**Research Proposal**

**PhD/Masters/Honours**

|  |  |
| --- | --- |
|  | |
| Faculty: | **Environment, Society and Design** |
|  | |
| Department/School: | **Informatics & Enabling Technologies** |
|  | |
| Degree: | **B.Science(Honours)** |
|  | |

|  |  |
| --- | --- |
|  | |
| Student Name: | **Richard (Rikki) Andrew Cattermole** |
|  | |
| Student ID Number: | **1126775** |
|  | |
| Address: | **52 Bethel Crescent**  **Christchurch**  **New Zealand** |
|  | |
| Email Address: | **alphaglosined@gmail.com** |
|  | |

|  |  |  |
| --- | --- | --- |
|  | | |
| Proposed Thesis or Dissertation Title: | Web routers: An explorative performance review | |
|  | | |
| Brief Description of Research: | Performance evaluation of data structures and algorithms for web routing as used in web servers and web service frameworks. | |
|  |  | |
| Due Date for Thesis or Dissertation Submission: | |  |
|  | | |

Table of Contents

[Web routers: An explorative performance review 1](#_Toc467892086)

[Performance evaluation of data structures and algorithms for web routing as used in web servers and web service frameworks. 1](#_Toc467892087)

[Table of Contents 2](#_Toc467892088)

[Contents 2](#_Toc467892089)

[Proposal Checklist 3](#_Toc467892090)

[Supervisor Team Details 4](#_Toc467892091)

[Main Supervisor Approval 5](#_Toc467892092)

[Further Approvals for this Proposal 6](#_Toc467892093)

[Web routers: An explorative performance review 7](#_Toc467892094)

[Introduction 7](#_Toc467892095)

[Literature Review 8](#_Toc467892096)

[The World Wide Web 8](#_Toc467892097)

[The Server 9](#_Toc467892098)

[Performance issues 11](#_Toc467892099)

[Current routing approaches 11](#_Toc467892100)

[Research Objectives/Questions 14](#_Toc467892101)

[Research context 14](#_Toc467892102)

[Method 14](#_Toc467892103)

[Timetable 16](#_Toc467892104)

[Health and Safety 16](#_Toc467892105)

[Budget 16](#_Toc467892106)

[References 17](#_Toc467892107)

# Proposal Checklist

*Copy (Ctrl+C) this ticked box*  *then highlight an answer box, and Paste (Ctrl+V) the ticked box in place of the answer box.* ***Sign and date using a pen.***

|  |  |
| --- | --- |
| 1. I have taken part in a Mutual Expectation Agreement (MEA) meeting with my supervisory team.   Date of meeting: |  Yes  4th/Aug/2016 |
| 1. I have given a seminar on my research proposal.   Date of seminar: |  Yes  No  / / |
| 1. I have booked (or completed) training in how to use the Endnote bibliography software. |  Yes  No |
| 1. I agree to consult with and obtain approval from the Human Ethics Committee BEFORE starting any empirical, lab, practical or field work, and I will comply with the conditions specified in the approval. |  Yes  NA |

**Student**  **Supervisor**

**Signature:**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **Signature:\_\_\_\_\_\_\_\_\_\_\_** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Date Proposal  
Submitted:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **Date:**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Supervisor Team Details

**Main Supervisor:**

|  |  |
| --- | --- |
| **Status:** | **Supervisor** |
|  | (Supervisor, Co-Supervisor) |
| Name: | Stuart Charters |
| Position Held: | Senior Lecturer |
| Email Address: | Stuart.charters@lincoln.ac.nz |

**Other Supervisors:**

|  |  |
| --- | --- |
| **Status:** | **Co-Supervisor** |
|  | (Associate Supervisor, Co-Supervisor) |
| Name: | Walt Abell |
| Position Held: | Senior Lecturer |
| Email Address: | Walter.abell@lincoln.ac.nz |
| Company & Postal Address (if External): |  |
|  |  |
| **Status:** |  |
|  | (Associate Supervisor, Associate Co-Supervisor, Advisor) |
| Name: |  |
| Position Held: |  |
| Email Address: |  |
| Company & Postal Address (if External): |  |
|  |  |
| **Status:** |  |
|  | (Associate Co-Supervisor, Advisor) |
| Name: |  |
| Position Held: |  |
| Email Address: |  |
| Company & Postal Address (if External): |  |

# Main Supervisor Approval

*I confirm the following:*

|  |  |
| --- | --- |
| 1. A full cost budget (itemised and listed by calendar year) has been prepared for this research project and is attached. I have examined the budget, and it appears to be suitable. |  Yes |
| 1. Equipment or facilities required for completion of the project are available or can be obtained (within contraints applying to purchase of capital equipment) using the project funding applied for or contracted. |  Yes  No   NA |
| 1. The outcome of this research project could produce commercialisable Intellectual Property (protectable by patent or not).   If YES, I have advised the Director of the Research & Commercialisation Office in writing (email), so that the Director can consider what further documentation may be necessary. |  Yes  No  Unsure   Yes |
| 1. I approve this proposal with the following comment: | |

Main Supervisor Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Further Approvals for this Proposal

**Other Supervisors:**

|  |
| --- |
| I \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, approve this proposal with the following comment:  Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
|  |
| I \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, approve this proposal with the following comment:  Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
|  |
| I \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, approve this proposal with the following comment:  Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
|  |

**Postgrad Convenor of Department/School:**

|  |
| --- |
|  |
| I \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, approve this proposal with the following comment:  Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
|  |

**Faculty Postgrad Convenor:**

|  |
| --- |
|  |
| I \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, approve this proposal with the following comment:  Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
|  |

# Web routers: An explorative performance review

# Introduction

The focus of research in recent times has been on the client side of web development, as defined by the Hyper Text Transfer Protocol (HTTP) specification (Fielding, et al., 1999). The main body of research has been to produce better ways to present information, and with the interaction to the user.

The focus has been on the server side has been in making dynamic content (e.g. address book entries based upon a query) more easily created and manipulated. Anecdotally there does not appear to be much development in this area. This has left certain technologies in use by the web servers with limited work done on them; an example of this would be the web router.

Web routers are the core technology that maps HTTP requests to the web service (code) that produces the response to be sent back to the client. A web router does not interact with the user directly, instead it is configured by descriptions (routes) of websites. Software developers implement a web service to produce dynamic content for a website. Dynamic content and static content (e.g. a websites logo) utilize a web server router to locate the resource handling code to produce a response to the HTTP request.

While the web router implementation is unseen its performance is crucial to the overall of the response times for a request. This research is to compare the performance of various web router implementations to recommend potential improvements.

# Literature Review

The World Wide Web

The World Wide Web was conceptualised in 1989 (World Wide Web Consortium, n.d.) since then there has been a large uptake in its usage by everyone across the globe to an estimate of 3.4 billion users as of October 8th of 2016 (Internet Live Stats, 2016). With every one of those users working with the standards of Uniform Remote Locator[[1]](#footnote-1), Hypertext Transfer protocol (Fielding, et al., 1999) and Hypertext Markup Language[[2]](#footnote-2) in some form or another.

During the early days, many different web browsers and servers were created, the majority of these have since long died off but the legacy that is the definition of each has not. As defined by the World Wide Web Consortium (W3C) (World Wide Web Consortium, 2014):

1. Web browser  
   A program which allows the display and execution of a web page for a user. Interacts with a web server to provide any data required. This is the most common form of client.
2. Web server  
   Retrieves files or resources from the file system or some form of backend such as a web application and sends them to the client as requested.
3. Web Server API or service  
   A standalone piece of software that will dynamically create content to send to a client. It communicates in some form to the web server to serve up content to the client.

These components are displayed in Figure 1 The web (HTTP) request + response cycle.



Figure 1 The web (HTTP) request + response cycle

With greater usage of the internet coinciding and with greater speeds during 1990s as predicted by Nielsen’s Law (Nielsen, 1998), companies and developers alike experimented with dynamic web pages allowing for user interactions not possible with static web pages alone. The Common Gateway Interface[[3]](#footnote-3) was created to allow for external program to be executed as part of the web page processing by a server. From this point on existing programming languages gained new uses that was not seen before, which helped to introduce other new programming languages. Some examples of this would be PHP (The PHP Group) and JSP (Java Server Pages) which has the primary purpose of dynamic page creation on each request by the client.

## The Server

Web servers and web (server side) APIs alike are a field of research that continues to introduce new areas of study for research in both academic and in informal capacities. Combined they share a very similar technology set, with only slightly different purposes and entry points. The web router resides on a server and is a required component for the operation of both web servers and web APIs.



Figure 2 General HTTP request + response processing activities

A web routers, primary goal is to map any incoming request from a socket to a function (code) to process it. The execution and processing of a request once mapped can be done on the server or a specific route handling code can be done in a separate process. The handler can be written in any language. This is shown in Figure 2 General HTTP request + response processing ; it is based upon HTTP 1.x.

The primary focus of web developers is upon coding the requests and manipulating of the response for the client side. For the server side the focus is upon handling the routes for a given purpose. There is little consideration by those who use an implementation and along with it, its performance. This can cause problems such as the time it takes to handle a request from getting it to responding to it back the client. The algorithms used in web routers (e.g. linked list, B-Tree) and data structures were originally created for database engines. In the context of a database they have been optimised and analysed for best performance. For a web router, these algorithms may be improved performance once they have been analysed and optimised for this use case.

Performance issues

The standard implementation of a web server has the following processing steps:

1. Asynchronous socket listener
2. Thread/Fiber router (choose the thread to execute the request handling in)
3. HTTP request processing
4. Routing to function call
5. HTTP response creation and return

In these steps there are a number of potential performance issues:

* Non-blocking asynchronous socket listening verses blocking synchronous socket listening
* Scheduling of the handling code upon a thread/fiber
* Blocking operations (e.g. communicating with the database, file reading/writing)

These have been worked on by developers since the beginning of the web, but comparatively little has been done in the routing component. Hence the focus of this study is on improving the performance of the web router.

## Current routing approaches

A web router maps incoming requests to a route handler (code). The approaches that are available to implement the routing can differ in the functionality and performance. These different approaches each have a different set of costs (memory and CPU) and the performance will depend upon the profile of the web traffic they are handling.

There is a variety of different methods used in implementing a web router. The main ones used are: tree graphs such as a Red-Black as implemented in Nginx[[4]](#footnote-4) or a single Regular Expression (regex) (Popov, 2014). A single regex can simplify the code required but will result in a limited functionality

Current implementations typically use the path from the HTTP header to perform lookups. Routes may contain constants as path segments, parameters or a “catch all” for all following path segments. Existing basic data structures can handle paths in this format. E.g. regular expressions are typically used to implement them. These can cover most cases by utilizing multiple instances of the router implementation. Each instance of the router handles a different HTTP method such as GET and POST.

Some servers support a feature known as rewriting. Rewriting is the process by which requests are modified into being another; however only internally. After a ‘rewrite’ of a request takes place it must be evaluated out as if it was a new request. The rules by which it can modify the request by can include the path, domain, time stamp, client IP address and any other HTTP request field. Most web routers do not implement this feature because of its complex nature however it is an add-on to many web servers.

In non-regex approaches, more information is stored in the data structure (HTTP method, port etc.). This allows the routing algorithm to use other conditions such as the HTTP request fields of User-Agent, Referer or Host. Support for this significantly increases the complexity of the implementation and limited research into this area was discovered in the creation of this proposal.

The data structure that the web router utilizes can take many forms including a list or a tree graph. These data structures are simple in design but have many optimisation opportunities such as cache locality for children in a tree graph which can improve performance by many magnitudes (Ross & Rao, 2000).

# Research Objectives/Questions

# Research context

The research proposal being presented here has the end goal of trying to improve the web routing performance which could make the world-wide web faster. In previous work a considerable amount of research has gone into making the web faster by focusing upon the total performance of the web request processing cycle with little investigation into the optimization of web router performance.

The overarching goal is to determine the performance of various web router algorithms for a given set of web request scenarios.

From this a set of sub-questions is formulated to help reach the overall research goal.

* What are the current performance metrics associated with the request/response cycle?
* What are common algorithms that web routers use and how do they relate to those used in other fields?
* What are the performance characteristics of commonly used algorithms to implement a web router given a range of routing scenarios as input?

The results of this investigation will allow for potential improvements in web router performance.

# Method

A benchmark harness will be created to execute tests to record the timings for routing of requests. The harness will provide a common interface to allow the timings to be gathered the same way for each web router implementation to be tested and compared.

The input data sets as per the benchmarking comprise two different set of information. The first is the routes to be stored into the relevant router data structure and the set of requests to be executed by the router implementation. The route will be preloaded before testing and will be fully optimised by the router before the execution of the benchmarking.

The design of the routes used in the data sets include: static paths “/my/path/goes/here” with variable number of parts “/part”, a variable number of variables “/my/path/:variable” and a catch all “/my/path/\*” for all values following the previous values. These will be combined into the forms: “/path/:vars/\*”, “/path/\*” and “/path/:vars” with path and vars being variable in number. The combination and complexity will be produced algorithmically for the purposes of getting as many corner cases as possible.

Different input sets will be generated by: how many path parts it contains, the number of parameters and how many have a catch all.

Each implementation is expected to run solely within a single thread. All input sets currently within memory prior to the execution and pre-start procedures having been executed. Each input set is executed in multiples of a ten giving: 10, 100, 1000 etc. The result for each multiple is then averaged for comparison with the same multiple for all the other inputs sets. Each test multiple will occur without any breaks (e.g. no thread sleeping) and will not include the time for preloading and optimisation.

The computer that will perform the benchmarking has the following specification with regards to parts being used for the analysis:

* CPU: Intel Xeon v3 E5-2630, 8 cores at 2.4ghz base frequency and 20mb cache
* Memory: DDR4 64GB at 3200mhz
* OS: Windows 10

Storage of results and input will be done within memory. Once results are fully generated and testing has concluded for an implementation test set, it will then be stored on a hard disk. This will prevent performance penalties associated with long term storage drives appearing in results.

Once the results have been gathered graphs comparing multiples for a test set as well as comparing a specific multiple between implementation for a test set. Particular attention will be paid to outliers within these graphs to determine problems of the implementation. Comparisons will also be made between test sets to determine how the type of web requests impacts performance.

# Timetable

|  |  |  |
| --- | --- | --- |
| **Task** | **Week of** | **Expected length of time** |
| Router implementation (1)  Flat array | December 5th | 1 week |
| Benchmark harness  Initial creation. This is expected to be a very rough implementation that does not get anywhere close to what is needed. | December 12th | 2 weeks |
| Router implementation (2)  Tree graph, no optimizations | January 2nd | 1 week |
| Router implementation (3)  Tree graph, optimizations | January 9th | 2 weeks |
| Router implementation (4)  Regex | January 16th | 1 week |
| Benchmark harness revision  Updates based upon what has been implemented with the routers. As well as include new features to make it more complete based upon what the routers are doing. | January 23rd | 2 weeks |
| Analysis upon router implementations performance | January 30th | 5 weeks |
| Writing of dissertation | March 6th | 11 weeks |
| Final feedback and revisions | May 22nd | 5 weeks |
| All work must be completed by this week. | June 26th – July 2nd |  |

# Health and Safety

No known health and safety risks exist for or by this research.

# Budget

A printing budget of $100 will be needed for final copies for submission of the dissertation.

No other expenses are expected.

# References

DigitalOcean. (2015, 28 1). *Apache vs Nginx: Practical Considerations*. Retrieved from DigitalOcean: Cloud computing designed for developers: https://www.digitalocean.com/community/tutorials/apache-vs-nginx-practical-considerations

Fielding, R., Irvine, U., Gettys, J., Frystyk, H., Masinter, L., Leach, P., . . . Compaq/W3C. (1999, June). *Hypertext Transfer Protocol -- HTTP/1.1.* Retrieved from Hypertext Transfer Protocol -- HTTP/1.1

Internet Live Stats. (2016). *Number of Internet Users (2016) - Internet Live Stats*. (Internet Live Stats) Retrieved 10 8, 2016, from Internet Live Stats: http://www.internetlivestats.com/internet-users/

Nielsen, J. (1998, April 5). *Nielsen's Law of Internet Bandwidth.* Retrieved from Nielsen Norman Group: UX Training, Consulting, & Research: https://www.nngroup.com/articles/law-of-bandwidth/

Popov, N. (2014, February 18). *Fast request routing using regular expressions*. Retrieved from http://nikic.github.io/2014/02/18/Fast-request-routing-using-regular-expressions.html

Ross, K. A., & Rao, J. (2000). Making B+- trees cache concious in main memory. *ACM SIGMOD international conference on Management of data* (p. 475486). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=335449

The PHP Group. (n.d.). *History of PHP.* Retrieved 10 8, 2016, from PHP: Hypertext Preprocessor: http://php.net/manual/en/history.php.php

Torres, A., Galante, R., Pimenta, M. S., & Martins, J. A. (2017). Twenty years of object-relational mapping: A survey on patterns, solutions, and their implications on application design. *Information and Software Technology, 82*, 1-18.

World Wide Web Consortium. (2014, March 14). *How does the Internet work*, 72360. Retrieved 10 8, 2016, from World Wide Web Consortium (W3C): https://www.w3.org/wiki/index.php?title=How\_does\_the\_Internet\_work&oldid=72360

World Wide Web Consortium. (n.d.). *Help and FAQ - W3C*. (W3C) Retrieved 10 8, 2016, from World Wide Web Consortium (W3C): https://www.w3.org/Help/#invention

1. https://tools.ietf.org/html/rfc3986 [↑](#footnote-ref-1)
2. https://www.w3.org/MarkUp/draft-ietf-iiir-html-01 [↑](#footnote-ref-2)
3. https://tools.ietf.org/html/rfc3875 [↑](#footnote-ref-3)
4. <https://trac.nginx.org/nginx/browser/nginx/src/http/ngx_http_file_cache.c?rev=953512ca02c6f63b4fcbbc3e10d0d9835896bf99> [↑](#footnote-ref-4)