



University of Stuttgart
Institute for Control Engineering
of Machine Tools and Manufacturing Units
(ISW)



Master's Thesis

Wait-free synchronization for inter-process communication in real-time systems

submitted by

Devrim Baran Demir

from Hamburg

Degree program

Examined by

Supervised by

Submitted on

M. Sc. Informatik

Prof. Andreas Wortmann

Marc Fischer

March 28, 2025

Declaration of Originality

Master's Thesis of Devrim Baran Demir (M. Sc. Informatik)

Address	Kernerweg 22, 89520 Heidenheim an der Brenz
Student number	3310700
English title	<i>Wait-free synchronization for inter-process communication in real-time systems</i>
German title	<i>Wait-free Synchronization für Interprozesskommunikation in Echtzeitsystemen</i>

I now declare,

- that I wrote this work independently,
- that no sources other than those stated are used and that all statements taken from other works—directly or figuratively—are marked as such,
- that the work submitted was not the subject of any other examination procedure, either in its entirety or in substantial parts,
- that I have not published the work in whole or in part, and
- that my work does not violate any rights of third parties and that I exempt the University against any claims of third parties.

Stuttgart, March 28, 2025

Kurzfassung

Vorhersehbare und korrekte Interprozesskommunikation (IPC) ist für Echtzeitsysteme von entscheidender Bedeutung, da Verzögerungen, Unvorhersehbarkeit oder inkonsistente Datenstände zu Instabilität und Ausfällen führen können. Traditionelle Synchronisationsmechanismen verursachen Blockierungen, die zu Prioritätsinversionen und erhöhten Antwortzeiten führen. Um diese Herausforderungen zu bewältigen, bietet die wartefreie Synchronisation eine Alternative, die den Abschluss von, wie der Austausch von Daten zwischen mehreren Prozessen, in einer begrenzten Anzahl von Schritten garantiert und so die Systemreaktionsfähigkeit und -zuverlässigkeit sicherstellt.

Diese Arbeit untersucht die Nutzung von wait-free Datenstrukturen für IPC in Echtzeitsystemen mit Fokus auf deren Implementierung in Rust. Das Eigentumsmodell und die strengen Nebenläufigkeitsgarantien von Rust machen es besonders geeignet für die Entwicklung latenzarmer und hochzuverlässiger Synchronisationsmechanismen. Diese Arbeit analysiert bestehende wait-free Methoden für IPC in Echtzeitsystemen und bewertet ihre Leistung im Vergleich zu herkömmlichen Synchronisationsmethoden.

Stichwörter: Echtzeitsysteme, wait-free Synchronisation, lock-free Synchronisation, Interprozesskommunikation, rust

Abstract

Predictable and correct Inter-Process Communication (IPC) is essential for Real-Time System (RTS), where delays, unpredictability or inconsistent data can lead to instability and failures. Traditional synchronization mechanisms introduce blocking, leading to priority inversion and increased response times. To overcome these challenges, wait-free synchronization provides an alternative that guarantees operation completion, such as the completion of data exchange between multiple processes, within a bounded number of steps, ensuring system responsiveness and reliability.

This thesis explores the use of wait-free data structures for IPC in RTS, focusing on their implementation in Rust. Rust's ownership model and strict concurrency guarantees make it well-suited for developing low-latency and high-reliability synchronization mechanisms. This work examines existing wait-free techniques for real-time IPC, and evaluates their performance against conventional synchronization methods.

Keywords: real-time systems, wait-free synchronization, lock-free synchronization, inter-process communication, rust

Contents

Kurzfassung	ii
Abstract	iii
1. Introduction	1
1.1. Motivation	1
1.2. Objective	2
1.3. Structure of the Thesis	3
2. Background	4
2.1. Real-Time Systems	4
2.2. Inter-Process Communication	5
2.2.1. Shared Memory	5
2.3. Synchronization	5
2.3.1. Mutual Exclusion	6
2.4. Wait-Free Synchronization	8
2.5. Lock-Free Synchronization	8
2.6. Rust Programming Language	8
2.7. State of the Art	8
3. Related Work	9
4. Methodology	10
5. Choosing optimal wait-free data structure	11
6. Results	12
7. Implementation	13
8. Conclusion and Future Work	14
Bibliography	15
List of Acronyms	18
List of Figures	19

Contents

List of Tables	20
A. Appendix	21

1. Introduction

1.1. Motivation

In modern manufacturing and automation, control systems must operate under strict timing constraints to function reliably. If a system fails to meet these constraints, unexpected delays can disrupt processes, leading to instability or even hazardous failures in safety-critical environments. For this reason, Real-Time Operating System (RTOS) and low-level programming languages like C, C++, and Rust are widely used to ensure predictable execution times.

Many real-time applications involve multiple tasks that must run concurrently and share resources efficiently. Without proper synchronization, problems such as data corruption or race conditions can occur leading to unpredictable behavior. Traditional synchronization methods are commonly used to manage access to shared resources and rely on blocking processes to share data through shared memory. However, these blocking mechanisms introduce significant challenges in real-time settings. Since traditional synchronization methods require processes to wait for resource availability, they can lead to increased response times, potential deadlocks, potential process starvations, and potential priority inversions. These delays are unacceptable in systems that require strict timing guarantees. [1]–[3]

Consider, for instance, a factory setting where a worker presses an emergency stop button to halt machinery. The system must react instantly, as any delay could lead to severe consequences, including equipment damage or injury. However, if the system experiences one of the discussed problems, it may become unresponsive at a critical moment. This highlights a fundamental issue with blocking synchronization techniques: while they ensure orderly access to shared resources, they can also introduce delays that are unacceptable in real-time environments.

To overcome these limitations, there is an increasing interest in wait-free and lock-free synchronization techniques. A wait-free algorithm guarantees that every operation completes in a finite number of steps, regardless of contention from other processes. This property ensures system responsiveness and predictability, which are essential for real-time applications. By eliminating blocking and contention-based delays, wait-free synchronization prevents priority inversion and ensures that high-priority tasks execute without interference. [1], [2], [4]

Lock-free synchronization, on the other hand, ensures that at least one process makes progress in a finite number of steps, but it does not guarantee that every process will complete its operation. While lock-free algorithms tend to be more efficient than wait-free

1. Introduction

ones, they allow the possibility of process starvation in high-contention scenarios. This can be problematic in real-time and heterogeneous computing environments, where some processes might be repeatedly delayed due to interference from faster processes. So we see that lock-free synchronization has a slightly different goal than wait-free synchronization, which is why in this work we will only focus on wait-free datastructures to stay in the scope of this work. [4]

Efficient synchronization mechanisms are particularly important in the context of IPC, which plays a crucial role in RTS. IPC allows processes to exchange data efficiently, but its performance is heavily influenced by the synchronization techniques used. Traditional IPC mechanisms, which often rely on blocking some processes, can introduce significant latency and reduce throughput. Wait-free data structures offer a promising alternative by ensuring that communication operations complete within predictable time bounds. However, selecting appropriate wait-free data structures and evaluating their performance in real-time environments remains a challenge. [5]–[8]

The Rust programming language provides useful features for implementing real-time synchronization mechanisms. Its ownership model and strict type system prevent data races and enforce safe concurrency without requiring traditional locking mechanisms. Additionally, Rust offers fine-grained control over system resources, making it a strong candidate for real-time applications that demand both low latency and high reliability. [9], [10]

The concepts and methods introduced here, including RTS, IPC, synchronization techniques and problems, wait-free synchronization, and the rust programming language are explored in greater depth in chapter 2.

1.2. Objective

The primary goal of this research is to find a wait-free data structure that can be used to implement a wait-free synchronization for IPC through shared memory in RTS using Rust. To do so, this study aims to:

- Identify and analyze existing wait-free synchronization techniques for IPC through shared memory for RTS.
- Implement and compare the performance of existing wait-free synchronization mechanisms for IPC through a shared memory for real-time scenarios with each other.
- Choose and analyze which wait-free data structure for IPC through shared memory in a real-time setting using Rust that is best suited.

By addressing these objectives, this work contributes to the field of wait-free synchronization for IPC in RTS by providing a practical solution with rust. The insights

1. Introduction

gained from this research can help improve the reliability and performance of real-time applications across various domains.

1.3. Structure of the Thesis

2. Background

To establish a clear foundation for the concepts and definitions introduced throughout this thesis, we provide a fundamental overview of the key topics relevant to this research. This includes an introduction to RTS, Inter-Process Communication (IPC), and synchronization techniques, with a particular focus on wait-free synchronization. Additionally, we examine the Rust programming language, as it serves as the primary development environment for this study. Furthermore, we explore existing synchronization methods in RTS to contextualize the motivation and contributions of this work.

2.1. Real-Time Systems

In RTS the correctness of the system does not only depend on the logical results of computations, but also on timing constraints. These systems can be classified into Hard Real-Time System (HRTS) or Soft Real-Time System (SRTS). HRTS have strict timing constraints, and missing a constraint is considered a system failure and may lead to a catastrophic disaster. The system must guarantee that every timing constraint has to be met. An use case would be industrial automation where all the machines and robotic modules have to communicate with each other as quick as possible to ensure no blockage of the manufacturing line. [11]

On the other hand, SRTS try to stick to the timing constraints as much as possible, but missing some timing constraints is not considered a system failure. Infrastructure wise SRTS are similar to HRTS, since it is still considered important to meet these timing constraints. An example would be a multimedia system where it would be considered fine if sometimes frames are dropped to guarantee the video stream. [11]

Sometimes these two systems appear in combination, where some functions have hard real-time constraints and some have soft real-time constraints. Krishna K. gives a good example in his paper where he describes that for the apollo 11 mission some components for the landing processes had soft real-time behavior and the rest still functioned with hard real-time constraints. [11]

Since the workfield of this thesis is within HRTS, the term RTS will be used synonymously with the terminology HRTS.

2.2. Inter-Process Communication

Now the processes used in a RTS also have to share information with each other so the system can function. So some kind of IPC is needed. IPC allows processes to share information with each other using different kind of methods. We will mainly focus on one method explained later. In general IPC is needed in all computing systems, because processes often need to work together (e.g. a producer process passes data to a consumer process). Lets take the brake-by-wire technology as example. Brake-by-wire is a technology for driverless cars where some mechanical and hydraulic components from the braking systems are replaced by wires to transmit braking signals, since there is no driver anymore to press on the braking pedal [12]. This of course requires different processes to share information together. In the context of this thesis this kind of communication requires strict timing constraints as stated as before, since any kind of delay or blockage would lead to fatal consequences. [13], [14]

To achieve IPC different kind of mechanisms are used. The focus for this work lies on shared memory, so that is also the mechanism we will look into.

2.2.1. Shared Memory

To achieve any kind of information sharing between processes, these processes will need to have access to the same data regularly. With a shared memory segment, multiple processes can have access to the same memory location. So all processes which are part of the IPC can read and write to this common memory space avoiding unnecessary data copys. With that processes exchange information by directly manipulating memory. This kind of IPC is particular useful for real-time applications, which handle large volumes of data or are required to quickly transfer data between sensors and control tasks. What is also important to know is that the section of the code, that programs these data accesses by different processes is called critical section. [14]–[17]

The problem with this is that the system somehow has to manage how the processes access the shared memory segment. This is mostly done by using different kind of synchronization techniques. Without any synchronization mechanism race conditions or inconsistent data can occur. [14], [15]

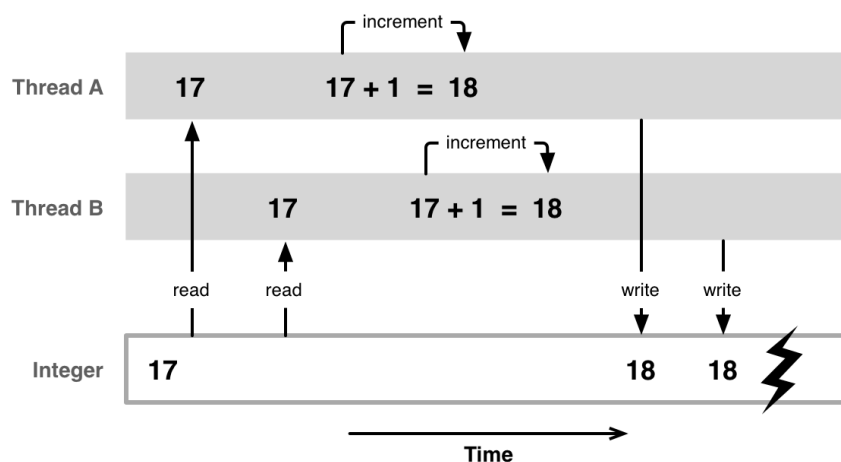
2.3. Synchronization

So as we see, synchronization is a crucial part of IPC in RTS, especially when processes communicate via shared memory. Communication through shared memory always has a risk of race conditions (when multiple processes acces the same data and cause unexpected behaviour) and data inconsistency if the processes are not properly synchronized. Tratitional synchronization techniques ensure mutual exclusion (only one process at a time uses shared resource) thus avoiding race conditions and ensuring data consistency. Race conditions happen when for example two processes write to the same resource.

2. Background

Lets take a single counter instance with value 17 as a shared resource in a shared memory region. If one has process p1 and preocess p2 that increment the number we would think that the end result is 19. But what could happen is that p1 reads the value 17 and then p2 reads the value 17. Now internally both processes increment that number to 18 and both processes would write 18 to that shared resource. To understand this example more in detail fig. 2.1 visualizes a race condition with threads. The difference between processes and threads is just, that threads are part of a process which can perform multiple tasks simultaneously within that process [18]. So the concept shown in fig. 2.1 can be used for multiple processes too.

Figure 2.1.: Race condition between two threads, which write to the same shared variable [19].



2.3.1. Mutual Exclusion

As discussed mutual exclusion does only allow one process or thread to access the shared resource at a time. This approach inherently relies on blocking some processes, which can lead to several issues, including deadlocks, process starvation, priority inversion, and increased response times. While explaining some of these problems, this paper will also give an superficial explanation on some non wait-free Sysnchronisation techniques. [2]

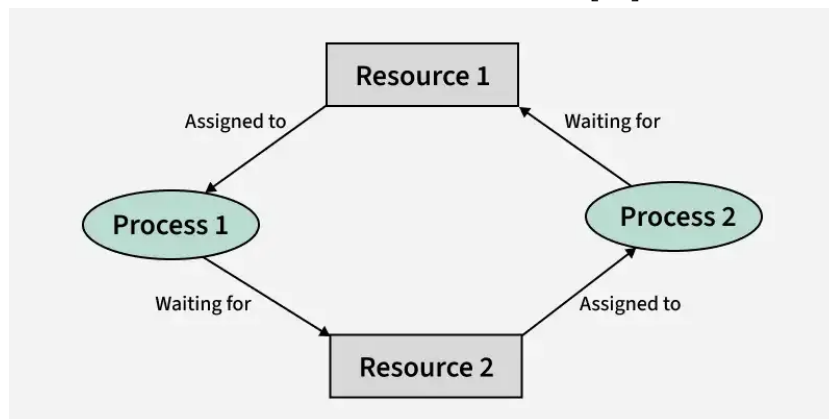
Deadlock

A deadlock happens when each process of a set of processes are waiting for a resource held by other processes in the same set. The processes which are in a deadlock wait indefinitely for the resources and never terminate. So the resources that these waiting processes hold are also not released and are also not available to other processes. As one

2. Background

can see, this leads to a situation where no process can make any process. To understand this better fig. 2.2 visualizes a deadlock with two processes. In this example Process 1 holds Resource 1 and waits for Resource 2, while Process 2 holds Resource 2 and waits for resource r1. So both processes are waiting for each other to release the needed resources. [20], [21]

Figure 2.2.: Deadlock between two processes, which wait for each other to release the needed resources [20].



Process Starvation

So will wait to enter the critical section. And a process that tries to entry the critical section, must eventually enter. Starvation is when that particular process will never enter the critical section. This usually happens when a synchronization method allows one or more processes to make progress while ignoring a particular process or processes. So this requires some sort of low and high priority processes. When we always have high priority processes available and some low priority processes, it might happen that these low priority processes will never be able to enter the critical section. This is a problem, since these low priority processes might be important for the system. [22]

Priority Inversion

What could happen too is that while a low priority process is accessing a shared resource, a high priority process is waiting for that resource. This is called priority inversion. [23]

Increased Response Times

So as we see traditional synchronization might lead to problems. But even if these problems do not occur we can see that mutual exclusion will always lead to waiting processes which leads to increased response times. This could lead to not meeting the

2. Background

timing constraints of a HRTS environment. So synchronization mechanisms are needed that do not introduce blocking.

2.4. Wait-Free Synchronization

Wait-free synchronization solves all the problems mentioned before. Wait-free synchronization guarantees that every process will complete its operation in a finite number of steps, regardless of contention from other processes. This means starvation or a deadlock is by definition not possible anymore. Since starvation or blocking cannot happen anymore priority inversion is eliminated too, because there is not anymore the possibility that a low priority process blocks the task of an high priority process. This ensures system responsiveness and predictability thus ability to meet timing constraints, which are essential for hard real-time applications.

2.5. Lock-Free Synchronization

2.6. Rust Programming Language

2.7. State of the Art

3. Related Work

4. Methodology

5. Choosing optimal wait-free data structure

6. Results

7. Implementation

8. Conclusion and Future Work

Bibliography

- [1] M. Herlihy, “Wait-free synchronization”, *ACM Trans. Program. Lang. Syst.*, vol. 13, no. 1, pp. 124–149, Jan. 1991. doi: 10.1145/114005.102808. [Online]. Available: <https://doi.org/10.1145/114005.102808>.
- [2] B. B. Brandenburg, *Multiprocessor real-time locking protocols: A systematic review*, 2019. arXiv: 1909.09600 [cs.DC]. [Online]. Available: <https://arxiv.org/abs/1909.09600>.
- [3] O. Kode and T. Oyemade, *Analysis of synchronization mechanisms in operating systems*, 2024. arXiv: 2409.11271 [cs.OS]. [Online]. Available: <https://arxiv.org/abs/2409.11271>.
- [4] A. Kogan and E. Petrank, “A methodology for creating fast wait-free data structures”, *SIGPLAN Not.*, vol. 47, no. 8, pp. 141–150, Feb. 2012. doi: 10.1145/2370036.2145835. [Online]. Available: <https://doi.org/10.1145/2370036.2145835>.
- [5] S. Timnat and E. Petrank, “A practical wait-free simulation for lock-free data structures”, *SIGPLAN Not.*, vol. 49, no. 8, pp. 357–368, Feb. 2014. doi: 10.1145/2692916.2555261. [Online]. Available: <https://doi.org/10.1145/2692916.2555261>.
- [6] M. M. Michael and M. L. Scott, “Simple, fast, and practical non-blocking and blocking concurrent queue algorithms”, in *Proceedings of the Fifteenth Annual ACM Symposium on Principles of Distributed Computing*, ser. PODC ’96, Philadelphia, Pennsylvania, USA: Association for Computing Machinery, 1996, pp. 267–275. doi: 10.1145/248052.248106. [Online]. Available: <https://doi.org/10.1145/248052.248106>.
- [7] H. Huang, P. Pillai, and K. G. Shin, “Improving Wait-Free algorithms for inter-process communication in embedded Real-Time systems”, in *2002 USENIX Annual Technical Conference (USENIX ATC 02)*, Monterey, CA: USENIX Association, Jun. 2002. [Online]. Available: <https://www.usenix.org/conference/2002-usenix-annual-technical-conference/improving-wait-free-algorithms-interprocess>.
- [8] A. Pellegrini and F. Quaglia, *On the relevance of wait-free coordination algorithms in shared-memory hpc: the global virtual time case*, 2020. arXiv: 2004.10033 [cs.DC]. [Online]. Available: <https://arxiv.org/abs/2004.10033>.

Bibliography

- [9] B. Xu, B. Chu, H. Fan, and Y. Feng, "An analysis of the rust programming practice for memory safety assurance", in *Web Information Systems and Applications: 20th International Conference, WISA 2023, Chengdu, China, September 15–17, 2023, Proceedings*, Chengdu, China: Springer-Verlag, 2023, pp. 440–451. doi: 10.1007/978-981-99-6222-8_37. [Online]. Available: https://doi.org/10.1007/978-981-99-6222-8_37.
- [10] A. Sharma, S. Sharma, S. Torres-Arias, and A. Machiry, *Rust for embedded systems: Current state, challenges and open problems (extended report)*, 2024. arXiv: 2311.05063 [cs.CR]. [Online]. Available: <https://arxiv.org/abs/2311.05063>.
- [11] K. Kavi, R. Akl, and A. Hurson, "Real-time systems: An introduction and the state-of-the-art", in Mar. 2009. doi: 10.1002/9780470050118.ecse344.
- [12] X. Hua, J. Zeng, H. Li, J. Huang, M. Luo, X. Feng, H. Xiong, and W. Wu, "A review of automobile brake-by-wire control technology", *Processes*, vol. 11, p. 994, Mar. 2023. doi: 10.3390/pr11040994.
- [13] *Inter process communication (ipc)*, <https://www.geeksforgeeks.org/inter-process-communication-ipc/>, 2025.
- [14] A. Venkataraman and K. K. Jagadeesha, "Evaluation of inter-process communication mechanisms", 2015. [Online]. Available: <https://api.semanticscholar.org/CorpusID:6899525>.
- [15] K. C. Wang, "Process management in embedded systems", in *Embedded and Real-Time Operating Systems*. Cham: Springer International Publishing, 2023, pp. 115–168. doi: 10.1007/978-3-031-28701-5_5. [Online]. Available: https://doi.org/10.1007/978-3-031-28701-5_5.
- [16] S. Mogare, A. Mahamune, D. Sathe, H. Bhangare, M. Deshmukh, and A. Ingale, "Message passing vs shared memory-a survey of trade off in ipc", *International Research Journal of Modernization in Engineering Technology and Science (IRJMETs)*, vol. 06, Nov. 2024.
- [17] M. A. Suleman, O. Mutlu, M. K. Qureshi, and Y. N. Patt, "Accelerating critical section execution with asymmetric multi-core architectures", in *Proceedings of the 14th International Conference on Architectural Support for Programming Languages and Operating Systems*, ser. ASPLOS XIV, Washington, DC, USA: Association for Computing Machinery, 2009, pp. 253–264. doi: 10.1145/1508244.1508274. [Online]. Available: <https://doi.org/10.1145/1508244.1508274>.
- [18] *Difference between process and thread*, <https://opensourceforgeeks.blogspot.com/2014/01/race-condition-synchronization-atomic.html>, 2025.
- [19] A. Thakur, *Race condition, synchronization, atomic operations and volatile keyword*. <https://opensourceforgeeks.blogspot.com/2014/01/race-condition-synchronization-atomic.html>, 2014.

Bibliography

- [20] *Introduction of deadlock in operating system*, <https://www.geeksforgeeks.org/introduction-of-deadlock-in-operating-system/>, 2025.
- [21] P. Chahar and S. Dalal, "Deadlock resolution techniques: An overview", *International Journal of Scientific and Research Publications*, vol. 3, no. 7, pp. 1–5, 2013.
- [22] P. A. Buhr, "Concurrency errors", in *Understanding Control Flow: Concurrent Programming Using $\mu C++$* . Cham: Springer International Publishing, 2016, pp. 395–423. DOI: 10.1007/978-3-319-25703-7_8. [Online]. Available: https://doi.org/10.1007/978-3-319-25703-7_8.
- [23] Y. Wang, E. Anceaume, F. Brasileiro, F. Greve, and M. Hurfin, "Solving the group priority inversion problem in a timed asynchronous system", *IEEE Transactions on Computers*, vol. 51, no. 8, pp. 900–915, 2002. DOI: 10.1109/TC.2002.1024738.

List of Acronyms

IPC Inter-Process Communication

HRTS Hard Real-Time System

SRTS Soft Real-Time System

RTOS Real-Time Operating System

RTS Real-Time System

List of Figures

2.1. Race condition between two threads, which write to the same shared variable [19].	6
2.2. Deadlock between two processes, which wait for each other to release the needed resources [20].	7

List of Tables

A. Appendix