TDT4900: Masters Thesis

CMR: A concurrent memory management system for Rust

Martin Hafskjold Thoresen

supervised by
Magnus Lie Hetland

May 31, 2018

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Sammendrag

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Acknowledgements

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Contents

List of Figures

Resolve These Things

Part I

Bing Bong

Chapter 1

Introduction

This is a thesis blabla
Something about memory management and parallel systems
bing bong about CMR
Motivation

Chapter 2

Background

2.1 Garbage Collectors

A *Garbage Collector* usually refers to an automatic subsystem that handles memory management without requiring programmer assistence. Many widespread language implementations, including Java, Python, and Go, use a garbage collector, although the internal details of each system varies greatly.

31/05 10:58
What types of GC are there?

The job of the garbage collector is to identify memory segments that are no longer used by the program. One way of doing this is to represent the program memory as a graph $G = (V, E)$ where V is all allocated memory segments and $(u, v) \in E$ if the region u contains an address that is inside the segment v. One consequence of this model is that memory addresses cannot be computed from other values.

31/05 10:58
More here

31/05 10:58
Discussion about GC in general?

2.2 Operating Systems

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.2.1 Virtual Memory

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit.

Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Memory Maps

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.2.2 Signals

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.2.3 Threads and Processes

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.3 Programming Languages

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.4 Concurrency

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.4.1 Common Patterns in Concurrent Programming

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.4.2 The ABA-Problem

2.5 Memory Reclamation

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.5.1 Hazard Pointers

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.5.2 Forkscan

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.6 Related Works

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2.6.1 Crossbeam

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit

blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Chapter 3

Rust

Rust is a new programming language focusing on safety, performance, and concurrency[12]. The official first stable release, Rust 1.0, was released in May 2015, and a new version of the language as well as the official compiler, `rustc`, is released every 6th week. The language is developed as an open source project on the version control platform GitHub[6] by over 2000 contributors as of May 2018[13]. The Rust project is organized into *teams*, such as the Core Team, the Compiler Team, and the Documentation Team. Many of the members of the Rust teams are Mozilla employees, and Mozilla officially sponsors the Rust project. The language has no formal specification, although all language changes are developed and documented through an Request For Comments (RFC) process. For a thorough introduction to Rust, see [15].

3.1 Introduction

Rust is a compiled language with a minimal runtime, similar to C and C++. `rustc` uses LLVM[2] as a compiler backend for code optimization and code generation. The performance of Rust code is very similar to that of C and C++[1]; variations are often due to the lack of stable features like SIMD support, or from different compile time information given to LLVM by either language.

Rust has many features from the ML family of programming languages, such as pattern matching and tagged enums, and a rich type system with type inference. Most notably, and unlike most other modern programming languages, Rust does not have a garbage collector. Despite this, Rust programs does not handle memory management manually; memory management is typically done statically at compile time by utilizing language features covered in the upcomming sections.

Rust uses `struct`s similar to C and C++ which can have *methods*, but it does not have inheritance. `Trait`s are similar to interfaces: they define methods and optionally an implementation, and `structs` *implement* the `Trait`. Traits can even be implemented for types that we have not defined ourselves, as long as we have defined the `Trait`. This is useful, since it means we can extend types from the standard library, or from other third party crates[1]. Important `Trait`s include `Deref` (the `*` operator), `Clone` (values that are clonable), and `Drop` (ran when a value is destroyed).

[1] A *crate* is a project unit, similar to a library

When an owned value leaves its scope, it is destroyed and its `Drop` method is ran. Primitive types, such as `char` or `u32` does not have a `Drop` implementation, but types which holds a resource, like allocated memory, often has. `String` and `Vec<T>` are common examples. `String` has a pointer to an internal buffer, which needs to be freed upon destruction in order not to leak memory. This `free` call is done inside `String::Drop`.

3.2 The Borrow Checker

A central concept in Rust is that of ownership. At any moment, an object has exactly one binding which *owns* the object. Ownership may be transfered ("*moved*", which is the default behaviour), or it may be *lended out*. Then the receiver is *borrowing* the binding. There are two types or borrows: immutable and mutable borrows. One of the reasons to differentiate between mutable and immutable borrows, is references in Rust can be either aliased, or mutable, but never both. That is, if there is a mutable reference to some object, then that reference has to be the *only* reference. This ensures that immutable references are never changed, which makes it simpler for the programmer to reason about the code since we get referential transparency, in addition to that it enables more compiler optimizations.

Borrowed objects are in effect *references* to some data, similiar to pointers or references in other programming languages. While Rust does have raw pointers (see Section 3.4), it is rarely used, and passing values by reference is prefered. The three types of ownership handling is shown in Fig. 3.1. In Fig. 3.1a we move `x`, so `x` is no longer usable after the last line, and an attempt to use it is caught as a compile time error: `error[E0382]: use of moved value: ‘x‘`. Since the caller of `foo` has "sent" the `Foo` to the function, it does no longer have to do any cleanup: this is now `foo`s responsibility.

Fig. 3.1b shows immutable borrow of `x`; the function `foo` may use the `Foo`, but it cannot mutate it. Fig. 3.1c shows a mutable borrow; now `foo` may mutate the `Foo`. Note that the binding `x` also needs to be mutable in order to borrow mutably.

(a) Ownership transfer

```
fn foo(f: Foo);
let x = ...
foo(x);
```

(b) Immutable Borrow

```
fn foo(f: &Foo);
let x = ...
foo(&x);
```

(c) Mutable Borrow

```
fn foo(f: &mut Foo);
let mut x = ...
foo(&mut x);
```

Figure 3.1: The three types of ownership handling.

Understanding the borrow checker is often a pain point for new programmers, and the period in which new Rust programmers learns an intuition about how to structure programs within these rules is often refered to as "fighting with the borrow checker".

3.3 Lifetimes

Lifetimes is the second important concept in Rust. The idea of lifetimes is to reason about the duration of the program execution in which some object is valid — its lifetime. By tracking the lifetime of all variables at compile time the Rust compiler is able to catch errors such as returning function local variable addresses. Section 3.3 shows an example function attempting to do this.

```
fn foo(_a: &i32) -> &i32 {
  let num: i32 = 420;
  let r: &i32 = &num;
  r }
```

Since Rust tracks the lifetime of all variables, it knows that the lifetime of `num` is the same as that of the function body. The lifetime of `r` is the same, as it is a reference to `num`. So when we try to return `r` in the last line of the function, Rust realizes that the lifetime of the reference we return ends its life at the end of the function; this is clearly not what we wanted, since it would make the returned reference dead on arrival. Compilation fails with the following error: `error[E0597]: ‘num‘ does not live long enough`.

Although Rust programmers may have to think about the lifetime of the variables, they seldom have to write lifetime annotated functions, due to *lifetime elison* — the compiler can ususally figure out the most general lifetime that fits the function. Functions may be annotated with explicit lifetimes, for instance if it takes multiple references in which the relative difference of the lifetimes of the references is important. `struct`s can also be annotated with lifetimes, and in fact is required to be so if any of its members are references. This is because the lifetime of the struct is bounded by the lifetime of its member variables.

```
struct Person<'a> {
  age: i32,
  name: &'a str }
```

Should we have a function that crates a new `Person` we might want to annotate it explicitly, if the function takes multiple references, but only one of these references is the `name` field:

```
fn make<'x, 'y>(f: &'y File, n: &'x str) -> Person<'x> { ... }
```

This way we can convey the information that the resulting `Peron` should live as long as `n`, but may outlive the file `f`.

3.4 Unsafe Rust

When talking about the Rust progrmaming language, one usually talks about a subset of Rust, called *Safe Rust*. In Safe Rust, there are no race conditions, mutable memory locations are never aliased, and all pointer accesses are valid. The real world, on the other hand, rarely offers these guarantees, and the unfortunate truth which Rust programmers must deal with is that in order to implement some of these safe abstractions we want (like `Vec`, `Mutex`, and `Box`), some unsafety is required. For this reason, Rust offers an escape hatch for some of its rules: *Unsafe Rust*.

The difference between Safe and Unsafe Rust is only four things. In Unsafe Rust one may: 1) dereference raw pointers 2) mutate statics 3) call `unsafe` functions 4) implement `unsafe` traits. One way of thinking about the unsafety of ones codebase is that there should be no undefined behaviour in safe code, no matter how the code looks like. In other words, it should be impossible to mess up so badly as to invoke undefiend behaviour without typing `unsafe`.

Dereferencing raw pointers is naturally `unsafe`, as it is not possible to statically guarantee that the address of the pointer is valid memory, or that the objects it points to is still alive, nor that mutation of that memory does not change an immutable reference some other place in the program. Mutation of `static` variables is also unsafe due to mutability of aliased references, and due to the lack of thread synchronization.

`unsafe` functions and traits are just a marker added to the function or trait, signaling that not all uses of this is guaranteed to be safe. As an example, the trait `Send` is a marker trait and types implementing `Send` may be sent across thread boundaries. While this is fine for most types, there are types which does not allow this. The reference counted pointer `Rc<T>` is an example, which is a pointer to a tuple[2] `(count, data)`. The `count` is incremented each time `.clone()` is called, and decremented when a variable is `Drop`ped. To understand why this cannot be send across thread boundaries safely, consider what happens if T_1 `.clone()` at the same time as T_2 `Drop`s it: the `count` field is written to twice without any synchronizationor atomic operations[3] — a race condition!

3.5 Concurrency

bing bong

3.5.1 Concurrency and Aliasing

One observation to make from the reference rules as presented in Section 3.2 is that since references are either aliased or mutable, then there can be no writes shared data between threads, in Safe Rust, even using atomics. While this is *technically* true, the Rust standard library uses `&T` and `&mut T` slightly different than "immutable" vs "mutable" in this context: `&T` means that the type may be shared between threads.

Take `AtomicUsize` as an example, a `usize` exposing atomic operations like `store`, `load`, and `compare_and_swap`, which signatures are shown in Listing 3.1.

Listing 3.1: Signatures for selected operations on `AtomicUsize`

```
fn load(&self, order: Ordering) -> usize;
fn store(&self, val: usize, order: Ordering);
fn swap(&self, val: usize, order: Ordering) -> usize;
fn compare_and_swap(&self, current: usize, new: usize, order: Ordering) -> usize;
```

Clearly, `AtomicUsize::store` modifies memory of the `usize`; despite this the function is `&self` and not `&mut self`, since the operation is allowed on variables which are shared between threads. This is a useful distinction, since we can have methods on `AtomicUsize`

[2] Not really, but for our purposes here we can pretend that it is.
[3] `Rc` does not use atomics for performance reasons, but `Arc` does, and it does implement `Send`.

that *is* `&mut self`, which then is only possible to invoke should the variable not have been shared between threads yet; we know this since this means that we have aliased mutable references, which is not allowed. For instance, `AtomicUsize::get_mut(&mut self) -> &mut usize` allows the underlying `usize` to be changed without any synchronization overhead.

3.5.2 Common Patterns

The standard librarys synchronization module `std::sync` contains primitives that most concurrent programs require, such as `Mutex`, `Channels`, `Condvar`, and `Atomics`. A common pattern in Rust is the Resource Allocation Is Initialization (RAII) pattern. The idea is that resources should be managed automatically when constructing and destructing an object. `Mutex` uses these ideas: `Mutex::lock` returns an `Result<MutexGuard>`, where the `MutexGuard` wrapps a mutable reference to the data that is protected by the `Mutex`. When the `MutexGuard` goes out of scope, its `Drop` implementation is ran, and the `Mutex` is unlocked.

It is common among Rust programmers to build abstractions over lower level primitives. For instance, a common pattern in parallel and concurrent programming is to have a *thread pool*, which is given work, and internally handles the thread synchronization and work division. Example usage of such an abstraction could be `let tp = ThreadPool::new(); tp.execute(|| ... );`. Since this can be implemented without any special compiler support, such crates are usually made as third party libraries.

Another example is data parallelism: given some collection of data we want to iterate over the elements and perform some operation on each element. The Rust library `rayon` offers exactly this: parallel iterators. Instead of writing `vec.iter()` to iterate over a `Vec` and then performing some operation on each element sequentially, with `rayon` we can write `vec.par_iter()`, and get data parallelism for free. The operation is then ran in parallel with any number of threads. Internally `rayon` uses a thread pool and work stealing to handle the division of labour among the threads.

09/05 12:32 Look at stuff from last semester. Move into std sec?

3.5.3 Memory Orderings

3.6 Nightly Rust

The Rust language and compiler follows a fixed release schedule, where a new stable version is released every six weeks. In addition to this there is the beta branch, which is the upcomming version, and the nightly version which is the most recent version, build daily from the `master` branch of the source tree.

The nightly version of the compiler allows users to opt in on *untsable* features: features that are partially or fully implemented, but which details are not yet committed to. These featuers includes new APIs in the standard library, new syntax, and new language features all together. As we have used multiple unstable features in CMR, we look at some of them in deatil.

3.6.1 Non-Lexical Lifetimes

The current implementation of lifetime checking in the compiler is *lexical*, meaning variables are live until they go out of scope, despite not being used. This is a limitation that one may want to get rid off. The feature Non-Lexical Lifetimes (NLL) lifts this requirement, and lets the lifetime of a variable last only until its last usage. Having this it is possible to seemingly break some of Rust rules, like aliased mutable references:

```
let mut v = vec![1,2,3];
let r1 = &mut v;
let r2 = &mut v;
```

This clearly violates one of the Rust rules, namely that we cannot have mutable aliased referenes. Yet, in this example we have two mutable references, `r1` and `r2`, to the same data. With NLL this will compile, as we do not use `r1` after having made `r2`, so its lifetime is implicitly ended right after its declaration. If we write `r1.push(1);` after `let r2`, we get the same error as without using NLL, since the lifetime `r1` overlaps with the lifetime of `r2`.

3.6.2 Trait Objects

When using traits in function signatures or structs we can either make the struct generic over some type that implements the trait, or we can use dynamic dispatch. As generics usually are implemented by copying the source code for the type for each invocation of a new type, it increases code size and compilation time. In addition, collections and similar structures cannot mix different types: a `Vec<SomeTrait>` cannot both contain elements of type `A` and `B`, even if both implements `SomeTrait`.

Dynamic dispatch is the other option. Now variables are *fat pointers*, containing both the pointer to the data type, and a pointer to a `vtable`[4], which contains information about the function addresses for that type, as shown in Fig. 3.2. The entry in the `vtable` is all functions for some trait. With this we can take any concrete type, and follow its vtable pointer, in order to find the implementation of some trait function for that type. In Fig. 3.2, both `Foo` and `Bar` implements some trait which have a function named `fnc`. By following the pointers from the stack, we get the data (left) and the function pointer (right). This way of implementing Trait Objects are usually not mandated by any standard, but it is popular across different language implementations nevertheless.

While trait objects offers greater flexibility in the usage of traits, the pointer jumping leads to worse cache behaviour which may have a large impact on performance, and important compiler optimizations like inlining is impossible.

3.6.3 Specialization

Specialization is a feature which allows multiple implementations of a trait for the same type, where the implementations are ordered by their specificity.

Assume we want to implement the trait `Debug` for a struct that is generic over some type `T: Struct<T>`. We might want to have different implementations of `Debug`

[4] the name `vtable` comes from the C++ world, where function on abstract types are called `virtual` functions

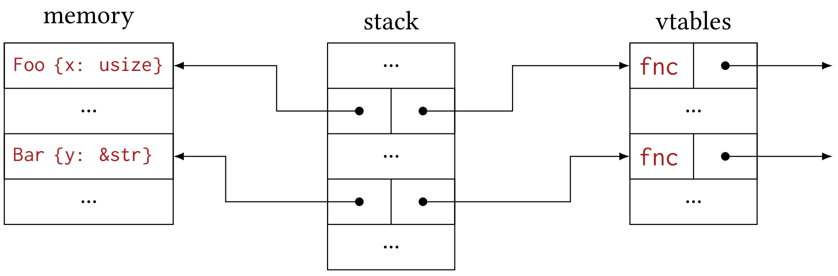


Figure 3.2: Illustration of memory when using Trait Objects.

depending on whether the generic parameter T implements Debug or not. Specialization makes this possible.

Listing 3.2: Using specialization to implement a trait twice.

```
impl<T> Trait for Struct<T> {
  default fn fmt(&self);
}
impl<T: Debug> Trait for Struct<T> {
  fn fmt(&self);
}
```

With only these two implementation it is clear which of the two we want for any type: if <T> implements Debug we want the second, and if it does not, we want the first. However, if we mix in yet another trait, Clone, such that we have a third implementation

```
impl<T: Clone> Trait for Struct<T> { ... }
```

it is no longer clear which implementation to use if T implements *both* Clone and Debug. The current implementation forbids such implemementations.

3.6.4 Allocators

The final nightly feature that we look at is *allocators*. It is not yet possible to change the default allocator in stable Rust, but a suggested API for creating new allocators and specifying the default system wide allocator for Rust programs is avaiable by opting in on the allocator feature. The feature defines a trait GlobalAlloc that defines functions analagous to malloc and free from libc, and a attribute #[global_allocator] to select which allocator we want to use.

The default allocator for Rust is jemalloc[8]. By using other external crates we can use either the default system allocator, or jemalloc wrapped in our own allocator. This can be useful if we want to do bookkeeping, gather statistics, or do any thread synchronization outside of the actual allocator we are using.

Listing 3.3: Custom allocators wrapping jemalloc and the system allocator

```
pub struct WrapJemalloc;
unsafe impl GlobalAlloc for WrapJemalloc {
    unsafe fn alloc(&self, layout: Layout) -> *mut Opaque {
        ⟨Do something before calling alloc⟩
        Jemalloc.alloc(layout) }
    unsafe fn dealloc(&self, ptr: *mut Opaque, layout: Layout) {
        ⟨Do something before calling free⟩
        Jemalloc.dealloc(ptr, layouer); } }

pub struct WrapSystem;
unsafe impl GlobalAlloc for WrapSystem {
    unsafe fn alloc(&self, layout: Layout) -> *mut Opaque {
        ⟨Do something before calling alloc⟩
        System.alloc(layout) }
    unsafe fn dealloc(&self, ptr: *mut Opaque, layout: Layout) {
        ⟨Do something before calling free⟩
        System.dealloc(ptr, layout); } }
```

Chapter 4

CMR

In this chapter we present a concurrent memory reclamation scheme called *CMR*. We define the problem of memory management carefully in Section 4.1, in order to get a complete understanding of which problem we set out to solve. In Section 4.2 we present an abstract overview of CMR, in order to get a high level understanding of the system as a whole, without having to think about technical or implementation details. Section 4.3 discusses the primitives and operations of CMR, and how they are used. Finally in Section 4.4 we argue for the correctness of the system as presented in this chapter. By reasoning about CMR without an implementation we later aim to show that the implementation (Chapter 5) fits the description of the system as we define it in this chapter.

4.1 Problem Definition

We start by defining some central concepts. Memory M is the set of all addresses in the address space of the machine. A *block* is a tuple (a, n) and represents the memory segment $[a, a + n)$. M is a disjoint set $M = A \cap F$ where A is the set of allocated blocks, and F is the remaining of the memory space. F needs not, and is almost never, a consecutive segments, but simply all memory that is outside any allocated block. We call such memory *invalid*, and all memory in an allocated block *valid*. We model the program memory as a graph $G = (A, E)$ where $(u, v) \in E$ iff there is a pointer in u pointing to an address in the segment v. That is, memory segments are the vertices, and pointers in the program are the edges. See Fig. 4.1 for a possible memory layout with a graph.

As most programs need memory blocks of dynamic size, allocation and deallocation, *"freeing"*, is commonplace. The problem of memory management is to know when it is *safe* to free a memory block. We want to avoid the following memory hazards:

Definition 4.1 (*use-after-free*)**.** Memory that was allocated and then freed is read.

Definition 4.2 (*invalid-read*)**.** Memory that has never been allocated is read.

Definition 4.3 (*double-free*)**.** A block is freed twice without being allocated in between.

use-after-free is the most hazarous of the three, as program behaviour is often undefined when freed values are read; in many language implementations undefined

```
a = Node { value = 4, next = null }
b = Node { value = 8, next = a }
list = [a, b, 3]
```

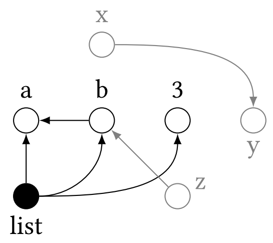


Figure 4.1: Code sample (left) with possible heap layout (right). If the black filled node is the only root, the black nodes are reachable, and the gray nodes are not.

behavious means that the entire program is illegal, and one cannot assume anything about its behaviour (see Section 2.3). The consequence of *use-after-free* usually ranges from reading values that are unchanged from the time the block was freed, to mutation of memory that has been reused.

invalid-access is the least frequent of the three, as it requires the programmer to conjure a pointer out of thin air, since it has never been allocated in the system. As with *use-after-free*, this is too is usually undefined, with similar consequences. Despite their similarities we choose to have *invalid-access* as a separate category, as pointer arithmetic may lead to these hazards.

double-free is technically not a memory hazard, as the operating system can check for the validity of pointers that are freed, although this is often not done in practice. It is not clear whether this is due to performance penalties of checking, or if it is primarily a legacy bevhaviour; POSIXs definition of `free` states that it is undefined behaviour to pass a non-allocated pointer to `free`[11].

We aim to show that CMR guarantees that neither of the three memory hazards are possible.

4.1.1 Shared Memory

Newer languages like modern C++ and Rust aim to avoid having the programmer manage memory manually, due to a long history of the consequences of memory hazards. For single threaded application, this may be considered a problem with suggested solutions. Rusts ownership model and lifetime tracking (Chapter 3), and similar methods from the C++ standard library, are proposed solutions. However, the ownership model does not handle shared memory functionally, as objects in shared memory might not have an owner responsible for its management. Despite not being a complete solution, having "solved" single threaded memory management turns out to be of great help.

We divide up A into two disjoint parts: *owned* and *shared* memory. Owned memory is all memory which management is already handled by some system, like Rusts ownership model or the smart pointers of C++. Shared memory is the memory in which the structures that is not modeled well by other constructs live, like the nodes in a linked list.

A key idea to recognize is that despite data being in Shared memory, they might themselves own data that is in Owned memory, like the binary tree in Fig. 4.2. The

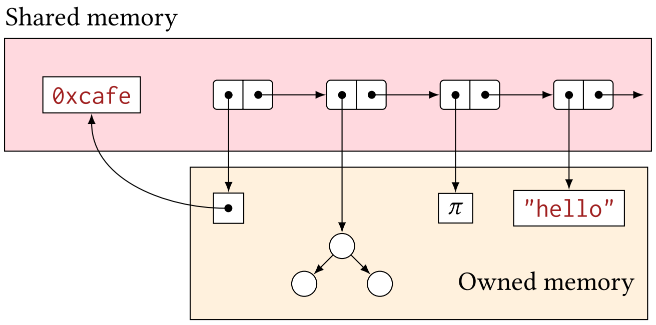


Figure 4.2: Example of memory layout showing owned memory (beige) and shared memory (red). Types in shared memory may contain pointers to owned memory, and vice versa.

destruction of a list node containing the binary tree will utilize the system for owned memory, and make sure that the binary tree is cleaned up properly. It does not matter if the list node itself resided in owned or shared memory. With this distiction we can reduce our problem space significantly, as we only have to worry about the small subset of A that is shared memory. Note that it is also possible to have the data types that are referenced from shared memory but stored in owned memory, like the pointer poining to `0xcafe` in Fig. 4.2. This includes pointers on a stack frame, but might also include a entry in a hashmap. It is these pointers that CMR aims to control.

4.2 Overview

CMR is based on the idea that if we have access to all roots in the system at an instant, finding the set of all reachable blocks R from the set of roots R_0 is simple: R is the transitive closure of "there is a pointer from x to y" on R_0. We call identifying R for *reachability analysis*. By then tracking all allocated blocks A, we can identify the set of unreachable block U by taking the relative complement of R in A: $U = A \setminus R$.

When performing the reachability analysis in a concurrent systems, simply following pointers while maintaining a frontier of unvisited blocks is not sufficient. Since there are multiple threads in the system, some other thread T' may come along and change pointers, causing reachable blocks to be observed as unreachable by the reclaiming thread, as shown in Fig. 4.3.

In order to handle mutation problems, we obtain a snapshot of the entire processes memory by freezing all threads, reading their roots, and forking the process. Thus, we have a snapshot of the entire memory of the program, in which we also have all roots. Now the reachability analysis is simpler, since there will be only one thread in the forked process, namely the garbage collecting thread. Fig. 4.4 shows pseudocode for both the reclaiming thread and the other threads for this procedure.

define "Rust memory" and shared memory

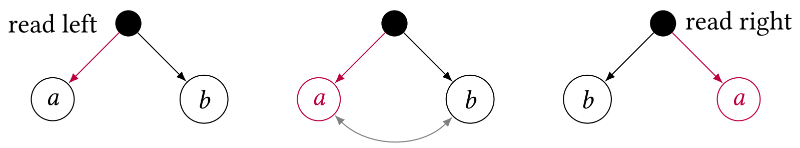


Figure 4.3: Mutation in the memory graph may lead to reachable blocks being observed as unreachable.

```
CMR():
    freeze_threads()
    wait_for_writes()
    read_guards()
    fork()
    unfreeze()
    for addr in unreachables():
        free(addr)
```

```
on_freeze():
    write_roots()
    register_done()
    wait_for_unfreeze()
```

Figure 4.4: Pseudocode of CMR. The leftmost code is for the thread that runs the reclamation pass, and the rightmost code is other threads in the system.

4.3 Primitives of CMR

The central data type to achieve this is `Guard`. A `Guard` is an object in which a pointer to shared memory is stored. All pointers to shared memory are stored in a Guard. By having access to all `Guard`s, CMR have access to all roots in the program at any instant. `Guard`s are similar to Hazard Pointers (HPs) (Section 2.5.1), except that no thread synchronization is performed when making new or updating existing `Guard`s, which reduces the overhead of CMR significantly compared to HPs.

CMR has four types that are essential to understanding how we manage safe access to shared memory: *Guard*, *Atomic*, *NullablePtr*, and *Ptr*. All types are generic over the type of the data they protect, which is omitted for brevity.

Definition 4.4 (Guard)**.** A *Guard* is an object that contains a *root* or $\perp$. The Guard is non-movable in memory. All roots are stored in Guards.

Definition 4.5 (Atomic)**.** An *Atomic* is a pointer type that provides safe concurrent access to its users.

Definition 4.6 (NullablePtr)**.** *NullablePtr* is an immutable pointer that may be $\perp$. It is obtained through a Guard. When a NullablePtr p is obtained from a Guard g, g is immutable thoughout the lifetime[1] of p.

Definition 4.7 (Ptr)**.** *Ptr* is an immutable pointer that may *not* be $\perp$. Its semantics are similar to that of NullablePtr, but the two are distinct types for simplification of the null-case. All accesses to shared memory is through a Ptr.

[1] We use the same meaning of lifetime as Rust (Section 3.3)

These four types are the building blocks of writing concurrent structures in CMR. All pointers are stored in an `Atomic`, and loading an atomic using a `Guard` yields a `NullablePtr`, which may be promoted to a `Ptr`, if it is not null. The `Ptr` provides access to the data the pointer points to. In addition to having the types, CMR defines a number of operations that acts on these types. We look at each type in turn, and define their operations.

4.3.1 Guard

A Guard can be constructed with the initial value of ⊥ with *make-guard*

$$\text{make-guard} :: () \longrightarrow Guard \tag{4.1}$$

It can also copy the value of another Guard with *copy-guard*.

$$\text{copy-guard} :: (Guard, Guard) \longrightarrow () \tag{4.2}$$

General usage of `Guard` is to construct the number of `Guards` one needs for some operation. These `Guards` are then used to load `Atomics` into.

4.3.2 Atomic

Atomic is a regular atomic pointer variable, supporing operations such as *store*, and *compare-and-swap*.

$$\text{store} :: (Atomic, NullablePtr) \longrightarrow () \tag{4.3}$$

$$\text{compare-and-swap} :: (Atomic, NullablePtr, NullablePtr) \longrightarrow NullablePtr \tag{4.4}$$

It is not safe to *load* an atomic, as there is no guarantee that the pointer read is protected by a guard. Instead, CMR defines *load-atomic*, which loads an Atomic into a Guard, and returns the value read as a NullablePtr:

$$\text{load-atomic} :: (Guard, Atomic) \longrightarrow NullablePtr \tag{4.5}$$

4.3.3 NullablePtr

The NullablePtr is just a convenience type in order to not have to handle the ⊥ case of all pointers. Whether the pointer is null or not can be checked:

$$\text{is-null} :: (NullablePtr) \longrightarrow bool \tag{4.6}$$

CMR also supports using the lower bits of a pointer to store extra information (a *tag*). This is useful for implementing deletion in linked lists, among other things. The tag is read with *tag*,

$$\text{tag} :: (NullablePtr) \longrightarrow int \tag{4.7}$$

and a new NullablePtr can be constructed with a given tag using *with-tag*.

$$\text{with-tag} :: (NullablePtr, int) \longrightarrow NullablePtr \tag{4.8}$$

The actual address is obtained through *addr*

$$\text{addr} :: (NullablePtr) \longrightarrow int \tag{4.9}$$

4.3.4 Ptr

Reconsider having NullablePtr in here. I think its mostly a rust impl thing, as we just want Deref on Ptr but must handle null as well.

Ptr may be used in the place of NullablePtr, since is it just a special case of it. All functions that take a NullablePtr can also take a Ptr.

4.3.5 Other actions

We also need a few other operations to make sure that the implementation of funcions are valid. For instance, in load-atomic there is a window in between reading the atomic and storing the pointer read it in the guard in which a reclamation pass may have happened. The higher order function *without_reclamation* makes sure that this is safe, by running the given function without a reclamation pass happening in between:

$$\text{without\_reclamation} :: (() \longrightarrow T) \longrightarrow T \tag{4.10}$$

As CMR controls memory allocations, it also defines its own allocation function:

$$\text{alloc} :: (Guard, T) \longrightarrow Ptr \tag{4.11}$$

4.4 Correctness

With these types and operations we are able to prove important properties of the system.

Theorem 4.8 (`Guard` is valid)**.** *If a* `Guard` *is not* $\perp$*, it points to valid memory.*

Proof. Since the `Guard` $g \neq \perp$, it has loaded its value from an Atomic a using `atomic-load`. We first show that a is itself in valid memory by induction: *Base case:* the Atomic resides in Rust memory, and is thus valid. *Inductive case:* the `Atomic` resides in shared memory, and thus accessed through a `Ptr` p. This `Ptr` is protected by a `Guard` $g_2 \neq g$, since g_2 is immutable throughout the lifetime of p, and g is being changed. g_2 is valid by induction, so the pointer value in a is reachable. This shows that the value read from the `Atomic` is valid.

Using `without_reclamation` (Eq. (4.10)), we make sure that the read of a and the store in g happens without a reclamation pass in between. Thus all valid pointers before the read is still valid after the store. After the store operation in g has completed, g protects v. Thus v is valid. □

Lemma 4.9 (`Ptr` is valid)**.** *The* `Ptr` *points to valid memory.*

Proof. The `Ptr` p is read from a `Guard` g and g is immutable throughout the lifetime of p so they have the same value. $p \neq \perp$, so this follows by Theorem 4.8. □

Lemma 4.9 is the most important result in this section, since it guarantees that accesses of the memory in a `Ptr` is valid. Thus, a memory access through a `Ptr` can not result in a invalid-read (Definition 4.2) or use-after-free (Definition 4.1) hazard.

Part II

Abcdef

Chapter 5

Implementation

In this chapter we look at the Rust implementation of CMR. We look at four things: how thread freezing is implemented (Section 5.2), how reachability analysis is done (Section 5.3), thread and process intercommunication (Section 5.1), and how the primitives and operations as defiend in Chapter 4 are implemented (Section 5.4).

5.1 The Reclaimer

There is only at most one reclaimer at any time. The first thing a potential reclaiming thread tries to do, is take the lock (R2). Then we take another lock that is used for allocation (R3) (see Section 5.5.1). This may block, but since we have not frozen any threads yet, any thread holding the lock will eventually release it, assuming it itself is not deadlocked. Next we initialize shared variables, such as the counters used by threads in their signal handler. After they are initialized, we freeze all threads (R5), and count how many threads were successfully frozen, which is used in (R6) where we wait on all threads to finish their job. When done, the data that the threads write out, their `Guards` and allocations, are read (R7).

Listing 5.1: Pseudocode of the work of the reclaimer

```
 R1 run_reclaim_pass() {
 R2     if reclaim_lock.lock().fail() { return }
 R3     lock_malloc()
 R4     init_shared_vars()
 R5     n = freeze_threads()
 R6     while sh_done_counter.load() != n { wait() }
 R7     (guards, allocs) = read_thread_datas()
 R8     memory_map = mmap()
 R9     write_marker(memory_map)
R10     if fork() == Child {
R11         rs = find_reachable()
R12         write_to_mmap(rs, memory_map)
R13         exit(0) }
R14     sh_frozen.store(false)
R15     send_to_background_thread(allocs, memory_map) }
```

rewrite

When `fork()`ing the child process continues the thread of the parent process that called `fork()`, such that it has access to everything that the original thread had. As such, we don't need to communicate from parent to child. However, the job of the child process is to run reachability analysis, and we do need its result. CMR uses *memory map* (Section 2.2.1) for IPC, which is set up before the fork. Since we need to know when the child is done writing its results, we write a *marker* word as the first word in the memory map. Then we `fork()`. The child process does the reachability analysis, and writes the result after marker in the memory map. When it is done writing, it overwrites the marker with the number of elements written.

In order to minimize the delay of the reclamation pass from the point of view of the user, we spawn a background thread which handles the part of the reclamation performed after we `fork()`. This ensures that no user thread has to wait for the child process to finish its reachability analysis before going back to running application code. After forking, the parent process unfreezes the other threads, and sends all necessary data to the background thread. Note that the `reclaim_lock` is not released, even though the thread is exiting the procedure. The lock is only released when the background thread is fully done with the reclamation pass.

5.2 Thread Freezing

Signals is a process communication mechanism used by POSIX compliant operating systems. POSIX threads (`pthreads`) also supports signaling for communication between threads in a process. We utilize this in order to implement thread freezing, by registering a signal handler for the signal `SIGUSR1`. Pseudocode for the signal handler CMR uses is shown in Listing 5.5. The signal handler is registered with the `sigaction` function, and threads are signaled with `pthread_sigqueue`.

Due to a complication in which a threads shutdown procedure is initiated without the signal handler being removed, we need to check if we are shutting down (`SH1`), as we do not have access to the thread local storage, which we need. If we are shutting down we will not participate in the rest of the handler, but clean up in order to avoid

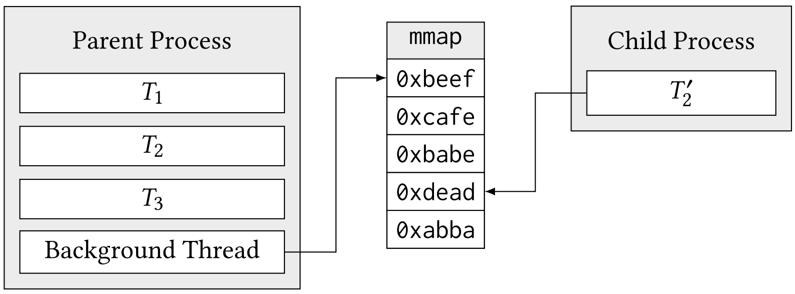


Figure 5.1: Illustration of IPC through a memory map. T_2 in the parent process is the reclaiming thread, so T_2 is the one thread in the child process. Both processes have access to the same memory in the memory map.

deadlocks (SH2). We use sh_enter_counter to keep track of how many threads are present in the signal handler; the reclaimer knows how many threads it succesfully signaled, so it knows how many threads to expect. (SH3) registers a threads presence, in addition to giving each thread a unique index in the range $[0, n)$, where n is the number of threads signaled. This is used in (SH4), where each thread writes out their guards and allocations into the global vector thread_datas. We then register that we have written our data (SH5), and wait for the reclaimer to unfreeze us (SH6).

Since we don't know that we have exited the handler after SH6, do we really need SH7??

The signal handler is registered by a thread with a call to the thread_activate procedure, which must be called before the thread is using CMR. The system has a fixed upper bound on the number of threads that may be active at any time, through the thread_datas vector, which is pre-allocated with a fixed size .

bench automation of this in Guard or something?

Since a thread handler may have registered it leaving by decrementing sh_enter_counter, but then preempted before leaving the handler, CMR uses pthread_sigqueue instead of the more standard pthread_kill, as pthreads may simply ignore a signal that is sent to a thread when that thread is already in the signal handler for a given signal. This is problematic, since the reclaimer keeps carefully track of where threads are in their handler, and waits for all threads to reach certain checkpoints. If the reclaimer is lead to believe that a thread just entered the handler when in fact it was just leaving, it will lead to a deadlock.

check this out

5.3 Reachability

Reachability analysis is done through a standard mark-and-sweep algorithm. The reclaimer collects all roots from all active threads, and looks through the memory of each data type that is pointer to by the roots. If any data type contains a Guard or Ptr, the address is marked as seen, and added to the frontier. When the frontier is empty, we have registered all reachable memory. Listing 5.3 shows the algorithm, which is effectively breadth-first search through the memory graph.

Finding pointers in arbitrary data types might involve significant work since the size of the data types can be arbitrarily large, in addition to that memory might not be initialized, and false positives. Instead of scanning through the memory block linearly, CMR has a Trait called Trace, which all data types that is in shared memory must implement. A type implementing Trace knows a bound on how many shared memory pointers it contains, and can write these out to a buffer. For instance, a Node in a single

should write somewhere why we can assume that the ptr is the start of a block

Listing 5.2: Pseudocode for the signal handler used by CMR

```
SH1 if is_in_cleanup:
SH2     clean_up_and_return()
SH3 id = sh_enter_counter.fetch_add(1)
SH4 write_out_data_to(thread_datas[id])
SH5 sh_done_counter.fetch_add(1)
SH6 while sh_frozen.load():
SH7     wait()
SH8 sh_enter_counter.fetch_sub(1)
```

Listing 5.3: The algorithm used to find all reachable memory blocks.

```
FIND-REACHABLE(ROOTS)
    Frontier = Roots
    Seen = Roots
    while mem = Pop(Frontier)
        for ptr in Ptrs(mem)
            if ptr not in Seen
                Push(Frontier, ptr)
            Insert(Seen, ptr)
    return Seen
```

linked list contains only one pointer, namely its next pointer, which is trivial to write out.

The implementation of this uses *Trait Objects* (Section 3.6.2), which involves dynamic dispatch. This solution is potentially expensive, as it may involve cahce misses in the I-cache, although the number of misses is limited by the difference in data types in shared memory, which normally is smaller than in Rust memory.

Listing 5.4: Definition of the `Trace` trait and a sample implementation for a linked list node. The impementation uses *specialization* (Section 3.6.3) as the implementation of `Node`s containing data that itself is `Trace` is different.

```rust
pub trait Trace {
    fn count(&self) -> usize { 0 }
    fn write(&self, &mut [TraitObject]) -> usize { 0 }
}
pub struct Node<T> {
    data: ManuallyDrop<T>,
    next: Atomic<Node<T>>,
}
impl<T> cmr::Trace for Node<T> {
    default fn count(&self) -> usize { 1 }
    default fn write(&self, slice: &mut [TraitObject]) -> usize {
        let p = unsafe { self.next.load(SeqCst) };
        if !p.is_null() {
            slice[0] = ptr::trait_object(p);
            1
        } else {
            0
        }
    }
}
```

`Trace` contains default implementations of the two functions, such that primitive types can easily implement it. `write` takes a buffer, writes all pointers to it as `TraitObject`s, and returns the number of objects written. `count` gives an upper bound on the number of pointers written. This is useful for collection types, like `Vec` or

HashMap, which also may contain pointers to shared memory.

Node is a standard node from a linked list, containing data, and a next pointer. The implementation of write loads the next pointer (T12), which is an unsafe operation, as there is no Guard protecting the pointer. This is safe in the context of the reclaimer since the memory will be freed at earliest when we finish the reachability analysis, and at that point we no longer read the memory. The implementation only writes out the pointer if it is non-null. While this is not required for CMR to function, it simplifies the logic in the reachability analysis.

5.4 Data Types

In this section we look at the concrete Rust implementation of selected data types and operations, and argue for their correctness.

Guard

The Guard is implemented as a single word, in addition to an empty type (the PhantomData) as Rust *requires* generic types to be used. Guards aren't normally constructed directly (), but rather declared with the guard! macro, which constructs it and calls Guard::register. An excerpt of the definitions of Guard is shown in Listing 5.5.

non movable types somewhere

Listing 5.5: Excerpt of Guards definitions

```
struct Guard<T> {
  ptr: usize,
  _marker: PhantomData<T>,
}
impl<T: Trace> Guard<T> {
    pub unsafe fn new() -> Self { Guard { inner: 0, _marker: PhantomData, } }
    pub fn copy_guard(&mut self, other: &Self) { self.inner = other.inner; }
    pub fn register(&mut self) {
        ROOTS.with(|r| { let mut v = r.borrow_mut();
                         v.push(GuardPointer::from_guard(self)) });
    }
    ...
}
macro_rules! guard {
    ($var:ident) => { let $var = unsafe { &mut $crate::guard::Guard::new() };
                      $var.register();
    }
}
```

rewrite, focus on register

Since we need to keep track of type information dynamically, we construct a GuardPointer which does exactly this, and pushes it onto a thread local Vec, ROOTS. It is also possible to construct a NullablePtr from the Guard. The guard! macro corresponds to make-guard (Eq. (4.1)). copy-guard (Eq. (4.2)) is shown in the Listing.

Using a macro for declaring Guards is a trick that serves two purposes: 1) the user needs to write unsafe in order to make a Guard without calling register on it, and 2) the user never gets a binding to a variable of type Guard, but only of type &mut Guard, which prevents moving the Guard, hence fulfulling the requirement of a Guard being non-movable from Definition 4.4.

In addition to the guard! macro for declaring new Guards, CMR also defines the guard *procedure*, which loads an Atomic into a Guard.

Atomic

Atomic is mainly a wrapper around Rusts AtomicPtr, although the internals differ slightly. CMR defines its own type so that we can defide on the return types of certain functions. We include the definition of the struct, as well as cas, the compare-and-swap operation, in which we utilize some Traits from the Rust standard library to convert between types.

write a little here

Listing 5.6: Excerpt of Atomics definitions (Trait bounds omitted for brevity)

```
pub struct Atomic<T> {
    data: AtomicUsize,
    _marker: PhantomData<T>,
}
impl<T> Atomic<T> {
    pub fn cas<'a, A, B>(&self, a: A, b: B, ordering: Ordering)
     -> Result<A, NullablePtr<'a, T>> {
        let (old, new) = (raw(a), raw(b));
        let ret = self.data.compare_and_swap(old, new, ordering);
        if ret == old { Ok(A::try_from(NullablePtr::new(ret)).unwrap()) }
        else { Err(NullablePtr::new(ret)) }
    }
    ...
}
```

NullablePtr

NullablePtr is used as the canonical pointer type in CMR, and all pointer like types are converted to NullablePtr using the From Into traits, which handles coversion between types. For instance, we implement From<*const T> for NullablePtr<T>. This way we can write functions that are generic over all types of pointers, so that the user of CMR does not have to handle these conversions themselves. This is used in the functions for pointer tagging, as well as the cas in Listing 5.6 (the types A and B). Listing 5.7 shows some of the free functions for pointer tags that are generic over different pointer types.

Listing 5.7: Implementation of pointer tagging functions

```
pub fn tag<'a, P, T: 'a>(p: P) -> usize where P: Into<NullablePtr<'a, T>> {
    let n: NullablePtr<T> = p.into();
    n.0 & ones(TAG_BITS) }

pub fn with_tag<'a, P, T: 'a>(p: P, tag: usize) -> P
where P: Into<NullablePtr<'a, T>> + TryFrom<NullablePtr<'a, T>> {
    let p = p.into();
    let n = (p.0 & !(ones(TAG_BITS))) | tag;
    P::try_from(NullablePtr::new(n)).unwrap_or_else(|_e| panic!("failed conversion")) }
```

`ones(k)` returns the bitmask with the `k` lower bits set, and `TAG_BITS` is a predefined number of bits allowed to use for tagging for any pointer. We convert from `P` to `NullablePtr` with `.into()` (TA2). In `with_tag` we need to use `TryFrom`, which is a conversion trait that may fail. In CMR `Ptr<T>` implements `TryFrom<NullablePtr>`, where the conversion fails if the `NullablePtr` is null. We assert that this failure should never happen (TA9) with the rationale that if we converted some type `P` into a `NullablePtr` and changed its tag, we should be able to convert back to `P`, even though the conversion is not always possible in general.

Common Operations

In addition to the types member function, we have defined two important free functions: `guard` and `alloc`. `guard` takes a `&mut Guard` and `&Atomic`, `load`s the atomic, and protects the pointer read in the `Guard`, without a reclamation pass happening in between — this is ensured by the `atomic` call.

```
pub fn guard<'a, T>(guard: &'a mut Guard<T>, a: &Atomic<T>) -> NullablePtr<'a, T> {
    without_reclamation(|| { let p = unsafe { a.load(SeqCst) };
                             guard.inner = ptr::raw(p);
                             p }) }
```

`alloc` is similar to `Box::new`, as it just wraps `cmr::alloc::alloc` in addition to protecting the pointer returned in the `Guard` passed in.

```
A1 pub fn alloc<T: Trace>(guard: &mut Guard<T>, t: T) -> Ptr<T> {
A2     let ptr = alloc::alloc(t);
A3     guard.inner = ptr::addr(ptr);
A4     alloc::register(ptr);
A5     ptr }
```

Note that we do not need consider a reclamation pass happening concurrently in `alloc`, despite there being a window in between the allocation and the pointers protection in the `Guard`. This is because CMR only frees *registered* addresses. Since we protect the pointer in the `Guard` (A3) before registering it (A4), the pointer will be protected when it is registered, and hence is subject for decallocation.

08/05 15:42
Should talk somewhere about the registering process

5.5 Complications

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

10/05 18:30
SignalVec?

10/05 18:30
Initialization stuff?

something something pitfalls

5.5.1 Locks in libc

In order to protect programmers from deadlocks, POSIX defines a subset of functions as *async-signal safe*, meaning they are safe to call from a signal handler. Functions that are async-signal safe includes `time()`, `open()`, and `mkdir()`, but it does *not* include `malloc()`. As such, allocation in signal handlers is not safe, and is a source of deadlock bugs. This itself was not a large problem for CMR, as its signal handler did not require any allocation. However, as threads are frozen by the reclaimer in a signal handler, it is also not safe for the *reclaimer* to call `malloc`, despite not being in a signal handler itself. This is because some thread may be in the process of allocating memory, and have aquired a lock internal to libc, right before being signaled. The thread is still holding the libc lock and is frozen in its signal handler by the consolidator, which prevents *all threads*, even those oblivious to CMR, from allocating.

This problem is not solved properly by CMR, but its effects are mitigated by wrapping the general allocator in Rust to go through yet another lock, the `alloc_lock`, which can be aquired by the reclamating thread (this is why we have (R3) in Listing 5.1). This prevents most allocations of deadlocking, but not all. Rust uses `pthreads` internally for thread handling on Linux, which allocates internally, both in `spawn` and `join`. The former may be circumvented by aquiring the allocation lock before calling it, but this is no solution for the latter, since the thread may depend on allocating before exiting.

deactivate/reactivate around joins?

add example or something here

Chapter 6

Methodology

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

6.1 Development

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

6.2 Testing

Testing is an important part of sfotware development. While formal methods have not yet made its way into the software development insdustry, simpler and more heuristic methods, like unit testing and integration testing, are widespread. However, as Edsger Dijkstra famously said: “Testing shows the presence, not the absence of bugs”[4]; while testing is useful to improve the quality of software, it is far from sufficient.

In the world of concurrent programming this is even less so. Many bugs are manifested through unfortunate[1] thread execution interleavings done by the scheduler. We try to reveal these interleavings by repeatedly running tests until our confidence that no such interleavings exists is sufficiently high. In addition, we run tests with tools such as Valgrind[17] and our own sanitizer (Section 6.2.1). Tests were also ran with and without compiler optimizations, as these optimizations often reveal yet more bugs.

6.2.1 Sanitizer

To automate validation of pointer reads we made a compile time feature[2] called `sanitize` that tracked all allocations, frees, and pointer reads. Allocations and frees were tracked in two `HashMap`s, `ALLOCATIONS` and `FREES`. On each new allocation, we insert it into the `HashMap` while asserting that it was not there previously. We also remove it from the frees map, in case it had previously been allocated and freed. Since we are using a custom pointer type, `Ptr`, checking the validity on each pointer access is possible, as shown in Listing 6.1.

```
pub fn alloc<'a, T: Trace>(t: T) -> Ptr<'a, T> {
    let addr = B::into_raw(B::new(t)) as usize;
    #[cfg(feature = "sanitize")] {
        let mut a = ALLOCATIONS.lock().unwrap();
        assert!(a.insert(addr));
        let mut f = FREES.lock().unwrap();
        f.remove(&addr); }
    unsafe { Ptr::new(addr) } }
```

Listing 6.1: Verifying all pointer accesses with `sanitize`

```
impl<'a, T> Deref for Ptr<'a, T> {
    type Target = T;
    fn deref(&self) -> &T {
        #[cfg(feature = "sanitize")] {
            let a = ::alloc::ALLOCATIONS.lock().unwrap();
            if !a.contains(&addr(self)) {
                let was_freed = ::alloc::FREES.lock().unwrap().contains(&addr(self));
                panic!("{:x} is not valid. Was is freed? {}", self.data, was_freed); } }
        unsafe { &*(self.as_raw()) } } }
```

6.3 Benchmarking

The benchmarks are ran with 5 second trials, where a function is ran repeatedly for the duration with any specified number of threads. The number of executions is counted for each thread and the total operations per second is reported. All threads run the same code, but they may have different thread local data. This is useful when benchmarking `HashMap::insert`, so that the threads can inserts values with different keys.

[1] Some would call interleavings that reveal bugs fortunate
[2] *features* are similiar to `#ifdef`s in C and C++

There are a number of pitfalls when it comes to benchmarking code. We discuss a few of them; [3] is a good resource for experimental testing of data structures.

Initialization of data structures should not be done on a single core as this creates a strong skew of data locality for that core, and other cores will have reduced performance due to the data locality. This is especially important on systems with multiple CPU sockets.

6.3.1 Trench

In order to more effectively benchmark threaded applications, an open source benchmarking library called `trench`[16] was developed. The library handles thread management and state for the runs of the benchmark. Trench supports both mutable thread local state and immutable shared state between all threads. For CMR this is useful since we can put the data strucutres we want to benchmark in the immutable shared state, as neither of the operations we want to test are `&mut self` (see Section 3.5.1). The user specifies the function to be benchmarked, the number of threads, and the states, and the duration of the benchmark, and `trench` handles the rest. The number of runs of the function specified during the given duration is measured. Listing 6.2 shows the benchmark for `HashMap::insert`. `RandomSource` allows us to pregenerate random numbers that we can insert into the hashmap, such that the random number generation itself is thread local, and is not included in the benchmarking loop.

Listing 6.2: `Hashmap::insert` benchmark using `trench`

```rust
fn hashmap_insert(num_threads: usize) {
    fn func(state: &HmState, local: &mut RandomSource<u64>) {
        state.hashmap.insert(local.next(), 0); }
    let b = trench::TimedBench::<HmState, RandomSource<u64>>::with_threads(num_threads);
    b.with_local_state(|l| { cmr::thread_activate();
                             l.gen_n(10_000_000); });
    let res = b.run_for(duration(), func);
    b.with_local_state(|_| cmr::thread_deactivate());
    println!("cmr::HashMap\tinsert\t{} ops/sec", fmt_thousands_sep(res.ops_per_sec)); }
```

The `with_local_state` function runs the closure on each thread in parallel; this is used both for initializing the local state, and for thread local initialization and destruction. The global and local states, `HmState` and `RandomSource`, implements the trait `Default`, so that we do not have to initialize it ourselves.

Chapter 7

Usage of CMR

In this chapter we look at usage code for CMR. We have implemented four data structures: a stack, a queue, a list, and a hashmap. We want to look closer at how the abstractions that CMR provides are used, and how difficult they are to use.

29/05 08:36 Write more here

7.1 Lock-free Stack

We begin by looking at an implementation of a conurrent stack, which is arguably the simplest concurrent data strucutre. The definitions of the `Stack` and `Node` structs and the two most important operations on a stack, `push` and `pop`, is shown in Listing 7.1. We look at each one in turn. Construction of the stack is omitted for brevity.

Push

`push` allocates the stack node itself, so it takes the value we want to push onto the stack (`ST5`). We start out by declaring two `Guard`s (`ST6`): one for the new node we allocate, and one for the head of the stack. We must protect the head of the stack, since the node may be removed after we read its address, and we would have a dangling pointer. Next we allocate a new `node` (`ST7`), which is done outside the retry loop so that we only have to allocate one time per call to `push`. Now we enter the retry loop, which we repeat until we succeed in changing the top pointer of the stack to our new node. The top

06/05 19:14 After writing 2.4.1, revisit this

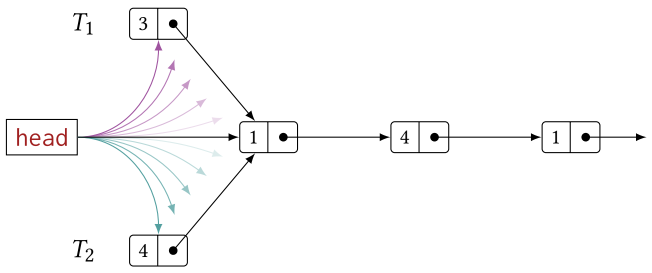


Figure 7.1: T_1 and T_2 both tries to swap the `head` pointer towards their node.

Listing 7.1: `Stack::push` and `Stack::pop`

```
 ST1 struct Stack<T> { top: SharedGuard<Node<T>>, }
 ST2 struct Node<T> { data: ManuallyDrop<T>, next: Atomic<Node<T>>, }
 ST3
 ST4 impl<T> Stack<T> {
 ST5     pub fn push(&self, t: T) {
 ST6         guards!(_new_top, _top);
 ST7         let mut new_top = cmr::alloc(_new_top, Node::new(t));
 ST8         loop { let top = cmr::guard(_top, &self.top);
 ST9                unsafe { new_top.get_mut().next = Atomic::new(top); }
ST10                if self.top.cas(top, new_top, SeqCst).is_ok() { break; } } }
ST11
ST12     pub fn pop(&self) -> Option<T> {
ST13         guards!(_top, _next);
ST14         loop { let top = cmr::guard(_top, &self.top).ptr()?;
ST15                let next = cmr::guard(_next, &top.next);
ST16                if self.top.cas(top, next, SeqCst).is_ok() {
ST17                    let node = unsafe { top.move_out() };
ST18                    return Some(node.data()); } } } }
```

node is read (ST8), and the `next` pointer of the new node is set to the head (ST9). If we succeed of chaning the `top` pointer of the stack to our new node, we break out of the loop and return (ST10). If not, we retry until we do.

06/05 19:14
write something on `unsafe` here

Pop

`pop` is similar to `push`. We declare two `Guard`s (ST13), but this time they are for the first and second node in the stack. We read the `top` pointer (ST14), and return from the function if it is `null` using the `?` Rust operator. We then read the next pointer of the node (ST15); here the `null` case is the same as the non-`null` case. We try to swap the head pointer from the first to the second node (ST16); if we fail we restart, and if we succeed we move out the `Node` from the `Guard`. This is an `unsafe` operation, as the type is copied out of its original place, effectively aliasing it. At last, the data is returned.

As an example of why reading and returning the node data is `unsafe` in the general case, consider two threads T_1 and T_2 using a `Stack<Box<T>>`. T_1 is looking at a node n, and T_2 is `pop`ping n from the stack, getting the `Box<T>` back from it. Now T_2 `drop`s the `Box`, which frees the pointer. If T_1 decides to look at the data in n, it will dereference a freed pointer, which is a use-after-free (Definition 4.1). Despite being `unsafe` in the general case, it is safe for the implementation of the stack as presented, since no operation on the stack looks at the data of a node, except in (ST17), where only one thread may be for any given node, since we succeed the `cas` operation.

06/05 19:14
fix this argument please

7.2 Lock-free Queue

06/05 19:14
refs here pls. Explain how it works?

The queue implemented is based on the well known Michael-Scott Queue from [10]. `push` is shown in Listing 7.2; `pop` is omitted due to its similarity with `Stack::pop`.

Listing 7.2: The push operation on a Michael-Scott Queue.

```
MS1  impl<T> MsQueue<T> {
MS2      pub fn push(&self, t: T) {
MS3          guards!(_new_node, _tail, _next);
MS4          let new_node = cmr::alloc(_new_node, Node::new(t));
MS5          loop { let tail = cmr::guard(_tail, &self.tail).ptr().unwrap();
MS6                 let next_ptr = &tail.next;
MS7                 let ptr = cmr::guard(_next, next_ptr);
MS8                 if ptr::addr(ptr) != 0 { let _ = self.tail.cas(tail, ptr, SeqCst); }
MS9                 else if next_ptr.cas(ptr::null(), new_node, SeqCst).is_ok() {
MS10                    let _ = self.tail.cas(tail, new_node, SeqCst);
MS11                    break; } } } }
```

We start out by declaring three Guards (MS3): one for the new node, one for the current tail, and one for the tails next node, which may be present. We load tail (MS5), and its next pointer (MS7). Since the Michael-Scott queue is always non-empty, we know that the head is non-null, and it is therefore safe to promote the NullablePtr to a Ptr using .unwrap(). If the next pointer is non-null the node we believed was the tail was not the tail after all. We try to swing tail from the node we read, to its next node (MS8) before restarting. If the tail was null we try to cas its next field from null to our new node (MS9). If we succeed, we cas the tail to our node and exit. If we fail, we restart. Note that we do not check the results of the the cas where we set the tail to the node we just inserted; if this operation fails, it just means that some other thread came along and noticed that tail was not the real tail, and cased it to the last node (MS8).

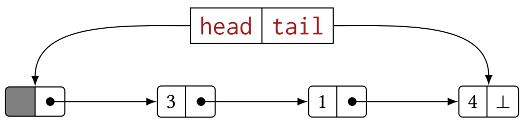


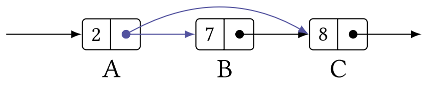
Figure 7.2: The Michael-Scott Queue. The first node in the queue is a sentinel node.

7.3 Lock-free List

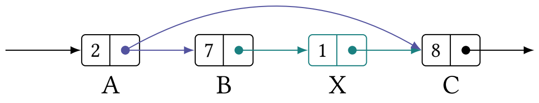
Michael introduced a concurrent List in [7]. The list is similar to the Stack from Section 7.1, but we support more operations than push and pop, including queries and removals, and insertions into the list. Definitions of List are very similar to that of the Stack and Queue, and is therefore omitted here.

07/05 22:22 insert more stuf here?

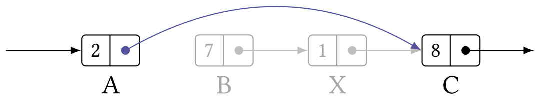
Having arbitrary insert and remove opens for a problem known as *double-remove*, shown in Fig. 7.3. Let there be two threads in the system T_1 and T_2, and let A, B, and C be three consecutive nodes in the list. If T_1 wants to remove the B node, there is a small window in which T_2 may insert a new node, X, between B and C. When T_1s cas operation succeeds — note that A.next was not touched by T_2 — it will accidently swing the pointer past the new node X without noticing. This is a variant of the ABA problem (Section 2.4.2).



(a) T_1 calls `cas(A.next, B, C)`.



(b) T_2 inserts X before the `cas` is ran.



(c) The `cas` succeeds. Both B and X are removed.

Figure 7.3: Double removal with `List::remove`.

A solution to this problem, as shown by Harris in [9], is to exploit memory alignment on modern CPU architectures: `struct`s are *aligned* in memory, meaning their address is a multiple of some power of two. This causes the least significant bits of their address to always be zero; bits that may be used for other purposes. We use the least significant bit in the `.next` field in a node for a *tag*, signaling that the node is logically deleted, and should not be acted upon. To see how this helps the problem as shown above, T_1 would start out the deletion process of B by calling `cas(B.next, C, with_tag(C, 1))`. Should this fail, T_1 can just read `B.next` again, and retry. When it succeeds, it may try to `cas A.next` over B to C. Now T_2 realizes that it should not insert X between B and C, since it reads the tag of B, realizing that it was deleted.

7.3.1 The Entry API

Many of the most interesting operations on the List involves iterating through the list. Due to the ownership and lifetime rules that Rust imposes, it may be tricky to implement typical iteration and juggle pointers around. For this reason, the API uses an indirection for iterating through the list: `Entry`.

An `Entry` is like a pointer into the list, which can `step()` to the next node, get a pointer to the `current()` node, `remove` the current node, insert a new node between (`insert_between`) two nodes, and find nodes which data satisfies arbitrary closures `Fn(T) -> bool`. Since there is some overhead in declaring a `Guard`, `Entry` contains two references to `Guard`s rather than the `Guard`s themselves. This makes construting a `Guard` nearly free. Another implication of this is that `Entry` is movable in memory (as `Guard` is not).

This indirection simplifies many operations, and we barely need to deal with lifetime and ownership issues, although it almost requires NLL (Section 3.6.1) to use.

Listing 7.3: Partial `Entry` API from the List implementation.

```
pub struct Entry<'a, T: 'a> {
    current: &'a mut cmr::guard::Guard<Node<T>>,
    previous: &'a mut cmr::guard::Guard<Node<T>>,
}
impl<'a, T> Entry<'a, T> {
    pub fn step(&mut self) -> Result<T>;
    pub fn current(&'a self) -> cmr::NullablePtr<'a, Node<T>>;
    pub fn previous(&'a self) -> cmr::NullablePtr<'a, Node<T>>;
    pub fn insert_between(&self, new_node: ptr::Ptr<Node<T>>) -> Result<T>;
    pub fn delete(&self) -> Result<T>;
    pub fn seek_with<F>(&mut self, f: F) -> Result<T> where F: Fn(&T) -> bool;
}
```

Listing 7.4: Implementation of `List::for_each` using the `Entry` API.

```
pub fn for_each<F: Fn(&T)>(&self, f: F) {
    guards!(_a, _b);
    let mut entry = self.entry(_a, _b);
    while let Some(ptr) = entry.current().ptr() {
        f(ptr.data());
        if entry.step().is_err() { break; } } }
```

7.4 Lock-free Hash Table

The hash table is a versatile and popular data structure. It is widely used due to its constant time operations, including queries, insertions, and removals.

Lock-free hash tables are interesting for the same reasons. Despite the interest, designing a concurrent hash table turns out to be a difficult problem. Blabl, resize Most hash tables are split up in *buckets*, such that the hash of the elements within a bucket share some property (eg. a common prefix). Increasing the number of buckets is knows as *resizing*, which makes sure that the number of elements in each bucket is limited; many algorithms and hash functions give bounds on the number of elements in each bucket.

7.4.1 Split-Ordered List

We start by describing the *Split-Ordered List*. Split-Ordered Lists were introduced in [14]. The nodes in the list is ordered by the *reverse hash* of the value in the node. In addition, the list contains *sentinel nodes*, which are the beginnings of the buckets in the hash table. By making the number of buckets $b = 2^k$ we can double b when the load factor is too high, and insert one more sentinel node between each of the nodes already present, effectively differentiating between one new bit of the reverse hash.

7.4.2 Hash Table

07/05 17:00 Clean up these sections. Maybe have only one section?

Using the Split-Ordered List we can implement a concurrent hash table by having an array of pointers to sentinel nodes, and a "size" of the bucket array. If a sentinel pointer is `null`, then the node is not yet in the table. When inserting a new element into the table, we first find the sentinel node that precedes the node we want to insert (the *parent*); this is known, since we know the ordering of the nodes in the list — the reverse hash. However, due to the resizing method, the parent may not have been inserted yet. If not, we can simply recurse on the insertion method, and insert the parent first. Then, we simply iterate through the list, and find the place in which our new node should be. Assuming a small load factor, this is a fast operation.

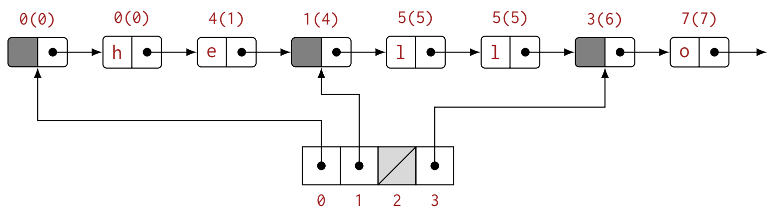


Figure 7.4: The Split-Ordered List. Node labels shows the `hash` and its reverse in parenthesis.

Fig. 7.4 shows the split-ordered list with a table size of 4. The nodes in the list are ordered by the reverse of their hash (shown in parenthesis). Given a node `n`, we find the sentinel node that should preceed it in the list by taking `hash(n) % table_size`. Note that this is not the reverse hash. For instance, inserting a node where `hash(n) == 7`, we look in bucket `7 % 4 == 3`, and iterate from sentinel node 3. Inserting a node where `hash(n) == 10`, we would get `bucket = 2`, which is `null`, so we need to insert the sentinel node first.

We look at *n* operations on the hash table: `contains`, `insert`, *and maybe more.*

7.4.3 Contains

Listing 7.5 shows the implementation of `HashMap::contains`. The implementation of utility functions such as `bucket_and_revhash` are omitted for brevity. We find the `parent` node (HC4), and use the `Entry` API from `List` (HC6) to look for the first node which hash and key is the same; if we encounter a node which hash is more than our node, we know that we have gone too far. `Entry::seek_with_opt` lets us break out of the search early by returning `None` (HC10). If we find a node with both the right hash and the right key, we return `Some(true)` from the closure (HC11), and `seek_with_opt` returns `Ok`. If we get back `Ok`, the search succeeded, so we return `true`, and `false` otherwise.

Listing 7.5: Implementation of `HashMap::contains`.

```
HC1 impl<K, V> HashMap<K, V> {
HC2     pub fn contains(&self, k: &K) -> bool {
HC3         let (bucket, rev_hash) = self.bucket_and_revhash(k);
HC4         let curr = self.get_or_insert_bucket(bucket);
```

```
        guards!(_curr, _prev);
        let mut entry = list::Entry::from_node_ptr(curr, _curr, _prev);
        entry.seek_with_opt(|data|
            Some(match data {
                &Entry::Value((h, ref key, _)) => {
                    if h > rev_hash { return None; }
                    else { h == rev_hash && k == key }
                }
                _ => false })
        ).is_ok() } }
```

7.4.4 Insert

HashMap::insert is more complicated, as there are multiple things that can go wrong, and that some operations require cleanup. Listing 7.6 shows the implementation of insert.

31/05 15:43 fix HI numbers here

Listing 7.6: Implementation of HashMap::insert.

```
HI1  impl<K, V> HashMap<K, V> {
HI2      pub fn insert(&self, k: K, v: V) {
HI3          let (bucket, rev_hash) = self.bucket_and_revhash(&k);
HI4          let curr: cmr::Ptr<_> = self.get_or_insert_bucket(bucket);
HI5          guards!(_new_node, _curr, _prev, _r1, _r2);
HI6          let node_data = Entry::Value((rev_hash, k, v));
HI7          let mut new_node = cmr::alloc(_new_node, list::Node::new(node_data));
HI8          'restart: loop {
HI9              let mut entry = list::Entry::from_node_ptr(curr, _curr, _prev);
HI10             let res = entry.seek_with(|e| match e {
HI11                 &Entry::Value((h, ref key, _)) => h >= rev_hash,
HI12                 &Entry::Sentinel(h) => h > rev_hash });
HI13             if let Err(list::Error::Empty) = res {
HI14                 ⟨End of list case⟩
HI15                 break 'restart; }
HI16             if res.is_err() { continue 'restart; }
HI17             if entry.insert_between(new_node).is_err() { continue 'restart; }
HI18             ⟨Remove other nodes with the same key⟩
HI19             break; }
HI20         self.increment_length(); } }
```

We start out by hashing the key, finding the reverse hash (HI3) and the bucket of the sentienl node, and a pointer to the node is acquired (HI4). We declare five (!) Guards (HI5), and alloc our new node (HI7). Next we make our entry from the sentinel (HI9), and find the correct place to put our new node (HI10). The new node is put before any other nodes with the same hash, but after the sentinel node, should their hashes be the same. We insert the new node in front of the old nodes so that other threads will see the most recently updated node first. The result of this operation has three cases: 1) we fail with Empty which means we got to the end of the list, and is handled by inserting the new node at the end of the list (HI14), 2) we fail with another failure case and restart (HI19), and 3) we succeed and actually insert our new node into the list (HI20), where we, again, restart upon failure.

After insertion we must check for other nodes with the same key, since there should only be one entry for any given key in the map (HI21). This is done by making a new `Entry` with the new node, `step`ping once, so that the `current` node is not our new node, and `delete()` any node that has the right key. When we hit a node which hash is more than our own, we are done (HI24). Since the map is generic over the key type `K`, comparing the type for equality might be expensive. Therefore, we check the hash before checking the key for equality (HI25).

`HashMap::insert` contains a single `unsafe` block: (HI14). This operation is `unsafe` since we are mutating `new_node`, which potentially could cause a race condition if the node was concurrently read by another thread. However, we know that it is not read by any other thread, since we have not been succesfull yet in inserting it into the list.

7.5 Crossbeam Integration

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Chapter 8

Results

In this chapter we look at the expermental results of the system as a whole. In Section 8.1 we look at the overhead of the operations done by CMR. In Section 8.2 we look at the performance of the data structuers implemented, with and without the overehead CMR, and compare them to alternatives in the Rust ecosystem, like Crossbeam[5] and data structures in the standard library wrapped in a `Mutex`.

8.1 Operations of CMR

The operations that CMR provides that are most interesting to look at is allocation (`cmr::alloc`) and guard initialization and destruction (`guard!`), as these operations are the only ones that have any significant overhead. Atomic loads pointer manipulations are mainly tricks of the type system to ensure the safey of the operations, and has no run-time overhead.

It is also interesting to look at the overhead of `fork()` calls, as this is an important part of the way CMR works.

The results are primarily from a . We have also ran the full benchmark suite on other machines. Interesting results are shown. The complete dataset of all results on all machines are listed in Appendix A

30/05 13:57 insert IST machine here

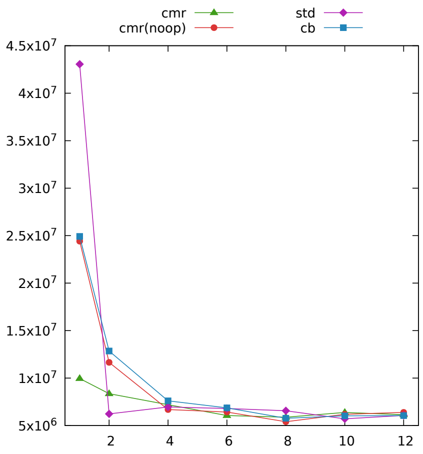
30/05 13:58 consider this after getting IST results

8.2 Data Structures

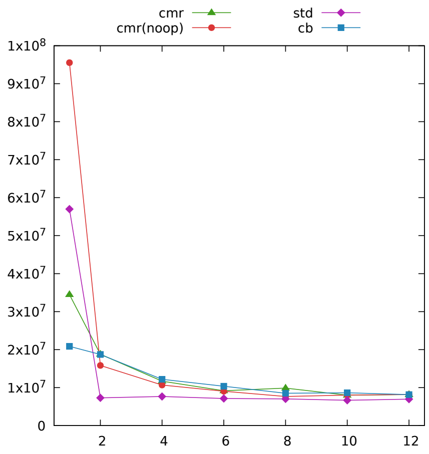
We look at all data structuers in the order that they were introduced in Chapter 7.

30/05 14:23 remove clearpages

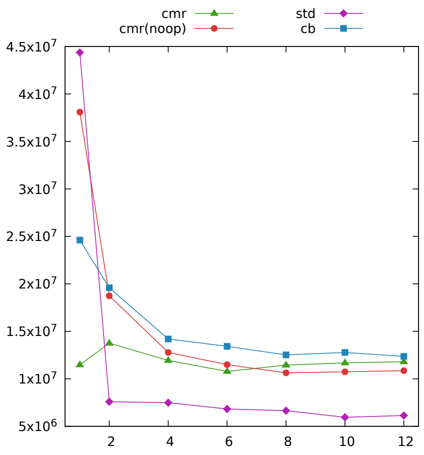
8.2.1 Stack



(a) `Queue::push`



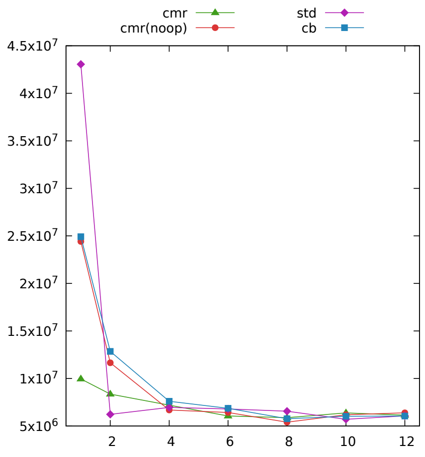
(b) `Queue::pop`



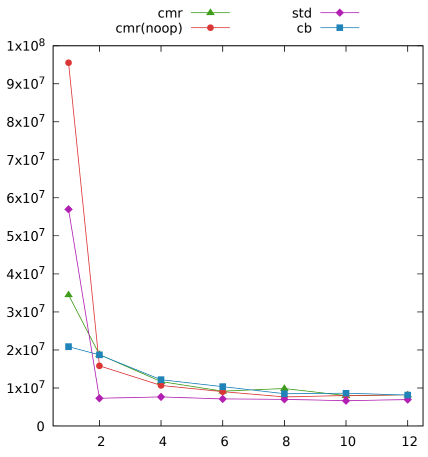
(c) `Queue` with 50% push, 50% pop

Figure 8.1: Queue performance on Gribb

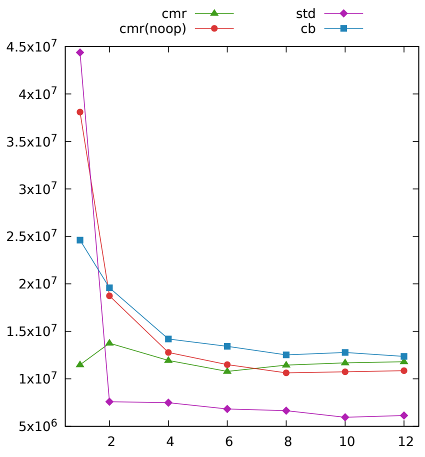
8.2.2 Queue



(a) `Queue::push`



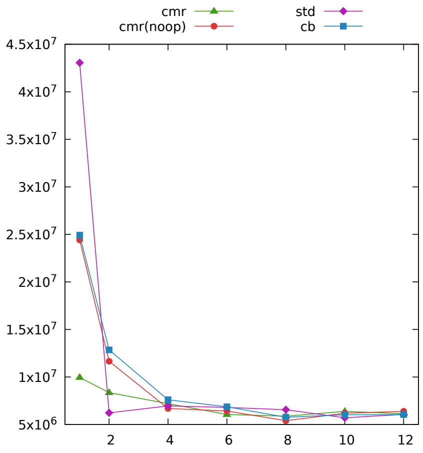
(b) `Queue::pop`



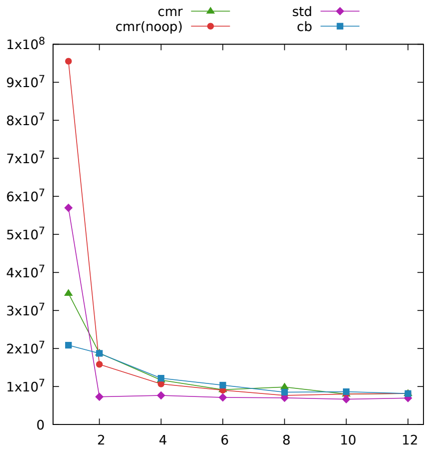
(c) `Queue` with 50% push, 50% pop

Figure 8.2: Queue performance on Gribb

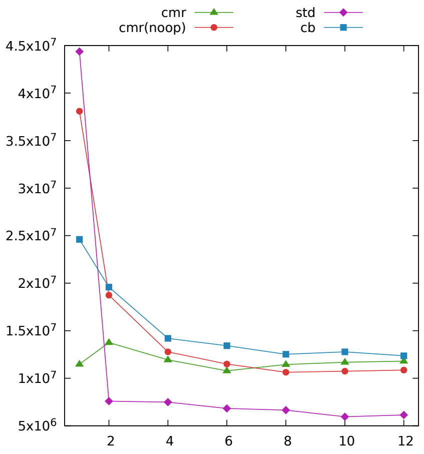
8.2.3 List



(a) Queue::push



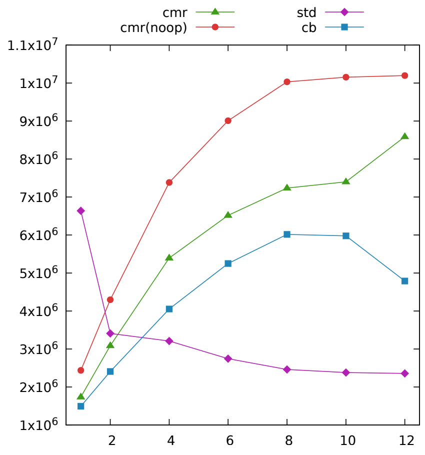
(b) Queue::pop



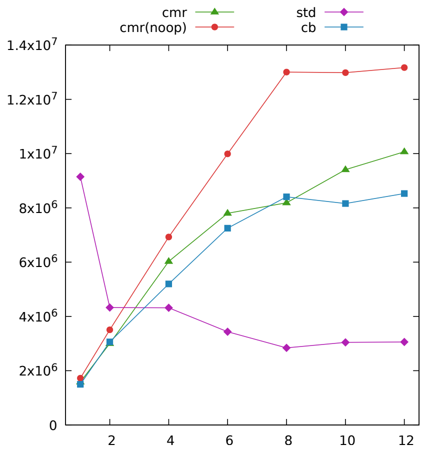
(c) Queue with 50% push, 50% pop

Figure 8.3: Queue performance on Gribb

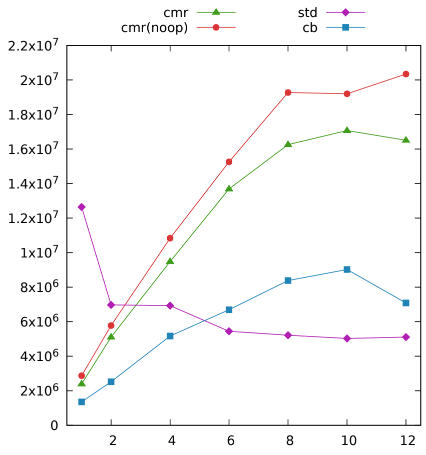
8.2.4 HashMap



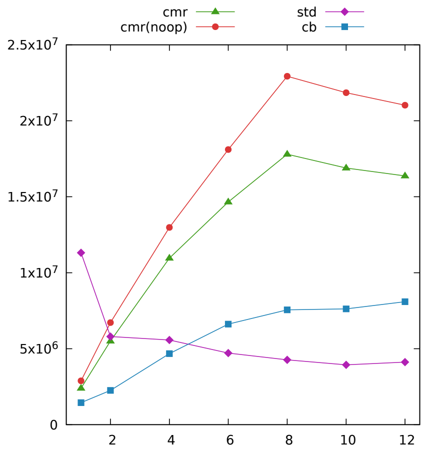
(a) `HashMap::insert`



(b) `HashMap::remove`



(c) `HashMap::contains`



(d) `HashMap` with 80% contains, 10% insert, 10% remove

Figure 8.4: HashMap performance on Gribb

8.3 Thirt Party Use

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Chapter 9
Conclusion

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

9.1 Is CMR Useful?

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

9.2 Alternatives

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit

blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

9.3 Closing Words

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Appendix A

Extended Results

Abbreviations

`pthreads` POSIX threads. 28, 29, 34

HPs Hazard Pointers. 22

IPC Inter-Process Communication. ix, 28

NLL Non-Lexical Lifetimes. 16, 42

RAII Resource Allocation Is Initialization. 15

RFC Request For Comments. 11

Bibliography

[1] Rust programs vs C gcc. https://benchmarksgame-team.pages.debian.net/benchmarksgame/faster/rust.html. [Online; accessed 31-May-2018].

[2] The LLVM Compiler Infrastructure. https://llvm.org, 2017.

[3] Brown, T. Good Data Structure Experiments are R.A.R.E. https://www.youtube.com/watch?v=x6HaBcRJHFY. [Online; accessed 15-May-2018].

[4] Buxton, J. N., and Randell, B. *Software Engineering Techniques: Report on a Conference Sponsored by the NATO Science Committee.* NATO Science Committee; available from Scientific Affairs Division, NATO, 1970.

[5] Crossbeam. https://github.com/crossbeam-rs/, 2017.

[6] GitHub. http://github.com/.

[7] Harris, T. L. A pragmatic implementation of non-blocking linked-lists. In *International Symposium on Distributed Computing* (2001), Springer, pp. 300–314.

[8] Jemalloc. http://jemalloc.net/, 2017.

[9] Michael, M. M. High performance dynamic lock-free hash tables and list-based sets. In *Proceedings of the fourteenth annual ACM symposium on Parallel algorithms and architectures* (2002), ACM, pp. 73–82.

[10] Michael, M. M., and Scott, M. L. Simple, fast, and practical non-blocking and blocking concurrent queue algorithms. In *Proceedings of the fifteenth annual ACM symposium on Principles of distributed computing* (1996), ACM, pp. 267–275.

[11] POSIX.1-2017. http://pubs.opengroup.org/onlinepubs/9699919799/.

[12] The Rust Programming Language. http://rust-lang.org/.

[13] The Rust Programming Language on GitHub. http://github.com/rust-lang/rust. [Online; accessed 31-May-2018].

[14] Shalev, O., and Shavit, N. Split-ordered lists: Lock-free extensible hash tables. *Journal of the ACM (JACM) 53*, 3 (2006), 379–405.

[15] The Rust Programming Language. https://doc.rust-lang.org/book/second-edition/.

[16] trench, the threaded benchmarking library. https://github.com/martinhath/trench, 2017.

[17] Valgrind. http://valgrind.org/, 2017. [Online; accessed 12-Nov-2017].