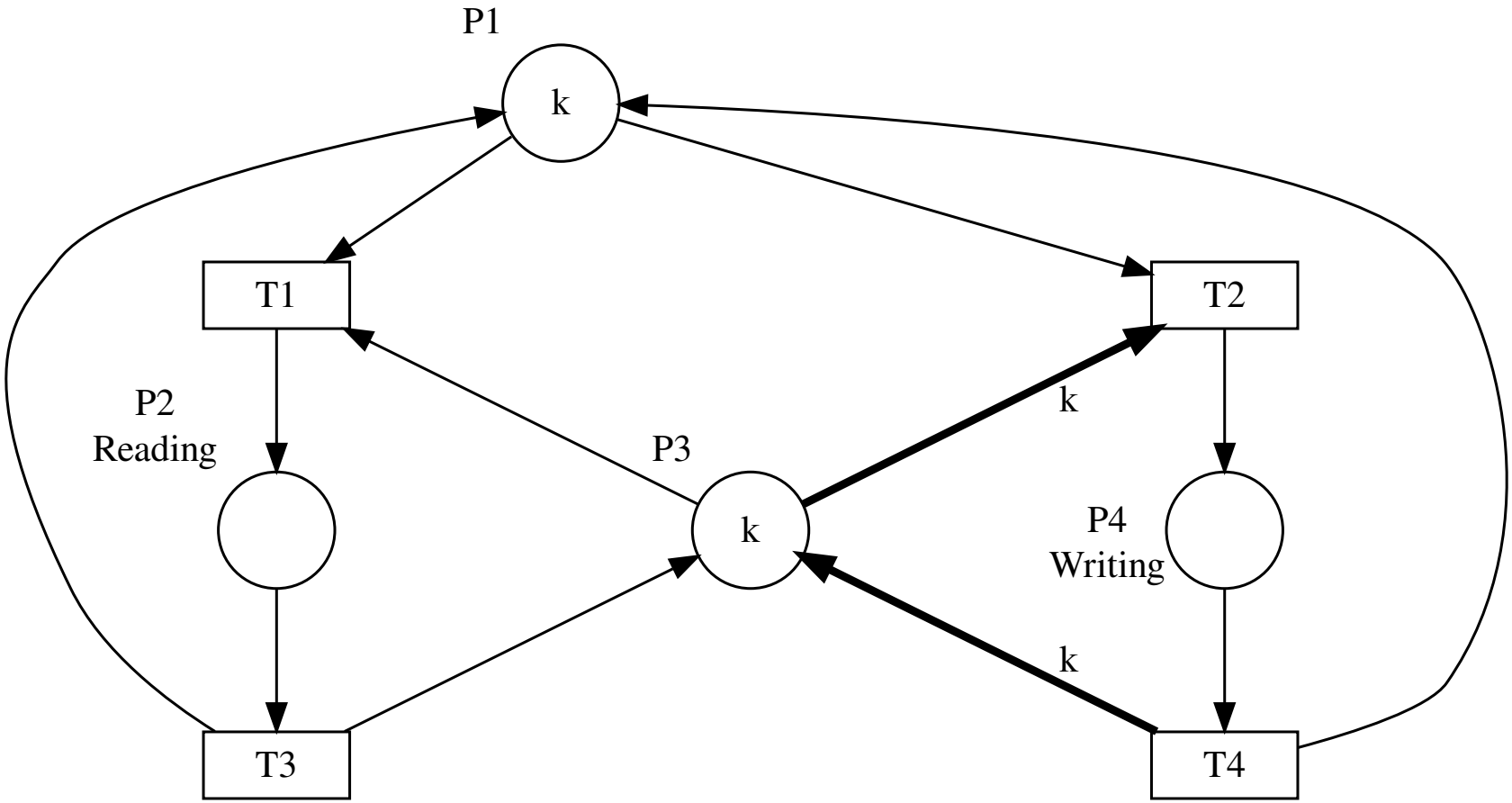
A simple protocol in which Process 1 sends messages to Process 2 and waits for an acknowledgment to be received before continuing. For simplicity, no timeout mechanism was included.



Synchronization control: Readers and writers

A Petri net system with k processes that either read or write a shared value.

- If one process writes, then no process may read.
- If a process is reading, then no process may write.
- There can only be zero or one process writing at any given time.



Agenda

Análisis de alcance

Petri nets can be analyzed using formal methods to conclude whether the net can reach a deadlock or not. There is a notion of liveness analogous to the one found in computer systems.

Análisis de alcance

Petri nets can be analyzed using formal methods to conclude whether the net can reach a deadlock or not. There is a notion of liveness analogous to the one found in computer systems.

Several model checkers are being developed and there is even a Model Checking Contest that takes place every year. State-of-the-art tools can handle Petri net models with more than **70 000 transitions** and **one million places**.

Análisis de alcance

Petri nets can be analyzed using formal methods to conclude whether the net can reach a deadlock or not. There is a notion of liveness analogous to the one found in computer systems.

Several model checkers are being developed and there is even a Model Checking Contest that takes place every year. State-of-the-art tools can handle Petri net models with more than **70 000 transitions** and **one million places**.

Translating source code to a Petri net has been done before for other programming languages [19, 20] and also for Rust [21, 22]. The difficulty lies in supporting more synchronization primitives than simple mutexes and translating code from real-world applications.

Agenda

Compilation stages in *rustc*

- Lexing: The source text is turned into a stream of atomic source code units known as tokens.

Compilation stages in *rustc*

- Lexing: The source text is turned into a stream of atomic source code units known as tokens.
- Parsing: The stream of tokens is converted into an **Abstract Syntax Tree**.

Compilation stages in *rustc*

- Lexing: The source text is turned into a stream of atomic source code units known as tokens.
- Parsing: The stream of tokens is converted into an **Abstract Syntax Tree**.
- High-level Intermediate Representation (HIR):
  - Desugar loops: `while` and `for` to simple `loop`.
  - Type inference: The automatic detection of a type of an expression.
  - Trait solving: Ensuring that each implementation block (`impl`) points to a valid trait.
  - Type checking.

Compilation stages in *rustc*

- Lexing: The source text is turned into a stream of atomic source code units known as tokens.
- Parsing: The stream of tokens is converted into an **Abstract Syntax Tree**.
- High-level Intermediate Representation (HIR):
  - Desugar loops: `while` and `for` to simple `loop`.
  - Type inference: The automatic detection of a type of an expression.
  - Trait solving: Ensuring that each implementation block (`impl`) points to a valid trait.
  - Type checking.
- Mid-level Intermediate Representation (MIR):
  - Checking of exhaustiveness of pattern matching.
  - Desugar method calls to function calls (`x.method(y)` becomes `Type::method(&x, y)`).
  - Add implicit dereferencing operations.
  - Borrow checking.

Compilation stages in *rustc*

- Lexing: The source text is turned into a stream of atomic source code units known as tokens.
- Parsing: The stream of tokens is converted into an **Abstract Syntax Tree**.
- High-level Intermediate Representation (HIR):
  - Desugar loops: `while` and `for` to simple `loop`.
  - Type inference: The automatic detection of a type of an expression.
  - Trait solving: Ensuring that each implementation block (`impl`) points to a valid trait.
  - Type checking.
- Mid-level Intermediate Representation (MIR):
  - Checking of exhaustiveness of pattern matching.
  - Desugar method calls to function calls (`x.method(y)` becomes `Type::method(&x, y)`).
  - Add implicit dereferencing operations.
  - Borrow checking.
- Code generation:
  - *rustc* relies on LLVM as a backend.
  - It leverages many optimizations of the LLVM intermediate representation.
  - LLVM takes over from this point on.
  - At the end, object files are linked to create an executable.

Agenda

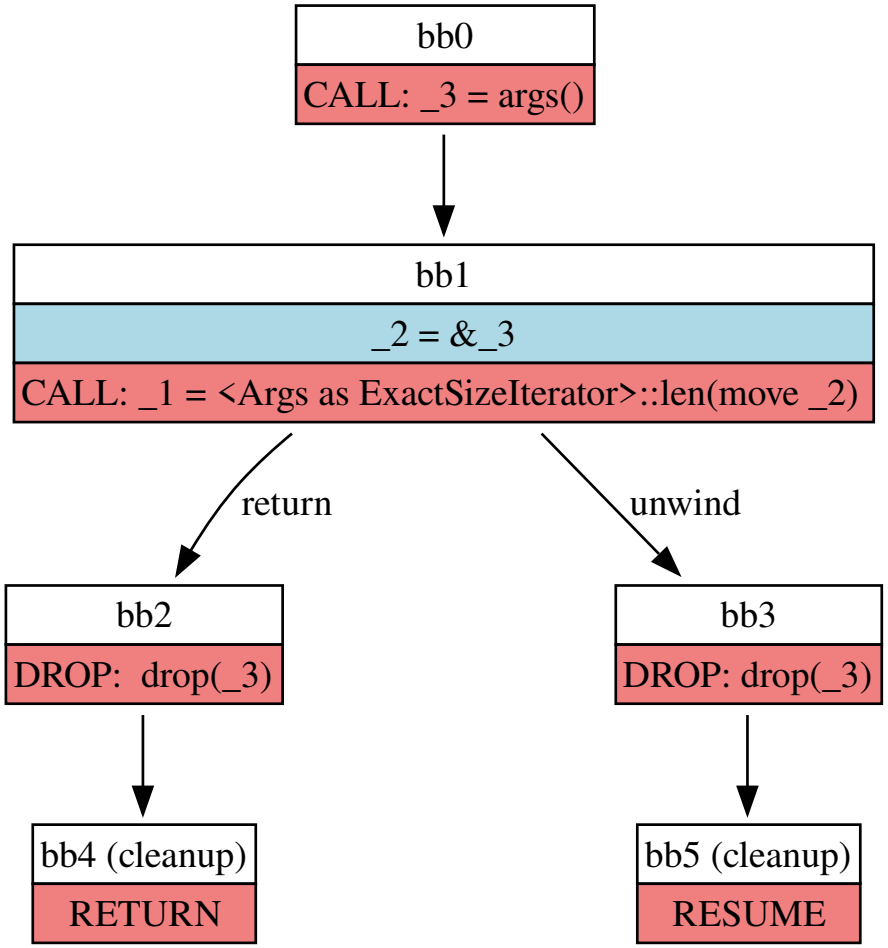
Hello World in MIR

BB means “basic block”. Each one is formed by statements and one terminator statement. The terminator statement is the only place where the control flow can jump to another basic block.

```
1   fn main() -> () {
2       let mut _0: ();
3       let _1: ();
4       let mut _2: std::fmt::Arguments<'_>;
5       let mut _3: &[&str];
6       let mut _4: &[&str; 1];
7
8       bb0: {
9           _4 = const _;
10          _3 = _4 as &[&str] (Pointer(Unsize));
11          _2 = Arguments::<'_>::new_const(move _3) -> bb1;
12      }
13
14      bb1: {
15          _1 = _print(move _2) -> bb2;
16      }
17
18      bb2: {
19          return;
20      }
21  }
```

MIR como un grafo que muestra el flujo de ejecución

The MIR is a form of control flow graph (CFG) used in compilers. In this form, the translation to a Petri net becomes evident.



Agenda

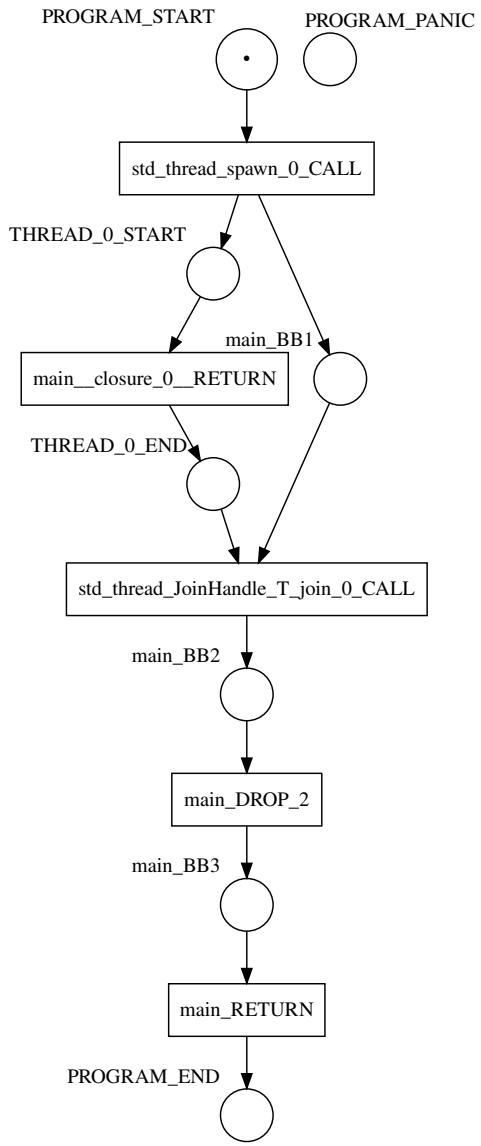
Programa de ejemplo

Let's consider a trivial program that spawns a thread that does nothing and immediately joins it.

```
1  fn main() {
2      let thread_join_handle = std::thread::spawn(move || {
3          // some work here
4      });
5      // some work here
6      let _res = thread_join_handle.join();
7  }
```

- `std::thread::spawn` should create an additional token that models the program counter of the second thread.
- The joining thread should wait until the spawned thread finishes.

Modelo de red de Petri de un thread



Agenda

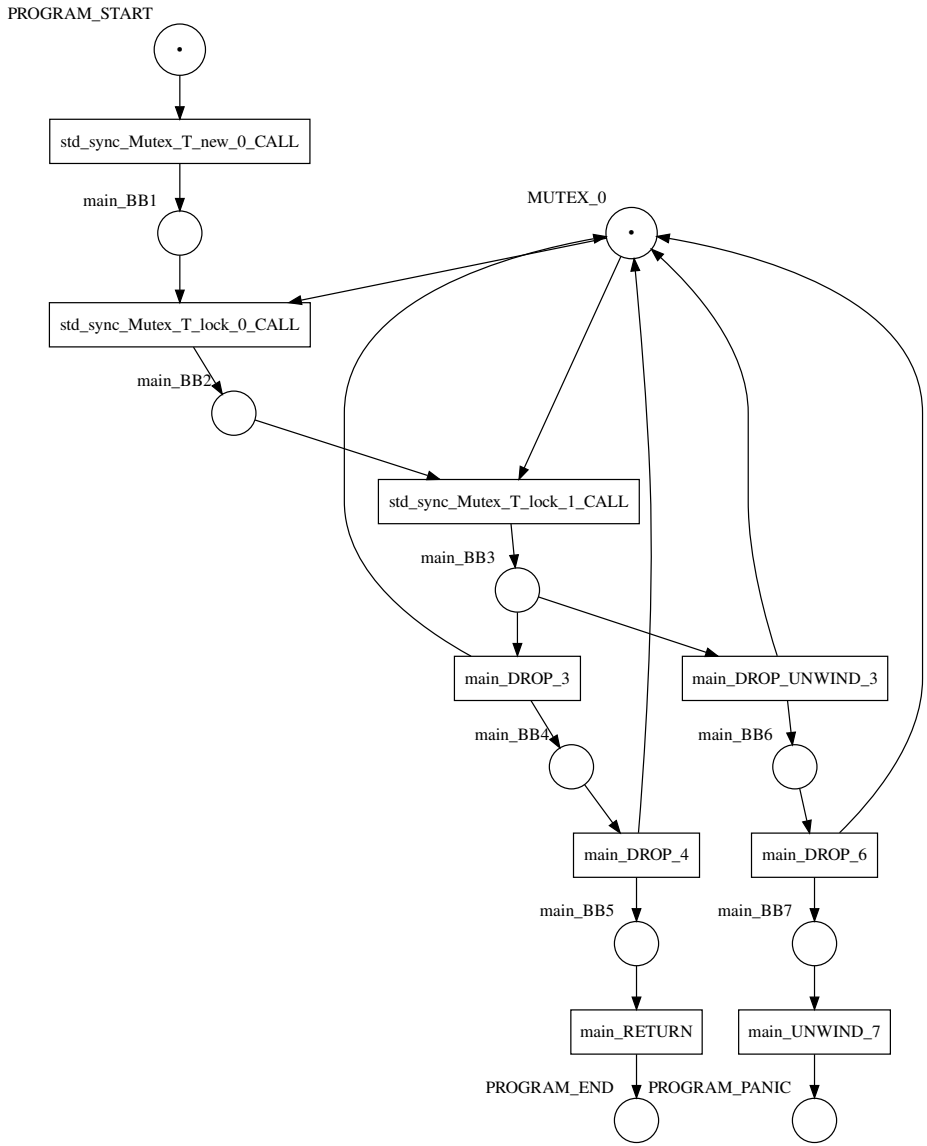
Programa de ejemplo

Consider a simple program that locks a mutex twice. The second lock operation will deadlock because the lock handle returned by the first call to `std::sync::Mutex::lock` is not dropped until it falls out of scope.

```
1  fn main() {
2      let data = std::sync::Mutex::new(0);
3      let _d1 = data.lock();
4      let _d2 = data.lock(); // cannot lock, since d1 is still active
5  }
```

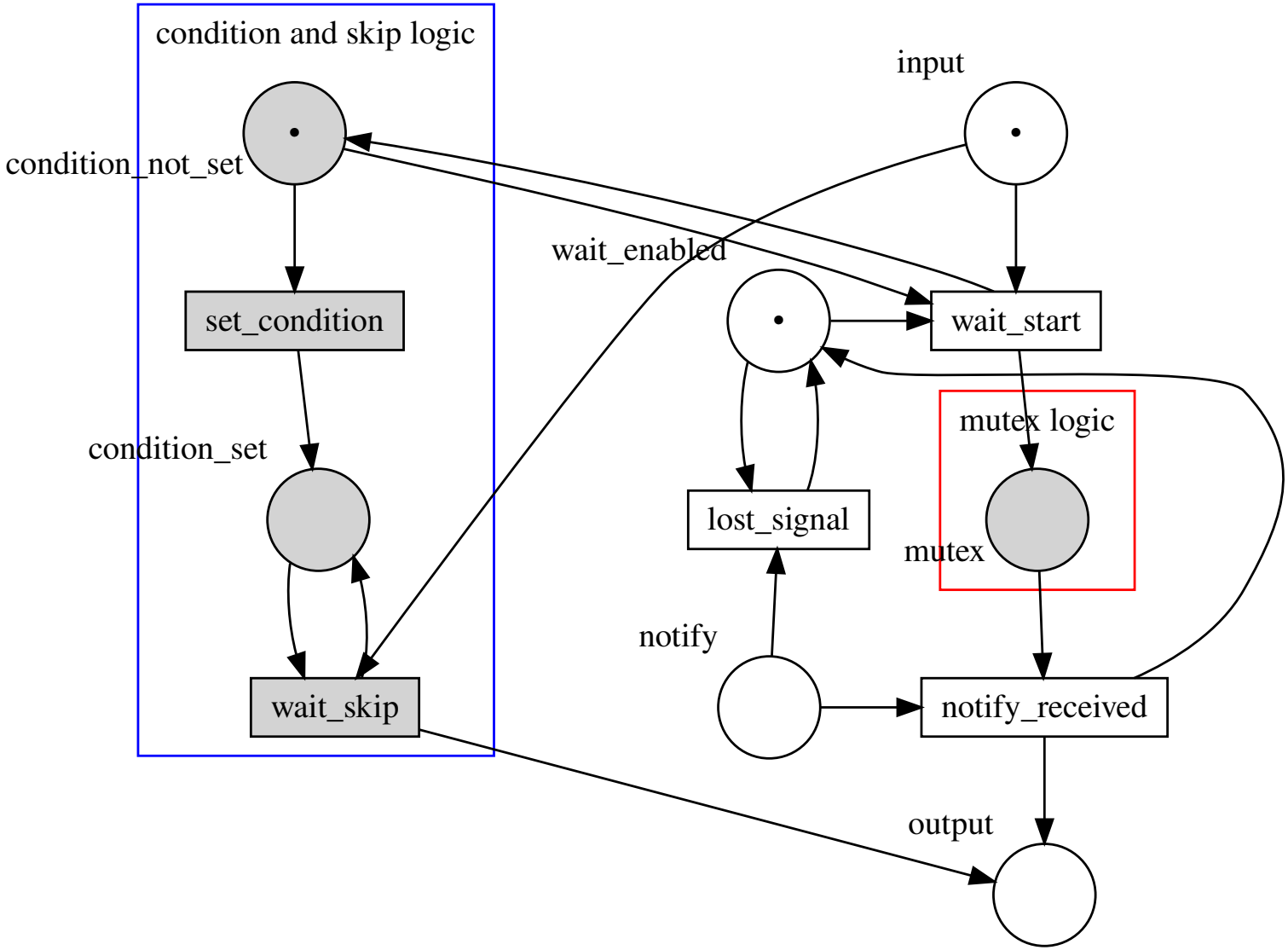
- There should be a single place that models the mutex.
- Locking the mutex is taking the token from the mutex place.
- Unlocking the mutex is setting the token back in the mutex place.

Modelo de red de Petri de un mutex



Agenda

Cómo modelar una *condition variable*



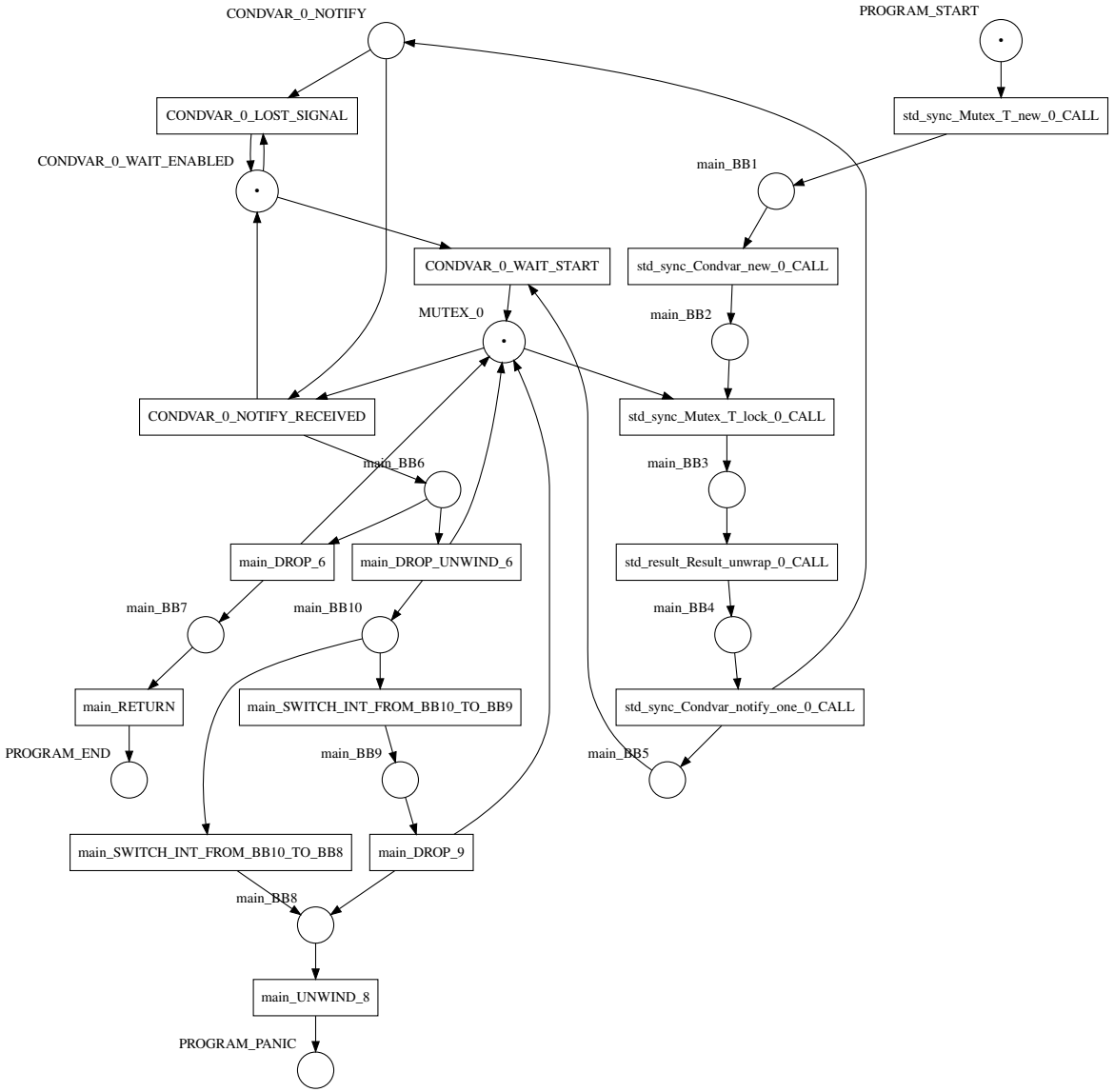
Programa de ejemplo

We have to use a very simple example program to keep the net small. In this case, the thread is trying to notify itself, which leads to a lost signal.

```
1  fn main() {
2      let mutex = std::sync::Mutex::new(false);
3      let cvar = std::sync::Condvar::new();
4      let mutex_guard = mutex.lock().unwrap();
5      cvar.notify_one();
6      let _result = cvar.wait(mutex_guard);
7  }
```

- The model for the condition variable should appear in the Petri net.
- The notify place should be set.
- But the signal gets consumed because `std::sync::Condvar::wait` was not called.

Modelo de red de Petri para el programa de ejemplo



Agenda

Bibliografía I


E. Stepanov, “Detecting Memory Corruption Bugs With HWASan.” https://android-developers.googleblog.com/2020/02/detecting-memory-corruption-bugs-with-hwasan.html, 2020.
Accessed on 2023-02-24.


J. V. Stoep and C. Zhang, “Queue the Hardening Enhancements.” https://android-developers.googleblog.com/2020/02/detecting-memory-corruption-bugs-with-hwasan.html, 2020.
Accessed on 2023-02-24.


The Chromium Projects, “Memory safety.”
https://www.chromium.org/Home/chromium-security/memory-safety/, 2015.
Accessed on 2023-02-24.


D. Hosfelt, “Implications of Rewriting a Browser Component in Rust.”
https://hacks.mozilla.org/2019/02/rewriting-a-browser-component-in-rust/, 2019.
Accessed on 2023-02-24.


P. Kehrer, “Memory Unsafety in Apple’s Operating Systems.”
https://langui.sh/2019/07/23/apple-memory-safety/, 2019.
Accessed on 2023-02-24.


M. Miller, “Trends, Challenges, and Strategic Shifts in the Software Vulnerability Mitigation Landscape.”
https://www.youtube.com/watch?v=PjbGojjnBZQ, 2019.
Accessed on 2023-02-24.


S. Fernandez, “A proactive approach to more secure code.”
https://msrc.microsoft.com/blog/2019/07/a-proactive-approach-to-more-secure-code/, 2019.
Accessed on 2023-02-24.

Bibliografía II


A. Gaynor, "What science can tell us about C and C++'s security." https://alexgaynor.net/2020/may/27/science-on-memory-unsafety-and-security/, 2020.
Accessed on 2023-02-24.


J. V. Stoep and S. Hines, "Rust in the Android platform." https://security.googleblog.com/2021/04/rust-in-android-platform.html, 2021.
Accessed on 2023-02-22.


J. Corbet, "The 6.1 kernel is out." https://lwn.net/Articles/917504/, 2022.
Accessed on 2023-02-24.


S. D. Simone, "Linux 6.1 Officially Adds Support for Rust in the Kernel." https://www.infoq.com/news/2022/12/linux-6-1-rust/, 2022.
Accessed on 2023-02-24.


Mozilla Wiki, "Oxidation Project." https://wiki.mozilla.org/Oxidation, 2015.
Accessed on 2023-03-20.


E. Garcia, "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/, 2022.
Accessed on 2023-02-24.


Y. Wu and A. Hauck, "How we built Pingora, the proxy that connects Cloudflare to the Internet." https://blog.cloudflare.com/how-we-built-pingora-the-proxy-that-connects-cloudflare-to-the-internet/, 2022.
Accessed on 2023-03-20.

Bibliografía III


J. Howarth, “Why Discord is switching from Go to Rust.”
https://discord.com/blog/why-discord-is-switching-from-go-to-rust, 2020.
Accessed on 2023-03-20.


The Rust Project Developers, “Rust case study: Community makes rust an easy choice for npm,” 2019.
https://www.rust-lang.org/static/pdfs/Rust-npm-Whitepaper.pdf.


R. Pereira, M. Couto, F. Ribeiro, R. Rua, J. Cunha, J. P. Fernandes, and J. Saraiva, “Energy efficiency across programming languages: how do energy, time, and memory relate?,” in *Proceedings of the 10th ACM SIGPLAN international conference on software language engineering*, pp. 256–267, 2017.


P. Albini, “RustFest Barcelona - Shipping a stable compiler every six weeks.”
https://www.youtube.com/watch?v=As1gXp5kX1M, 2019.
Accessed on 2023-02-24.


K. M. Kavi, A. Moshtaghi, and D.-J. Chen, “Modeling multithreaded applications using petri nets,” *International Journal of Parallel Programming*, vol. 30, pp. 353–371, 2002.


A. Moshtaghi, “Modeling Multithreaded Applications Using Petri Nets,” Master’s thesis, The University of Alabama in Huntsville, 2001.


T. Meyer, “A Petri Net semantics for Rust,” Master’s thesis, Universität Rostock — Fakultät für Informatik und Elektrotechnik, 2020.
https://github.com/Skasselbard/Granite/blob/master/doc/MasterThesis/main.pdf.


K. Zhang and G. Liua, “Automatically transform rust source to petri nets for checking deadlocks,” *arXiv preprint arXiv:2212.02754*, 2022.

Agenda

Recursos para aprender Rust

- The Rust Book: Available online and locally with the default Rust installation.
- Rust by Example: Another official book with a more practical approach.
- Rustlings: Small exercises to get you used to reading and writing Rust code!
- Comprehensive Rust: A three-day Rust course developed by the Android team.
- Take your first steps with Rust: A simple course on Microsoft Learn.
- Rust Programming Course for Beginners by freeCodeCamp.org.
- No Boilerplate: A Youtube channel mainly dedicated to topics connected with Rust. Some ideas were used for this presentation.

Simuladores en línea de redes de Petri

- A simple simulator by Igor Kim can be found on https://petri.hp102.ru/. A tutorial video on Youtube and example nets are included in the tool.
- A complement to this is a series of interactive tutorials by Prof. Wil van der Aalst at the University of Hamburg. These tutorials are Adobe Flash Player files (with extension `.swf`) that modern web browsers cannot execute. Luckily, an online Flash emulator like the one found on https://flashplayer.fullstacks.net/?kind=Flash_Emulator can be used to upload the files and execute them.
- Another online Petri net editor and simulator is http://www.biregal.com/. The user can draw the net, add the tokens, and then manually fire transitions.

¿Preguntas?

Links

Tesis	https://github.com/hlisdero/thesis
Herramienta	https://github.com/hlisdero/cargo-check-deadlock
Presentación	https://github.com/hlisdero/thesis/tree/main/presentation_es
Crate publicado	https://crates.io/crates/cargo-check-deadlock