University of Dublin

TRINITY COLLEGE

An Experimental Comparison of Concurrent Data Structures

Mark Gibson

B.A.(Mod.) Computer Science

Final Year Project April 2014
Supervisor: Dr. David Gregg

School of Computer Science and Statistics

O'Reilly Institute, Trinity College, Dublin 2, Ireland

DECLARATION

I hereby declare that this project is entirely my own work and that it has not been submitted as an exercise for a degree at this or any other university

_____ _____

Mark Gibson Date

Acknowledgements:

Firstly, I would like to thank Dr. David Gregg for providing the inspiration for this project and giving me the opportunity to undertake it. This project would not have been possible without his input, support and optimism.

My second reader Melanie Bouroche for the time taken to review this project.

Fionnuala, Jo, Dave and Michael for their support through this project and my degree as a whole.

Contents

1 Introduction

1.1 My Work:

This purpose of this project is to determine and compare the differences between concurrent data structure implementations and investigate if the performance of these implementations is maintained across different architectures.

To do this I have implemented three concurrent data structures, a ring buffer, linked list and a hash table. Each data structure has several variations with regards to how they operate, such as the utilisation and placement of different pointers.

I have implemented each data structure with a mixture of locked and lock free algorithms. Among the locks used are a simple pthread mutex lock [reference], a compare-and-swap lock [reference] and a ticket lock [reference]. For the lockless algorithms I use the C++ 11 atomic library [reference] which contains the necessary atomic operations to implement lockless algorithms.

I have gathered data from these three data structures using varying thread counts and other variables, such as list or table size. The data structures are compared on a total of three different architectures to determine whether the performance of the algorithms is robust across architectures or not.

1.2 Context:

There has been much work done in the area of implementing concurrent data structures, with the area of concurrent programming expanding rapidly corresponding with the rise in multicore machines [reference]. However, despite all this research and work into concurrency, there is still a lack of data on the comparisons of different locking algorithms on these data structures. We are still unsure of the performance of different locking methods

and whether lockless algorithms are always preferred over locked alternatives. Hence, this project hopes to shed some light on the area by taking three concurrent data structures and testing them with several locking strategies to deduce if there is any correlation between certain algorithms and the relevant data structure's performance.

2 Background & Literature Review

2.1 What Motivated You?

I had been introduced to the idea of concurrency in the third year of my degree and it had piqued my interest. The solutions that concurrency provided for such computing problems as the memory and power wall to me seemed quite elegant. I saw the potential that this technology had and so I took another module based on concurrency in my final year so that I may learn about it in a greater depth. This proved useful to my understanding and so when it came to choosing a project for my final year, I decided to combine my new found interest in concurrency with data structures.

The problem that presents itself is that there seems to be little data comparing the performance of concurrent data structures using different locking algorithms. There is plenty of work done with regards to designing concurrent data structures [Moir et al. 2001] and implementing them [Herlihy 1993], but when it comes to practical implementation I have a difficult time in finding relevant work done in this area.

Hence, I am hoping to add to what little has been done in this area by performing my tests and analysis.

2.2 Locked & Lockless Programming

A lock in terms of computer science is a synchronisation mechanism which is used to control access to a resource in an environment that contains more than one thread of execution [reference]. Locks works by allowing only one thread to 'own' it and only the thread who owns the lock can access the resource [reference]. While this is a convenient way of ensuring mutual exclusion [reference], it does not scale with an increased amount of computing power or threads [reference], as only one thread can access the resource at any given time.

The term *starvation* when referring to multithreaded applications is the situation where a process is perpetually denied resources and as a result will never complete its assigned task, this can take place due to a poorly designed scheduler or in our case where several threads are competing for a lock. If a thread can never acquire a lock then it will never complete its task and so is subject to starvation [reference].

The term *lockless* in computer science when referring to a non-blocking algorithm [reference] equates to an algorithm where threads that are competing for a shared resource do not have their execution indefinitely postponed by mutual exclusion. These algorithms, while not using locks such as a mutex, still use atomic instructions as a means to protect a shared resource [reference].

The term *lock free* when discussing non-blocking algorithms means that individual threads are allowed to starve, but progress is guaranteed on a system wide level. At least one of the threads will make progress when a program's threads are run for sufficiently long.

The term *wait free* when discussing non-blocking algorithms represents the strongest non-blocking guarantee of progress. It guarantees both system wide progress and starvation freedom for the threads [reference]. Every operation has a bound on the number of steps the algorithm will take before the operation completes [reference]. It is for this reason that all wait free algorithms are also lock free [reference].

An *atomic instruction* is an operation that completes in a single step relative to other threads. When an atomic instruction is performed, much like a transaction in a database [reference], it will complete entirely in one step or will not do anything. Without this guarantee of completion, lockless programming would not be possible, as there would be no way, other than using a lock such as a pthread mutex, to protect a shared resource used by multiple threads [reference].

A *concurrent data structure* in computer science is a data structure that has been designed and implemented for use by multiple threads [reference]. As a result they are significantly more difficult to design and verify than sequential data structures, due to the asynchronous nature of threads. However, this added complexity can pay off as concurrent data structures can be very scalable if the shared resources of the data structure can be properly protected and utilised by the threads working on it [reference].

2.3 Sources Used

2.3.1 The Art of Multiprocessor Programming

When it came to researching what work had been done before now, I initially looked towards "The Art of Multiprocessor Programming" by Maurice Herlihy and Nir Shavit. It covers much of the current state of multiprocessor programming, detailing some of the various problems that are encountered with concurrent programming, such as the Producer-Consumer problem [reference] and the ABA problem [reference].

It then delves into the foundations of shared memory [reference] and the basics of multithreaded programming, detailing the spin lock and the issue of contention, where many threads vie for control of the lock [reference]. It then goes through several data structures, such as the linked list and hash table, describing the different aspects of design and implementation and the problems one can face when attempting to implement locked and lockless forms of these data structures.

Considering how closely this book follows my own work, I draw on it heavily throughout the design and implementation phases of my project.

2.3.2 Designing Concurrent Data Structures

"Designing Concurrent Data Structures" by Mark Moir and Nir Shavit, goes into depth on the processes required to successfully design a concurrent data structure, both in the general sense, talking about issues like blocking and non-blocking techniques [reference], performance and verification techniques [reference] while also going into detail for a range of specific data structures, like *stacks*, *queues*, *linked lists* and *hash tables*.

2.3.3 Implementing Concurrent Data Objects

"A Methodology for Implementing Highly Concurrent Data Objects" by Maurice Herlihy goes through the process of implementing a concurrent data structure, highlighting the issues with the conventional techniques of relying on critical sections [reference]. Instead he suggests using a lockless approach, going into detail on the differences between lock-free and wait-free approaches.

2.3.4 Experimental Analysis of Algorithms

"A Theoretician's Guide to the Experimental Analysis of Algorithms" by David S. Johnson discusses the issues that can arise when attempting to analyse algorithms experimentally, where he goes over several principles which he feels are essential to properly and accurately analysing algorithms ranging from using efficient implementations [reference] to ensuring comparability [reference]. In addition, he also goes over ideas and techniques of presenting data [reference] which I found to be most interesting.

2.3.5 A Lock-Free, Cache Efficient Shared Ring Buffer

"A Lock Free, Cache-Efficient Shared Ring Buffer for Multi-Core Architectures" by Patrick P. C. Lee et al. goes into great depth and detail for designing and implementing a high performance, concurrent ring buffer by attempting to optimise cache locality when it comes to accessing control variables used for thread synchronisation.

2.3.6 Resizable Scalable Concurrent Hash Tables

"Resizable, Scalable, Concurrent Hash Tables" by Josh Triplett et al. focuses on presenting algorithms for both shrinking and expanding a hash table while at the same time retaining wait-free concurrency.

3 Method: (Fat)

3.1 What do you have to do?

My work is as follows; firstly, I design and implement the three concurrent data structures. This involves adding both locked and lockless modes of operation to the data structures to allow them to be used by multiple threads.

Secondly, I run these data structures and gather data on their performance. This data is based on the number of iterations performed by each program per second, the number of threads running concurrently and size of the data structure in question. I run these data structures on one or two additional machines to investigate if the data structure's performance carries over to other architectures.

Finally I gather this data together and analyse it for anything of interest. To aid in the collection and analysis of this data as accurately as possible I use tools such as Perf [Perf Wiki. Available: https://perf.wiki.kernel.org/index.php/Main_Page]. Perf measures hardware performance counters [reference] and records such things as idle CPU cycles, cache misses and branches taken. I use this data to analyse the performance data I gather to explain why certain locks outperform others for example.

3.2 Approach

My approach to this project is a modular one. I select a data structure I am interested in. I then research the chosen data structure and investigate what work has been done on designing and implementing the data structure concurrently. After this, I implement the data structure before gathering data from it. I graph and analyse the data I have gathered and generate several conclusions about the data structure and the relevant locking algorithms. Once this is done I choose another data structure and so the process continues as I build my project up piece by piece.

3.2.1 Recap of locked & lockless

As mentioned previously in the 'Background' section of the project, Locked algorithms use mutexes or other such constructs to provide a lock. Threads must acquire this lock to enter the critical section. Once a thread finishes in the critical section it releases the lock allowing another thread to enter. All threads without the lock are blocked, forming a bottleneck of execution [reference]. I am implementing my locked variations as follows: I implement different locks such as pthread mutex, test-and-set etc. by using the pre-processor to define variables which impact the path of execution [reference]. For example, when I implement the pthread mutex lock I use the –D option on the g++ command line to define a macro. In the case of the pthread mutex lock this macros is "LOCKED" just as the test-and-set lock's macro is "TAS".

Lockless algorithms are defined by their use of atomic instructions to ensure thread-safe execution such as compare-and-swap which allows an object to atomically check if it contains a value and if so to swap it out for another value [reference]. Lockless algorithms do not use locks and so system wide throughput is guaranteed. To ensure that the locked and lockless algorithms are equal I wrap each atomic instruction which can fail in a while loop so that if it fails then it is forced to try again. I do this because with the locked variations, if a thread is trying to add an item to a linked list for example it will always succeed unless the list is full. However, if a thread attempts to do this with an atomic instruction and it fails then the node may not be added and so the work done by the two threads would not be equal.

3.2.2 List of locked modes

Below is a list of the different locks I use throughout the project. Each heading gives the name of the lock and the macro that I use to define it is given in brackets.

3.2.2.1 Pthread Mutex Lock (LOCKED)

This lock is composed of a pthread mutex [reference]. I choose it as I think that it is a simple lock to implement. In my opinion it also provides an excellent baseline for me to test other locks against due to its simplicity.

3.2.2.2 Test-and-test-and-set lock (TTAS)

The *test-and-test-and-set* lock [Herlihy et al, 2008, pg. 144] lock works by using the C++ 11 atomic exchange instruction [Atomic Operations Library. Available: http://en.cppreference.com/w/cpp/atomic] to atomically set and unset a lock. Each thread repeatedly checks the lock, if they find that it is equal to one then they sleep for a specified amount of time using the sleep instruction [reference]. After they wake up, the thread then checks the lock again to see if it is equal to one, if so then it starts the loop again, if not then it sets the lock to one, acquiring it, and enters the critical section [Herlihy et al, 2008, pg.22]. Upon finishing, the thread then sets the lock to zero, releasing it and the process continues

on. I implement this algorithm as it is an efficient implementation of a spinlock as the sleep instruction stops the cpu traffic from becoming overwhelming [Herlihy et al, 2008, pg.147].

3.2.2.3 Test-and-test-and-set-no-pause lock (TTASNP)

This locked mode is the *test-and-test-and-set* lock but without the sleep instruction after the second while loop. I added this as I am interested in the effect that the sleep instruction has on the performance of the lock when compared to the normal *test-and-test-and-set* lock.

3.2.2.4 Test-and-test-and-set-relax lock (TTAS_RELAX)

This is near identical to the normal *test-and-test-and-set* lock but with one difference, the sleep instruction is replaced by the intrinsic _mm_pause() which is designed to reduce the performance impact that repeated thread polling can have on bus traffic and the CPU's pipeline [reference]. I add this variation as, like with the *test-and-test-and-set-no-pause* lock, I am curious as to how the change affects the lock's performance and if the intrinsic gives this mode an advantage over the sleep instruction.

3.2.2.5 Test-and-set lock (TAS)

I implement a *test-and-set* lock because it is somewhat less sophisticated when compared to the *test-and-test-and-set* lock [Herlihy et al, 2008, pg. 144]. It is more basic because, while the regular *test-and-test-and-set* lock tells a thread to sleep after it has failed to acquire the lock; the *test-and-set* lock does no such thing and simply allows the thread to continue polling. This can lead to a dramatic increase in the amount of bus traffic between the CPUs in the machine and therefore can result in a loss in performance when compared to the *test-and-test-and-set* lock [Herlihy et al, 2008, pg.145].

3.2.2.6 Test-and-set-with-pause lock (TASWP)

Like with the *test-and-test-and-set* lock, I add a sleep instruction to the *test-and-set* lock to investigate what, if any difference it has on the lock's performance.

3.2.2.7 Test-and-set-relax lock (TAS_RELAX)

This lock is identical to the *test-and-set-with-pause* lock, except that instead of a sleep instruction it used the intrinsic _mm_pause(). I add this lock as I am curious as to how this lock compares to the *test-and-set* lock in terms of performance.

3.2.2.8 Compare-and-swap lock (CASLOCK)

The next lock I implement is a lock based on the atomic instruction, *compare-and-swap* which takes an object and attempts to change the value it has. If the object contains an expected value, then this value is replaced with the new value, else the object remains unchanged [Herlihy et al, 2008, pg.113]. This can then be placed within a loop, where threads continuously poll until one of them acquires the lock successfully and breaks free into the critical section. This can create a lot of bus traffic however, similar to that of the *test-and-set* lock and so I add an exponential back off, similar in style to the *test-and-test-and-set* lock, where a thread, upon failing to acquire the lock sleeps, but with each failed attempt, sleeps for a progressively longer time up to a defined maximum.

3.2.2.9 Compare-and-swap-no-delay lock (CASLOCKND)

This lock is the same as the *compare-and-swap* lock but where it has an exponential back-off to try and reduce bus traffic this lock does not have that. As with previous lock variations, I am interested to see how the lack of a back-off impacts the performance of this lock compared to the regular *compare-and-swap* lock.

3.2.2.10 Compare-and-swap-relax lock (CASLOCK_RELAX)

This lock takes the exponential back-off present in the regular *compare-and-swap* lock and replaces it with the intrinsic _mm_pause(). This is done to compare the variations of the *compare-and-swap* lock and see how they perform compared to one another.

3.2.2.11 Ticket lock (TICKET)

The final type of lock I add is a ticket lock, where each thread is given a ticket, and they are allowed to enter the critical section whenever their ticket is being served [Herlihy et al, 2008, pg.32]. This lock performs very poorly once the number of threads exceeds the number of CPU cores, as due to the queue like nature of the threads when using the ticket lock, if a thread is de-scheduled as it is in the critical section then the entire queue is held up as a result, leading to a significant drop in performance [reference]. As with the *test-and-test-and-set* lock, if a thread polls and finds that it is not its turn in the queue yet, it sleeps, where the amount of time sleeping is proportional to how far back in the queue the thread is, so if the thread is relatively close to the top of the queue it will sleep for less than if it was near the bottom of the queue.

3.2.2.12 Ticket-relax lock(TICKET_RELAX)

As with the previous locks, I compare the impact of the sleep instruction on the ticket lock by replacing it with the intrinsic _mm_pause and comparing the two with regard to performance.

3.3 Data Structures

3.3.1 Ring Buffer

For my first data structure I decided to go for a circular FIFO queue. I chose this due to its relative simplicity when compared to other data structures and I felt that it would give me an opportunity to get to grips with the atomic libraries I would be working with, as well as give me a chance to finalise how I will be collecting data from the data structures.

The previous implementation I decide to base mine on is located in “Designing Concurrent Data Structures” by Mark Moir and Nir Shavit. In it they describe the design for a concurrent queue which utilises a head and tail pointer to allow for parallel execution. In it they use a dummy node to prevent deadlock. My locked implementation differs as I only use one lock to control the front and back of the queue. I do this as I want my implementation to be even simpler. The reason for this is that the ring buffer is more of a testing stage where I can easily implement my different locking methods and test them out. At the end of my project if I have enough time I will come back to the ring buffer and implement a more advanced locking algorithm but for now this I what I need.

My lockless implementation differs as I do not use a dummy node, but instead the producer and consumer threads look ahead to see if the next node is being used.

At this point in the project I consider implementing assembly versions of the locks I have. http://locklessinc.com has several implementations in assembly such as a *ticket* lock and *test-and-set* lock. I attempt to implement the assembly locks, which is successful. I now compare the assembly locks against the C++ locks I already have. I discover that the difference in performance is negligible and so I decide to stay with my C++ implementations as I feel more comfortable using and modifying them.

3.3.1.1 Locked

For the locked variation I decided to go for a simple locking strategy where if a thread wished to interact with the buffer that it would acquire a lock, perform its operation and release the lock. This approach would only allow one thread to access the buffer at any one time and so would hopefully provide a nice contrast to the lockless implementation.

While implementing the different locked modes I came across the idea of implementing the locks in assembly, something which had already been done for some of the locks [Spinlocks and Read Write Locks. Available: http://locklessinc.com/articles/locks/]I decided to compare the performance of some of the locks I had already written to their assembly counterparts. If it was the case that the assembly implementations proved to have an advantage over the C++ versions then I would switch to them in order to procure more accurate results. Hence, I integrated them into the buffer and compared them to their C++ implementations to try and identify a performance difference. After comparing the locks, I found the difference in performance to be negligible between them and so decided to stick with the C++ implementation of the locks, as I found them easier to work with.

3.3.1.2 Lockless

For my lockless implementation of the ring buffer, I decided to implement a single producer – single consumer model. To push, the front of the buffer is taken and the index after it is examined. If the back of the buffer is not pointing there, then an item is pushed to the front of the buffer, and the index after it becomes the new front. Alternatively, to pop, the back of the buffer checks that it does not share the current index with the front of the buffer, and only then will it remove an item from the buffer.

I found this to be a good introduction to the C++11 atomic library as I was able to get to grips with declaring atomic variables and calling the library's functions, such as std::atomic_fetch_add which atomically increments a value by a given amount [reference].

3.3.2 Linked List

For my next data structure, I implement a singly linked list. I choose a linked list because I have experience with implementing and testing linked lists both sequentially and concurrently.

The implementation I choose to base mine on is found in "The Art of Multiprocessor Programming" by Herlihy and Shavit. It is simple and it follows conventional designs which I am already familiar as mentioned previously. My lockless implementation does diverge slightly however, while Herlihy and Shavit use a *find* function to obtain the necessary details to add or remove a node I instead choose to implement this step as part of the *add* or *remove* functions due to problems I have encountered in the past with using pthreads and calling nested functions.

3.3.2.1 Singly Linked List

This variation of the linked list contains one class, the Node class which is used to make up the linked list. This class had two attributes and a constructor function. The first attribute, key, represents the value assigned to the node. The second attribute, next represents a pointer of type Node which is used to point to the next node in the linked list. Finally, the constructor takes two parameters, a value and a pointer of type Node and assigns them to their respective attributes within the node. This variation is implemented in such a way as to be ordered, so that the smallest values are at the head of the list and that there are no duplicate values in the list.

The head of the list, a pointer of type Node, is not part of any class as I decided to not add a List class for this variation as I encountered problems with calling the pthreads.

This variation contains three functions, Add, Remove and printList. Add works by randomly generating a value using the rand() function [reference]. It then creates a node using this key and attempts to add it to the list of nodes. To begin, it gets the current value of the head variable. If the head is equal to NULL then a list does not yet exist, so it sets up the list. If the list already exists but the node that has just been generated has a smaller value than the one at the head of the list, then the new node is inserted in front of the head of the list and the head is changed to the new node. If the list exists and the node to be added is not smaller than the head of the list then the list is traversed by getting a copy of the head pointer and repeatedly assigning the value of each node's next pointer. In this sense it can move down the list, checking the values of each node as it goes. If it finds a node that is larger than the new node in terms of key value then it inserts the new node before the larger node. It does this by pointing the new node's pointer to the larger node and by getting the node previous to the larger node and assigning its next pointer to the new node. If the end of the list is reached, marked by a node's next pointer being equal to NULL then the new node is simply added onto the end of the list, by assigning the next pointer of the last node in the list to the new node, making it the last node in the list.

The Remove function works in that it first generates a random number which will act as the value that it will search for and try to remove from the list. Firstly it takes a copy of the head of the list and checks if it is equal to NULL. If it is then there are no nodes in the list to remove. Alternatively if the key of the head of the list is equal to the key that the function is searching for then it will point the head to the next node in the list and remove the now isolated node. If neither of these cases is true then the list is traversed until either the node is found or the end of the list is reached. If the node is found in the list then the next pointer of the node before it is changed so that it points to the node that the node to be deleted points to, effectively removing that node from the list.

The printList function is relatively simple compared to the Add and Remove functions. It simply takes a copy of the head of the list and traverses the list until it reaches the end. For each node it prints out their key value followed by a comma.

3.3.2.1.1 Locked

The locked version of this variation would be similar to the locked version of the ring buffer I implemented, where any attempt to act on the list would require a thread to acquire the lock, which it would then release once it had completed its work. Since this was a locked variation I did not have to declare the head of the list as an atomic variable, so I was able to simply declare it as volatile which prevents the compiler from optimising any code that it is a part of

[reference]. I declared the key attribute of the Node class to be volatile for the same reasons, along with any function level variables that dealt with the head or Node.

For the Add function I added in all the different locking modes to acquire the lock before the key value was randomly generated and added in the unlocking code at the end of the function, ensuring that only one thread could access the body of the function at any one time.

The same was done for the Remove function, the acquiring and releasing code was added before the key generation and after the body of the function respectively.

The printList function did not need to have any locking code added as it only called in the main function once the threads had finished their work and had been terminated.

3.3.2.1.1 Lockless

I decided to base my lockless implementation of this variation of the linked list on the atomic instruction 'compare_exchange' and its associated functions from the C++ 11 atomic library [reference]. To do this I would need to declare at least one atomic variable to call the necessary functions so I chose the head pointer of type Node. I did this because having an atomic head pointer would allow me to atomically change the head of the list. For this implementation I decided to remove the volatile keywords from the code and see if it made a difference to the validity or the performance of the data structure.

For the Add function, the code was somewhat smaller in size than the locked version as I did not need to add the different locking modes. Instead, the head is atomically loaded into a variable which is then checked to see if it is equal to NULL. If so then the atomic head pointer is changed from NULL to the new node that was created beforehand. This is done using the std::atomic_compare_exchange_weak function which acts as an atomic compare-and-swap instruction [reference], else if the head needed to be changed to another node than the atomic function would be called again, instead swapping the value of head from the old node to the new node.

It was at this point that I came across a point of interest in the code. I was unsure how to proceed with writing the code for atomically traversing the list so I decided to implement it serially and see what happened. I then ran the code several times and to my surprise the list it created was ordered with no duplicates and appeared to work locklessly for all intents and purposes. I repeated the procedure for the Remove function which was designed identically to the Add function, with atomic instructions for dealing with the head but serial code for dealing with list traversal and the results were the same.

To try force an error from my implementation I changed the maximum list size to five and ran it. Such a small list should have encountered a lot of contention considering the number of threads acting on it and yet no errors were found in the lists that were produced.

3.3.2.2 Doubly Linked Buffer

3.3.2.2.1 Locked

After implementing the singly linked list and observing some of the data that was being gathered I saw that once the list started to get long, past 1,000 nodes in length, the performance dropped off significantly. I concluded that it was due to the time being spent by the threads traversing the list looking for an insertion point or a node to delete.

I felt that this was not optimal, as the size of the list was interfering with the comparison of the locking algorithms. Hence, I decided to remove the traversal issue all together and implemented a multi-consumer, multi-producer linked list buffer. This worked by always adding and removing from the head and tail respectively. There was no traversal of the list necessary and while this did mean that the list would no longer be ordered or free from duplicates, it was my opinion that this would provide clearer data from the locking algorithms.

To implement this I added a tail variable of type Node * which I declared using the volatile keyword, similar to the head variable. The tail would be used by having it point to the end of the list, recording where the end of the list and giving a location for the threads to remove nodes from. However, to implement this I realised that I would need to add a second pointer to the Node class, prev, as if the tail was pointing to the end of the list then whenever a node was removed the tail would need some way of then pointing to the previous node in the list.

In one sense this simplified the implementation as the code required for traversing the list was no longer required; all that was needed was code to set up the list if no node existed and to add/remove from the head and tail respectively.

3.3.2.2.2 Lockless

To implement the lockless version of the doubly linked buffer I started off by declaring the new tail pointer as an atomic object. This would allow me to atomically remove objects from the end of the list as the atomic head pointer allowed me to add things onto the front of the list.

Adding objects involved generating the node to be added then atomically switching the head pointer from what it was pointing at to the new node being added. Since the list no longer had to be traversed, the process of adding a node lucklessly became much simpler.

Removing a node was much the same as adding a node but in reverse, where the tail pointer was atomically switched to point to the previous node in the list using the old tail's prev pointer to become the new tail of the list. The old tail was then discarded.

3.3.2.3 Singly Linked Buffer

It was only after I had finished implementing the doubly linked buffer that I realised that I did not need the second pointer for each node if I simply rearranged the placement of the head and tail pointers. If I swapped the head and tail pointers around then I would again only need one pointer per node to implement the data structure. It works as follows: the tail pointer would keep track of the oldest node in the list. Whenever a new node was added, the last node to be added would then be pointed to this node and the head pointer would move to the new head. It was in essence, flipping my initial implementation around but that small change reduced the complexity and size of the data structure as now, each node again only needed to store one pointer and all the code that was added to deal with the second pointer could be removed.

In terms of implementation it was very similar to the doubly linked buffer with the only real differences being that there were no longer any references to a Node's prev pointer as that had been removed and the references to the head and tail would be mixed up as they had switched position and function with this latest implementation. This was the case with both the locked and lockless variations of the data structure.

3.3.3 Hash Table

I choose a hash table as my third and final data structure to implement because it is a data structure that I have always had an interest in. In addition, I will be able to utilise my work on linked lists by using them to represent the buckets in my hash table.

As to which implementation I am to base mine on, I again choose one from "Designing Concurrent Data Structures" by Moir and Shavit. In it they describe the design of a concurrent hash table which uses lockless linked lists as buckets, exactly how I wanted to design mine. I also draw from "Reizable, Scalable, Concurrent Hash Tables" by Triplett et al when I implemented the resize function.

3.3.3.1 Locked

For the locked version of the hash table, I decided to have two implementations. The first implementation involved locking the entire hash table with a lock whenever a thread wanted to interact with the table. The second implementation differed from the first in that there was no global lock, but instead each bucket had its own lock. So whenever a thread wished to interact with a specific bucket, it would obtain the bucket's lock and perform its work, in this way it allowed for the absence of a global lock and instead had a more modular approach which I would then compare to the first locked implementation. To ensure that each iteration of the program is equal I create the hash table at the start of each iteration and delete it at the end so that each iteration works with the same data structure.

3.3.3.1.1 Global Lock

The premise for the globally locked hash table was simple, I wanted a baseline to compare my other two implementations on, the lockless and lock per bucket variations. In addition, I felt that it would be useful to get the add and remove functions working and tested in this implementation before moving onto more advanced variations.

As this was a baseline implementation, I decided to go for a very basic locking strategy, where a lock was acquired before a thread interacted with the table at all, and that the lock was global, in that only one thread could interact with the hash table at any given time, any other thread that attempted to interact with the table would be blocked.

3.3.3.1.2 Lock Per Bucket

This variation of my locked implementation of the hash table would be different in the sense that instead of threads acquiring a global lock, where only one thread would be able to access the table at any one time, each list in the table, or bucket, would have its own lock. In this way, multiple threads could work on the hash table at any given time and that they would acquire a lock for the bucket they were about to interact with so a thread would only be blocked if it attempted to interact with a bucket that another thread was already interacting with.

I felt that this implementation was more complex than the globally locked variety, I ran into some trouble when I attempted to implement the rest of the locking modes besides the basic pthread mutex lock, though I discovered that it was because I had mixed up a reference to one bucket's lock with another. After I had corrected this I was able to implement the rest of the locked modes, TAS, CAS, TICKET etc with no further delays.

3.3.3.2 Lockless

For designing the lockless hash table I made the following decisions based on research done with regards to lockless hash tables; it would be a closed addressing hash table, each index in the table would point to a linked list, so any collisions would result in a node being added onto the relevant list. Finally, it would have a coarse-grained resize function, which involved transferring the lists or buckets to a new, larger table [reference].

To represent the buckets I decided to use the lockless linked list I had already implemented, as I had already tested it when I was collecting the data from it and it would save me time. I decided to go with my FIFO buffer implementation of the linked list to eliminate traversing the buckets as an issue. I gave each bucket two atomic variables, a head and tail pointer, which would reduce the time spent adding/removing nodes and would ensure that I could do it atomically through the use of the C++11 atomic library. This would be the only use of the atomic library; the hash table itself did not have any atomic variables.

After I had implemented the data structure I ran into two points of interest. The first was that as the program ran, it would sometimes post extremely low results for one of the iterations, usually the iteration using four threads in total. To try and discern the cause I added in a counter that tracked the failure rate of the atomic instructions in both the add and remove functions, but this proved to not be the cause of the problem as the resulting values I was getting were both quite low, no more than fifty failures per iteration and these did not correlate to the drop in performance I was observing. I decided to put it aside for a while, with the intention of returning and utilising the tool perf to try and find the cause of the performance drop.

The second point of interest I encountered was that the program occasionally caused a segmentation fault while it was running at high thread counts, around 32 threads or more, though sometimes it occurred at lower counts such as 8. As the problem's frequency seemed to increase at higher counts my first thought was that it might be a contention issue. After reviewing the code, I noticed that I was accessing the hash table a lot during both the add and remove function calls in the form of "htable->table[hash]". I believed that this may be the cause of the segmentation faults, as if a thread was halfway through an add, another thread may change the value of the hash among other things, leading to a segmentation fault. I tried to solve this by instead passing the bucket reference to a variable, tmpList which I would then use in the computation. In addition, I added several more checks into my code, checking that tmpList still pointed to the place it was supposed to and that if it was not then abort the operation and try again. To test to see if the problem had been fixed by my changes I set it to run twenty times, one after another, with the intention that if a segmentation fault would appear, indicating that the problem had not been fixed , that it would in these conditions. Luckily, this was not the case and my implementation seemed to be working correctly.

3.3.3.3. Resizing

To keep search times constant, I had to add in functionality to allow my hash table to resize itself when buckets got too full [reference]. I decided to implement a locked resize function first, which involved going through each bucket and rehashing each key. Then the key would be transferred to the new table, based on its new hash. I was able to write this part of the implementation serially, as it is only called inside the add function, where at which point a lock will have already been obtained, making the need for additional locks irrelevant.

For my lockless implementation I investigated several potential methods, one of which involved leaving the keys where they were and forming new lists from them by dynamically creating each bucket [reference, concurrent hash table]. Another option from the same paper was to resize the table in place, where the current table was made bigger and the keys rehashed. Yet another option was to incrementally resize the table, where all adds started to add to another table, with remove and contain calls checking both tables and only switching to the new table when all the keys had been transferred from the old [reference]. I decided to try and implement the first solution and see how I got on. I immediately ran into problems with segmentation faults as I was unable to implement a necessary amount of atomicity to stop the threads from interfering with each other. This problem persisted for the two other solutions I attempted, each was plagued by segmentation faults which I was unable to get rid of. In the end I had to settle for using a lock, similar to my locked implementation, where only one thread was allowed access to resize the table.

To compensate for my inability to implement a lockless resize function, I planned to test my implementations with a large initial table size. I hoped that this would minimise the need for the table to resize and so have the smallest impact on the performance, allowing me to compare the locked and lockless algorithms almost purely based on what I had written already, the add and remove functions.

3.3.3.4 Contains Function

Before I began testing my hash table I decided that I wanted it to replicate a real world hash table as closely as possible. To do this I would need to add in a contains function, a function that took key and searched for it in the hash table [reference]. I would need to implement this functionality in all three of my hash table variations. The implementation itself was relatively easy, I randomly produced a key, got its hash and then retrieved the bucket associated with that hash. Once I had that I then iterated through the bucket until I had either found the key or I reached the end of the bucket.

3.3.3.5 Tracking Search Results

As a means to record positive and negative hash table searches I added in two variables, pSearches and nSearches to represent the total number of positive and negative searches each time the program ran. I did this because I planned to utilise these when I was testing the table to see if there were any correlations between the number of successful/unsuccessful searches and the table performance

3.3.3.6 Choose function

With the addition of the contains function in my hash table, I now encountered something which I had not done so far in the project. Whereas with the ring buffer and linked list there were just two functions, the hash table now had three. I could no longer simply assign half of the threads to adding and half to removing items. I had to come up with a better solution. An

additional concern was that I wanted to replicate the function call ratios for hash tables, which are about 90% contains calls, 9% add calls and 1% remove calls [reference Art of Multiprocessor…]. In the end I decided to implement the choose function.

The choose function would be relatively simple, now, whenever a thread was spawned, it would call the choose function, instead of calling the add or remove function. Inside the choose function, a number would be randomly generated, initially I used the modulo operation to cap the number at 100 and then used an if-else block, where if the number was greater than 9 then the thread would call the contains function, else if it was greater than 0 it would call the add function, else it would call the remove function. After testing I found that this replicated the function call ratios I had encountered earlier, though I decided to change the cap of 100 to 128, so that the compiler would streamline the operation [reference], and hence, I changed the values in the if-else block to correspond to it.

4 Experiments & Evaluation: (Fatish)

-Evaluate locked vs lockless

-Evaluate differences between different locked modes, sleep vs cpu relax for example

4.1 Evaluation Strategy

To analyse the data that I gather from the three data structures I implement in this project I am following the strategy outlined below.

Firstly, I graph the data based on two main factors, the number of iterations per second generated by the algorithm being tested and the number of threads that were created for each iteration of the algorithm. This allows me to graph the performance of each algorithm as the number of threads being generated increases. I have decided on 128 as being the maximum number of threads spawned as I originally thought that this was twice the number of core stoker has which would be 64. However, I now know that stoker only has 32 cores, though I have decided to stick with the 128 figure.

To measure the iterations per second I have to first record how long the program takes to finish. The system time is gotten at the start of the main function in each program; each thread is then created and run. Once all the threads are finished the system time is gotten again, this is the stop time. The start and stop time are then used to calculate the running time and this is then divided by how many seconds each thread is allowed run for which produces the iterations per second for each algorithm.

I have chosen one second to be the length that each thread should run for. I chose this amount of time as this is the case for many experiments of this type [reference] and makes the calculation for iterations per second trivial.

Each program starts by generating one thread which then works until one second has passed. The thread then terminates and the program restarts and generates two threads.

This process repeats, the thread count doubling every time until 128 threads are spawned at once. They then work for one second, at which point the program finishes.

In addition to the iterations per second and thread count I am varying the size of the different data structures to investigate whether the size of the relevant data structure plays a role in the performance of the locked and lockless algorithms.

To ensure that I am collecting accurate data, I am making sure to only collect data from each machine when CPU load is low so as not to jeopardise the data. In addition, while I initially collected data by running each algorithm only once I felt the variance between the different iterations, while small, was not negligible and so I changed my method to instead run each algorithm 7 times and to get the median of each data set produced.

I have chosen the median over the average as I feel that it gives a better representation of the data [reference]. I calculate the median for each algorithm after it has run 7 times before moving onto the next algorithm. I ensure that the calculation of the median does not impact on the performance of the algorithms as it is done outside of the timed sections of the program.

To speed up the time it takes to gather results I have written several bash scripts to automate the defining of the different locked modes of operation. Without them I would need to manually enter each program and define/undefine each mode of operation each time I run the tests and that would take up a lot of time. The script compiles the program multiple times, defining a different mode of operation each time. I do this by using the –D option in the g++ compiler. It then runs the code which prints out the data before moving onto the next mode.

To further increase the quality of my implementations I use the –O3 flag on the g++ compiler to turn on several optimisations supported by the compiler [reference].

All code is written in C++ and compiled using g++ 4.7.2.

4.1.1 System Overview

4.1.1.1 Stoker

Stoker is a multicore machine owned by the School of Computer Science and Statistics. It has four processors, each of which has eight out-of-order, pipelined, superscalar cores. Each of these cores has two-way simultaneous multithreading [reference].

The architecture s Intel Ivy Bridge EX 22nm and each of its 32 cores runs at 2.00 GHz [reference].

4.1.1.2 Cube

Cube is a multicore machine owned by the Internet Society in Trinity College Dublin. It has 8 processors, with each using two-way simultaneous multithreading. [https://wiki.netsoc.tcd.ie/index.php?title=Cube].

The architecture is Gainestown 45nm and each of its 16 cores runs at 2.27 Ghz [reference].

4.1.1.3 Local Machine

Local Machine is a multicore machine owned by myself, Mark Gibson. Its architecture is Sandy Bridge 32nm. It has four cores running at 3.30 Ghz each [reference].

4.1.2 Hardware Performance Counters

As mentioned previously I am using hardware performance counters as part of my evaluation of the data structures to help me determine why certain implementations perform the way they do. Below are some of the counters I am using and what they record:

Cycles records the number of CPU cycles used by the program.

Cache References records the number of times that the cache was referenced during the execution of the program.

Cache Misses records the number of times that the cache was referenced but returned a cache miss [reference]. I commonly use this as a ratio, depicting what percentage of cache references reported misses.

Branches Taken records how many branches were taken during a program's execution [reference].

Branch Misses records how many branches were not predicted successfully. Used in a similar way to cache misses to provide a ratio of how many branches were misses out of all that are taken.

Stalled Frontend Cycles records how many CPU cycles were wasted in the first stages of the CPU's pipeline, the fetching and decoding of instructions. Stalls happen when the pipeline is waiting for a value to be read from memory among other things [reference].

Stalled Backend Cycles records how many CPU cycles were wasted in the final stages of the CPU's pipeline, the execution of instructions. Stalls happen when the pipeline is waiting for a value to be read from memory among other things [reference].

4.2 Ring Buffer

4.2.1 Evaluation

Apart from the iterations per second and thread count I vary the size of the buffer to investigate how this affects the locked and lockless variation. The starting size of the buffer is 128; I initially had this at 100 but I decided to change it to be a power of two to minimise the effect that the modulo operation may have on the program's performance since the compiler reduces it to a bitwise AND operation [reference].

4.2.2 Results & Analysis

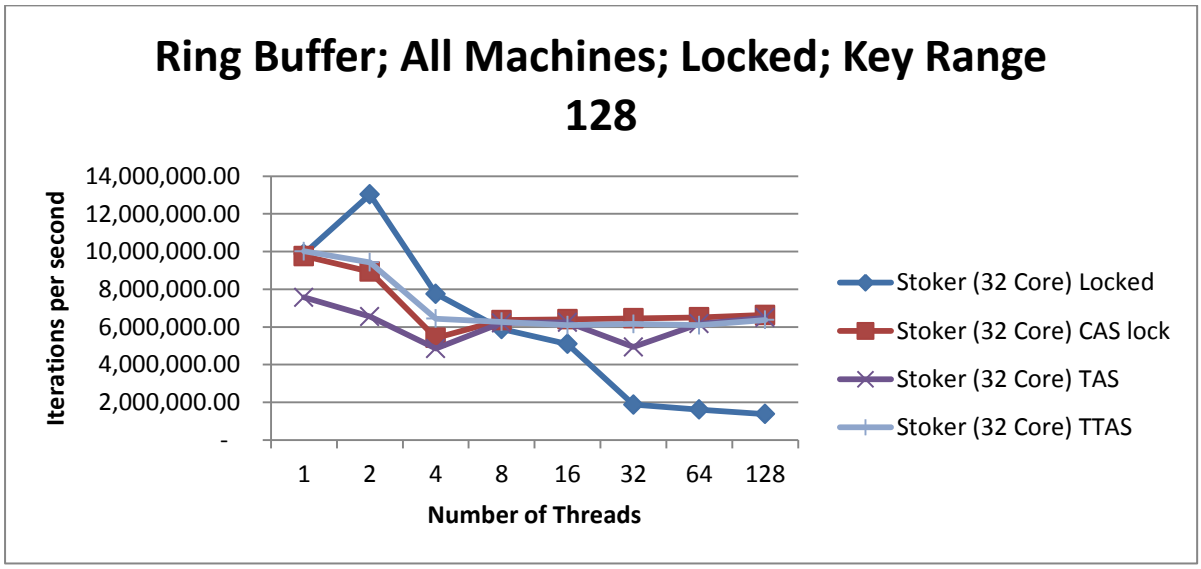
4.2.2.1 Lock Comparisons

I begin my evaluation by first focusing on the performance of the different locks before moving onto the lockless. Since the lockless implementation is *single-producer-single-*

consumer and the locked implementations are *multiple-producer-multiple-consumer* I am unable to compare the two and so I am interested in how the locks perform in relation to each other.

Note that while I was performing my tests, the *compare-and-swap-no-delay* lock and the *compare-and-swap-relax* lock began exhibiting issues where they would randomly throw a segmentation fault. This took me by surprise as no such issues had arisen during implementation. To attempt to fix the problem I added several memory barriers to the code and removed the –O3 flag when I was compiling the code but to no avail. As I did not have the time to delve into the problem further I had to abandon the two locks for the remainder of the ring buffer evaluation.

The Graph below represents the best four locks; the "Locked" mode represents a simple *pthread mutex* lock, "CAS lock" represents a *compare-and-swap* lock, "TAS" represents a *test-and-set* lock and "TTAS" represents a *test-and-test-and-set* lock. These four locks have the best performance consistently across the three machines.



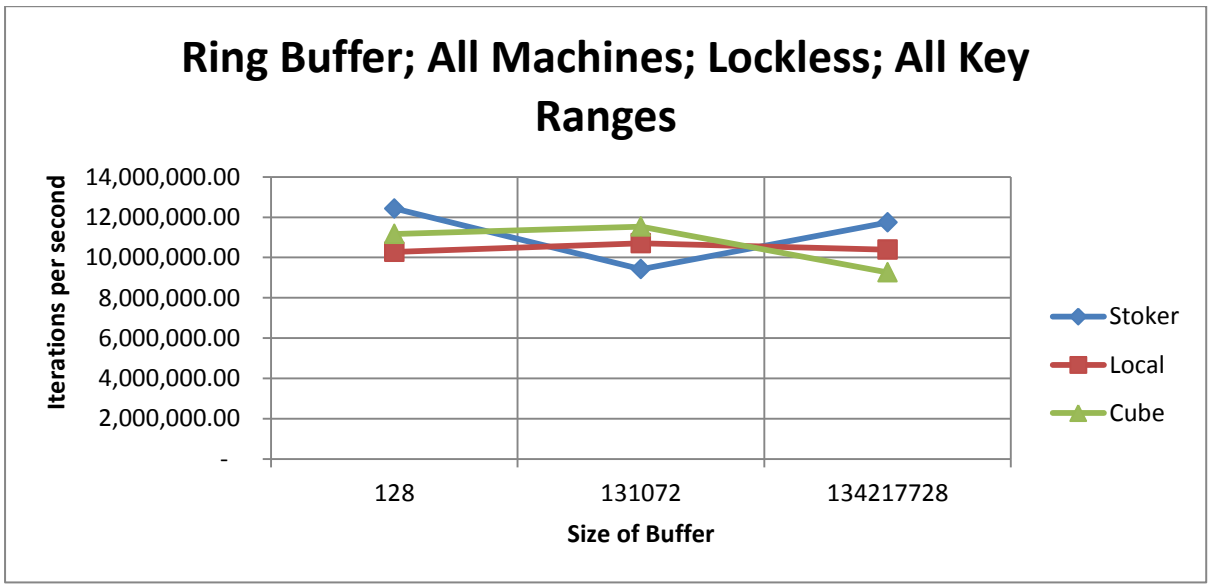
To determine as to why these four locks do quite well I compare two of them, the *pthread mutex* lock and *test-and-test-and-set lock* against a lock that does not perform as well, the *test-and-set-relax* lock which is similar to the *test-and-set* lock but it has a cpu_relax instruction after each thread attempts to acquire the lock in an attempt to reduce bus traffic. The hardware performance counter data is as follows:

	pthread mutex	*test-and-test-and-set*	*test-and-set-relax*
Cycles	2,102,076,100,442.00	76,583,801,948.00	2,589,400,577,492.00
Cache Misses %	38.35	37.37	92.12
Branch Misses %	0.09	0.06	0.33
Stalled Frontend Cycles %	95.10	67.82	99.31
Stalled Backend Cycles %	64.99	46.33	92.63

From the table we can see that *test-and-set-relax* has a ratio of cache misses to cache references of over twice that of both the *pthread-mutex* lock and the *test-and-test-and-set* lock. In addition the lock has five times more branch misses with regard to the total amount of branches it took and has a greater ratio of misses for front and backend cycles. From this it can be seen why the *test-and-set-relax* lock does so badly while the *pthread-mutex* lock and the *test-and-test-and-set* lock do relatively well when compared to it.

4.2.2.2 Lockless Comparison

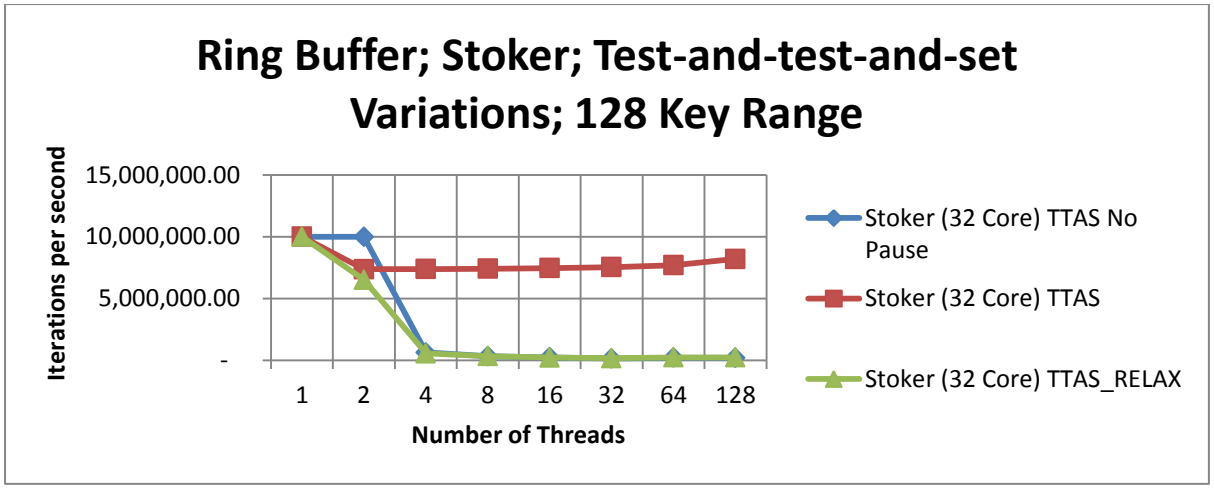
Since the lockless ring buffer implemented is a *single-producer-single-consumer* data structure it cannot be compared to the locks. The implementation only spawns two threads in total, one to push items onto the queue and one to pop the off. As a result the lockless implementation can only be compared to itself. For this I run it across the different architectures with different sizes to see how it performs. From the resulting table below we can see that the size of the buffer has a minor impact on the performance of the lockless implementation:



As seen above there does not appear to be a correlation between the maximum size of the buffer and the performance of the lockless algorithm, though we can see that Stoker has the best performance at the first and last size it is by no means a clear winner in terms of performance.

4.2.2.3 Test-and-test-and-set Variations

I want to see investigate the relative performance of the different *test-and-test-and-set* locks I have implemented. I have graphed the three variations below. It can be seen that the *test-and-test-and-set* lock using a sleep instruction has the best performance of the three. The *test-and-test-and-set-no-pause* seems to have a slight performance boost when the thread count is two while the *test-and-test-and-set-relax* lock is inferior for all thread ranges.



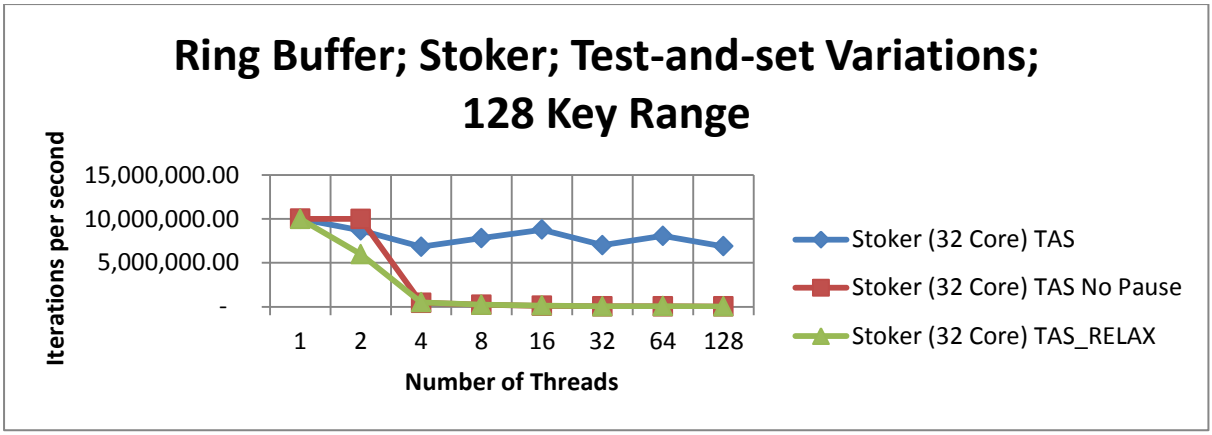
Analysing the data I gather from the hardware performance counters the reason for the regular *test-and-test-and-set* lock is clear, as can be seen from the table below:

	Cycles	Cache References	Cache Misses	Stalled Frontend Cycles
Test-and-test-and-set	7.92E+10	1.39E+07	4.71E+06	5.41E+10
Test-and-test-and-set-no-pause	2.39E+12	3.75E+08	2.34E+08	2.21E+12
Test-and-test-and-set-relax	2.41E+12	3.09E+08	1.99E+08	2.35E+12

From the table it can be seen that the regular *test-and-test-and-set* lock uses far less CPU cycles than the other two variations. In addition, it has far fewer cache references, and out of those misses a far smaller proportion than the other two variations. The *test-and-test-and-set-no-pause* lock has the largest number of cache misses which is likely due to its constant polling, which constantly invalidates cache lines, causing more bus traffic and more misses as a result [reference]. Finally the *test-and-test-and-set* lock has the lowest proportion of stalled frontend cycles meaning that it wastes the fewest CPU cycles. From the data above it can be clearly seen why the regular *test-and-test-and-set* lock has the best performance out of the three variations.

4.2.2.4 Test-and-set Variations

I turn my attention now to the *test-and-set* lock, of which I have implemented three variations. These are the *test-and-set –no-pause* lock, the *test-and-set* lock and the *test-and-set-relax* lock. Below is a graph showing their relative performances:



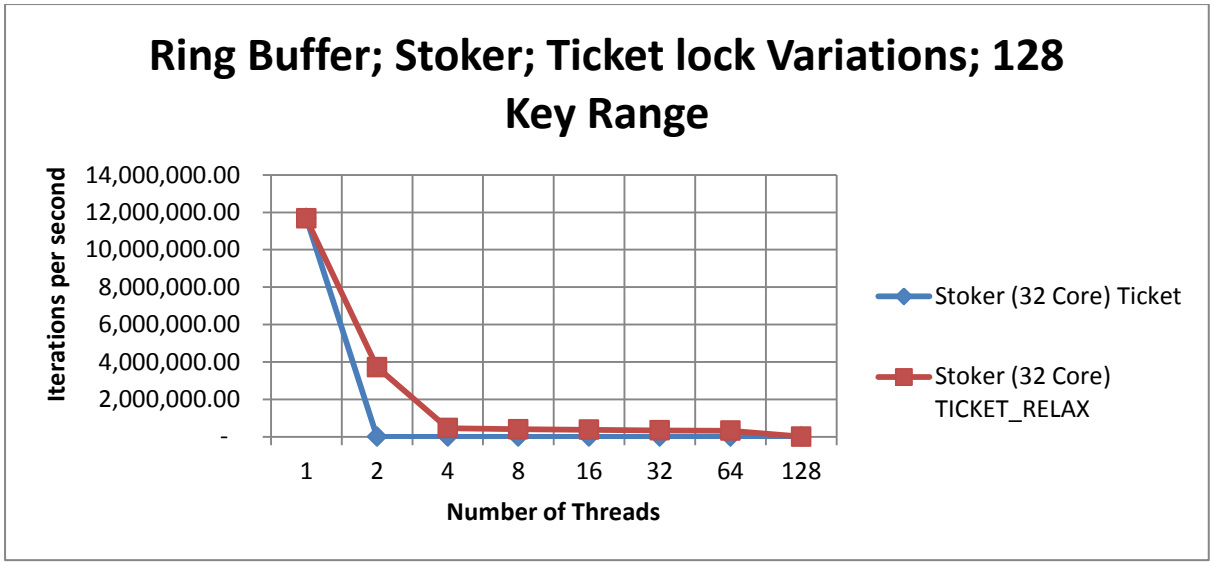
It can be seen from the graph that the *test-and-set* lock is the best performing lock once the thread count reaches four and onwards. Below is data from the hardware performance counters:

	Cycles	Cache references	Cache Misses	Stalled Frontend Cycles	Stalled Backend Cycles
Test-and-set	8.98E+10	6.51E+06	1.33E+06	5.93E+10	4.50E+10
Test-and-set-no-pause	2.66E+12	2.05E+08	1.84E+08	2.65E+12	2.58E+12
Test-and-set-relax	2.67E+12	2.25E+08	2.07E+08	2.65E+12	2.48E+12

From the data it can be seen that the *test-and-set* lock utilises far fewer CPU cycles and has a far lower proportion of stalled cycles both on the front and backend. Both the *test-and-set-no-pause* and *test-and-set-relax* locks utilise their CPU cycles extremely poorly, with both locks having 95% and over stalled cycles. Couple this with a far lower proportion of cache misses and it is clear why the *test-and-set* lock comes out on top in terms of performance out of the three locks.

4.2.2.6 Ticket Lock Variations

The final lock I wish to investigate for the ring buffer is the *ticket* lock and its variation, the *ticket-relax* lock. It is known that ticket locks perform poorly once the number of threads exceed the number of cores and I want to confirm that and see if the *_mm_pause()* intrinsic affects this in any way. Below is the graph detailing the two locks' performance:

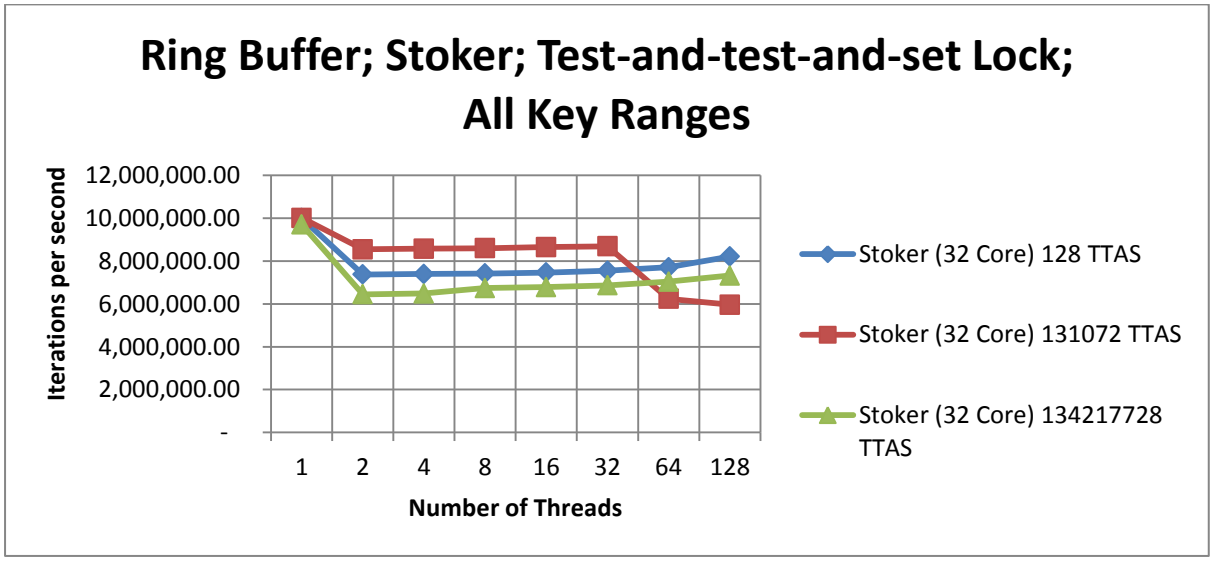


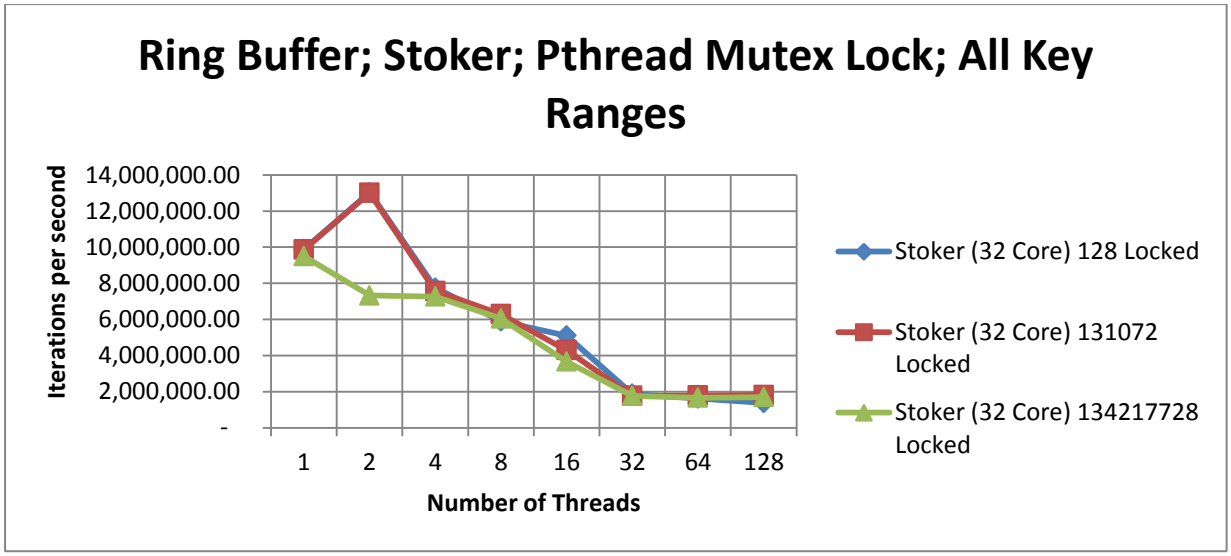
From the graph it can be seen that both locks drop sharply in performance, though somewhat earlier than expected considering that Stoker has 32 cores. The hardware performance counter data can be seen below:

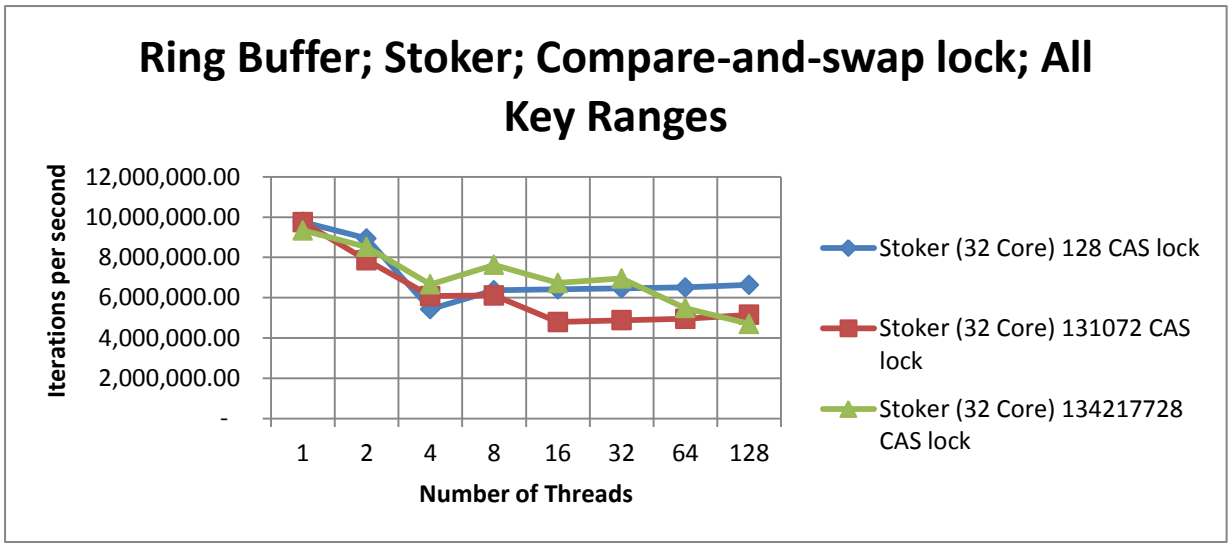
4.2.2.7 Does size affect performance?

It has already been shown in the *lockless* comparison that the size of the buffer has little to no effect on the performance of the *lockless* implementation. Hence, I will now evaluate the *pthread mutex*, *test-and-test-and-set* lock and the *compare-and-swap* lock to investigate if this property carries over to the locked implementations.

From the three graphs below it can be seen that while there are some performance differences between the buffer sizes used, it is inconclusive as to whether or not the size of the buffer directly affects the performance of the implementation.

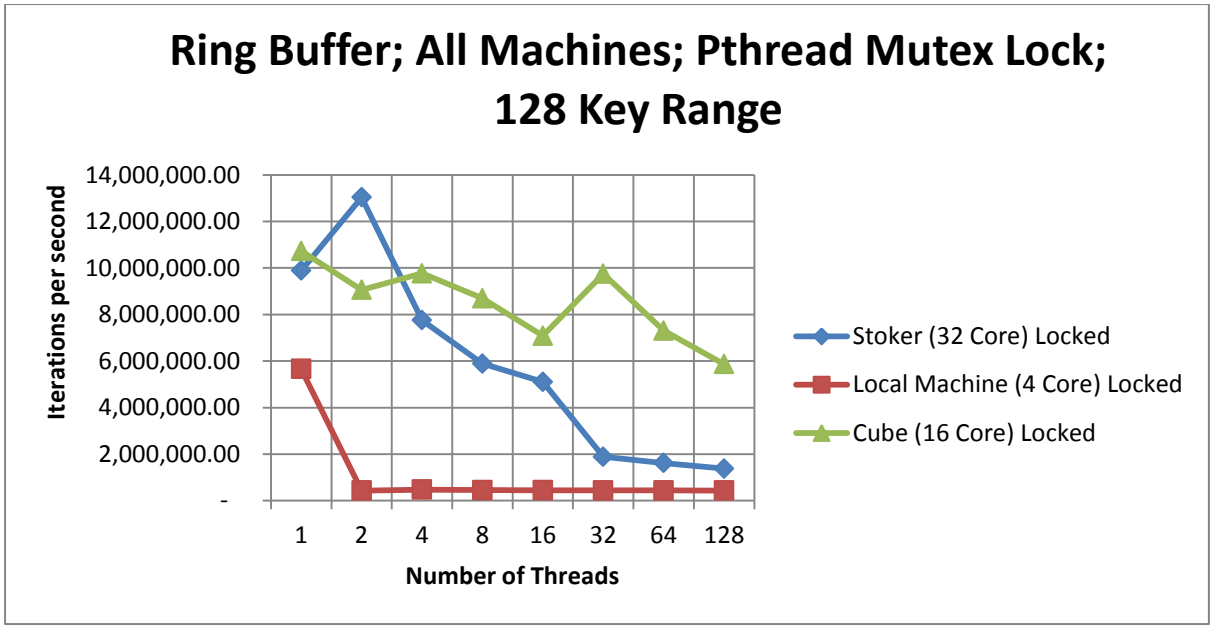


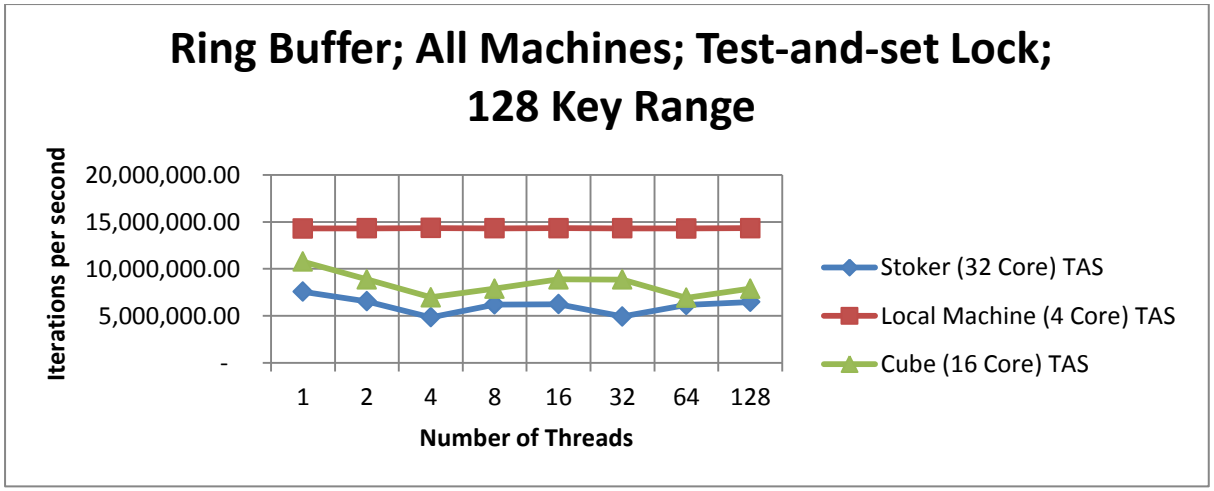


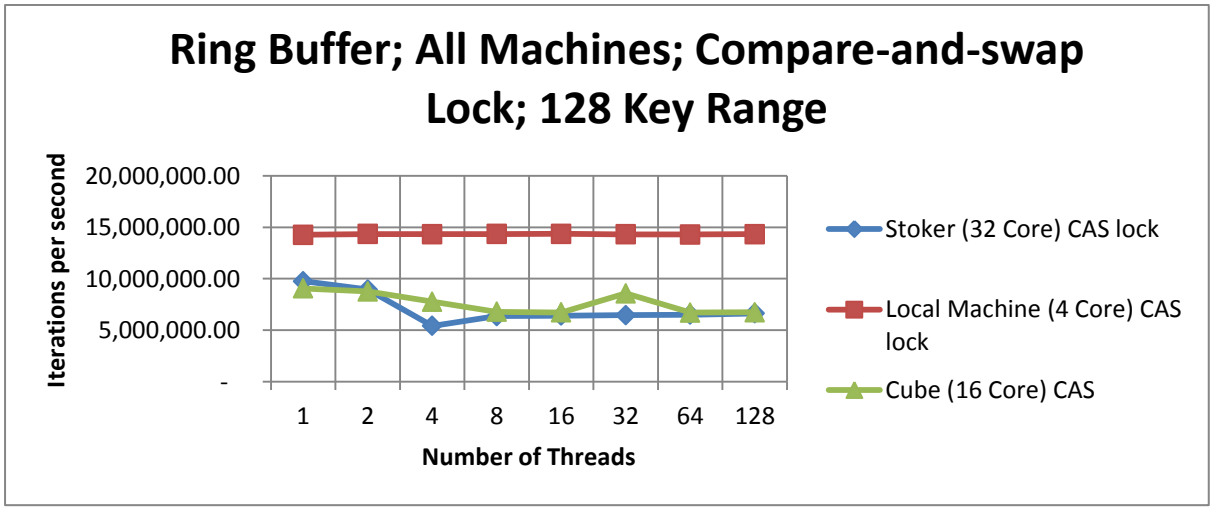


4.2.2.8 Is the ring buffer robust across architectures?

For the final piece of evaluation for the ring buffer I will now investigate whether or not the locked and lockless implementations maintain their relative performances across multiple architectures. The *pthread mutex* lock, *test-and-set* lock and *compare-and-swap* lock implementations will now be run across all three architectures. Below are a series of graphs, each representing one lock over the three architectures of Stoker, Cube and Local.







From the graphs above it can be seen that some locks are more robust across architectures than others. For example, the *test-and-set* lock's performance has a similar shape across

the three architectures while the *pthread mutex* lock is quite different on the three architectures.

4.3 Linked List

4.3.1 Singly Linked List

4.3.1.1 Evaluation

For the singly linked list I vary it by changing the maximum size of the list to investigate if it has any effect on the performance of the locked and lockless algorithms. This is represented by the variable KEY_RANGE which I use with the modulo operation and the rand() function [reference] to produce key values for the nodes in the list. Since this list is ordered and there are no duplicates allowed, the value of KEY_RANGE is the largest value a node can have and since no nodes are generated with a higher value, this acts as a hard cap on the maximum length of the list.

I initially set out to test the list using the values 100, 100,000 and 1,000,000,000, however, as mentioned previously, to minimise the cost of calling the modulo operation so often I changed them to powers of two, namely 128 (2^7), 131072 (2^17) and 134217728 (2^27) so that the compiler will replace the modulo calls with a bitwise AND [reference] to minimise the performance impact of the instruction.

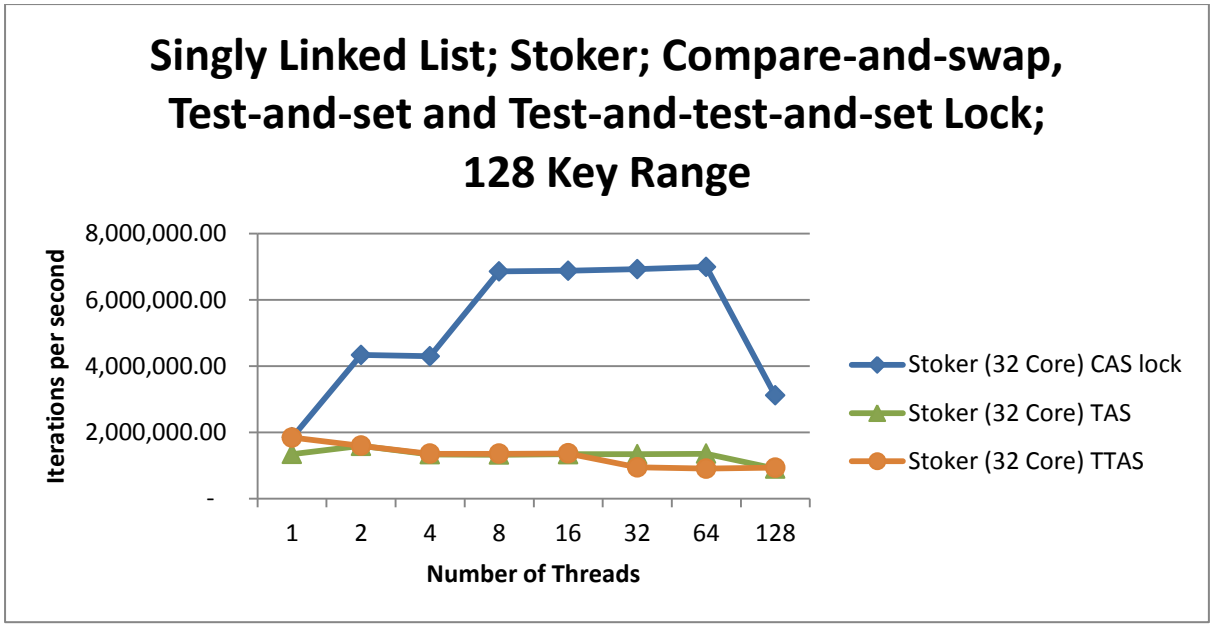
4.3.1.1 Results & Analysis

4.3.1.1.1 Locked Comparison

I start evaluating the singly linked list with a maximum allowed size of 128 nodes. Again, as with the ring buffer I start off by comparing the different locks against each other.

As with the ring buffer, both the *compare-and-swap-no-delay* and *compare-and-swap-relax* have run into issues, throwing segmentation faults where there was no issue before. Not having the time to discern why I exclude them from the rest of the evaluation.

Out of the locks it is the *compare-and-swap*, *test-and-set* and *test-and-test-and-set* which perform the best, with the *compare-and-swap* lock vastly outperforming the other two as seen below:



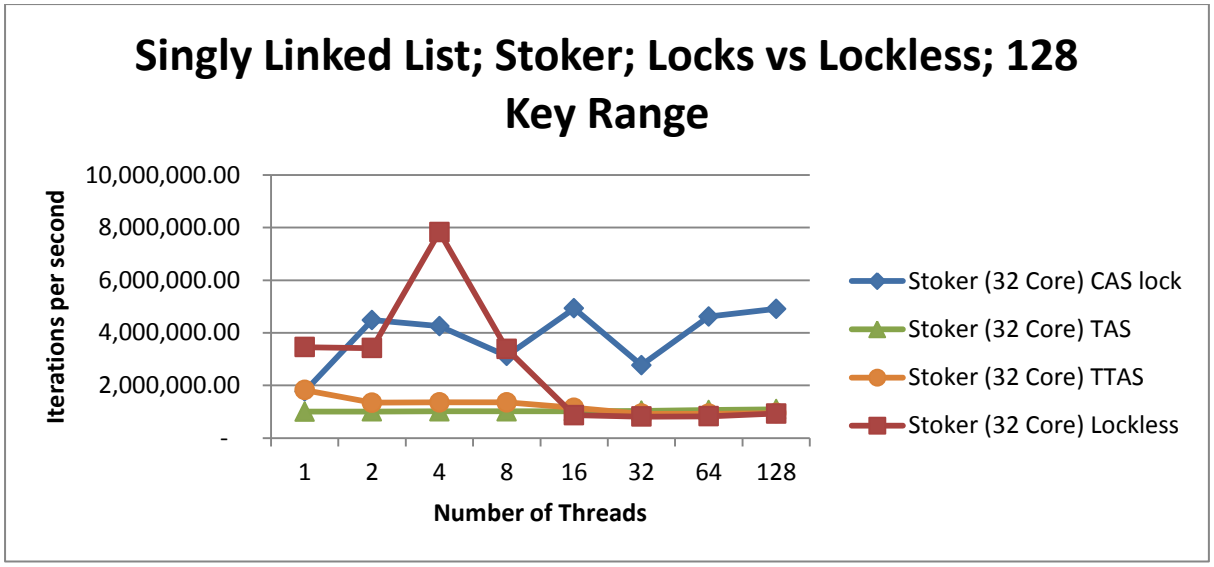
If the hardware performance data is examined it is clear why the *compare-and-swap* lock has such excellent performance in relation to the other two locks.

	Cycles	Cache References	Cache Misses	Branches	Branch Misses
Compare-and-swap	7.93E+10	2.53E+07	1.82E+07	1.37E+10	5.08E+07
Test-and-set	7.38E+10	3.80E+07	3.03E+07	1.30E+10	8.77E+07
Test-and-test-and-set	8.26E+10	3.70E+07	2.95E+07	1.46E+10	8.50E+07

Surprisingly, for such a large gap in performance, the *compare-and-swap* lock does not stand out from that data gathered. It has a similar amount of CPU cycles compared to the other two locks along with a comparable amount of cache misses. In fact the *test-and-test-and-set* lock has twenty percent fewer cache misses then the *compare-and-swap* lock. None of the data that I gather should make the *compare-and-swap* lock stand out as it does so this comparison requires further investigation.

4.3.1.1.2 Locked vs Lockless Comparison

I now compare the top performing locks from the previous section with the lockless implementation. The results of the comparison are shown below:



It can be seen that the lockless implementation performs very well at the earlier thread counts, especially at a thread count of four, however it then dips sharply, following the performance of the

4.3.1.1.3 TTAS

4.3.1.1.4 TAS

4.3.1.1.5 CASLOCK

4.3.1.1.6 Robust across machines?

4.3.2 Doubly Linked Buffer

4.3.2.1 Evaluation

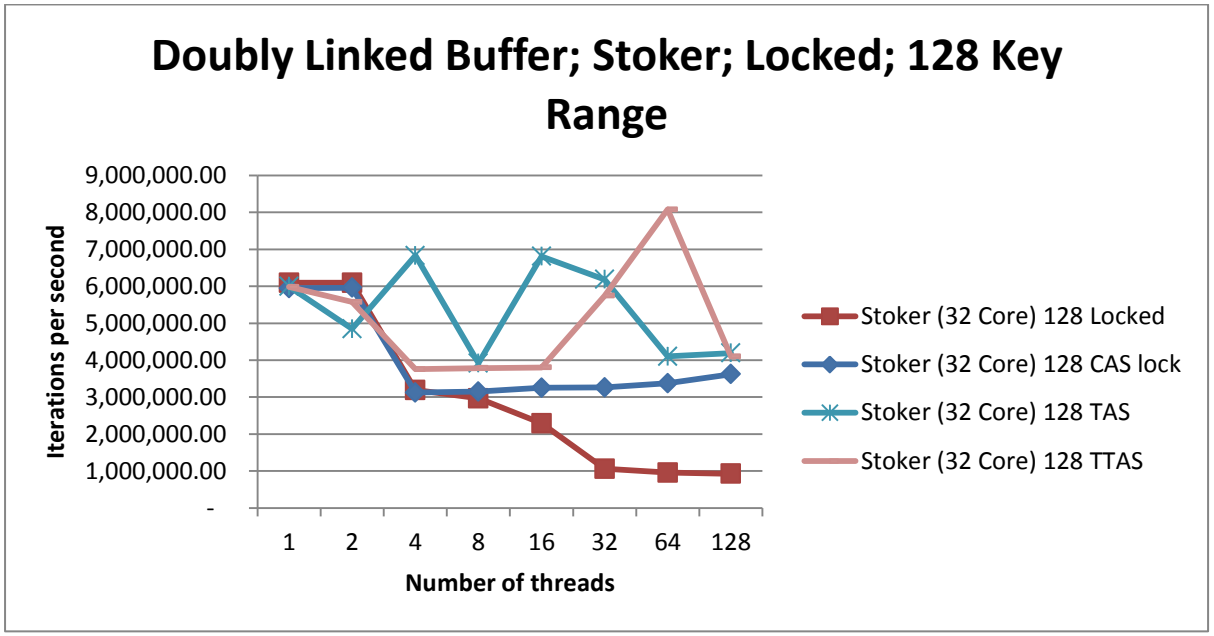
This variation of the linked list differed from the singly linked list due to the fact that this list is neither ordered nor does it prevent duplicates from being added to the list. In addition, nodes are added and removed from the head and tail respectively so that threads no longer have to spend any time searching the list, turning it into a buffer like object. This variation was implemented so that the locked and lockless versions could be compared as closely as possible, removing the randomness of the singly linked list where a thread may insert a node at the head of the list or may have to travel the full length based on the node that was randomly created.

As with before, the initial size to be tested is 128 with the size increasing up to 131072 to investigate if size impacts this version of the linked list at all.

4.3.2.1 Results & Analysis

4.3.2.1.1 Locked Comparison

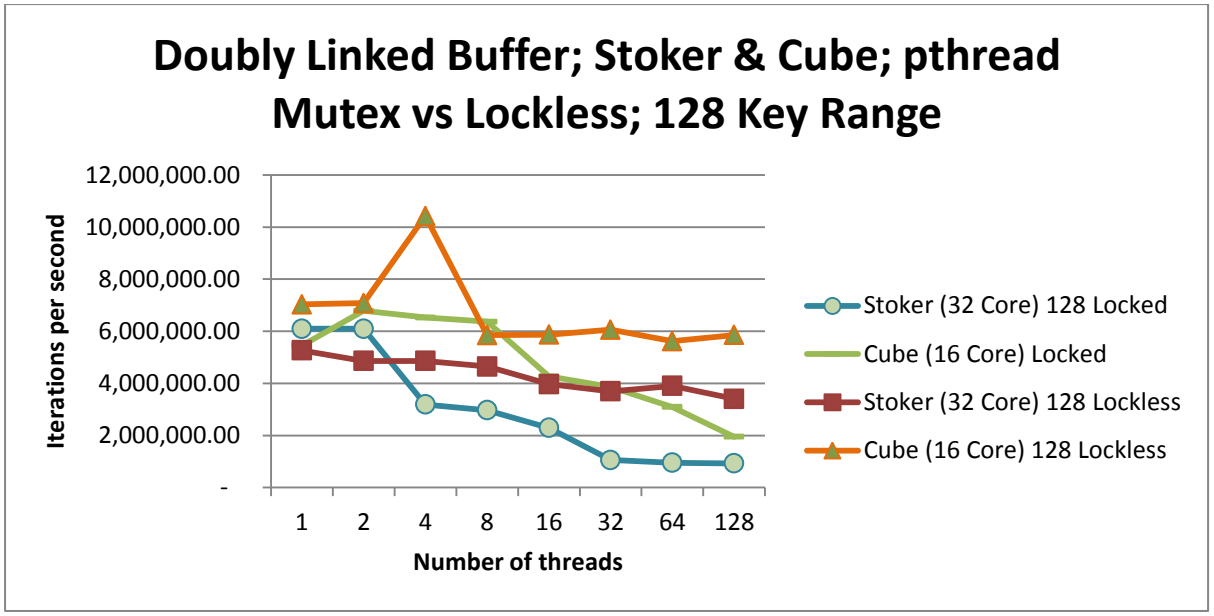
The four best locked modes of operation on Stoker were the pthread Mutex Lock, CAS lock, TAS and TTAS lock. In general, the TAS, TTAS and CAS lock along with their varieties did well across all machines.



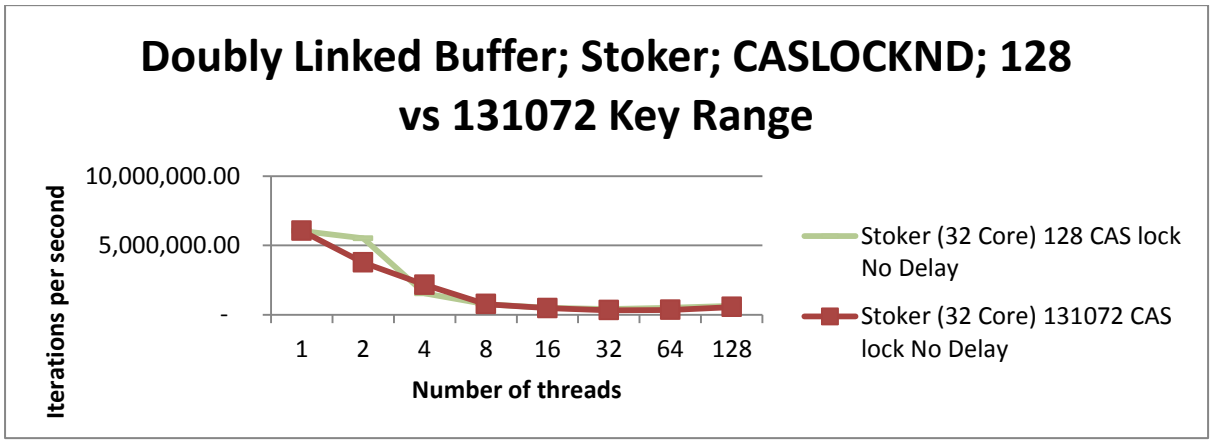
4.3.2.1.2 Lockless Comparison

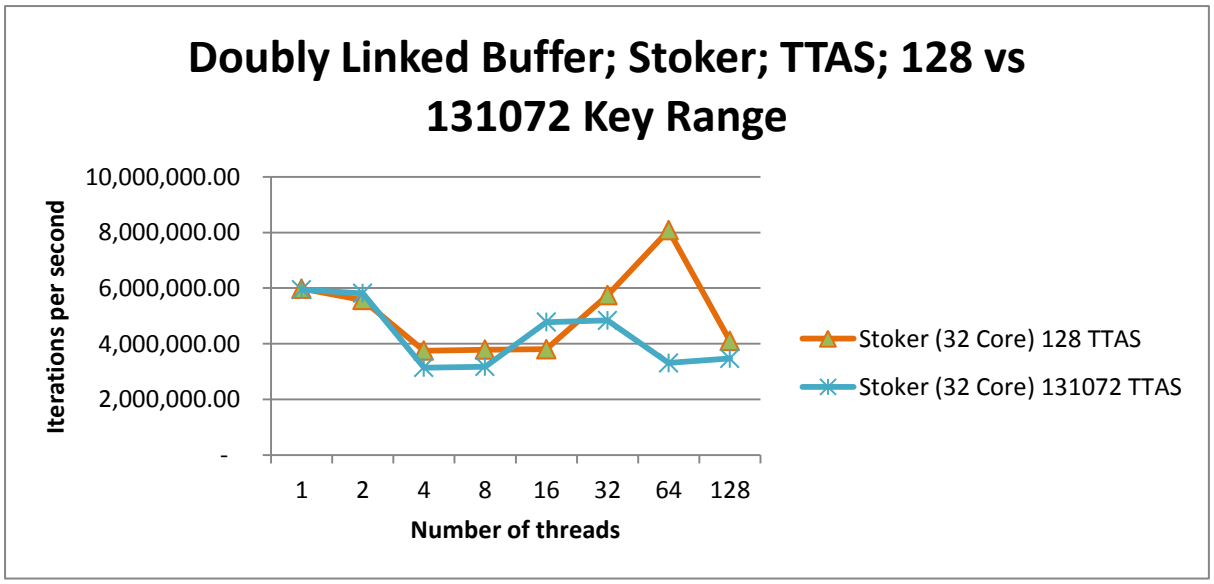
4.3.2.1.3 Locked vs Lockless Comparison

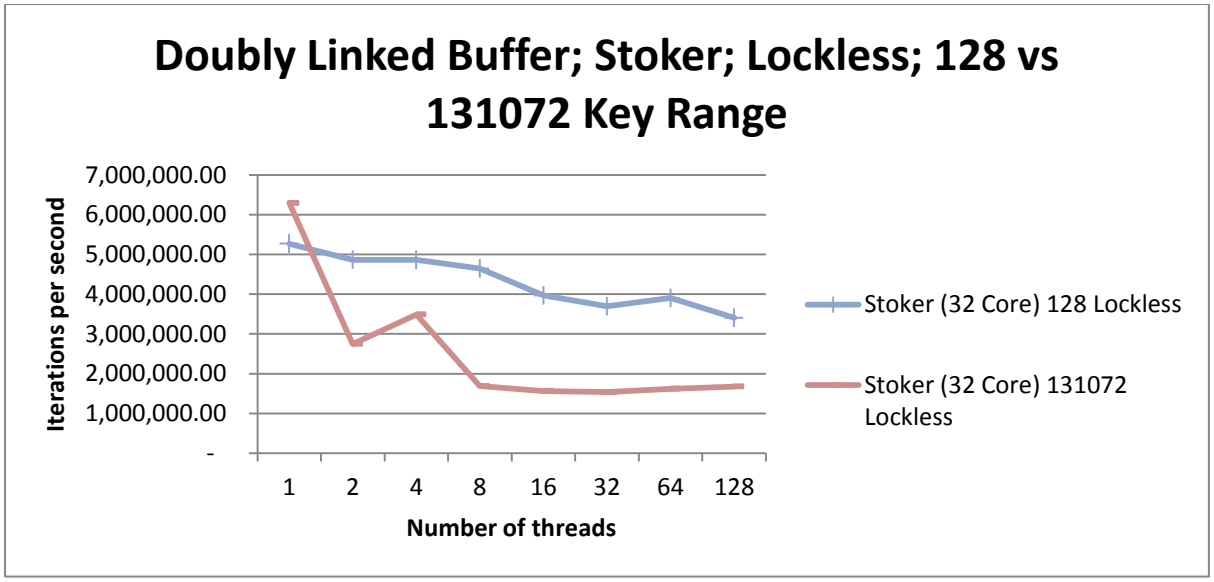
The lockless implementation did well against the locks with all machines reporting results to match or exceed the results from the best performing locks, the CAS, TAS and TTAS locks, especially around the early thread counts



I then ran the tests again, but this time used a size of 131072 to see if the size of the doubly linked buffer had an impact on the performance of the locks and on the lockless implementations:







From the above graphs it can be seen that for some modes of operation the size of the buffer makes no difference, as in the first graph comparing CASLOCKND. However in the subsequent two graphs we can see a performance difference, where the TTAS lock has a spike in performance at 64 threads and where the lockless version seems to have better overall performance than the version with the larger buffer.

4.3.3 Singly Linked Buffer

4.3.3.1 Evaluation

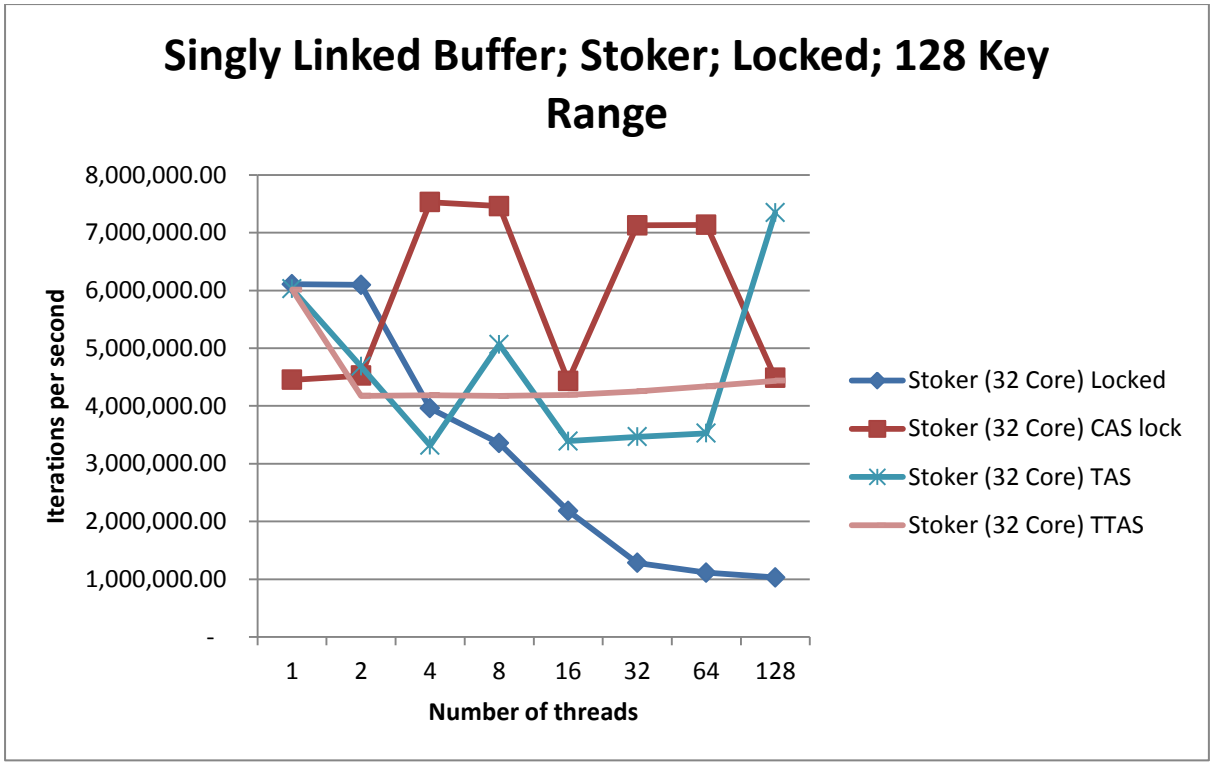
The evaluation for the singly linked buffer is much the same as for the doubly linked buffer. I began testing with a size of 128 on the locked modes of operation to compare them before moving onto to testing and comparing the lockless version against them.

The singly linked buffer is similar to the doubly linked buffer in that nodes are continuously added onto the head and removed from the tail except in this variety the placement of the head and tail pointer are switched and thus eliminating the need for each node to have a second pointer. I am interested in seeing if this lighter implementation affects the performance of either the locked and lockless versions.

4.3.3.1 Results & Analysis

4.3.3.1.1 Locked Comparison

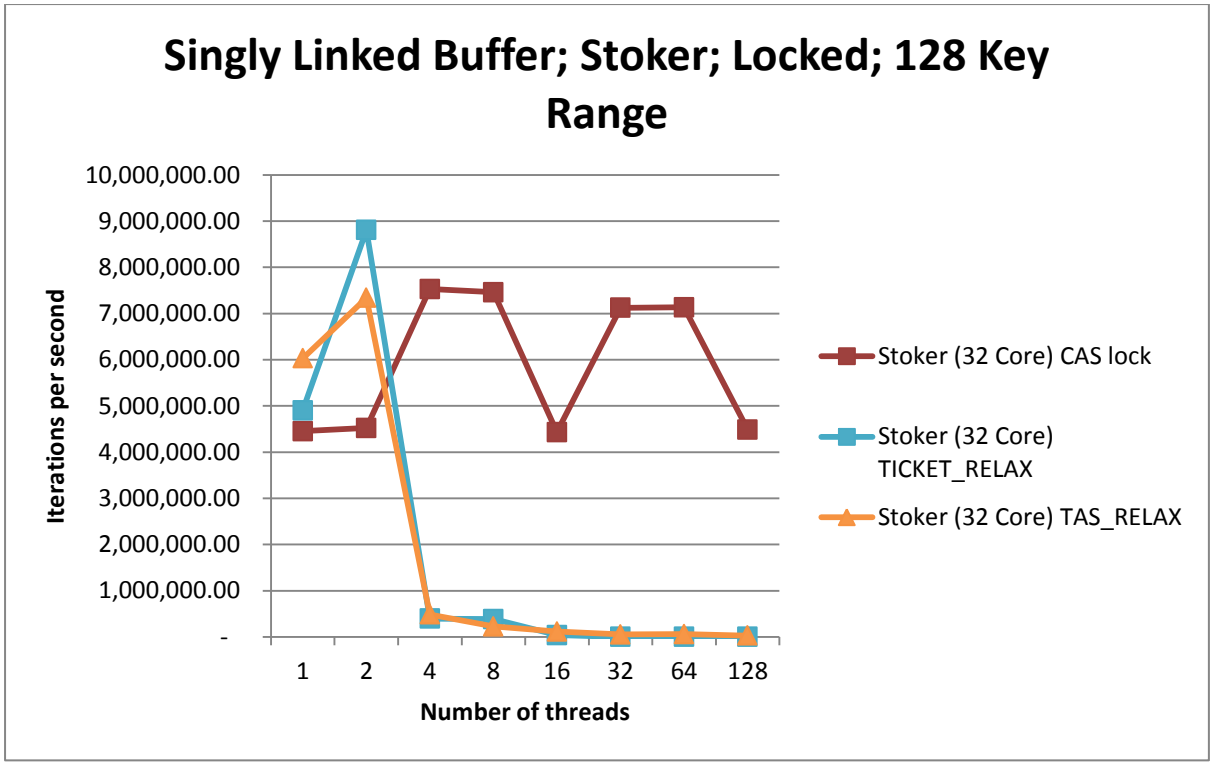
I ran all the locks on stoker, with TAS, TTAS and CASLOCK coming out as the locks with best performance at medium to high thread counts.



Counter	CAS Lock	TAS Lock	TTAS Lock	Pthread Mutex Lock
Cycles	119,375,736,696	2,653,918,458,528	90,884,965,440	2,094,344,583,826
Ratio of cache references to misses (%)	72.98154683	91.02209543	91.31047094	44.43437786
Ratio of branches taken to branch misses (%)	0.034478659	0.346017034	0.02636733	0.084829484
Ratio of frontend cycles to stalled cycles (%)	69.62426661	99.51888271	59.5599505	93.91710949

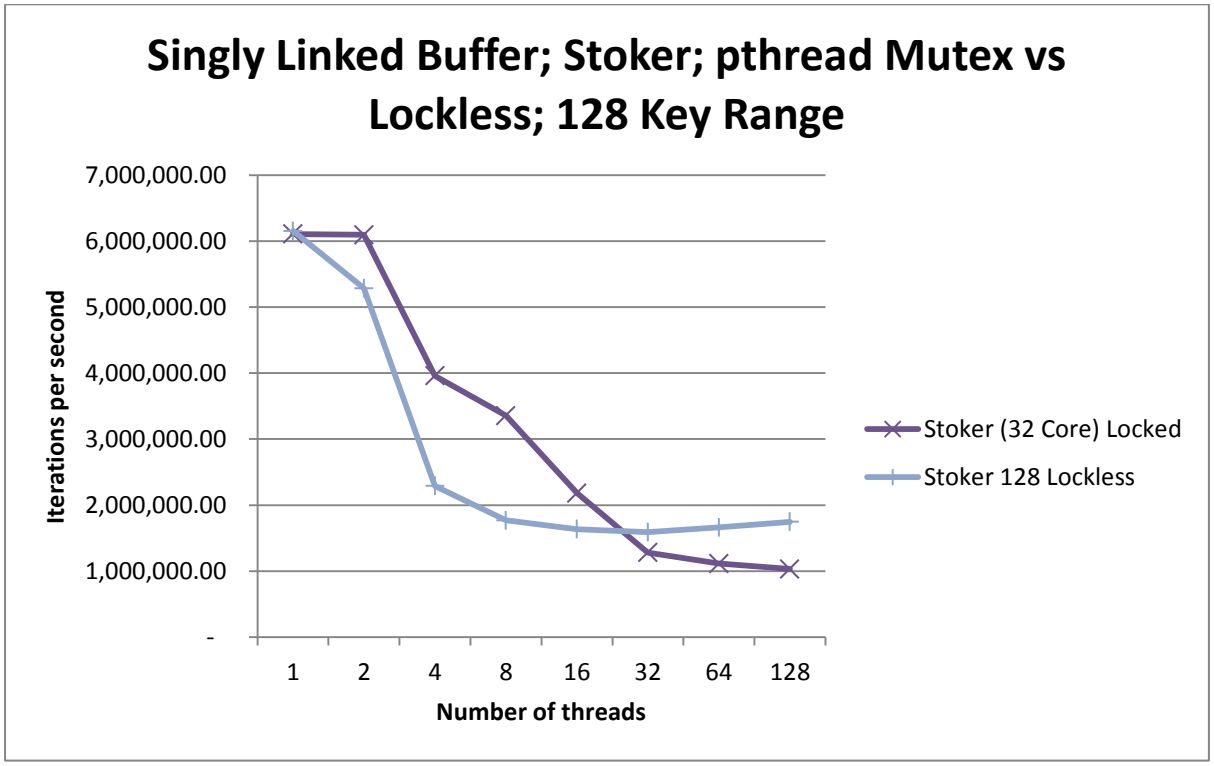
From the above table we can see that the TTAS lock had far less CPU cycles than the other locks but it wasted the fewest, shown by the stalled frontend cycles.

For low thread counts, the best three locks were TICKET_RELAX, TAS_RELAX and the CASLOCK again, though both the TAS_RELAX and TICKET_RELAX fall off sharply at four threads while CASLOCK continues to perform well until the 128 thread count.



4.3.3.1.2 Locked vs Lockless Comparison

Below is a graph comparing the pthread mutex lock and lockless implementation of the singly linked list; we can see that the lockless implementation actually beats the lockless version in performance up until the thread count of 32 and at that point on the lockless version outdoes the lock.



Counter	Stoker Lockless	Stoker pthread Mutex Lock
Cycles	2,223,499,181,754.00	2,094,344,583,826.00
Cache References	806,570,160.00	595,195,796.00
Ratio of cache references to misses (%)	67.93237305	44.43437786
Ratio of branches taken to branch misses (%)	0.09188388	0.084829484
Ratio of frontend cycles to stalled cycles (%)	93.58296982	93.91710949

As you can see, the hardware performance counters report that the two implementations perform relatively similarly, with around the same ratio of stalled cycles and misses branches. The lockless variation does pull away with the number of cycles and it recorded 20% more cache references, so even that it missed a larger portion of its cache references, it still performed well enough to outdo the pthread mutex lock at higher thread counts when contention was higher.

4.4 Hash Table

4.4.1 Evaluation

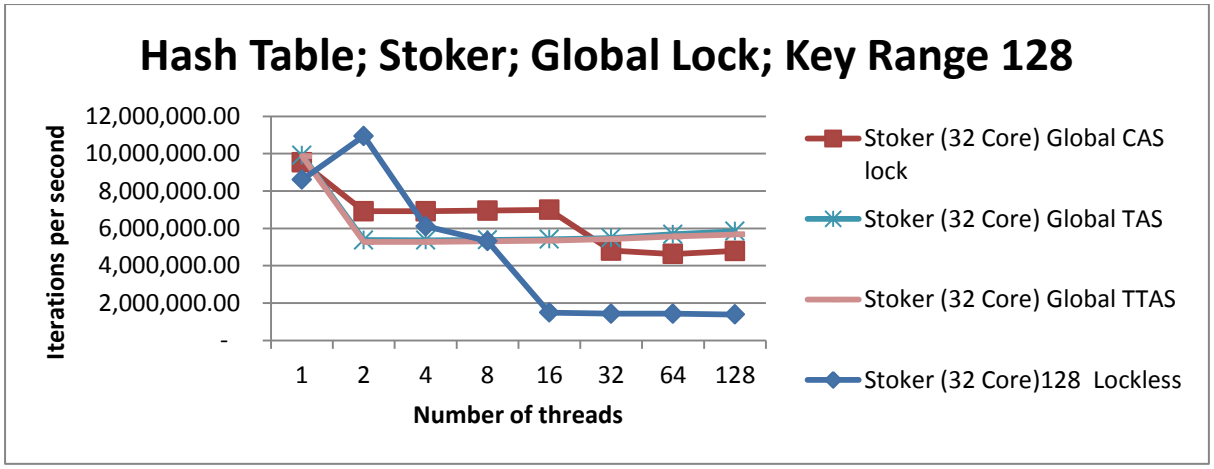
For the hash table I have two locked variations and the lockless variation. As discussed in the method section of the report, the globally locked version uses a single, global lock to grant mutual exclusion. Lock per bucket is more granular and gives each bucket its own lock so multiple threads can interact with the table but only on separate buckets. Finally the lockless version does not use any locks so multiple threads can interact with the same bucket.

To begin evaluation, I start with an initial table size of 128, with no resizing of the table. I then increase the table size to investigate how it impacts the performance of the three variations. After that I turn on resizing for the same purpose and finally I test the variations on Cube and my Local Machine to see whether the variations' performance changes with the architecture it runs on.

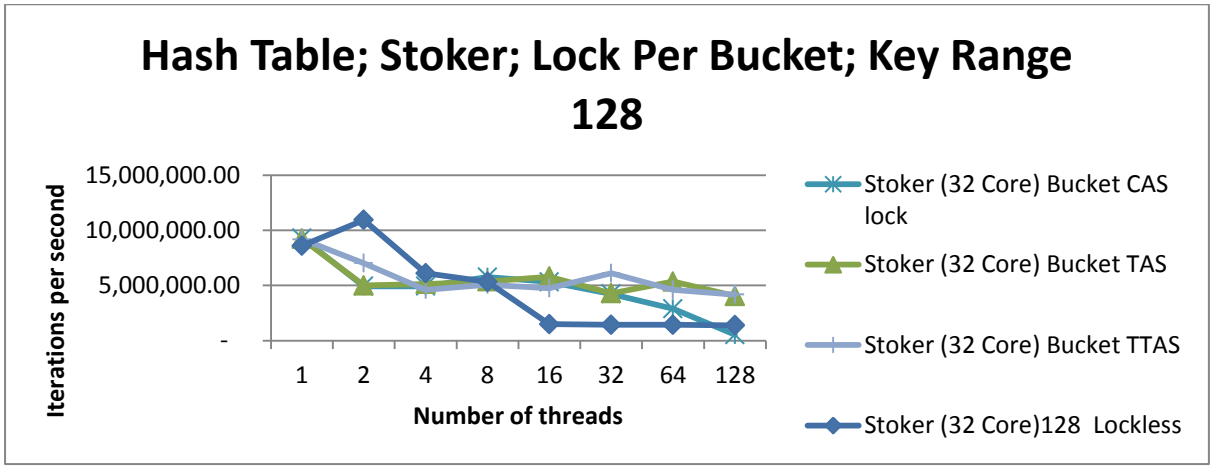
4.4.2 Results & Analysis

4.4.2.1 How do variations measure up?

Below are the top three best performing locks using the global lock variation along with the lockless variation of the hash table.



Below are the top three best performing locks using the lock per bucket variation along with the lockless variation of the hash table.



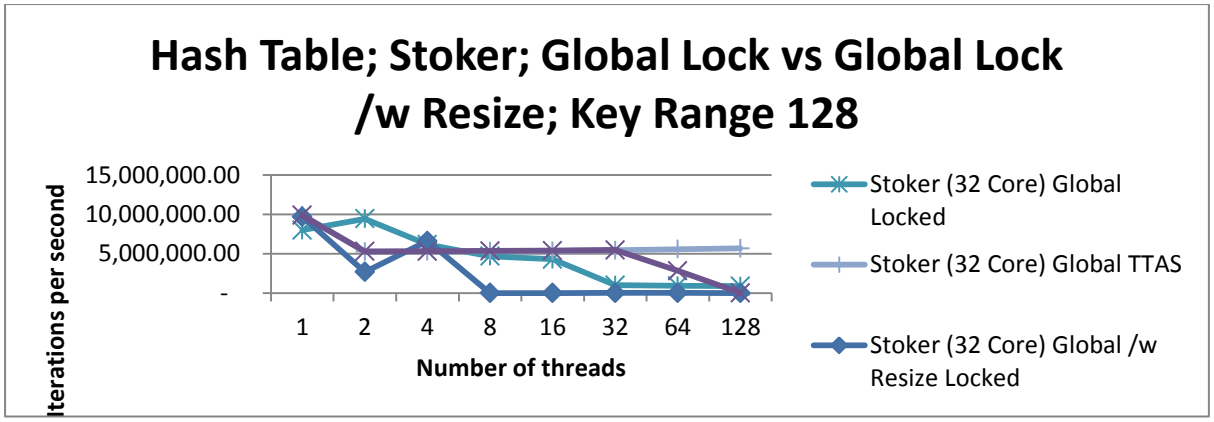
From the graphs, it is clear that the lockless algorithm does well with a lower thread count compared to the two locked variations but falls off from eight threads onwards

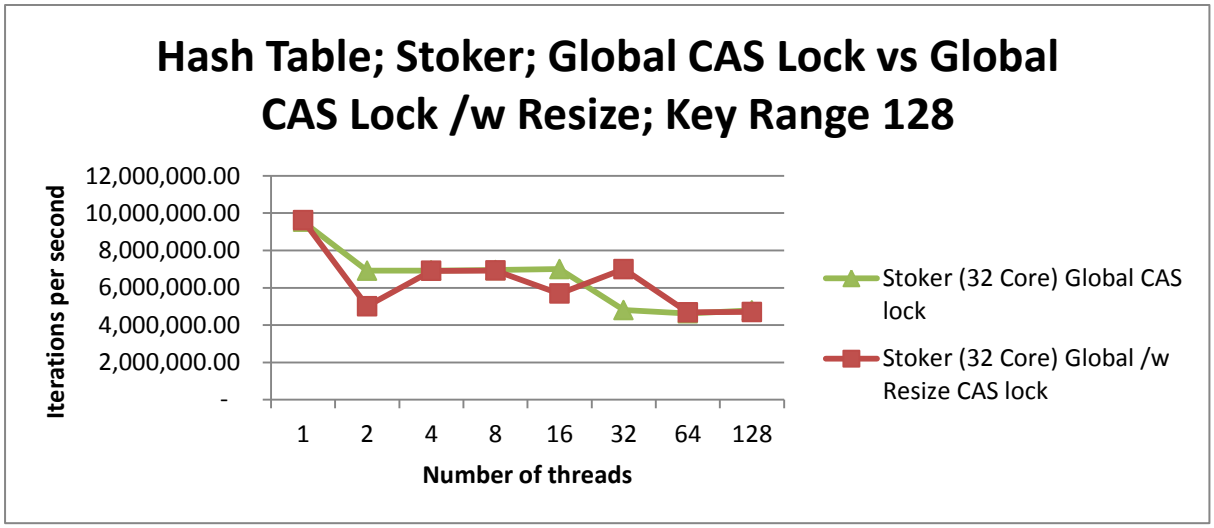
4.4.2.2 What impact does resizing have?

To investigate the impact resizing has on the performance of the three variations I compare each variation with itself, putting the data gathered with no resize functionality beside the data with resize functionality enabled.

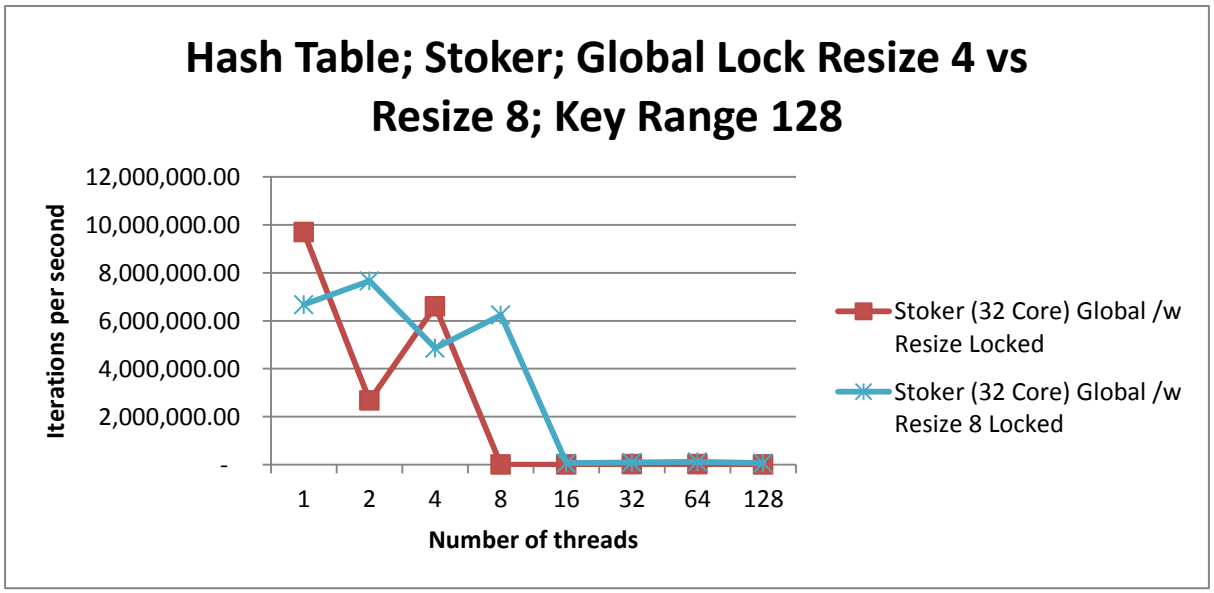
I choose the maximum length of a list allowed before resizing occurs to be four, as chosen by Herlihy and Shavit [reference]. To compare, I then change this to eight to observe the differences between the two.

Above it can be seen that the resize functionality causes a noticeable drop in performance whenever it is required. For the pthread mutex lock with resize it can be seen that it drops sharply, at thread count 2 and thread count 8 at which point it stays very low. For the TTAS lock with resize it appears to perform almost identically to the regular TTAS lock until thread count 32 at which point it can be seen that it drops sharply and continues to do so as it get to the 128 thread count.





Relative to other locks, the CAS lock with resize performed quite well when compared to its resize-less version. A drop is seen at thread count 2, 16 but apart from those two points it performs equally if not better than the regular CAS lock.



The graph above shows the pthread mutex lock with resize functionality. Blue is that with a maximum list length of 4 while red is a maximum list length of 8. It can be seen that the

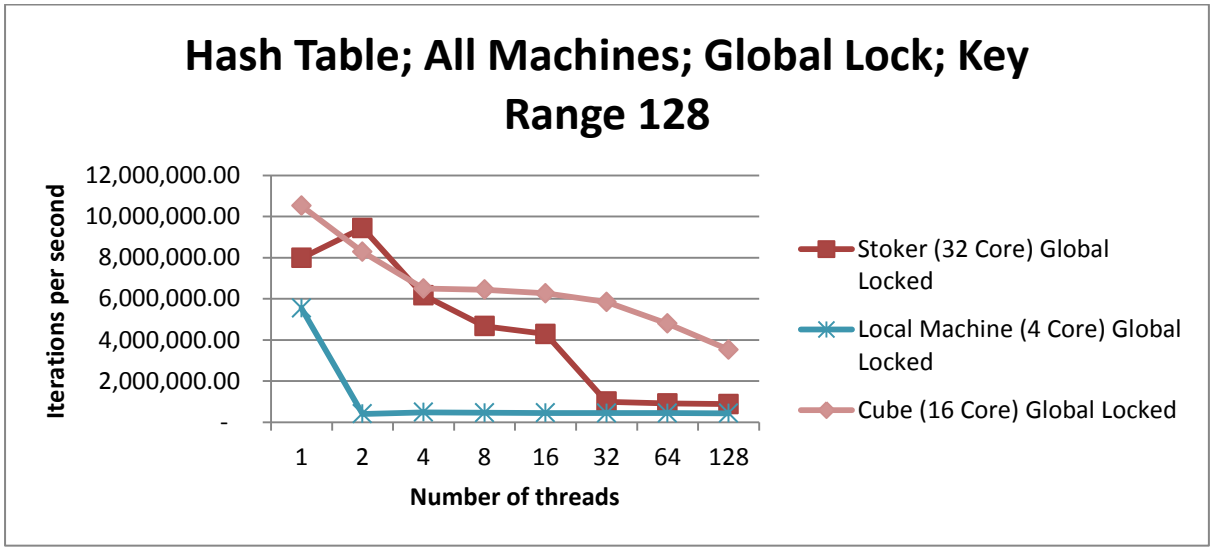
increase in maximum list length has not real impact on the performance of the lock except that it will perform better for slightly longer before dropping due to the added time required generating a bucket containing 8 nodes as opposed to 4.

From these results it can be seen that the resize functionality does impact the performance of the implementations, though some locks appear to deal with it better than others.

4.4.2.3 How does the size of the table affect performance?

4.4.2.4 Do the variations' performances changes with different architectures?

From the graph below we can see that the Global Lock is not robust across architectures with its performance changing distinctly.



5 Afterword: (Thin)

5.1 Conclusions

Relate to questions asked in introduction and if they have been answered and what new questions have arisen

This purpose of this project is to determine and compare the differences between concurrent data structure implementations and investigate if the performance of these implementations is maintained across different architectures.

5.1.1 Ring Buffer

I have concluded that the best locks for my implementation of the ring buffer are the *compare-and-swap*, *test-and-set* and *test-and-test-and-set* locks based on their performance across the range of thread counts I use.

The regular *test-and-test-and-set* lock using a sleep instruction is the best performing out of itself and the two other variations, *test-and-test-and-set-no-pause* and *test-and-test-and-set-relax* locks.

The same goes for the *test-and-set* lock which out-performed its two variants, *test-and-set-no-pause* and *test-and-set-relax*.

The results gathered to determine if the size of the buffer affects the performance of the implementation is inconclusive. While the buffer size appears to affect the performance of some locks, such as the *test-and-set* lock it seems to have no impact on other locks such as the *pthread mutex* or the *compare-and-swap* lock.

Finally, I have concluded that while most locks do not retain their performance across architectures, this is not the case for some. Again, the *test-and-set* lock behaves similarly across the three architectures but others are not so robust.

5.1.2 Linked List

5.1.3 Hash Table

5.2 Future Work

How my work could be improved and developed. What disadvantages are there in my approach (lack of memory management) etc.

6 Bibliography & Appendix: (Thin)

6.1 References:

Herlihy, Shavit. 2008. The Art of Multiprocessor Programming.

N/A (17/07/2013). *Atomic Operations Library.* Available: http://en.cppreference.com/w/cpp/atomic. Last accessed 29/01/2014

usleep(3) – Linux man page. Available: http://linux.die.net/man/3/usleep. Last accessed 29/01/14

Michael Brady. 2013. Concurrent Systems II.

(10/02/2007). Circular Buffer. Available: http://c2.com/cgi/wiki?CircularBuffer. Last accessed 29/01/14

Lockless Inc. Spinlocks and Read Write Locks. Available: http://locklessinc.com/articles/locks/ . Last accessed 29/01/2014

(20/02/14). [Perf Wiki. Available: https://perf.wiki.kernel.org/index.php/Main_Page].

Moir, Shavit. 2001. Concurrent Data Structures

Herlihy, 1993. A Methodology for Implementing Highly Concurrent Data Objects

6.2 Appendix: