Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

SCHOOL OF COMPUTER SCIENCE AND STATISTICS

TITLE OF WORK

FIRSTNAME LASTNAME

SUPERVISOR
NOVEMBER 25, 2025

Adapted from a template created by Prof. Michael Brady,
School of Computer Science, TCD
(remove line 45, title.tex)

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

M.SC. STATISTICS AND SUSTAINABILITY

Declaration

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

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at `http://www.tcd.ie/calendar`.

I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at `http://tcd-ie.libguides.com/plagiarism/ready-steady-write`.

Signed: _____ Date: _____

Abstract

A short summary of the problem investigated, the approach taken and the key findings. This should not be more that around 400 words.

The must be on a separate page.

what's the title for our title abstract one page five paragraphs area and digital twin project research questions two paragraphs how to solve them paragraph to implement and evaluate main findings one paragraphs expanding the abstract

introduction literature review design implementation evaluation conclusion

Acknowledgements

Thanks Mum!

You should acknowledge any help that you have received (for example from technical staff), or input provided by, for example, a company.

Contents

List of Figures

List of Tables

Nomenclature

A	Area of the wing	m^2
B		
C	Roman letters first, with capitals…	
a	then lower case.	
b		
c		
Γ	Followed by Greek capitals…	
α	then lower case greek symbols.	
β		
ϵ		
TLA	Finally, three letter acronyms and other abbreviations arranged alphabetically	

If a parameter has a typical unit that is used throughout your report, then it should be included here on the right hand side.

If you have a very mathematical report, then you may wish to divide the nomenclature list into functions and variables, and then sub- and super-scripts.

Note that Roman mathematical symbols are typically in a serif font in italics.

1 Introduction - Chapter

1.1 Section

1.1.1 Subsection

SSH

SSH (1) :
SSH -> transport layer -> secure, low-level transport protocol.
Provides strong encryption, crypto based host authentication and integrity protection.
Authentication -> host-based authentication. This does not perform user authentication.
Simple and flexible -> to allow parameter negotiation and to minimize number of round trips. Key exchange method, public key algo, symmetric encryption and others are negotiated.
Most env -> 2 round-trips needed for full key exchange, server authentication, service request and accept. Worst case - 3 round trips.

Connection setup : Works over any clean 8-bit binary-transparent transport. Transport should protect ssh connections against transmission errors. Client Initiates.

tcp/ip - listens to port 22(ssh).

After connection establishment - client and server must send an identification string.

Identification string : SSH-protoversion-softwareversion SP comments CR LF

Protoversion - 2.0, comments - optional, if comments is used, SP should be used to separate softwareversion and comments. Identification - terminated by a single Carriage Return(CR) and single Line Feed(LF). NULL characters must not be sent. Max length - 255 characters including CR and LF.

Server - may send other lines of data before sending version string. Each line should be terminated by a CR and LF. Such lines should not begin with SSH- and should be encoded in ISO-10646 UTF-8. Client must be able to process such lines. If they are displayed, control character filtering should be used. Primary use is to allow TCP-Wrappers to display error message before disconnecting.

Protoversion and softwareversion - consist of printable US-ASCII characters, with exception of whitespace and minus sign. Softwareversion - used to trigger compatibility extensions and to indicate capabilities of implementation.
Example : SSH-2.0-billsSSH_3.6.3q3\CR\LF
Key exchange will begin immediately after sending this. All packets will use the binary packet protocol.

TLS

TLS (2)
Transport Layer Security Protocol - primary goal - provide secure channel b/w 2 communicating peers. Only req - reliable, in-order data stream.

Properties of secure channel :
Authentication - server side always authenticated, client side optional.
Authentication - via asymmetric crypto like RSA, Elliptic Curve Digital Signature Algo (ECDSA) or Edwards-Curve Digital Signature Algo (EdDSA) or symmetric

pre-shared key (PSK).
Confidentiality - Data sent over the channel is only visible to endpoints. TLS does not hide the length of data it transmits, though endpoints may pad TLS records.
Integrity - Data sent cannot be modified by attackers without detection.

TLS has 2 primary components:
Handshake protocol - authenticates communicating parties, negotiates crypto modes and parameters, establishes shared keying material. Resists tampering by attackers.
Record protocol - uses parameters established by handshake to protect traffic. Divides traffic into records independently protected with keys.

TLS - application protocol independent. Does not specify how protocols add security with TLS, how to initiate handshaking or interpret certificates. Left to protocol designers running on top of TLS.

SMTP

SMTP (3)
Objective - transfer mail reliably and efficiently.
SMTP - independent of transmission subsystem and requires only reliable ordered data stream channel.
Important feature - capability to transport mail across multiple networks, referred to as “SMTP mail relaying”.

Network examples: mutually-TCP-accessible hosts on public Internet, TCP/IP Intranet, LAN, WAN.

A process can transfer mail to another using relay or gateway processes. There can be multiple intermediate relay hosts.

SMTP client with a message establishes a two-way channel to an SMTP server. SMTP is responsible for transferring mail to servers or reporting failure.

How messages are presented and domain identifiers are determined is a local matter. Domains may be final or intermediate destinations. SMTP clients that forward all traffic blindly or do not maintain retry queues may conform to the spec but are not fully capable. Fully-capable SMTP implementations support queuing, retrying, and alternate address functions.

1.2 Cluster of Re-Used Keys

The research paper titled *"Cluster of Re-Used Keys"*(4) surveys long-term cryptographic public keys used for TLS and SSH protocols across hosts in ten countries. The hosts examined were primarily those running SMTP, indicating an interest in measuring the security of email and related services. The primary finding is that key re-use is widespread across multiple IP addresses and even across different Autonomous Systems (ASes).

1.2.1 Definition of a Cluster

A **cluster** is defined as a set of IP addresses where each host shares at least one public key with another host in the same set.

1.2.2 Key Re-Use Observations

From a scan of 18,268 hosts in Ireland, approximately **53%** of hosts running a cryptographic service were using public keys that were also observed on another IP address. Out of 54,447 host/port combinations running cryptographic protocols in the Irish scan, only **36%** (20,053 entries) used unique keys—highlighting substantial key re-use.

The scan also identified a total of **1,437 clusters**, with the largest cluster containing **1,991 hosts**.

1.2.3 Security Risks of Key Re-Use

Key re-use introduces several undesirable security and privacy dependencies among cluster members:

Masquerade A breach on any one host allows an attacker to impersonate any other host within the same cluster.

Increased Leak Risk The probability of a private key leak increases significantly; compromise of one host exposes all other hosts sharing the same key.

Credential Capture An attacker masquerading as a legitimate service can capture sensitive credentials such as IMAP or SMTP passwords.

Web Origin Policy Breach If clustered hosts belong to different web origins, a compromise allows attackers to steal HTTP cookies, bypassing the browser's origin policy.

1.2.4 Causes of Key Sharing

Common reasons behind widespread key re-use include:

- A single host being assigned multiple IP addresses
- Redundant mirrored hosts
- Cloned virtual machines containing pre-installed host keys
- Large-scale use of wildcard certificates
- Vendors shipping products with default key pairs
- Operational laziness or misconfiguration by service operators

1.3 Cloud Platform Comparison for Internet-Scale Scanning

1.3.1 AWS and Azure Implementations

Component	AWS EC2 Implementation	Azure VM Implementation
Instance Type	Use network-optimized instances such as C-series (Compute Optimized) or M-series (General Purpose) with high network performance. C5n instances are ideal for network-intensive tasks.	Use high-throughput instances such as the F-series (Compute Optimized) or Mv2-series (Memory Optimized), which provide high networking speeds.
Operating System	A minimal Linux distribution (e.g., Ubuntu Server, Amazon Linux) is preferred. Scanning tools run optimally on Linux.	Same: a minimal Linux distribution is recommended for compatibility and performance.
Tool Installation	Install high-speed scanning tools like ZMap (for rapid port discovery) and ZGrab or `tls-scan` for application-layer handshakes and public key collection.	Same toolchain: ZMap for SYN-ACK scanning, followed by ZGrab or `tls-scan` to perform TLS/SSH handshakes and extract certificates or host keys.

1.4 ZMap Network Scanning Tool

ZMap is a fast, stateless, single-packet network scanner designed for Internet-wide network surveys (5). On typical hardware, ZMap is capable of scanning the entire

public IPv4 address space on a single port in under 45 minutes. For example, it can send a TCP SYN packet to every IPv4 address on port 25 to identify potential SMTP servers.

ZMap supports GNU/Linux, macOS, and BSD, and currently implements probe modules for TCP SYN scans, ICMP probes, and DNS queries.

1.4.1 Performance and Limitations

ZMap transmits packets as quickly as the host network interface allows. Since it intentionally does **not** implement congestion control, two main risks arise:

Target Network DoS Scanning a small subnet at an excessively high rate may unintentionally cause a denial-of-service (DoS). It is recommended not to run ZMap at speeds of 1 Gbps or higher when scanning small networks; instead, keep rates below 10 Mbps to avoid overwhelming the target.

Source Network Overload ZMap may also overload the source network. Some switches and routers cannot handle high-speed traffic consisting of many small packets.

ZMap scans in-scope IP addresses in a random order to reduce localized impact on target networks. The load experienced by a subnet depends on both the configured sending rate and the size of the target IP range.

Potential consequences of overloading a network include:

- network administrators blocking the scanning IP address,
- routers silently dropping scan traffic,
- disruption of local users due to bandwidth starvation.

1.4.2 Execution Model

By default, ZMap uses four threads, unless the host machine has fewer than four CPU cores. The recommended number of threads for ZMap is:

$$T_{\text{required}} = T + 2,$$

where T is the number of sending threads.

1.4.3 Basic Usage

To perform an initial scan that sends TCP SYN packets to all IP addresses in a subnet `xx.xx.xx.xx/xx` on port 80 at a send rate of 128 packets/s, use:

```
sudo zmap -p 80 -r 128 xx.xx.xx.xx/xx
```

ZMap produces two types of output:

1. a list of IP addresses that responded, and
2. periodic scan status messages printed every second.

To save scan results to a CSV file, use:

```
sudo zmap -p 80 -o output.csv -r 128 xx.xx.xx.xx/xx
```

2 Literature Review

2.1 Materials

2.2 Synthetic Procedures

2.2.1 Parameters Varied

2.3 Characterisation Techniques

2.3.1 A

2.3.2 B

2.3.3 C

2.3.4 D

3 Experimental Methods

3.1 Materials

3.2 Synthetic Procedures

3.2.1 Parameters Varied

3.3 Characterisation Techniques

3.3.1 A

3.3.2 B

3.3.3 C

3.3.4 D

4 Results and Discussion

5 | Conclusions and Future Work

6 Figures, Tables, Referencing

It is very important to properly refer in the text to any figures, tables or previously published work that you are discussing. Adequate and consistent referencing is one of the criteria which will be used to assess your project report.

6.1 Figures

Graphs, pictures and other images should be included in your report as a numbered, captioned figure. An example is given in Figure 6.1.

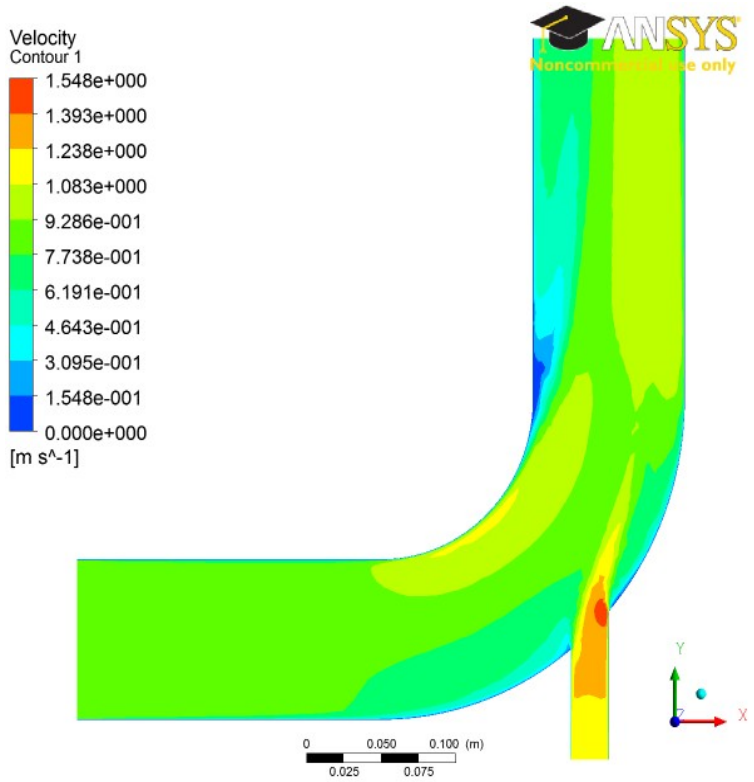

Figure 6.1: Velocity distribution on the mid-plane for an inlet velocity for case 1.

The figure and caption should be centred. The figure numbering starts at 1 at the beginning of each chapter. The caption should provide a brief description of what is being shown. The figure should appear in the document after it is referred to in the text. No figure should be included which is not referred to in the text. Ensure that the size and resolution of images imported from software are sufficient to read any text.

6.2 Tables

Tables are an important way of displaying your results. Table 6.1 is a sample table, adapted from the Master/Doctoral Thesis template at `http://www.latextemplates.com/cat/theses`, which was generated with this code:

```
\begin{table}[b]
\caption{The effects of treatments X and Y on the four groups studied.}
\label{tab:treatments}
\centering
\begin{tabular}{l l l}
\toprule
\textbf{Groups} & \textbf{Treatment X} & \textbf{Treatment Y} \\\midrule
1 & 0.2 & 0.8\\
2 & 0.17 & 0.7\\
3 & 0.24 & 0.75\\
4 & 0.68 & 0.3\\
\bottomrule\\
\end{tabular}
\end{table}
```

Tables are numbered in the same way as figures. Typically tables also have a short caption, but this is not universally true. The number and caption appear above the table, not below as with figures. Again, no table should appear in the report which has not been referred to in the text. Tables should come after they are discussed in the

text. The exact formatting of the table depends somewhat on the content of the table, but in general, the text in the table should be the same font and size as the main text.

6.3 Equations

All equations should be numbered sequentially. Do not restart the numbering at the beginning of each chapter. Unlike figures and tables, you may not need to refer to every equation in the text. You should take care to format equations properly. Do no simply try to use plain text. Use the equation layout facilities. An example of how equations should appear is shown in Equation 1. Here is the code for it:

```
\begin{equation}
\textrm{div}(\underline{u}) = \frac{\delta u}{\delta x} + \frac{\delta v}{\delta y} +
        \frac{\delta w}{\delta z} = 0
\label{sampleequation}
\end{equation}
```

$$\text{div}(\underline{u}) = \frac{\delta u}{\delta x} + \frac{\delta v}{\delta y} + \frac{\delta w}{\delta z} = 0 \tag{1}$$

6.4 Referencing published work

It is important to give appropriate credit to other people for the work that they have shared through publications. In fact, you must sign a declaration in your report stating that you understand the nature of plagiarism. As well as avoiding plagiarism,

Table 6.1: The effects of treatments X and Y on the four groups studied.

Groups	Treatment X	Treatment Y
1	0.2	0.8
2	0.17	0.7
3	0.24	0.75
4	0.68	0.3

citing results or data from the literature can strengthen your argument, provide a favourable comparison for your results, or even demonstrate how superior your work is.

There are many styles to reference published work. For example, the parenthetical style (which is also called the *Harvard style*) uses the author and date of publication (e.g. "Smith and Jones, 2001"). There is also the Vancouver style (or the *citation sequence style*), which is used in this document. In the Vancouver style, the publications are cited using bracketed numbers which refer to the list in the References section at the end of the report. The references are listed in the order that they are cited in the report. A variant is *name sequence style*, in which the publications are referenced by number, but the list is arranged alphabetically. The following paragraph shows the use of the Vancouver style:

> Several studies have examined the sound field around tandem cylinders generated by flow(**?** **?**), while other investigations have focused on the effect of an applied sound field on the flow(**?**). Papers from conference proceedings(**?**), books(**?**) and technical reports(**?**) can be dealt with in the same style.

The Vancouver style has the advantage that it is a little more compact in the text and does not distract from the flow of the sentence if there are a lot of citations. However, it has the disadvantage that it is not immediately clear to the reader what particular work has been referenced.

It actually does not matter which particular referencing style is used as long as three important considerations are observed:

- the referencing style used throughout the document is consistent;
- all material used or discussed in the text is properly cited;
- nothing is included in the reference list that has not been cited.

This template has a suitable referencing style already set up – you should use it and

use the built-in BibTeX system to manage your references. See above for examples of how to cite a reference and look in the `sample.bib` file to see BibTeX references. Remember Google Scholar and other search engines will give you BibTeX references for lots of academic publications. Otherwise, you can easily make up your own based on the examples in that file.

7 | LaTeX

LaTeX, or more properly "LaTeX 2_ε", is a very useful document processing program. It is very widely used, widely available, stable and free. Famously, TeX, upon which LaTeX is built, was originally developed by the eminent American mathematician Donald Knuth because he was tired of ugly mathematics books (**?**). Although it has a learning curve (made much less forbidding by online tools and resources – see below), it allows the writer to concentrate more fully on the content, and takes care of most everything else.

While it can be used as a word processor, it is a *typesetting* system, and Knuth's idea was that it could be used to produce beautiful looking books:

> *LaTeX is a macro package which enables authors to typeset and print their work at the highest typographical quality, using a predefined, professional layout.*[1]

LaTeX has great facilities for setting out equations and a powerful and very widely supported bibliographic system called BibTeX, which takes the pain out of referencing.

Three useful online resources make LaTeX much better:

(1) An excellent online LaTeX environment called "Overleaf" is available at `http://www.overleaf.com` and runs in a modern web browser. It's got this template available – search for a TCD template. Overleaf can work in conjunction with Dropbox, Google Drive and, in beta, GitHub.

[1] This is from **?**). Did we mention that you should minimise your use of footnotes?

(2) Google Scholar, at `http://scholar.google.com`, provides BibTeX entries for most of the academic references it finds.

(3) An indispensable and very fine introduction to using LaTeX called *"The not so short introduction to LATEX 2ε"* by **?**) is online at `https://doi.org/10.3929/ethz-a-004398225`. Browse it before you use LaTeX for the first time and read it carefully when you get down to business.

Other tools worth mentioning include:

- `Draw.io` – an online drawing package that can output PDFs to Google Drive – see `https://www.draw.io`.

Bibliography

[1] Tatu Ylonen and Chris Lonvick. RFC 4253: The Secure Shell (SSH) Transport Layer Protocol. https://datatracker.ietf.org/doc/html/rfc4253, 2006.

[2] Eric Rescorla. RFC 8446: The Transport Layer Security (TLS) Protocol Version 1.3. https://datatracker.ietf.org/doc/html/rfc8446, 2018.

[3] John Klensin. RFC 5321: Simple Mail Transfer Protocol (SMTP). https://datatracker.ietf.org/doc/html/rfc5321, 2008.

[4] Stephen Farrell. Clusters of re-used keys. Cryptology ePrint Archive, Paper 2018/299, 2018. URL https://eprint.iacr.org/2018/299.

[5] Zakir Durumeric, Eric Wustrow, and J. Alex Halderman. Zmap: Fast internet-wide scanning and its security applications. In *USENIX Security Symposium*, 2013.

A1 Appendix

You may use appendices to include relevant background information, such as calibration certificates, derivations of key equations or presentation of a particular data reduction method. You should not use the appendices to dump large amounts of additional results or data which are not properly discussed. If these results are really relevant, then they should appear in the main body of the report.

A1.1 Appendix numbering

Appendices are numbered sequentially, A1, A2, A3... The sections, figures and tables within appendices are numbered in the same way as in the main text. For example, the first figure in Appendix A1 would be Figure A1.1. Equations continue the numbering from the main text.