T802 Research Project

MSc in Advanced Networking

James Bensley

A3255083

12th September 2013

Protocol deployment limits and inefficiencies in ISP networks

Abstract

Throughout this research project, the efficiency, reliability and practicality of popular networking protocols are examined and analysed. Given the age of the most common protocols used on networks and the Internet today; are they still suitable considering the now unprecedented scale of the Internet, can they support the increasing demands of modern day networks for millisecond convergence and fault recovery, do they efficiently transport bandwidth orders of magnitude larger than ever believed possible? These are the kind of questions this research project looks at, studying the original designs and use purposes of Ethernet, IPv4 and TCP (amongst others), comparing them to their modern superseding revisions and increased operational complexity. It also looks at the network services they carry and the demands of present day networks, investigating where the protocols fall short of their requirements, how we can detect this, and how these scenarios can be rectified or mitigated. This is achieved by first compiling a baseline from protocol standards, guidelines and deployment statistics. Research on known existing protocol issues is brought under investigation to assist in looking for protocol design under sight and efficiency limits. The open views of the networking industry are then added in for a more qualitative insight. Experimental research builds upon these ideas to ratify them and create conclusive data. Conclusions are drawn from the research findings which provide a clearer picture on the current operational status of the key networking protocols and their optimal deployment scenarios.

Contents

List of figures

List of tables

Glossary

Term	Definition
ARPA/DARPA	(Defence) Advanced Research Projects Agency – A military research agency in the USA that developed early computer networks
DF Bit	Don't Fragment Bit – A binary marker on an IP packet to signal the data contained can not be fragmented
DSCP	Differentiated Services Code Point – A marker within an IP packet that describes the quality of service a packet should receive
Ethernet	A software protocol for communication between neighbouring devices
ICMP	Internet Control Messaging Protocol – Control messages sent and received between two communicating hosts to manage connectivity
IEEE	Institute of Electrical and Electronics Engineers – A professional association dedicated to technical innovation and standardisation
IP	Internet Protocol – A software protocol for addressing and transportation of data between devices
IPSEC	Internet Protocol Security – A suite of mechanisms used to encapsulate IP traffic to provide various security measures
Jitter	The undesired deviation from true periodicity of an assumed periodic signal
MPLS	Multi-Protocol Label Switching – An encapsulation method for data communication to allow traffic behaviour engineering
MSS	Maximum Segment Size – The largest amount of data TCP can communicate in a single packet
MTU	Maximum Transmission Unit – The maximum amount of data transportable in a layer 2 Ethernet frame, excluding headers
OSI Model	Open Systems Interconnect Model – A design model used for inter-system communication
PDU	Protocol Datagram Unit – The relative name for a data unit sent by any protocol at any layer of the OSI model
PMTUD	Path Maximum Transmission Unit Discovery – A technique using ICMP packets to discover the MTU on the network path between two remote devices
QoS	Quality of Service – A traffic prioritising scheme to ensure important data is delivered during network congestion
RFC	Request For Comments – Public documents that request comments and scrutiny before progressing to become standards or accepted best practices
RTT	Round Trip Time – The time taken for data to be sent from one device to another, and back again
RWIN	Receive Window Size – The maximum amount of data a receiver of a TCP connection can receive without acknowledging the sender
TCP	Transmission Control Protocol – A software protocol for reliable communication of data between devices
UDP	User Datagram Protocol – A software protocol for unreliable communication between devices

Acknowledgements

I would like to thank Eddie Bennett, my tutor and supervisor throughout this project. Without his instructions, explanations, criticism and expert reviews I would not have been able to plan this project with any sense of order and direction, let alone execute and complete it. Additionally, extended special thanks are owed to my two editors in chiefs, colloquially known as ‘Mother’ and ‘Father’. Without these two experts of written English this project could not have been written clearly or concisely.

Additionally I am thankful and appreciative to all those who filled out my public survey, in particular author Jon Day for taking the time to reach out to me.

1. Introduction

1.1 Background to the problem/issue

The earliest recorded deployment of Ethernet I have found in a commercial environment dates to 1976, by its co-inventor Robert Metcalfe *(Metcalfe, Boggs, 1976)*. IP and TCP shortly followed: in 1980 the first IPv4 RFC was published *(DARPA IPTO, 1980)*. Over 30 years after the initial adoption of these now fundamental networking protocols, despite their revisions over the decades, I pose the question; *Do they hinder performance and efficiency of the present day networks they support, considering the changes in networking demands, design and operation, over the years?*

Many networking protocol related issues are already common knowledge. Table 1.1 below shows the lower layers of the ISO's OSI Model. It lists potential issues to networking that can arise for common protocols at each layer. It is not a comprehensive or complete list, yet it shows a significant number of issues related to deployment, configuration, interoperability and unintended use; that exist for common networking protocols:

Table 1.1: ISO OSI model lower layer issues

OSI Layer	**Potential Issue**
2 – Data Link	MTU size mismatch, PMTUD failure, unused or invalid Class of Service markings, Ethernet loops, Auto MDIX mismatch, IEEE802.1ad/ah tunnelling overhead, interface buffer depths, in-order delivery (Ethernet frames), label stacking (for MPLS), load-balancing (for MPLS and Ethernet)
3 – Network	Max packet fragment size, unused or invalid DSCP/QoS markings, jitter, high RTT, buffer burst limits, unused or invalid DF bit making
4 – Transport	MSS limit (for TCP), RWIN size (for TCP), window scaling (for TCP), bandwidth delay product, congestion avoidance algorithm efficiency, out of order packets, bufferbloat

The frequency of potential issues and the severity of impact are both circumstantial to the network services these protocols are providing for. Some common and severe issues include path MTU discovery (Luckie, Cho, Owens, 2005) causing data fragmentation and retransmission (for the user this manifests as slow or seemingly no connectivity), and bufferbloat (Cerf, Jacobson, Weaver, Gettys, 2011) causing packet delays and jitter (again manifesting as slow or seemingly no connectivity for the user).

These examples above can reduce the quality of a connection to the point where it appears that there is no longer connectivity at all. In such a scenario drastic and immediate action is often taken. Some issues though, can happen just infrequently enough to be difficult to diagnose, and yet often enough to be distressing for valued network customers. The protocol overhead of the IEEE802.1ad standard ("QinQ tunnels", typically used for connectivity reselling) or VPNs over DSL (for secure remote access), can cause intermittently poor application performance. This means that network services are mostly accessible but slower or less responsive, which impacts user experience or employee efficiency. NAT session time-outs and UDP and TCP keep-alive timers (used to keep track of active connections passing through a network) can cause periodic issues. When these timers expire prematurely, a working network application can cease to function. These are indications that protocols are being stretched to their operational limits.

A further understanding of how my opening question might indicate the potential benefits that can be achieved, and the coverage of those benefits to networking as a result, can be better demonstrated by examining the time line in Figure 1.1 below. Figure 1.1 shows a time line illustrating the growth of the Internet wherein the same few protocols are shown in repeated use. Events underneath the X axis time line relate to the Internet infrastructure and depict early networks forming; protocol adoption and the growth of the global routing table. This raw data format does not however examine how widespread protocol adoption is, nor the vastness of the Internet. Events above the X axis line are taken from a minuscule portion of networks that have increased to colossal proportions, yet still only represent a few crossroads of the Internet today. From these statistics we can surmise the inarticulate magnitude of the Internet today when compared to the original networks of the late 1960's and early 1970's, when Ethernet, IP and TCP were first conceived and deployed.

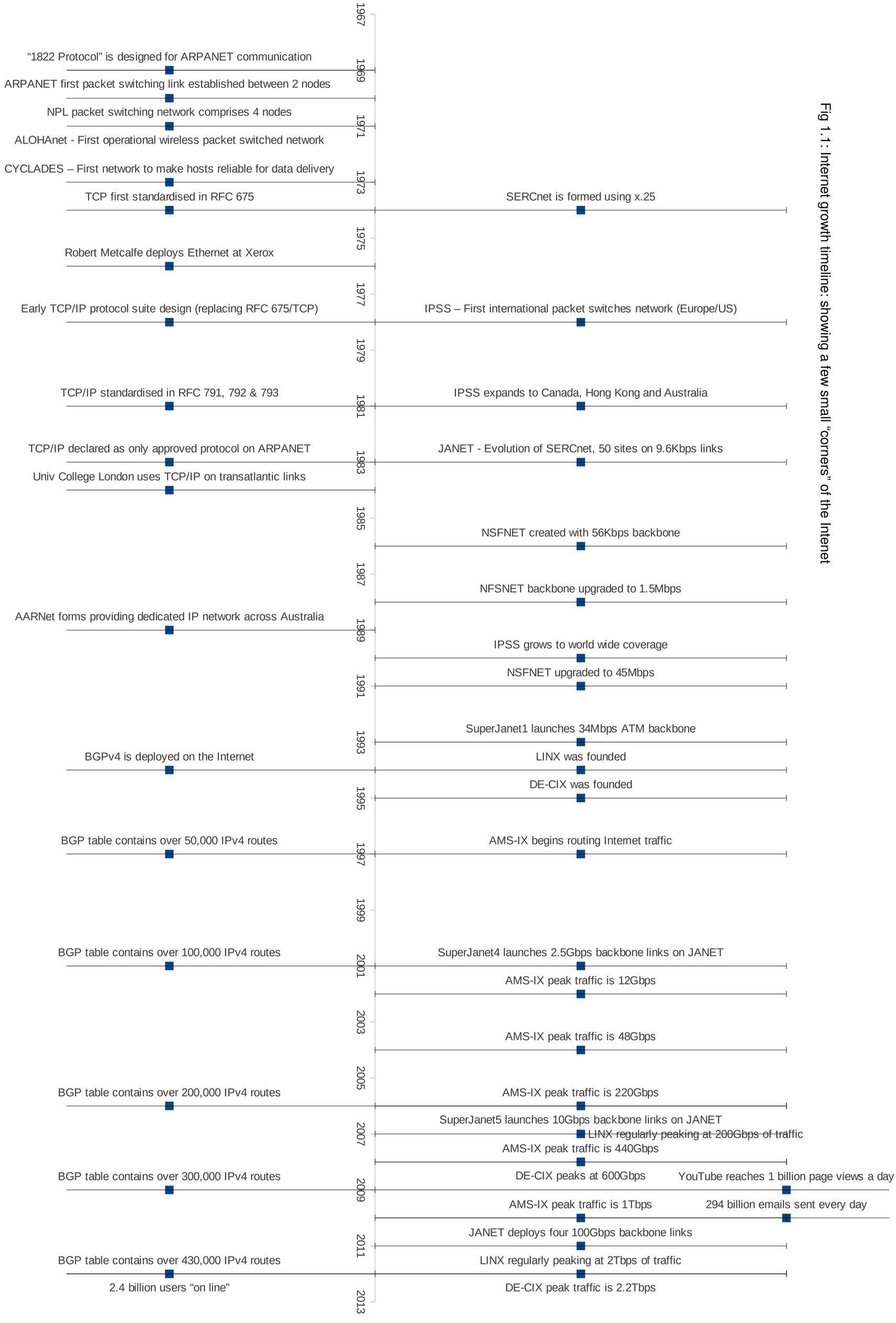


Fig 1.1: Internet growth timeline: showing a few small “corners” of the Intenet

1.2 Justification for the research

The same protocols are reused to meet almost all networking requirements; their functionality is fixed but their usage requirements change. Considering the age of the most commonly used protocols and how prevalent their deployment is, I question if the original designs allow for global adoption and adaptation as required for today's Internet driven world?

In his 2008 book Patterns in Network Architecture - A Return to Fundamentals, on the subject of protocol efficiency and design, John Day, a 1960's ARPA junior who had a hand in the early computer networks and the Internet states;

I have often said, only half jokingly, that "the biggest problem with the ARPANET was we got too much right to begin with." Meaning that for a project for which there had been no prior experience, for which there was considerable doubt it would even work, there was some brilliant work and some brilliant insights to the point that it was "good enough," and there was no overwhelming need to address the problems it did uncover…As one would expect with any first attempt, some were mistakes, some things were unforeseen, some shortcuts were taken, some areas went unexplored, and so forth.
(Day, 2008a).

There has been a considerable amount of research on protocol shortcomings, and there are a considerable amount of protocols to research. In most cases this research has focused on individual protocols, with a very specific subject scope, often overly technical in delivery, and often difficult to understand and act upon. This research will be presented in a precise and detailed manner, yet easily interpretable, covering multiple issues, as an overarching networking guidelines document. This will provide a better understand of the common networking protocols, and to also find and improve service degradations on networks they manage or maintain.

There are financial benefits to be gained from the research for commercial networks like ISPs. An example of this would be a network inefficiently delivering customer data (reconfiguring interface buffers for example could avoid packet drops at congestion points) or by allowing them to serve more customers (introducing jumbo frames into the network core to pass more customer data). An end-to-end QoS model for example can be used to ensure minimum service level guarantees for their users; additionally it becomes a marketable product for the service provider *(Xiao, Ni, 1999)*.

With a growing demand for voice and video communication over the Internet *(Cutler, S. 2012)*, the end business users' telecommuting or teleconferencing experience would benefit from improvements made to their connectivity. This can affect employee and user productivity.

For domestic users of an ISP, the UK government is working towards the delivery of 'broadband' Internet connectivity to all home users, but only at a 2Mbps minimum (*BBC. 2012a, BBC. 2012b*), so the quality of service they receive is paramount. Any improvements in congestion, queuing or latency issues that arise in typical TCP and UDP flows for example, would be beneficial to home users. Real-time entertainment services such as Netflix and YouTube are now the single largest sources of Internet traffic within Europe and North America for fixed line (non-mobile) Internet access customers *(Sandvine, 2012, p17)*. This means that at the network edge, advancements in protocol operations like bufferbloat, optimal QoS schemes and higher access layer connectivity speeds are becoming more beneficial than ever, and more in-demand than ever.
This all shows that end user related protocol issues are frequent and their severity

ranges from minor to majorly service affecting. Whilst I focus on all parts of the network (not just the access layer, or aggregation layer, distribution, core etc) I consider any improvements in the access layer to be more widely beneficial, as this affects the majority of end users.

Ensuring a high calibre of network user experience has strategic benefits for UK and European networks relating to continental and global commerce, technical leadership, and industry social status. This was shown by Kaufmann *(2012)*, when networking played a pinnacle role in the UK hosting of the 2012 Olympics. Akamai Technologies Inc streamed 9,300 years of video in two weeks and peaked at 873Gbps of concurrent traffic. This has possibly guaranteed their position on future content delivery contracts, and likely forced the BBC iPlayer R&D and Akamai engineering departments to create some market leading innovations. Despite being partial speculation, if true, how much of their efforts could have been saved by protocol improvements?

1.3 Definitions

The following list details key concepts and their meanings, discussed throughout this research project;

- Protocol efficiency: this terminology describes the ability of a protocol to effectively transporting data between two network hosts, with regards to time taken and computing resources used, when operating within optimum conditions.

- OSI model layers: each layer of the OSI model provides a different function, which work in tandem to provide end-to-end connectivity between to network hosts. Different protocols operate at different layers of the connectivity model, to provide the variety of required functions.

- Optimum operating conditions: protocols can perform more or less efficiently depending on the conductions under which they must operate.

1.4 Scope and outline

The research is focused on the software protocols that operate on networks. Hardware is an equally complicated and vast topic; it could form a research project of its own. With adequate funding software does not usually produce the same limitations hardware does. Hardware can be replaced, upgraded, or decommissioned. Despite this there is some overlap into hardware. Section 2.1 below discusses who, what, when, where, why and how with regards to the context of the research topic in detail.

This research project covers issues that affect typical present day networks running the newest versions of the most common networking protocols, at the lower layers of the ISO OSI Model (ISO/IEC7498-1). These are networks that process Ethernet at layer 2, IPv4 and IPv6 at layer 3, and TCP, UDP and ICMP at layer 4. These are identified empirically as the most common protocols by Dhamdhere *(2003)*, Labovitz *(2011)* and Adhikari et al *(2012)*. Also on topic atop these protocols shall be; The smart edge, QoS, NAT, and VoIP.

I have not included any research on the physical layer (OSI layer 1), concentrating my research on networking experiences that occur over wired copper and fibre mediums. Mobile and wireless technologies warrant entire research projects of their own. Security implications from running certain protocols are also out of the scope of operational performance, due to the size of the subject matter.

When looking at the effects of inefficient protocol operations, I do not detail the outcome for a specific application. Also I do not demonstrate the detection and identification of specific protocol inefficiencies, through demonstration of a given method or technique. This is due to the vastness and variation of techniques possible. I detail the protocol issue and the effect on the upper OSI layers only. Pursuing application specific outcomes is off topic for the lower layer protocols.

1.5 Project outline

The remainder of this dissertation is structured as follows;

Chapter 2 discusses some of the protocol issues explored during the research, describing the context in which they occur and their technical effects. It continues on to discuss business effects and operational impacts on networks, detailing how this differs at the various OSI layers.

Chapter 3 details the project aim, objectives, research questions and required data sets. A description of the methods and techniques chosen to meet these various requirements follows. Procedures used to carry out these methods and techniques are also chosen and evaluated. A description of the data gathering processes follows the procedural review. Finally any ethical issues the research project may encounter are considered.

Chapter 4 begins with a summary of results from the data gathering process. It then presents an initial analysis of the data findings and finally presents an interpretation of the data in line with the research aim and objectives.

Chapter 5 presents an evaluation and conclusion of the data gathered in relation to the research aim and objectives. It also looks at the wider implications of the data gathered within the knowledge market and the significance of its findings. It continues on to discuss potential future work in the same topic as well as other potential work in related topics. It concludes with reflection and criticism of this dissertation.

2. Research Definition

2.1 The practical problem/issue

I theorise that a combination of the specific hardware device on which a protocol operates, the specific software protocol in use, and the pairing of the hardware and software combination, is what determines the type and severity of degraded service that can be delivered to a network application or user. Additionally, to a certain extent this will include where the issue can occur and the frequency of occurrence. The following points explain the problem in context detailing who is affected, the issues raised, additionally when, where, why and how they occur;

- From a business perspective the issues should be obvious in extreme cases such as streaming media not working during peak hours on a network because of congestion, or the breakdown of voice and video on conference calls because of NAT. The audience of affected users of the issues is almost anyone that uses a computer network, not just the obvious commercial or academic uses. Less obviously, public file mirrors for example need to be able to saturate their uplink's capacity to be more efficient at their purpose, which may require an alternative configuration from the default or standard. Transactional operations may require millisecond delays across a network, which might not be sustainable when using traditional path based (link state or distance vector) routing across the network topology, when compared to constraint based routing (based on live protocol measurements from across the network).

- The concept of a protocol being inefficient or susceptible to deployment issues is not new, and much research has already taken place on this subject. At the network edge, research into severe issues is regularly published which shows what protocol inefficiencies, design blunders, and practicality short comings are prominent. TCP performance over a 3G connection *(Chan, Ramjee, 2005),* which if used with streaming media for example is often unwatchable, is shown to be a result of the design specifications for TCP's reliability mechanisms. Deploying NAT devices in the path of VOIP clients *(Khlifi et al, 2006)* typically causes one way audio, or a complete loss of registration. NAT has contributed to the longevity of the IPv4 addressing scheme, but as it is deployed two fold with Carrier Grade NAT deployments rising, its overuse today is now a source of end-to-end protocol connectivity breaks *(Maennal et al, 2008)*.

- For around 35 years, researchers, academics and engineers have discovered various issues like the above and produced research to demonstrate this. Their research has shown that protocol issues are experienced far and wide across all areas of a network. The following is a list of RFC standards and their year of publication that define the minimum operational requirements (excluding extended functionality), just for TCP: RFC793 1981, RFC1122 1989, RFC1323 1992, RFC2581 1999, RFC2675 1999, RFC2873 2000, RFC2988 2000, and RFC3168 2001 *(Duke et al, 2006)*. Networks today are operating twenty four hours a day, three hundred and sixty five days per year, with most ISPs boasting a continually staffed Network Operations Centre for non-stop support and fault resolution. The time scales for when problems can arise in a network has now become "anytime", and using the RFC documents listed above as a rough guide, work to improve protocols is seemingly never ending. This research focuses on the most up to date implementations of networks and their protocols.

- Similar issues often affect many users because they are the result of networks having a large and varying scope of access mediums and technologies (PPPoA, PPPoE and L2TP over DSL and DOCSIS, 3G and 4G/LTE, Wi-Fi, and so on), meaning they occur regularly in today's heterogeneous networks. The core of a network typically remains unaffected due to industry best practices dictating that QoS, content filters, access management, and security filtering; all take place at the access layer of the network (the "smart" edge). This means that issues affecting end users typically occur at the smart edge due to the complexity there. This research project excludes the physical layer/access medium due to the wide and varying range of layer 1 technologies. It focuses mainly on the smart edge.

- John Day *(2008b)* reflecting on his 35 years as a computer scientist and Internet pioneer questions, why we have these issues today. He concludes that early researchers and engineers not pursuing the upper layers of the OSI Model more thoroughly removed the driver that would have forced them to "*complete the Internet architecture, force a more complete addressing architecture, force a greater attention to security, more requirements on the lower layers, and probably many other things*". He states that early networking growth through demand alone was not enough to drive more detailed protocol stress testing and research. Before networking hardware and the software protocols atop could be stretched to breaking point, Moore's Law ensured that processing and capacity improvements easily evaded such scenarios *(Moore, 1965)*

- Researchers and engineers extensively examine how problems arise and how they affect networks. The end result is often a modification or extension of a protocol to overcome identified issues. These modifications are mainly being deployed at the access layer of a network. It is here that the flexibility of a protocol is stretched to its limits. An example of protocol modification is the various TCP congestion avoidance algorithms developed to react to packet loss in different ways other than TCP's original design. Kurose and Ross *(2000)* described such modifications like TCP Tahoe, TCP Reno and TCP Vegas. As well as protocol improvements, due to the severity of some issues, complete protocol alternatives exist such as Myrinet and Infiniband. These are used as switching fabric technologies within data centres to consistently achieve lower latency and higher throughput than traditional Ethernet (IEEE 802.3). Both are shown as superior to Ethernet by Larsen et al. *(2009)*.

2.2 Existing relevant knowledge

Some of the problems in Table 1.1 can occur in default configurations of network devices, which means network devices that are deployed without being configured or "tuned" for their specific purpose *(Cisco Systems, 2009)*. Some of the issues might be more frequent to specific situations such as satellite links or wireless LANs *(Huang, Chien, 2004)*. Additionally some issues could simply be the result of protocol or device design under sight *(Touch, Perlman, 2009)*. This indicates there is few or no lower layer networking protocols that are suitable for all areas of networking requirements at the application/user layer.

It is widely accepted that the most common networking protocols forming the majority of network traffic today are TCP, UDP, ICMP, IPV4 and Ethernet (and increasingly IPv6). This is backed up by the efforts of Team Cymru *(2013a)*. These are the protocols that are repeatedly used to support almost all application/user layer activities. As shown below in my findings relating to each layer and these protocols, this introduces many problems at higher networking layers.

Layer 2 Issues:
A High Performance Computing deployment paper by Larsen et al. *(2009)* focuses down into the end hosts' inner workings. The authors inform the reader that "latency is orthogonal in definition to throughput" *(Larsen et al, 2009, p558)*. They are wisely stating that these are two separate issues that should be addressed as such. This paper includes some thorough information on Ethernet alternatives (Myrinet and Infiniband). The authors clearly show that both alternatives are more efficient than common place Ethernet in their HPC environment. They discuss a key feature of Ethernet which is MTU size and how this can affect latency and throughput when a mismatch occurs between two devices, due to fragmentation.

A paper by Elmeleegy, Cox, Eugene *(2009)*, discusses the count to infinity issue of Ethernet convergence when using the Rapid Spanning Tree Protocol. Part of its main focus is the lack of a TTL field in Ethernet headers. Over time this has become more troublesome. Initially, having no TTL field wasn't a major issue. Network loops were being prevented at layer 3; but today we have large scale layer 2 networks with many redundant links. An example of layer 2 scaling that was hindered, partly by Ethernet issues, was the London Internet Exchange switching from large scale layer 2 infrastructure to VPLS *(Cobb, 2012).*

This research shows that the Ethernet protocol has various issues if deployed inappropriately, including throughput inefficiencies and scaling flaws, among others. Configuration care is required on the hardware running these protocols and measurements are required to ensure operating efficiency and reliability.

Layer 3 Issues:
An article by Wu et al. *(2005)* compares TCP and UDP performance over IPv4 and IPv6. IPv4 was shown to be more efficient than IPv6 for both TCP and UDP transmission (demonstrated as having a higher throughput at fixed transmission speeds). One of the reasons Wu et al. identified for this was packet overhead. IPv6 resolves issues such as the IPv4 address shortage, and improves end-to-end QoS, but the headers are twice the size (IPv4 are at least 20 octets and IPv6, at least 40 octets). This requires double the processing power by networking equipment, which can also increase latency. An additional "design" burden from IPv6 adoption is device memory exhaustion. IPv6 routes use more memory in routing tables because of the longer address and mask pairs.

As shown by Floyd and Jacobson (*1993*), the ECN (Explicit Congestion Notification) bits can be set within IP headers at layer 3, to inform neighbouring devices of traffic congestion ahead. ECN itself has since been improved upon with AQM (Active Queue Management) to act more probabilistically than ECN, which is harsh in traffic management. AQM has in turn been revised and there are now many variations, RED (Random Early Detection) is one of the most popular and effective.

At layer 3, efficiency issues continue to rise. The level of service degradation experienced here is equal to that at layer 2 (more latency can be added in, poor congestion control affects almost all TCP flows, the overhead of TCP flows reduces throughput). Also as shown by the CAIDA report *(CAIDA, 2010a)* and Dhamdere *(2003)*, the ratio of UDP to TCP traffic is always growing. A possible sign of network changes such as increased stability and lower packet loss?

Layer 4 Issues:
Two recurring common topics in existing research are TCP congestion and TCP protocol overhead. When DCTCP (Data Centre TCP) performance is compared to TCP, Alizadeh et al. *(2010)* showed that TCP queue lengths fluctuate due to interface buffer depths, thereby causing additional delay. Also, the infamous RED algorithm will "cause wide oscillation in queue length" *(Alizadeh et al, 2010, p9)*. These issues have an effect on latency of data delivery (jitter in particular).

Jacobson and Nichols *(2012)* are tackling the long standing Active Queue Management problem. Their work on the new CoDel algorithm can be deployed right across the internet edge to benefit a majority of end users. They have shown that their AQM algorithm is more effective than the traditional Random Early Detection algorithm, and the well known Tail Drop algorithm, and the additional benefits CoDel can have.

All of this research into layer 4 issues, like most investigations into layer 4, targets TCP. As the most used protocol on the Internet *(Team Cymru, 2013a)*, TCP is under constant research scrutiny and protocol revision. It is evident from the amount of research available that TCP has both design inefficiencies and implementation inefficiencies. As before, this fourth layer of protocols has more impact on user experience than the previous layer.

All the existing research and knowledge shows that protocol related issues are widely accepted as a problem and continual work is happening to investigate and mitigate them. It also shows that research already carried out has achieved a positive outcome providing significant improvements for user experiences and operational efficiency of networks.

2.3 Aims, objectives, tasks and deliverables

This research is targeted at Internet and Telephony Service Providers (ISPs/ITSPs) who deliver data and voice and video services over converged IP infrastructure. The protocol focus is on the latest revisions of those above, discussing bleeding edge issues or existing issues that remain unresolved or insufficiently resolved.

I believe these criteria ensure the research can benefit the highest number of network users possible. Improvements to “The Internet” could be beneficial to everyone.

I have a single comprehensive research aim, which is as follows;

Propose methods for identifying and removing limitations to networking protocols used in the delivery of data across a modern network. The outcome will be to more efficiently serve network applications running atop of those protocols.

Table 2.3.1: Research objectives, questions and data sets: Fulfilling the data sets shall answer the research questions. The question output meets the research objects.

Research Aim: *Propose methods for identifying and removing limitations to networking protocols used in the delivery of data across a modern network. The outcome will more efficiently serve network applications running atop of those protocols.*		
Objective	**Question**	**Data Set / Deliverable**
O1. Identify what networking protocol inefficiencies are and how they can be recognised	Q1. What is considered inefficient communication of data?	D1. Present a table of identified popular network services and protocols, and their optimal operating considerations to maximise their productivity
	Q2. What is the effect of inefficient data communication?	D2. Present a table of the effects inefficient data communication has on identified services and protocols
O2. Research potential mitigation strategies suggesting additional ideas	Q3. What is a satisfactory mitigation, and what does it consist of?	D3. Create a matrix of required improvements that will bring a network service to an acceptable working level, alongside network inefficiencies to show their relationship
	Q4. What are the mitigations for the identified network inefficiencies?	D4. Produce a list of mitigation techniques identified through primary and secondary research

3. Proposed methodology

3.1 Methods and techniques chosen

To meet the requirements of my research questions in Table 2.3.1 I have chosen the methods and techniques shown in Table 3.1.1 below. The start to finish research work flow is shown here in Figure 3.1.1. It shows a breakdown from research aim, into objectives, with objectives broken down into research questions, each question is fulfilled by a given data set or deliverable, which comes from a research method, which is conducted using an appropriate research technique:

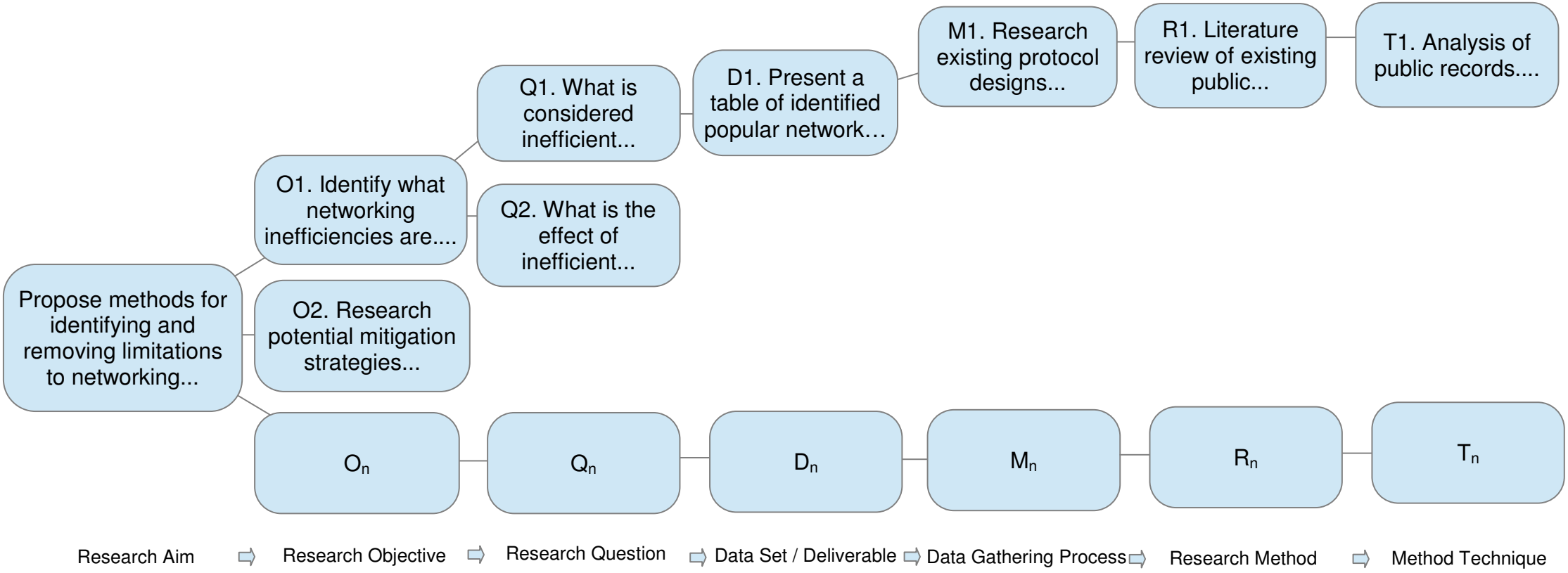


Figure 3.1.1: Work flow to fulfil the research aim; Working from left to right across the vertical columns, to satisfy the requirements of each research objective, and from top to bottom working down through each row (research objective) one at a time to fulfil the overarching research aim

Table 3.1.1: Research methods and techniques shown here are used to meet data requirements from Table 2.3.1. They follow the workflow set out in Figure 3.1.1 and provide an agenda for the research to follow:

Data Set Required / Data Gathering Method	**Proposed Research Method**	**Proposed Research Technique**
D1. / M1. Research existing protocols and their designs and deployment best practices	R1. Literature review of existing public standards documents and existing research	T1. Analysis of public records for protocol knowledge baseline
D2. / M2. Research existing literature on poor protocol performance outcomes	R2. Review research of existing literature and public documents	T2. Use secondary material for information gathering and statistics collecting
D2. / M3. Conduct a survey gathering communication issues experienced by network operators	R3. An empirical method using a survey for information discovery	T3. A basic survey distributed as a questionnaire amongst industry professionals electronically, and amongst colleagues and peers as an unstructured interview
D2. / M4. Conduct experimental research to measure poor protocol performance	R4. Experimental abstracted research	T4. Laboratory research to reliably and repeatedly gather data
D3. / M5. Review and compare protocol guidelines and best practices	R5. Research of existing literature and public documents	T5. Analysis of public records for information gathering and statistics collecting
D4. / M6. Research existing literature and protocol guidelines for detection and mitigation techniques	R6. Review research of existing literature and public documents	T6. Use secondary material for information gathering
D4. / M7. Conduct experimental research to demonstrate and suggest additional techniques	R7. Experimental abstracted research	T7. Laboratory research to reliably and repeatedly gather data

3.2 Justification

3.2.1 Accepted and Rejected Methods and Techniques

My three main techniques for data gathering are research review of existing public data, surveys of industry professionals asking for the specific data I require, and laboratory testing with networking equipment to interrogate gathered data.

- Research reviewing and secondary research allows me to create a knowledge baseline and build ideas on top, backing them up with statistics. This is achieved by a combination of re-interpretive reviewing to gather data, and some theoretical reviewing to synthesise new ideas from the existing data. To achieve this, I reviewed and analysed public records such as protocol standards and best practice guides, company records such as R&D department publications and conference talks on project experiences, and secondary material such as existing research papers and journal articles. By successfully exploring all these avenues, research reviewing has provided a large amount of data and knowledge for my research. When executing these techniques, M1, 2, 4 & 5 from Table 3.1.1 above, are completed back-to-back in a continuous process. This saves time, when compared to a discontinuous approach of staggered researching reviewing.

- Surveys allow me to gain additional data on specific ideas I chose; to fill in knowledge gaps or to extend or challenge established ideas. A focus group is difficult to organise and arrange (participants are required to all be available at the same time, and possibly place), and a Delphi style group discussion wouldn't yield as much data validity and accuracy (participants can start to tangent from the subject matter for example, or argue even). Structured interviews are again, difficult to organise and even harsher on time constraints. An online survey has the best ratio of efficiency and data accuracy in my opinion. The survey method allowed me to create a questionnaire with open and closed questions as I required, gaining very specific quantitative information, and more qualitative information, as I needed. A questionnaire is both a discovery process, to see what “the norm” is within the industry, and a cross-sectional technique looking at present day issues to attain cutting edge data. A survey is also a time efficient data gathering technique for my research

- Laboratory testing allowed me to compare and contrast the survey gathered data with my review research data, challenging conflicts or alignments between the two. I can carry out data gathering experiments in the expected home of a phenomena under investigation, as well as in an abstracted environment to observe alternate outcomes. This provided me with a triangulation approach between my main research methods, cross-referencing data for final evaluation. A technique like modelling and simulation is far too time consuming and over specific for my list of aims and objectives. I have made quick, easily repeatable, accurate tests for data validation. Observation and measurement test scenarios could have been possible, but it could fail to provide a sufficient data set size or validity in the given time period. A laboratory scenario allows me to efficiently repeat tests until enough data is gathered. This testing has allowed me to verify and challenge as required, discrepancies or alignments between literature and research review, or survey gathered data.

3.2.2 Method and Techniques

Table 3.2.2.1 below lists the research techniques used to satisfy the data set requirements from Table 3.1.1. It then shows the justifying attributes of each technique selected. Table 3.2.2.1 shows that each research technique meets the requirements of construct validity, internal validity, external validity and data validity, noting any additional details that might be needed to ensure this:

Research Technique	Construct Validity	Internal Validity	External Validity	Data Validity	Additional Justification
R1. / T1. Analysis of public records for protocol knowledge baseline	Fully met	Fully met: Avoid selection bias by scrutinising statistics and look for impartial bodies	Fully met	Fully met	Protocol specifications and existing publications on measured communication performance will provide an easy to scrutinise baseline
R2. / T2. Use secondary material for information gathering and statistics collecting	Fully met	Fully met	Fully met: Ensure criticality of reviewing of literature	Fully met	Existing public research on communication degradations shall provide a partial indication of networking issues
R3. / T3. A basic survey distributed as a questionnaire amongst industry professionals electronically, and amongst colleagues and peers as an unstructured interview	Fully met	Fully met: Caution is required to ensure a wide distribution and appropriate target audience	Fully met	Fully met: Distribute a test survey first to refine data gathering accuracy	A survey shall allow me to gather further information after T2 on issues, to discover additional information not found in T2 that is more qualitative.
R4. / T4. Laboratory research to reliably and repeatedly gather data	Fully met	Fully met: Caution is required to ensure accuracy and reliability of results	Fully met: Make available raw data set of results	Fully met	Laboratory research shall confirm or deny networking issues, and potentially extend researched issues identified in T2 and T3, verifying and completing the research output up to this point R4/T4

Table 3.2.2.1 continued:

R5. / T5. Analysis of public records for information gathering and statistics collecting	Fully met: Ensure the external literature data is relatable	Fully met	Fully met	Fully met	A contrasting evaluation here of protocol guidelines would allow me to find any trends in best practices or recurring operational targets
R6. / T6. Use secondary material for information gathering	Fully met: Ensure the external literature data is relatable	Fully met	Fully met	Fully met	Cross referencing documentation and research on fault mitigation techniques with protocol guidelines will form a strong sub-set of mitigation techniques
R7. / T7. Laboratory research to reliably and repeatedly gather data	Fully met	Fully met: Caution is required to ensure result accuracy and reliability	Fully met: Make available raw data set of results	Fully met: Ensure experiments stay within the scope of the research	Testing of compiled mitigation techniques from above T3 & T4 to form a new compilation of the most effective and efficient to deploy will require methodical laboratory driven testing

3.3 Research procedures

3.3.1 Procedure review

Research objectives, questions, tasks, methods, and techniques are outlined above to complete the research aim. At the lowest level of the research requirements, the following research procedures tie these research components together to fulfil the required data sets outlined in Table 2.3.1. In Table 3.3.3.1 below I have taken columns from Table 2.3.1 and Table 3.1.1, and added a column to explain and justify how the research procedures fulfil each data requirement:

Table 3.3.3.1: Procedural explanations for research methods

Data Set / Deliverable	**Method**	**Procedural Explanation**
D1. Present a table of identified popular network services and protocols, and their optimal operating considerations to maximise their productivity	M1. Research existing protocols and their designs and deployment best practices, include metrics of acceptable performance	This baseline research was partially completed already, as part of the research proposal and justification. By simply continuing existing research I formed a more defined starting point, from which the research project launched
D2. Present a table of the effects inefficient data communication has on identified services and protocols	M2. Research existing literature on poor protocol performance outcomes M3. Conduct a survey gathering communication issues experienced by network operators M4. Conduct experimental research to measure poor protocol performance	After gathering data through review research on protocol operational statistics and performance reports, a survey of network operators on their protocol failure identifying and troubleshooting experiences, and experimental research to back up these findings; I am able to cross reference this data into a table, which with some background explanation and evaluation leads to the completion of one research objective: “O1. Identify what networking protocol inefficiencies are and how they can be recognised”
D3. Create a matrix of required improvements that will bring a network service to an acceptable working level, alongside network inefficiencies to show their relationship	M5. Review and compare protocol guidelines with best practices	This comparative research of published information and my findings from D2. above, merge to produce a data matrix of protocols and services, identified issues, and what is required to mitigate or resolve the issues

Table 3.3.3.1 continued:

D4. Produce a list of mitigation techniques identified through primary and secondary research	M6. Research existing literature and protocol guidelines for detection and mitigation techniques M7. Conduct experimental research to demonstrate and suggest additional techniques	A combination of review research for tried and tested techniques, and experimental research to further the data gathered, shall culminate in a data set which coupled with D3 & M5 fulfils another research objective: “O2. Research potential mitigation strategies suggesting additional ideas ”

3.3.2 Survey Data Gathering Procedures

I created an on-line submittable survey using the free online service provided by Google Inc, called Google Docs, located here: http://docs.google.com. The public URL of my running survey is: https://docs.google.com/forms/d/1lqigAHYHEgLLHr2kifiyBwgJ9Nw5AFS6d_XVXfhKkTw/viewform.

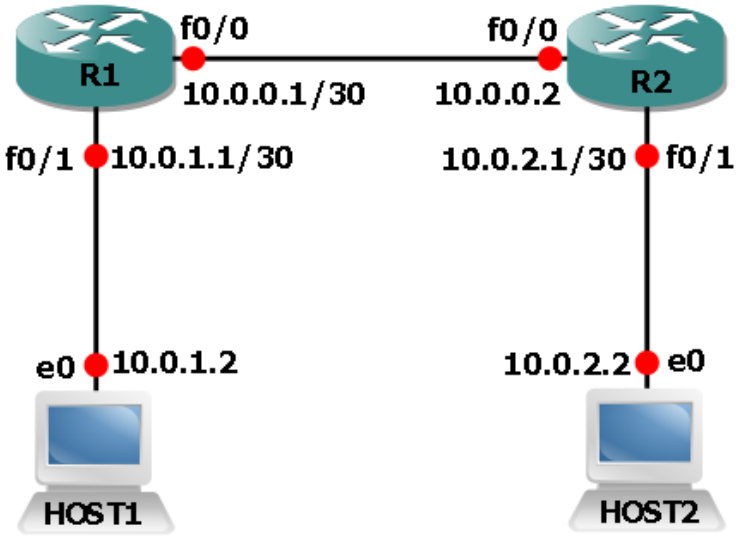
The survey questions and online communities the survey was distributed to are listed in Appendix B: Distributed Survey Questions. After creating the survey I sent it out the URL to the email mailing lists in Appendix B, and waiting for the responses. Additionally I directly emailed or asked in person colleagues and peers, to fill out the survey.

All the questions are open ended questions, to open the scope of input, except for question 1. I designed the questions to be open ended to get the opinions of the industry at large, which does allow for off-topic responses to be submitted. To limit this I supplied example answers with the questions to help guide the participants with their answers. Question 1 was a closed question, with multiple options and a single choice: “Q1 - At which lower OSI model layer do you think you most commonly experience issues on your network?” This is because it is a straightforward ‘which of the following list’ question. In question 1 I included the option of “Layer 1 / Physical Layer”. Having decided that Layer 1 technologies are outside the scope of this project, I added this answer to ratify the appropriateness of my decision, and to see if further research would benefit from including layer 1, or perhaps warrant a separate project entirely focused on layer 1 technologies.

3.3.3 Experimental data gathering procedures

For my laboratory based experimental research I set up a small network topology on which I can carry out various kinds of tests. I used the virtual network simulation software GNS3, to allow myself to easily change the topology if required. Detailed network topology and configuration details are shown in Appendix D: Laboratory Topology and Results.

Figure 3.3.3.1: GNS3 Laboratory Topology



For my experimental data gathering I used GNS3 to emulate both routers (Cisco 7206) and network hosts (Linux PCs) shown in Figure 3.3.3.1 above. Using the built-in tools I am able to perform various tests and take measurements. GNS3 includes the packet capturing software Wireshark, which allows me capture packets between hosts and routers, or between routers. The Linux OS version used (Linux Micro Core 4.0.2) includes various tools such as Netcat to send and receive raw data in TCP and UDP packets, which I capture in Wireshark. Using these tools together I generated network traffic and measured its propagation across the network. This allowed me to record the outcome of different traffic types traversing the network, to highlight the differences in protocol behaviour when carrying the same data payload.

3.3.4 Ethical Issues

I took precautions regarding various ethical considerations for this research, relating to bias data gathering or evaluating, impartial research sources, and similar. Table 3.3.4.1 below details the ethical issues and preventative measures I employed to prevent them from occurring:

Table 3.3.4.1: Ethical issues, responsibilities and preventative measures

Ethical Issue	**Ethical Responsibility**	**Preventative Measures**
Informed consent – Participants of my survey could breach a confidentiality agreement or disclosure agreement if they are unclear how the data they provide will be used	Participants of the survey I conducted need to know the reason it is being conducted, how I intend to reference the data, how I intend to distribute it and how I will keep their details confidential	Participants were informed in advance of participation through an email, of my intended actions. This requires them to provide a non-repudiable form of verification of their consent
Openness and integrity – Participants of my survey must be satisfied that their data is being used as they intended it to be and unaltered	Data provided by participants can be manipulated or misinterpreted for the benefit of the research	Collected data was be available to participants in its original format and in the format it is present in so they can verify its dissemination
Confidentiality – Participants who provided data for my survey must remain anonymous to protect them	Participants could be providing data that contradicts an employers moral rule set or publicly stated opinion	All data regarding individual identity has been anonymised referencing them only by a non-descript identifier as required, such as participant number
Data protection – Data collected on survey participants must be kept securely	Unauthorised dissemination of participants personal data could put the individual at personal and professional risk	All data is kept in a secure, password protected file (that is stored in a locked environment)
Vendor Bias – Experimental practical research conducted on products of a single or multiple vendors must be clearly explained	Research outcomes could be interpreted as a result of the specific vendors products used, be that positive or negative, which could be misleading if all the same vendor equipment used, or adversely non-interoperable vendors	I highlight that my experimental research is not vendor specific and I ensure I test generic theories and not ones that are proprietary technology influenced for example, or only occur due to mixing manufacture components

4 Analysis and Interpretation

4.1 Summary of data collected

The tables throughout section 4.1 represent the raw data collected to fulfil each of my required data sets (D1 –D4) in answer to my research questions (Q1 – Q4). Some of the data has been summarised to remain concise, where this occurs the full data is included in the appendices.

Table 4.1.1 below shows the data gathered to fulfil data set D1: Present a table of identified popular network services and protocols, and their optimal operating considerations to maximise their productivity. The most popular lower layer protocols in use today are listed alongside protocol deficiencies that exist when their optimal operating conditions are not met. These are listed as optimal deployment and operating considerations, which should be adhered to when deploying these protocols to aid them in efficient operation. At the end of the table are the most popular application protocols that are typically found to run over a modern day network. Some potential issues for layer 7 connectivity are listed, in relation to the deployment considerations for the lower layer protocols, to show their impact.

Table 4.1.1: Optimal operating conditions for protocols

OSI Layer	**Protocol / Application**	**Optimal deployment and operating considerations**
2 – Data Link Layer	Ethernet	• Does not support frame loss [1] • Does not support duplicate frames [2] • Does not support lossless transmission [3] • Does not support multipath [2] • Does not support network loops [2] • Does not support out of order frames [2] • Does not support duplicate MAC addresses
3 – Network Layer	IPv4	• Does not support datagram loss [4] • Does not support duplicate datagrams [4] • Does not support lossless transmission • Does not support multipath [4] • Does not support out of order datagrams [4] • Does not support duplicate IP addresses
	IPv6	• Does not support datagram loss [5] • Does not support duplicate datagrams [5] • Does not support lossless transmission • Does not support multipath [6] • Does not support out of order datagrams [6] • Does not support duplicate IP addresses [5]
4 – Transport Layer	TCP	• Significant performance decrease over high delay networks [7] • Conflicts with excessive queuing buffers [8] • Significant overhead, especially for short lived connections [9]

	UDP	• Does not support duplicate packets [10] • Does not support lossless transmission • Does not support out of order packets [10]
	ICMP	• Does not support out of order packets [11] • Does not support multipath [11]
7 – Application Layer	DNS, FTP [12], HTTP, HTTPS, ICMP [13], NTP, P2P, RDP, RTP/RTCP, SMTP.	• Some application protocols can't manage duplicate datagram units (relying on lower layer protocols) • Some application protocols can't detect lost or dropped data units (relying on lower layer protocols) • Some application protocols are sensitive to high delay and/or jitter • Some application protocols require end-to-end IP connectivity without NAT • Some application protocols don't support IPv6 or support IPv6 to a lesser extent than IPv4

[1]. "Support" throughout this table means that a failure scenario isn't accounted for by the protocol itself and continues operation unaware and/or without corrective actions.
[2]. These issues are implied by the current design revision of the protocol, *IEEE (2002)*. There are no frame sequence numbers, or a TTL value, or verification of frame reception for example, to assist with these potential issues.
[3]. The difference here between lossless transmission and frame loss, is that during congestion there is no queuing support or flow control.
[4]. These issues are implied by the current design revision of the protocol. There are no datagram sequence numbers or verification of datagram reception for example, to assist with these potential issues. IPv4 is defined in the following RFCs; *RFC791, (1981). RFC1349 (1992). RFC2474 (1998). RFC3168 (2001). RFC6864 (2013).*
[5]. These issues are implied by the current design revision of the protocol. There are no datagram sequence numbers or verification of datagram reception for example, to assist with these potential issues. IPv6 is defined in the following RFCs; *RFC1883 (1995). RFC2460 (1998). RFC5905 (2007). RFC5722 (2009). RFC5871 (2010). RFC6437 (2010). RFC6564 (2012).*
[6]. *A* function to rectify this issue is currently in the proposal stages of standardisation; *Wijnen, B. Lucent Technologies. (2003). RFC3539 Textual Conventions for IPv6 Flow Label. Network Working Group*
Amante, S. Level 3. Carpenter,B. University of Auckland. Jiang S. Huawei. Rajahalme, J. Nokia Siemens Networks. (2011). RFC6437 IPv6 Flow Label Specifications. Internet Engineering Task Force (IETF).
[7]. A solution for the issue of TCP retransmission timeouts already exists, but has only recently begun to be moved towards standardisation for implementation, in RFC5682 *(2009)*.
[8]. Gettys and Nichols (2011) have shown how severe and widespread this problem is ("very" is the adverb applied to both issues).
[9]. The research of Zhang et al *(2000)* has shown that bursty and short lived TCP connections suffer from slow-start issues. Although the research by Wei et al *(2006)* is inconclusive, they are also drawing the same conclusion. *Zhang, Y. Qiu, L. Cornell University. Keshav, S. Ensim Corporation. (2000). Speeding Up Short Data Transfers - Theory, Architectural Support, and Simulation Results. Technical Report. Cornell University, Ithaca, NY, USA.*
Wei, D. Low, S. EAS, Caltech. Cao, P. CS Stanford. (2006). TCP pacing revisited. In Proceedings of IEEE INFOCOM, 2006.
[10]. These issues are implied by the design of the protocol. Additionally it's worth noting, that no revisions have been made to the protocol since its initial publication. UDP is defined in RFC768 *(1980)*.
[11]. These issues are implied by the current design revision of the protocol. ICMP is defined in RFC792 *(1981)*. Some issues such as source quench messages over multipath are currently being addressed however *(RFC6633, 2012)*.
[12]. There has been a decline in measured FTP traffic according to Dhamdhere *(2003)*, and it does not show at all on the top ten results form Cymru *(2013b)* or Cymru *(2013c)*. One reason Dhamdhere suspects this is "Possibly due to shift from active to passive mode FTP, because of an increase in packet filtering firewalls." - I mention FTP due to its historical precedence.
[13]. ICMP is not an application itself: here this represents measurement and analysis tools that use ICMP such as 'ping' and 'traceroute', but also network management features like ICMP redirect, source quench and time-to-live exceeded packets.

Table 4.1.2 below lists the effects of inefficient protocol operation. It continues on to show how this affects the application layer of the OSI model. Each optimal deployment and operating consideration previously listed in Table 4.1.1 is extrapolated into a network operating and efficiency effect that is the result of neglecting the deployment and operating consideration. This is a logical abstraction from the protocol problem; application specific issues are off topic here, it is enough to simply note if an end-to-end connectivity issue can exist. The data in Table 4.1.2 fulfils required data set D2: Present a table of the effects inefficient data communication has on identified services and protocols.

Table 4.1.2: Effects of inefficient data communication:

OSI Layer	**Protocol / Application**	**Network operating and efficiency effect**	**Application layer disruption and service degradation**
2 – Data Link Layer	Ethernet	• Higher layers must detect frame loss, and request retransmission increasing the resource usage for the same data • Duplicate frames are forwarded and processed (increasing the resource usage for the same data) • Out of order frames can waste processing resources as they are not expected and will trigger retransmissions, so wasting resources further • Loops can cause 100% link and device utilisation (causing full connectivity loss) and packet duplication ○ Layer 2 services have unreliable communication ▪ Poor horizontal scaling efficiency for larger workloads ❖ Hosts with the same MAC address on a shared broadcast domain will disrupt each others connectivity	• Waste resources repeating processes for duplicate data or create errors due to unexpected duplicate data • Applications may delay whilst waiting for data to arrive that was not transmitted, or continue execution and receive data unexpectedly, if they don't communicate with the lower layers ○ Varying delay causes inconsistent results that applications may struggle to recognise and compensate ○ Time sensitive applications may not operate correctly during high delay periods ▪ Bandwidth could become heavily contented, limiting connectivity ▪ Latency can increase during congestion periods ❖ Data will be lost between applications, or unexpected data for another host can be received
3 – Network Layer	IPv4	• Higher layers must detect datagram loss, and request retransmission increasing the resource usage for the same data	• Waste resources repeating processes for duplicate data or create errors due to unexpected duplicate data

		• Duplicate datagrams are forwarded and processed increasing the resource usage for the same data • Out of order datagrams can waste processing resources as they are not expected and will trigger retransmissions, so wasting resources further ○ Layer 3 services have unreliable communication ▪ Poor horizontal scaling efficiency for larger workloads ❖ Hosts with the same IP address on a shared subnet will disrupt each others connectivity	• Applications may delay whilst waiting for data to arrive that was not transmitted, or continue execution and receive data unexpectedly, if they don't communicate with the lower layers ○ Varying delay causes inconsistent results that applications may struggle to recognise and compensate ○ Time sensitive applications may not operate correctly during high delay periods ▪ Bandwidth could become heavily contented, limiting connectivity ▪ Latency can increase during congestion periods ❖ Data will be lost between applications, or unexpected data for another host can be received
	IPv6	• Higher layers must detect datagram loss, and request retransmission (increasing the resource usage for the same data) • Duplicate datagrams are forwarded and processed increasing the resource usage for the same data • Out of order datagrams can waste processing resources as they are not expected and will trigger retransmissions, so wasting resources further ○ Layer 3 services have unreliable communication ▪ Poor horizontal scaling efficiency for larger workloads ❖ Hosts with the same IP address on a shared subnet will disrupt each others connectivity	• Waste resources repeating processes for duplicate data or create errors due to unexpected duplicate data • Applications may delay whilst waiting for data to arrive that was not transmitted, or continue execution and receive data unexpectedly, if they don't communicate with the lower layers ○ Varying delay causes inconsistent results that applications may struggle to recognise and compensate ○ Time sensitive applications may not operate correctly during high delay periods ▪ Bandwidth could become heavily contented, limiting connectivity

			▪ Latency can increase during congestion periods ❖ Data will be lost between applications, or unexpected data for another host can be received
4 – Transport Layer	TCP	• Resources are wasted as packets are retransmitted due to timers being exceeded • Excessive buffering causes flows to stall and suffer jitter ○ A lack of resources for connection tracking can cause connections to be dropped ○ Connection state tracking consumes resources even though the application may be tracking the connection state	• Applications may delay whilst waiting for data to arrive that was not transmitted, or continue execution and receive data unexpectedly, if they don't communicate with the lower layers ○ Connectivity can be lost or reset if the connection state information is removed ○ Resources are wasted when applications performing low level connection tracking don't interact with the lower layers directly
	UDP	• Duplicate packets are forwarded and processed increasing the resource usage for the same data • Out of order packets can waste processing resources as they are not expected and will trigger retransmissions, so wasting resources further ○ Layer 4 services have unreliable communication	• Waste resources repeating processes for duplicate data or create errors due to unexpected duplicate data • Applications may delay whilst waiting for data to arrive that was not transmitted, or continue execution and receive data unexpectedly, if they don't communicate with the lower layers ○ Varying delay causes inconsistent results that applications may struggle to recognise and compensate ○ Time sensitive applications may not operate correctly during high delay periods
	ICMP	• Can cause invalid readings on reporting tools like 'ping' or 'traceroute'	• Application output can be incorrect • Applications may not produce consistent results

7 – Application Layer	DNS, FTP, HTTP, HTTPS, ICMP, NTP, P2P, RDP, RTP/RTCP, SMTP.	• Some application protocols are sensitive to high delay and/or jitter • Some application protocols require end-to-end IP connectivity without NAT • Some application protocols do not support IPv6 or support IPv6 less	• Varying delay causes inconsistent results that applications can struggle to compensate for • FTP and NTP are just two protocols that embed the end host's IP into the payload of a protocol datagram • Connectivity can cease completely in some applications if forced over IPv6

Some application layer issues are repeated throughout the lower networking layers in Table 4.1.2 above. This is because applications can run directly over lower layers.

Table 4.1.3 below is a matrix showing the optimal deployment and operating considerations from Table 4.1.1 above, with a deployment best practice to mitigate the issue. Table 4.1.3 below fulfils the requirements of data set D3: Create a matrix of required improvements that will bring a network service to an acceptable working level, alongside network inefficiencies to show their relationship.

These logical mitigation techniques are formulated from the original design standards and research used to formulate Table 4.1.1, grouping together the issues listed in Table 4.1.2, and additionally, the results from the survey in Appendix C. We can see in the survey results for example that Ethernet loops are the second most encountered issue, so maintaining a single loop free forwarding path is an essential mitigation technique.

The following improvement and best practice groups are defined, to apply to each protocol layer issue in Table 4.1.3. Their individual meanings are as follows;

- Maintain lower layer health:
  This means the layer below the current one is the cause of this issue. For example, IPv4 not supporting datagram loss would not be an issue if the layer below (Ethernet) did not allow datagram loss by tracking Ethernet frame loss.

- Maintain a single loop free forwarding path:
  At each layer this means to ensure only one active path between any two points on the network. With Ethernet for example this represents the design ideology of having no loops, for IPv4 & 6 this means only one path in the routing tables unless equal cost multi-path routing is used and all routers are aware.

- Optimal queuing, scheduling, and flow control, for QoS:
  This represents issues that could be resolved by deploying even basic QoS or queuing technique to prioritise more sensitive flows over lesser ones.

- Careful administration of protocol planning and deployment:
  Here issues relating to configuration and human error, and management are combined.

Table 4.1.3: A matrix of improvements: These will bring a network service or protocol to an acceptable working level

	Improvements and best practices	Maintain lower layer health	Maintain a single loop free forwarding path	Optimal queuing, scheduling, and flow control, for QoS	Careful administration of protocol planning and deployment
OSI Layer & Protocol / Application	**Optimal deployment and operating considerations**				
2 – Data Link Layer – Ethernet	Ethernet does not support frame loss	**x**			
	Ethernet does not support duplicate frames		**x**		
	Ethernet does not support out of order frames		**x**		
	Ethernet does not support network loops		**x**		
	Ethernet does not support lossless transmission			**x**	
	Ethernet does not support multipath		**x**		
	Ethernet does not support duplicate MAC addresses				**x**
3 – Network Layer – IPv4	IPv4 does not support datagram loss	**x**			
	IPv4 does not support duplicate datagrams		**x**		
	IPv4 does not support out of order datagrams		**x**		
	IPv4 does not support lossless transmission			**x**	
	IPv4 does not support multipath		**x**		
	IPv4 does not support duplicate IP addresses				**x**

Table 4.1.3: continued

	Improvements and best practices	Maintain lower layer health	Maintain a single loop free forwarding path	Optimal queuing, scheduling, and flow control, for QoS	Careful administration of protocol planning and deployment
OSI Layer & Protocol / Application	**Optimal deployment and operating considerations**				
3 – Network Layer – IPv6	IPv6 does not support datagram loss	**x**			
	IPv6 does not support duplicate datagrams		**x**		
	IPv6 does not support out of order datagrams		**x**		
	IPv6 does not support lossless transmission			**x**	
	IPv6 does not support multipath		**x**		**x**
	IPv6 does not support duplicate IP addresses				
4 – Transport Layer – TCP	TCP does not support high delay networks			**x**	
	TCP conflicts with excessive queuing buffers			**x**	
	TCP can introduce significant computing overhead				**x**
4 – Transport Layer – UDP	UDP does not support duplicate packets		**x**		
	UDP does not support out of order packets		**x**		
	UDP does not support lossless transmission			**x**	
4 – Transport Layer– ICMP	ICMP does not support out of order packets		**x**		
	ICMP does not support multipath		**x**		

The classification of improvements above in Table 4.1.3 is determined by the mitigation technique used to implement that improvement, and the issues those techniques mitigate. '*Optimal queuing and scheduling…*' for example, in practice requires QoS classification, marking, and queuing techniques. These same techniques in turn mitigate both the mentioned issues 'TCP does not support high delay networks' and 'TCP conflicts with excessive queuing buffers'. Both issues are grouped under 'Optimal queuing and scheduling…'

Below in Table 4.1.4 is a list of mitigation techniques identified through review research and experimental research to fulfil the required data set D4: Produce a list of mitigation techniques identified through primary and secondary research. The protocol issues in Table 4.1.3 above that where marked under column 'Careful administration of protocol planning and deployment' are not carried into Table 4.1.4 below. Those issues are related to management, planning, and configuration. They are derived from issues such as human error that occur when setting up a new device or link for example. These issues and mitigation techniques are not included in this research as they are not operational issues of the protocol itself.

Table 4.1.4: Mitigation techniques identified through research

OSI Layer & Protocol / Application	**Deployment and operating improvement and best practice**	**Mitigation and management techniques/technologies that can be deployed**
2 – Data Link Layer – Ethernet	Maintain a single loop free forwarding path	• Spanning Tree Protocol (IEEE 802.1D-2004) [1] • Rapid-STP (IEEE 802.1w) [1] • Multiple-STP (IEEE 802.1s) [1] • Shortest Path Bridging (IEEE 802.1aq) [1] • Transparent Interconnect of Lots of Links (RFC6327) [1] • Link aggregation (IEEE 802.1AX-2008)
	Optimal queuing, scheduling, and flow control, for QoS	• Class of Service (IEEE 802.1p/Q) • Ethernet flow control (IEEE 802.3x) • Data Centre Bridging, comprising of Priority-based Flow Control (IEEE 802.1Qbb), Enhanced Transmission Selection (IEEE 802.1Qaz) and Data Centre Bridging eXchange (IEEE 802.1Qaz) • Congestion Notification (IEEE 802.1Qau) • Ensure excess capacity in layer below
3 – Network Layer – IPv4 & IPv6	Maintain a single loop free forwarding path	• Use Equal Cost Multipath routing instead of Unequal Cost Multipath routing • Use per-flow ECMP routing instead of per-packet ECMP routing • Use of constraint based routing decisions (MPLS-TE or RSVP-TE)

		• Avoid NAT (RFC2663 section 1, 2 & 7 [2]) – IPv4 only • Flow labels can improve multipath consistency (RFC6437) – IPv6 only • Ensure excess capacity in layer below
	Optimal queuing, scheduling, and flow control, for QoS	• Ensure support of ECN (RFC3168 & RFC6040) • Implement DiffServ (RFC2474) • Ensure PMTUD functionality (RFC1191, RFC1981 & RFC4459)
4 – Transport Layer – TCP, UDP & ICMP	Maintain a single loop free forwarding path	• Ensure lower layer protocols are loop free
	Optimal queuing, scheduling, and flow control, for QoS	• Ensure PMTUD functionality (RFC4459 & RFC2923) • Ensure support of ECN (RFC3168 & RFC5562) • Active Queue Management such as RED, WRED or CoDeL • Ensure excess capacity in layer below
3 & 4 – Network and Transport Layers	Efficient end-to-end operation	• Ensure ICMP traffic is not filter anywhere on the network

[1] Only the most appropriate of these needs to be chosen for any given layer 2 broadcast domain

[2] As described in: Srisuresh, P. Holdrege, M. Lucent Technologies. (1999). RFC2663 IP Network Address Translator (NAT) Terminology and Considerations. Network Working Group.

In Table 4.1.4 above, there is an additional mitigation and management technique that should be adhered to and configured network-wide. ICMP traffic should not be filtered at any point, unless it is strictly necessary to do so. One reason for filtering specific ICMP message types is that it can be more beneficial to network operations, compared to having unfiltered ICMP traffic on a network. An example of this is filtering ICMP source quench messages. These are being phased out of operation and should not been seen on live public networks. Allowing them is a security risk.

Using experimental data gathering techniques I was able to show the importance of ICMP messages, by replicating the point at which fragmentation must be signalled to an end host by its default gateway, which has been sending data using TCP or UDP. The raw data for this is shown in Figures D.2 to D.7 in Appendix D.

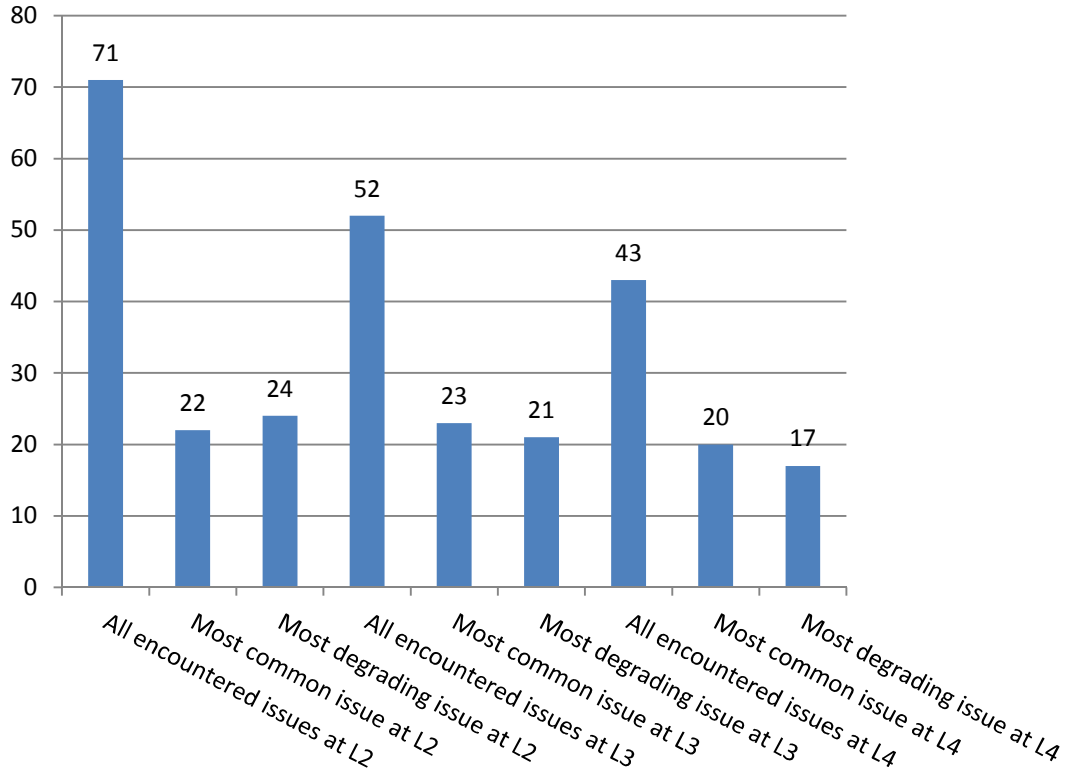
This is one of the many situations in which ICMP messages are required, or when it is more beneficial to receive them, than not. Also shown, is when the MTU on a link changes, such as the link between R1 and R2 in the experiment, the routers have to perform the fragmentation and reassembly. This can significantly increase the work load for a router.

4.2 Data Analysis

Table 4.1.1 above shows that the various protocols of today's networks have multiple pitfalls, most noticeable is the crossover between the layers of the OSI model, with a lack of multipath support for example, existing at every layer in different forms.

Figure 4.2.1 below visually clarifies this crossover statement above, which is seen in Table 4.1.1 and backed up by the survey data in Appendix C; Survey question number two had the highest number of responses of all survey questions (71 responses). This question asked for a list of all issues experienced at layer 2. It is an important relation; more layer 2 issues are reported than any other layer, and layer 2 Ethernet has the most entries in Table 4.1.1 and 4.1.2. All layers sit on top of layer 2, so layer 2 issues crossover into all higher layers.

Figure 4.2.1: Number of responses per question

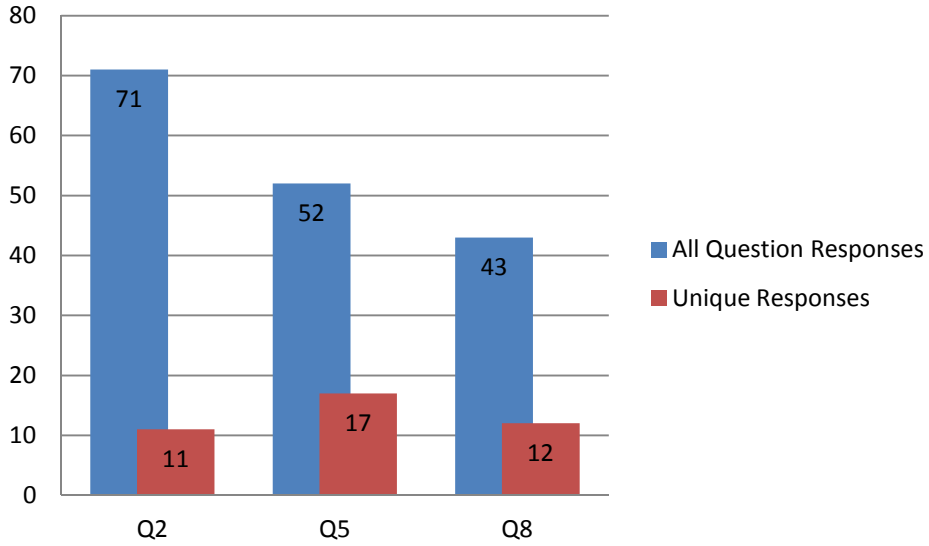


The survey data in Appendix C and Table 4.1.3 also shows another relationship between the layers of the OSI model. This one I find surprising: questions two, five and eight of the survey ask "What issues have you experienced at layer 2/3/4", Figure 4.2.2 shows the highest number of unique responses was to question five, and this is the highest percentage of unique responses to all responses for that question;

Percentage of unique responses to:

- Question 2: 15.49%
- Question 5: 32.69%
- Question 8: 27.90%

Figure 4.2.2: Number of unique responses per question



There is a problem with this statistic that is a result of the questionnaire design. The open ended questions mean that ambiguous or off topic answers can be submitted. The most commonly encountered issue at layer 3, according to the survey data was “IGP failure”. This isn’t a problem with any layer 3 protocol. This is likely to be an issue with either a layer 3 routing protocol, or a configuration error.

After taking ambiguous and irrelevant answers into account by removing them, and redrawing the graphs Figure 4.2.1 and Figure 4.2.2, the following are produced:

Figure 4.2.3: Number of relevant responses per question

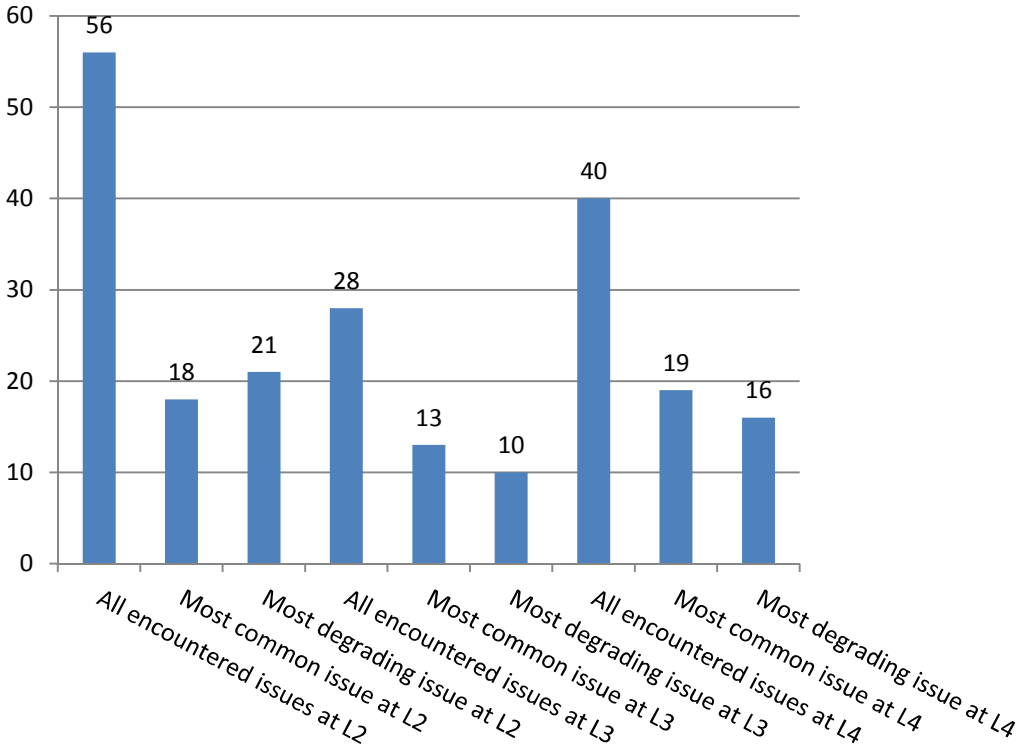
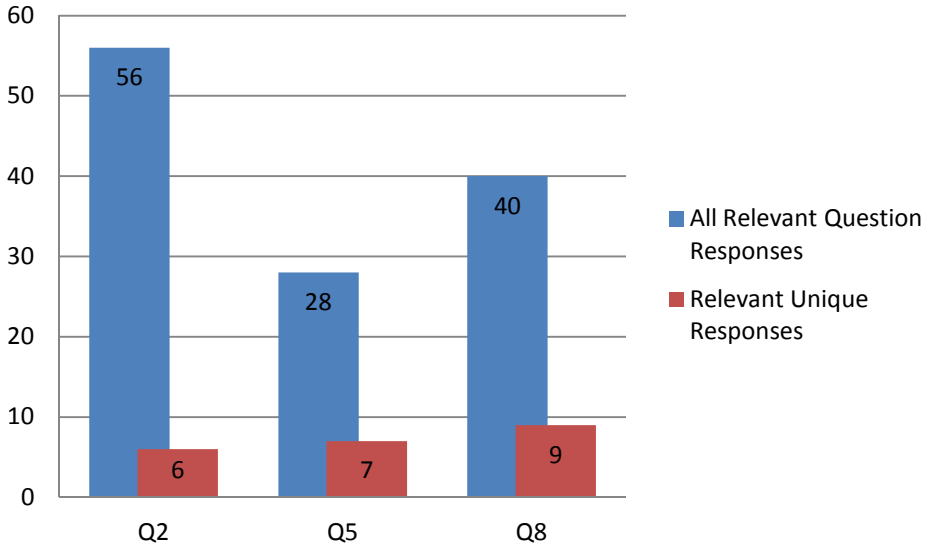


Figure 4.2.4: Number of relevant unique responses per question



These graphs now confirm the data I was expecting. The number of all layer 3 issues, and unique layer 3 issues, are fewer than for layer 4. Also, as I expected, the number of unique responses is highest for layer 4. As I mentioned earlier in this research project, most existing research has been around layer 4, in particular TCP.

In addition to the initially deceptive results above, explained up to this point regarding which layers more protocol related issues are experienced at, the findings from my experimental data gathering research have highlighted a partial management and partial technical problem.

The issue of ICMP filtering is a cross over of configuration and management planning, which I previously stated as not relevant to this research, and the behaviour of layer 3 and 4 protocols.

In the experiment, ICMP was required to signal to the end host to fragment the data it was transmitting, when payloads larger than the IPSEC MTU would permit. A flaw with this is that it only happens once. After reducing the MTU on the link between routers, the end hosts remain unaware. This increases processing requirements for the routers because they have to perform fragmentation and reassembly, on behalf of the end hosts.

Also we see how important ICMP messages are in this situation because the MTU available is much lower than the end host can safely assume. How low the MTU is reduced in this example scenario is shown below in Table 4.2.1. The link between routers R1 and R2 in the experiment can be used to represent an Internet WAN connection of various types.

Table 4.2.1: Diminishing MTU sizes for typical end user connections

Network Scenario (Connection Type)	WAN MTU Size (octets)	Protocol Overhead Breakdown	Max Segment Size (TCP & UDP)
Experimental Topology – Representing an Ethernet WAN such as EFM (IEEE 802.3ah-2004)	1500	IPv4 header = 20 TCP = 32 Total = 52	1500 – 52 = 1448
Typical ADSL/2+ office (PPPoA with VC/MUX)	1478	IPv4 = 20 TCP = 32 Total = 52	1478 – 52 = 1326
ADLS/2+ office with IPSEC tunnel	1478	IPv4 header = 20 IPSEC header = 30 Original IPv4 header = 20 Original TCP header = 32 Total = 102	1478 – 102 = 1376

4.3 Interpretation of results

The results of the data gathering process can be interpreted in relation to the research aim and objectives, starting with the first objective, O1: Identify what networking protocol inefficiencies are and how they can be recognised.

The list below is extracted from Table 4.1.1 and details the main protocol inefficiencies that exist across the lower layers of the OSI model that cause sub-optimal communication due to networking protocol operating inefficiencies. Issues relating to Protocol Datagram Units could exist at single or multiple OSI layers:

- No detection or mitigation of PDU loss
- No detection or mitigation of duplicate PDUs
- No mechanism for lossless PDU transmission or reception
- No support for multipath transmission or reception
- No detection or mitigation of a looping PDU path
- No detection or mitigation of out of order PDUs
- No detection or mitigation of duplicate source or destination PDU address

The findings presented in the list above derived from Table 4.1.1, which fulfils data set D1, forms the answer to the first question towards objective O1, Q1: What is considered inefficient communication of data?

The issues in the list above can cause data to traverse a network in a sub-optimal manner which can have an undesired effect on application operation at higher layers. These undesired effects are listed below and have been extracted from Table 4.1.2:

- Layers without PDU loss detection must rely on higher layers to detect the loss, and request retransmission, increasing the resource usage for the same data
- Layers without PDU loss detection must rely on higher layers to detect the loss, transmission of data directly over that layer is unreliable

- Layers that track loss must consume additional processing resources to do so
- When duplicate PDUs are forwarded and processed, there is increased processing and resource usage for the same data
- Processing resources are wasted when PDUs are received out of order, this can trigger retransmissions resulting in further wasting of resources
- PDU path loops waste link and device processing resources
- Without multipath there is poor horizontal scaling efficiency for larger workloads
- Duplicate addresses on sent or received PDUs leads to wasted processing resources on any duplicate receiver and possible retransmission of the PDU

This second list derived from table 4.1.2, which answers data set D2, forms the answer to the second research question of objective O1, Q2: What is the effect of inefficient data communication?

Together the two lists above show what protocol inefficiencies exist and the effect they have, to recognise when they are happening. This fulfils research objective O1.

The second research objective is O2: Research potential mitigation strategies suggesting additional ideas. The list above detailing undesired effects of protocol related issues, defines what a satisfactory mitigation would achieve as each issues is resolved. What a mitigation method would consist of is identified in Table 4.1.3. Table 4.1.3 fulfils data set D3: Create a matrix of required improvements that will bring a network service to an acceptable working level, alongside network inefficiencies to show their relationship. The data fulfilling D3, along with the list above, together answer research question Q3: What is a satisfactory mitigation, and what does it consist of?

Objective O2 also includes research question Q4: What are the mitigations for the identified network inefficiencies? The information in Table 4.1.4 is the technique for achieving the mitigations method from Q3. Thus, Table 4.1.4 which fulfils data set D4, answers research Q4.

These answers to the two research questions combine to meet research objective O2.

Collectively this forms a data set that answers the initial research aim: Propose methods for identifying and removing limitations to networking protocols used in the delivery of data across a modern network. The outcome will more efficiently serve network applications running atop of those protocols.

5. Conclusions

5.1 Conclusions

This research project has shown that there are various deficiencies within the design and operation of common networking protocols, which can cause them to inadequately serve the networking applications running atop those protocols. Lower layer protocols can fail to provide the connectivity requirements upper layer protocols demand of them.

The research has also shown that some issues are related to planning and administration, for example, duplicate host addresses or deploying redundant links in a loop intolerant topology. However, multiple operational and technical issues were identified, verified, and mitigated, that are not new to the networking industry, but aren't normally evaluated collectively. The OSI model for network communications, by its very definition of stacked protocol layers, means that the negative effects of any given layer can overlap and affect a neighbouring layer's operational reliability.

Inline with the original aim and objectives, the data gathered during this project has produced a variety of operating attributes that are considered characteristics of inefficient communication of data across a network. These include, for example, a lack of detection or mitigation of PDU loss, or out of order PDUs, by a network protocol. Next, it defined what the affects of protocols performing with these characteristics are on the communication of data across a network, for the upper layer applications.

The primary research methods focused on what a satisfactory mitigation strategy would need to achieve, and then identified mitigation methods that can meet those achievements. The experimental data gathering phase, although time consuming, could have continued on for a lot longer gathering additional relevant data. Despite this, enough reliable data was gathered too inform the reader of both research requirements: required mitigation achievements and mitigation methods. This takes the data gathering process to completion in line with the project aim and objectives.

Despite completing this process, the research hasn't produced ground breaking data. However, it has produced a collection of data that has not traditionally been viewed in unison, with the advocacy that it should be. Networks are formed of many layers, not just in relation to the OSI model. These layers should be built individually due to complexity but also evaluated as a whole, in order to determine design and implementation success with respect to the networks' objectives.

The resulting research data presented is a comprehensive solution: it is informative to network operators of almost all networks. This is in line with the project design, and it is important to stress this fact. MPLS at layer 2.5 has not been explored, nor has ATM or PPP. This is because they are not the most common networking protocols in use today, although MPLS is rapidly gaining popularity along with old fashioned PPP. This is partially due to the rapid deployment of ADSL and mobile broadband.

The data collected in this research project is intended to provide best practice guidelines, not a mandatory working regime. Network operators can use the data gathered and knowledge presented to ensure networks operate with a certain minimum level of efficiency. The severity of some operating deficiencies is very high, which means their significance is obvious, for instance providing congestion tolerant data paths. Some issues however, are less severe on application connectivity, so they can remain undetected.

ICMP filtering as shown by the experimental research performed is not usually connectivity affecting, in its degrading effects to network operations until a more sever issue occurs, when troubleshooting becomes more difficult. ICMP is used to identify and diagnose issues, so filtering this traffic can make fault resolution slower. For commercial networks fault resolution typically occurs at the worst possible time, which is during a service outage, thereby preventing network staff from keeping a low MTBF (Mean Time Between Failures) and MTTR (Mean Time To Resolve).

Of the various issue identified and mitigated, neither the cause nor resolution to any were specific to a particular hardware manufacturer, or software developer. Data submission through the on line survey was open to the public, but no patterns were found highlighting certain network types as being more susceptible than others, to specific issues. The areas of a network in which the explored issues can occur were also broad, there is no specific trend towards the core, aggregation, boarder, or access layers of networks.

The research highlighted various issues with the common networking protocols at the lower layers of the OSI model. The data presented showed that there are many issues for each protocol, rather than one or two protocols being significantly more failure prone than the others. This is an important observation because it is a strong argument towards all these aging protocols no longer being fit for their original purpose.

Over the years, modifications have been made to the common protocols investigated, to improve their efficiency. In extreme cases of protocol deficiency new replacement protocols have been developed, standardised and deployed. Examples of this include Data Centre Bridging, Infiniband and Myrinet to replace Ethernet. QUIC *(Roskind, 2013)* shows that this is still happening in the present day, a UDP replacement is being deploy across the Internet by Google Inc this year.

These alternate protocols are not nearly as prevalent across networks as the main protocols researched within this project. This is an indication for the opposing argument to the above statement, that the protocols investigated are still fit for their purpose, as they show no sign of a major deterioration in market share.

Protocol issues have been shown to occur in all networks types, at all network locations, and under a variety of operating conditions. Their impact varies from minor to major, and their frequency of occurrence is also diverse. Every connection is a potential fault location, so great care must be taken through a network.

5.2 Further Work

The data gathered here has not covered every possible issue in the most common networking protocols. Despite this, I now have the opportunity to recompose the data gathered during this research project more succinctly, into a separate document, and disseminate it as networking best practice guidelines. This would break down into four basic points, which are the research questions used to form this project:

- What is considered inefficient communication of data?
- What is the effect of inefficient data communication?
- What is a satisfactory mitigation, and what does it consist of?
- What are the mitigations for the identified network inefficiencies?

I theorise that this could form the basis of an ongoing project that could expand to include security recommendations, planning and administration best practice. This information would be most beneficial if available on the Internet for instant access, and more importantly, allow contribution from industry members as they tackle previously undocumented situations.

The target audience of a project like this would be anyone who is building a new network from scratch, or someone making a change to a live network, to reduce the overall number of networks connected to the Internet that are inefficient.

Many of the issues explored in this research cause connectivity problems, with some resulting in a full loss of all connectivity. There is another topic that can cause negative effects on a network, just as minor or severe as inefficient communication, and is also prevalent across most networks today: human error. An excellent white paper by Juniper Networks *(2008)* shows this, and these two topics could cross over significantly.

Various issues I have raised can occur after a network is fully deployed, due to human error, such as configuration mistakes. More research could be performed on the common causes of human error and the effects, and then merged with my research and existing research on the optimal performance of networks. I would expect to see significant overlap of the subjects with relations to topics such as efficiently upgrading networks, where protocols are adapting to live scenario changes. Also, as capacity or redundancy is added to a network, research could be conducted on how to tie this in with ideal protocol operating conditions to maximise the affectivity of adding new connectivity.

5.3 Implications and reflections of the work presented

There has been a shift in the protocol usage ratio away from TCP, towards UDP *(CAIDA 2010a)*, and in application usage with web based services now forming the largest proportion of all Internet traffic *(Labovitz, 2011, and Adhikari et al, 2012)*. There are two dominating web based protocols, HTTP(S) services and NetFlix *(Sandvine, 2012)*. This increase in web based protocol usage also means an increase in short lived connections, as HTTP communication is typically connectionless, which also shifts towards smaller average packet sizes *(CAIDA, 2010b)*.

Some bleeding edge technologies are emerging that comply with this trend. One example is Google Inc developing their own version of the UDP protocol, called QUIC *(Roskind, 2013)* to carry HTTP sessions between clients using their Chrome browser and their own customised web servers. Their products are almost entirely web based, so speeding up their delivery to improve business seems like an obvious motivation here.

Rewriting a protocol is not a readily undertaken task. It is generated by a high level of demand to improve network efficiency. It shows that various aspects of networking efficiency are still very important despite ideas such as Moore's Law, which has brought faster than 100Gbps links between large hardware devices or routers smaller than PCs capable of forwarding 80Gbps *(Moore, 1965)*. Speed is not enough to improve the quality of network services, even though UDP is considered faster than TCP because it is connectionless. According to Google, it is neither fast enough or efficient enough, in their QUIC design specification and rational.

The data presented in this research provides the knowledge to guarantee at least a minimum level of efficiency to a network. Much more research could be performed here, in particular experimentation to firstly discover additional improvements, and secondly improve on existing resolutions. Such experiments are costly and time consuming to perform on an ad-hoc basis, reducing their viability. Some issues such as loops in Ethernet based networks are not only recognised, and mitigated, their mitigation techniques are even standardised. There are many propriety and non-proprietary standards that are variations of the Rapid Spanning Tree protocol for this one issue. Adversely, the mention of the TCP congestion avoidance algorithm FQ CoDel during the research, is now only just leaving the initial testing stages of the laboratory, and progressing into beta testing on live networks. Further investigation in this area could warrant a research project of its own.

References

Adhikari, V. Zhang, Z. University of Minnesota. Guo, Y. Hao, F. Varvello, M. Hilt, V. Steiner, M. Bell-Labs Alcatel-Lucent. (2012). *Unreeling Netflix - Understanding and Improving Multi-CDN Movie Delivery*. Presented at IEEE INFOCOM 2012, Orlando, Florida. Available: http://www-users.cs.umn.edu/~viadhi/netflix.pdf, Last accessed 15th June 2013.

Albey, J. Afilias. Savola, P. CSC/UNET. Neville-Neil, G. Neville-Neil Consulting. (2007). RFC5905 Deprecation of Type 0 Routing Headers in IPv6. Network Working Group.

Alizadeh, M. Greenberg, A. Maltz, D. Padhye, J. Patel, P. Prabhakar, B. Sengupta, S. Sridharan, M. (2010). DCTCP: Efficient Packet Transport for the Commoditized Data Center. Microsoft Research Website: Microsoft. Available: http://research.microsoft.com/apps/pubs/default.aspx?id=121386, Last accessed 11th of November 2012.

Almquist, P. (1992). RC1349 Type of Service in the Internet Protocol Suite. Network Working Group.

Amante, S. Level 3. Carpenter, B. Univ of Auckland. Jiang, S. Huawei. Rajahalme, R. Nokia Siemens Networks. (2011). RFC6437 IPv6 Flow Label Specification. Internet Engineering Task Force (IETF).

Appenzeller, G. (2005). Sizing Router Buffers. Stanford University.

Arkko, J. Ericsson. Bradner, S. Harvard University. (2010). RFC5871 IANA Allocation Guidelines for the IPv6 Routing Header. Internet Engineering Task Force (IETF).

BBC. (2012a). UK broadband aided by planning permission rule changes. Available: http://www.bbc.co.uk/news/technology-19520294, Last accessed 13th Jan 2013.

BBC. (2012b). UK cities divide up fast broadband cash. Available: http://www.bbc.co.uk/news/technology-19651311, Last accessed 13th Jan 2013.

CAIDA (The Cooperative Association for Internet Data Analysis). (2010a). Analyzing UDP usage in Internet traffic. Available: http://www.caida.org/research/traffic-analysis/tcpudpratio/. Last accessed 15th June 2013.

CAIDA (The Cooperative Association for Internet Data Analysis). (2010b). Packet size distribution comparison between Internet links in 1998 and 2008. Available: http://www.caida.org/research/traffic-analysis/pkt_size_distribution/graphs.xml, Last accessed 16th June 2013.

Cerf, V. Jacobson, V. Weaver, N. Gettys, J. (2011) BufferBloat: What's Wrong with the Internet?. Interviewed by Vint Cerf [in person] Association for Computing Machinery, Vol 9, Issue 12. December 7th, 2011.

Chan, M. Ramjee, R. (2005). TCP/IP Performance over 3G Wireless Links with Rate and Delay Variation. *Wireless Networks*. 11 (1-2), p81-97.

Cisco Systems. (2009). Troubleshooting Buffer Leaks. Available: http://www.cisco.com/en/US/products/hw/iad/ps397/products_tech_note09186a00800a7b85.shtml, Last accessed 3rd Nov 2012.

Cobb, D. (2012). LINX - The World's First Juniper IX, presented at NANOG 54 2012, San Diego, 5-7 February. Available: http://www.nanog.org/meetings/nanog54/presentations/Wednesday/Cobb.pdf, Last accessed 11th March 2013.

Cutler, S. (2012). 40% of all lines are expected to be SIP by 2016, but what is SIP Trunking? Available: http://www.elitetele.com/news/read/40-of-all-lines-are-expected-to-be-sip-by-2016-but-what-is-sip-trunking, Last accessed 13th Jan 2013.

Day, J. (2008). Patterns in Network Architecture - A Return to Fundamentals. Boston: Pearson Education. Preface - xiv.

Defense Advanced Research Projects Agency Information Processing Techniques Office (1980). RFC760 DOD STANDARD INTERNET PROTOCOL. IETF Working Group: Information Sciences Institute University of Southern California.

Deering, S. Cisco. Hinden, R. Nokia. (1998). RFC2460 Internet Protocol, Version 6 (IPv6) Specification. Network Working Group.

Deering, S. Xerox PARC. Hinden, R. Ipsilon Network. (1995). RFC1883 Internet Protocol, Version 6 (IPv6) Specification. Network Working Group.

Dhamdhere, A. (2003). Internet Traffic Characterization CS8803. PowerPoint presenation, Georgia Tech College of Computing. Available: http://www.cc.gatech.edu/~dovrolis/Courses/8803_F03/amogh.ppt, Last accessed 15th June 2013.

Duke, M. Boeing Phantom Works. Braden, R. USC Information Sciences Institute. Eddy, W. Verizon Federal Network Systems. Blanton, E. Purdue University Computer Science. (2006). RFC4614 - A Roadmap for Transmission Control Protocol (TCP) Specification Documents.

Elmeleegy, K. Cox, A. Eugene, T. (2009). Understanding and Mitigating the Effects of Count to Infinity in Ethernet Networks. EEE/ACM Transactions on Networking. 17 (1), p186-199.

Floyd, S. Jacobson, V. Random Early Detection gateways for Congestion Avoidance. IEEE/ACM Transactions on Networking, V.1 N.4, August 1993, p. 397-413.

Gettys, J. Alcatel-Lucnet. Nichols, K. Pollere Inc. (2011). Bufferbloat: Dark Buffers in the Internet. Queue, Volume 9 Issue 11. p40.

Gont, F. SI6 Networks. (2012). RFC6633 Deprecation of ICMP Source Quench Messages. Internet Engineering Task Force (IETF).

Huang, R. Chien, A. (2004). Benchmarking High Bandwidth-Delay Product Protocols. Department of Computer Science University of California, San Diego.

IEEE 802 Working Group. (2001). IEEE Standard 802-200. IEEE Standard for Local and Metropolitan Area Networks: Overview and Architecture.

IEEE. (2002). IEEE 802-2001 Standard for Local and Metropolitan Area Networks: Overview and Architecture – Section 6.3.2.3 The Spanning Tree Protocol. Available: http://grouper.ieee.org/groups/802/secmail/pdfYD89wBpRqH.pdf, Last accessed 15th August 2013.

Information Sciences Institute. (1981). RFC791 Internet Protocol – DARPA Internet Program Specification. University of Southern California.

Jacobson, V. Nichols, K. (2012). Controlling queue delay. Communications of the ACM. 55 (7), p2-50

Juniper Networks. (2008). What's Behind Network Downtime? Available: http://www-05.ibm.com/uk/juniper/pdf/200249.pdf, Last accessed 9th September 2013.

Kaufmann, K. (2012). "How does AKAMAI Delivered The Olympics 2012", Presented at Netnod Autumn Meeting 2012, Stockholm, Sweeden, Piperska Muren. 9-10 October. Available: http://www.netnod.se/sites/default/files/Akamai-olympics.pdf, Last accessed 7th March 2013

Khlifi, H. Grégoire, JC. Phillips, J. (2006). VoIP and NAT/firewalls: Issues, traversal techniques, and a real-world solution. *IEEE COMMUNICATIONS MAGAZINE*. 44 (7), p93-99.

Krishan, S. Ericsson. (2009). RFC5722 Handling of OVerlapping IPv6 Fragments. Network Working Group.

Krishnan, S. Ericsson. Woodyatt, J. Apple. Kline, E. Google. Hoagland, J. Symantec. Bhatia, M. Alcatel-Lucent. (2012). RFC6564 A Uniform Format for IPv6 Extension Headers. Internet Engineering Task Force (IETF).

Kurose, J. Ross, K. (2000). 3.7 TCP Congestion Control (3rd ed). In: Kurose, J. Ross, K, Computer Networking - A Top-down Approach Featuring the Internet. United States: Addison Wesley. P239.

Labovitz, C. (2011). Internet Traffic Evolution 2007 – 2011. Presented at Global Peering Forum 6, 4-7th April, 2011
. Available: http://*www.monkey.org/~labovit/papers/gpf_2011.pdf*, Last accessed 15th June 2013.

Larsen, S. Huggahalli, R. Parthasarathy, S. Kulkarni, S. (2009). Architectural Breakdown of End-to-End Latency in a TCP/IP Network. International Journal of Parallel Programming. 33 (6), p566-571.

Luckie, M. Cho, K. Owens, B. (2005). Inferring and Debugging Path MTU Discovery Failures, paper presented at the Internet Measurement Conference 2005, Berkeley, 12-21 October. Available: http://conferences.sigcomm.org/imc/2010/papers/p102.pdf, Last accessed 5th March 2013.

Maennel, O. Bush, R. Cittadini, L. Bellovin, S. (2008). A Better Approach than Carrier-Grade-NAT. New York: Department of Computer Science, Columbia University. Available at: http://hdl.handle.net/10022/AC:P:29597, Last accessed 29th June 2013.

Metcalfe, R. Boggs, D. (1976). Ethernet: Distributed Packet Switching for Local Computer Networks. Communications of the ACM. 19 (7), p395-404.

Moore, E. (1965). Cramming More Components onto Integrated Circuits. Electronics. 38 (8), p114–117.

Nichols, K. Baker, F. Cisco Systems. Blake, S. Torrent Networking Technologies. Black, D. EMC Corporation. (1998). RFC2474 Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers. Network Working Group.

Postel, J. ISI. (1980). RFC768 User Datagram Protocol.

Postel, J. ISI. (1981). RFC792 Internet Control Message Protocol. Network Working Group.

Ramakrishnan, K. TeraOptic Networks. Floyd, S. ACIRI. Black, D. EMC. (2001). RFC3168 The Addition of Explicit Congestion Notification (ECN) to IP. Network Working Group.

Rekhter, Y. Cisco Systems. Moskowitz, B. Chrysler Corp. Karrenberg, D. de Groot, G. RIPE NCC. Lear, E. Silicon Graphics, Inc. (1996). RFC1918 Address Allocation for Private Internets. Network Working Group.

Roskind, J. Google Inc. (2013). QUIC Design Document and Specification Rational. Available: https://docs.google.com/document/d/1RNHkx_VvKWyWg6Lr8SZ-saqsQx7rFV-ev2jRFUoVD34/preview?sle=true, Last accessed 9th September 2013.

Sandvine. (2012). Global Internet Phenomena Report 2H 2012. Available: http://www.sandvine.com/downloads/documents/Phenomena_2H_2012/Sandvine_Global_Internet_Phenomena_Report_2H_2012.pdf, Last accessed 8th March 2013.

Sarolahti. P. Nokia Research Center. Kojo, M. University of Helsinki. Yamamoto, K. Hata, M. NTT Docomo. (2009). RFC 5682 Forward RTO-Recovery (F-RTO): An Algorithm for Detecting Spurious Retransmission Timeouts with TCP. Network Working Group.

Savola, P. CSC/FUNET. (2006). RFC4459 MTU and Fragmentation Issues with In-the-Network Tunnelling. Network Working Group.

Schutte, W. Warman, P. (2012). 2012 Country Summary Report - US. Newzoo. Available: http://www.newzoo.com/wp-content/uploads/us_summary_deck_new2.pdf, Last accessed 15th June 2013.

Sinha, R. Papadopoulos, C. Heidenmann, J. (2007). Internet Packet Size Distributions - Some Observations. Technical Report ISI-TR-2007-643, Colorado State University, Information Sciences Institute, May, 2007.pdf

Srisuresh, P. Holdrege, M. Lucent Technologies. (1999). RFC2663 IP Network Address Translator (NAT) Terminology and Considerations. Network Working Group.

Team Cymru Research NFP. (2013a). IP Protocol Trends. Available: http://www.team-cymru.org/stats/charts.swf?library_path=/stats/charts_library&xml_source=%2Fstats%2Fdata%2Fprotocols.xml, Last accessed 11th March 2013.

Team Cymru Research NFP. (2013b). Top Ten TCP Ports. Available: https://www.team-

cymru.org/stats/charts.swf?library_path=/stats/charts_library&xml_source=%2Fstats%2Fdata%2Ftop_ports-tcp.xml, Last accessed 26th August 2013.

Team Cymru Research NFP. (2013c). Top Ten UDP Ports. Available: https://www.team-cymru.org/stats/charts.swf?library_path=/stats/charts_library&xml_source=%2Fstats%2Fdata%2Ftop_ports-udp.xml, Last accessed 26th August 2013.

The Open University (2013), T802 Research Project, Supplementary thoughts on research aim, objectives and research questions, Milton Keynes, The Open University.

Touch, J. Perlman, R. (2009). RFC5565 Transparent Interconnection of Lots of Links (TRILL): Problem and Applicability Statement. Network Working Group.

Touch, J. USC/ISI. (2013). RC6864 Updated Specification of the IPv4 ID Field. Internet Engineering Task Force (IETF).

Wu, T. Chao, H. Tsuei, T. Li, Y. (2005). A measurement study of network efficiency for TWAREN IPv6 backbone. International Journal of Network Management. 15 (6), p411-419.

Xiao, A. Ni, L. (1999). Internet QoS: A Big Picture. IEEE Network. 13 (2), p8-18.

Appendix A: Extended Abstract

Research background and motivation

Of the many available protocols that can be used to provide data connectivity across a network, there are a select few that have become the protocols of choice on almost all modern day networks. These are software protocols that provide and govern communications between two or more networked devices. Since the standardisation and wide spread adoption of the ISOs OSI model of network connectivity, communication between devices is built upon multiple layers of protocols stacked on top of each other. This research project brings under investigation the most common protocols that typically operate at each of the lower layers of the OSI model, on a modern day network.

The most common protocols in use are TCP, UDP and ICMP at layer 4 (the transport layer), IPv4 and IPv6 at layer 3 (the network layer), and Ethernet at layer 2 (the data-link layer).

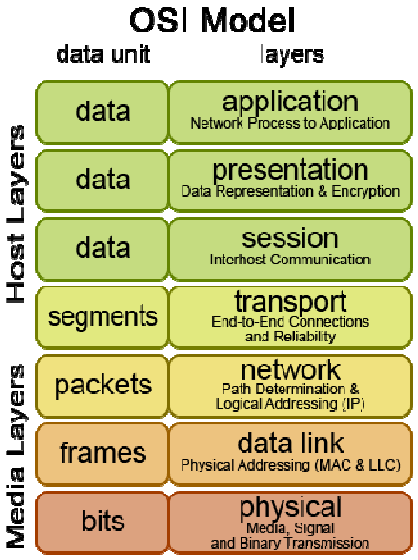


Figure A.1: Dino Korah. 2006. OSI Model - The figure here outlines the data units in the various layers of the OSI model. Available: http://commons.wikimedia.org/wiki/File:Osi-model.png, Last accessed 10th September 2013.

Some of these common protocols were invented nearly 40 years ago. This was during the initial birth of the Internet at a pioneering time for computer networking, when businesses and governments started to become interested in the applications of computer networking, and in funding networking research. In the late 1960s and early 1970s computers and networking equipment had a fraction of the computing power that the latest mobile phones have today. The speed of the inter-device connections in early test networks was less than that of a typical dial-up Internet connection popular during the 1990s. Today, such a connection is considered extremely slow, with typical domestic consumer connectivity being tens or hundreds of times quicker.

Since these software protocols were first written, they have been revised and updated to improve their performance and efficiency. They have been modified to include new features and remove security flaws. Despite this, networking demands and designs have changed to such an extent over the years; I argue that these protocols are no longer suitable anymore. Are they inefficient when run on today's networks, as they were prototyped on networks with a higher percentage of data loss? Are they unfit for large scale operations, due to their design around networks a fraction of the size of corporate networks today? It is time these protocols were re-examined with a focus on their efficiency.

The following questions were formulated to examine issues regarding protocol efficiency;

- What is considered inefficient communication of data?
- What is the effect of inefficient data communication?
- What is a satisfactory mitigation, and what does it consist of?
- What are the mitigations for the identified network inefficiencies?

Aim and objectives

The research aim below is achieved by answering the questions above;

Propose methods for identifying and removing limitations to networking protocols used in the delivery of data across a modern network. The outcome will more efficiently serve network applications running atop of those protocols.

This research aim is broken into two research objectives that consist of the first two, and last two research questions respectively;

1. Identify what networking protocol inefficiencies are and how they can be recognised

2. Research potential mitigation strategies suggesting additional ideas

Each research question is answered by gathering data. The compiled data sets provide the information required to fulfil the research objectives. By answering the research questions I was able to fulfil the research aim.

Methodology

Initially, review research was conducted to gather a knowledge baseline that identified key research aspects for research questions one and two: What is considered inefficient data communication? What is the effect of inefficient data communication?

After this initial research phase a survey was distributed as a questionnaire online, and amongst peers and colleagues. This provided qualitative answers from members of the industry, allowing access to years of experience with minimal effort. Finally, empirical data gathering was executed using experimental laboratory based research to produce results that could then be used in a triangulation process, to challenge the quantitative review research data and the qualitative survey data.

Further review research was later carried out to clarify the impact of all the previously gathered data, to establish where in the knowledge market the research results fitted.

Results

An important result is that most issues are experienced within layer 4 of the OSI model, due to the dominance of TCP over other layer 4 protocols like UDP or SCTP *(CAIDA 2010a)*:

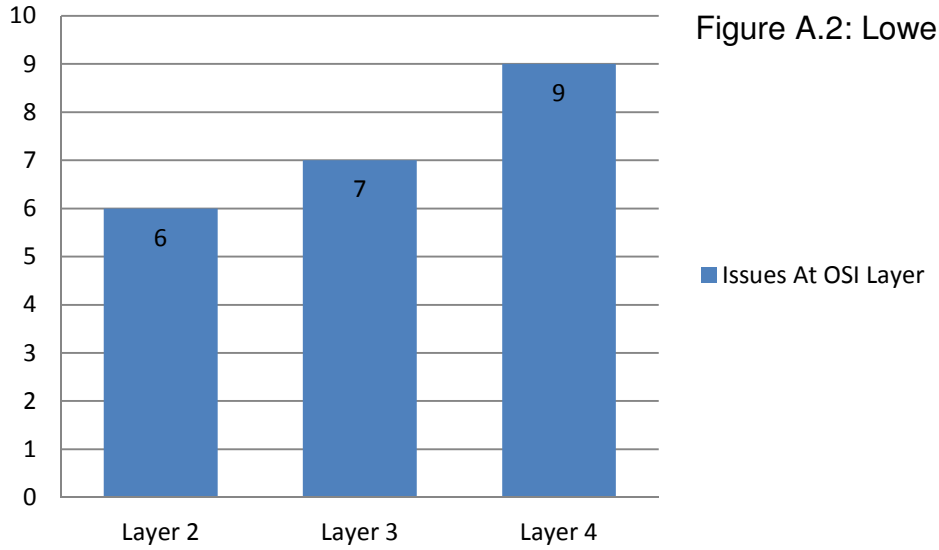


Figure A.2: Lower layer unique issues count

Also shown in the research was the importance of ICMP messaging for efficient protocol operation, due to the likelihood of variance in path MTU:

Table A.1: Diminishing MTU sizes

Network Scenario (Connection Type)	**WAN MTU Size (octets)**	**Protocol Overhead Breakdown**	**Max Segment Size (TCP & UDP)**
Experimental Topology – Representing an Ethernet WAN such as EFM (IEEE 802.3ah-2004)	1500	IPv4 header = 20 TCP = 32 Total = 52	1500 – 52 = 1448
Typical ADSL/2+ line (PPPoA with VC/MUX)	1478	IPv4 = 20 TCP = 32 Total = 52	1478 – 52 = 1326
ADLS/2+ line with IPSEC tunnel	1478	IPv4 header = 20 IPSEC header = 30 Original IPv4 header = 20 Original TCP header = 32 Total = 102	1478 – 102 = 1376

Analysis and Conclusions

The research highlighted various issues with the common networking protocols at the lower layers of the OSI model. The data presented showed that there are many issues for each protocol, rather than one or two protocols being significantly more prone to failure than the others. This is an important observation because it is a strong argument against aging protocols being no longer fit for their original purpose.

Over the years, modifications have been made to the common protocols investigated improving their efficiency. In extreme cases of protocol deficiency new replacement protocols have been developed, standardised and deployed. An example of this is Data Centre Bridging replacing Ethernet. These protocols are not as prevalent across networks as the main protocols researched. This is an indication against the above argument, that the protocols investigated are still fit for purpose as they show no sign of a deteriorating market share.

What is definitively shown is that the protocols are flawed, but mitigation strategies and methods exist to ensure no operating deficiency is experienced when running communications over these protocols.

Appendix B: Distributed Survey Questions

This survey was distributed online using Google Docs. The URL for the survey form is https://docs.google.com/forms/d/1lqigAHYHEgLLHr2kifiyBwgJ9Nw5AFS6d_XVXfhKkTw/viewform

Q1 - At which OSI model layer do you think you most commonly experience issues on your network? *
Please select only one
- Layer 1 / Physical Layer
- Layer 2 / Data Link Layer
- Layer 3 / Network Layer
- Layer 4 / Transport Layer

Q2 - Which issues have you experienced at OSI layer 2 (Such as Ethernet loops, MTU sizing, Out of order frames, Excessive headers QinQinQ/PBB, Encapsulation mismatch, etc)?

Q3 - Of the OSI layer 2 issues you have mentioned, which one have you encountered most?

Q4 - Of the OSI layer 2 issues you mentioned, which one is has been the most service degrading in your experience?

Q5 - Which issues have you experienced at OSI layer 3 (Such as routing loops, max fragment size exceeded, PDV/jitter, DF bit ignored, DSCP ignored or no QoS, etc)?

Q6 - Of the OSI layer 3 issues you have mentioned, which one have you encountered the most?

Q7 - Of the OSI layer 3 issues you mentioned, which one has been the most service degrading in your experience?

Q8 - Which issues have you experienced at OSI layer 4 (Such as MSS exceeded, TCP window scaling issues, bufferbloat, out of order packets, elephant/long fat networks, etc)?

Q9 - Of the OSI layer 4 issues you have mentioned, which one have you encountered the most?

Q10 - Of the OSI layer 4 issues you mentioned, which one has been the most service degrading in your experience?

Email Address (optional field to validate your response)

The survey was sent out to the following public mailing lists;

nanog@nanog.org – North American Network Operators Group

uknof@lists.uknof.org.uk – UK Network Operators Forum

uknot@uknot.org – UK Network Operating Technicians

end2end-interest@postel.org – end-to-end research and design principals

irtf-discuss@irtf.org – Internet Research Task Force: General discussion list

Appendix C: Distributed Survey Gathered Data

Table C.1 below contains the raw data produced by my distributed survey. The answers to each question are grouped together and a total count for each answer is given. In question 2 for example, 20 people said MTU sizing issues where a problem for them. Also, as indicated by the third column, due to the questions being open some answers can be written in multiple ways. MTU sizing issues includes the answers given that were written as “MTU mismatch” and “MTU sizes”. This repeats throughout the results.

Question	Responses & Count	Further Notes
Q1 - At which lower OSI model layer do you think you most commonly experience issues on your network? **Responses**: 25	Layer 1 / Physical Layer : 16 Layer 2 / Data Link Layer: 4 Layer 3 / Network Layer: 3 Layer 4 / Transport Layer: 2	
Q2 - Which issues have you experienced at OSI layer 2 (Such as Ethernet loops, MTU sizing, Out of order frames, Excessive headers QinQinQ/PBB, Encapsulation mismatch, etc)? **Responses**: 71 **Unique Answers**: 11 **Relevant Responses** *: 56 **Relevant Unique Answers** *: 6	MTU sizing (too small or mismatched): 20 * Ethernet loops: 18 * Encapsulation mismatch: 11 Encapsulation compatibility issues: 8 * Out of order frames: 5 * Broadcast storms: 3 * Duplicate MAC addresses: 2 * Corrupt headers: 1 Duplex mismatch: 1 Layer 2 segregation techniques failing: 1 VTP Failure: 1	MTU sizing includes the responses “MTU”, “MTU mismatch”, “MTU size issues” and “MTU sizes” Ethernet loops include the responses “loops” and “loops caused by users…” Encapsulation compatibility issues includes the responses “Excessive headers QinQinQ/PBB”, “Ethertype nesting..”, “Ethertype mismatched on emulated Ethernet…” and “Framing issues on emulated Ethernet…)” Broadcast storms include the responses “Multicast/Broadcast storms” and “The size of the broadcast domain...” Duplicate MAC address includes the responses “Failure of NIC drivers” and “MAC repetitiveness”
Q3 - Of the OSI layer 2 issues you have mentioned, which one have you encountered most? **Responses**: 22 **Relevant**	Ethernet loops: 9 * MTU sizing: 6 * Encapsulation mismatch: 2 Broadcast storms: 1 * Corrupt headers: 1	MTU sizing includes the responses “MTU”, “MTU mismatch”, “MTU size issues” and “MTU sizes” Duplicate MAC address includes the responses “Failure of NIC drivers” and “MAC repetitiveness”

Responses *: 18	Duplex: 1 Duplicate MAC addresses: 1 * Layer 2 segregation techniques failing: 1 Out of order frames: 1 *	
Q4 - Of the OSI layer 2 issues you mentioned, which one is has been the most service degrading in your experience? **Responses**: 24 **Relevant Responses** *: 21	Ethernet loops: 13 * MTU sizing: 6 * Broadcast storms: 1 * Encapsulation mismatch: 1 Layer 2 segregation techniques failing: 1 Out of order frames: 1 * Poor configuration: 1	MTU sizing includes the responses “MTU”, “MTU mismatch”, “MTU size issues” and “MTU sizes”
Q5 - Which issues have you experienced at OSI layer 3 (Such as routing loops, max fragment size exceeded, PDV/jitter, DF bit ignored, DSCP ignored or no QoS, etc)? **Reponses**: 52 **Unique Answers**: 17 **Relevant Responses** *: 28 **Relevant Unique Answers** *: 7	IGP failure: 12 No QoS: 10 * PDV/Jitter: 7 * DF bit ignored: 5 * Max fragment size exceeded: 3 * Asymmetric routing: 2 DDoS attacks: 2 Security restrictions: 2 ACL configuration issue: 1 Duplicate IP address: 1 * EGP Failure: 1 ICMP blocking/discarding: 1 IGP/EGP inconsistencies: 1 Next hop resolution: 1 Packet loss: 1 *	IGP Failure includes the responses “Routing loops”, “Blackholing”, “Route flapping”, “Routing protocols mess up” No QoS includes the responses “DSCP ignored”, “QoS issues”, “dscp write failure”, “dscp read failure”, “dscp/ip prec scheduling failure” DF bit ignored includes the response “DF bit set” Security restrictions includes “Firewalls being too clever …”, “Packet filtering problems” EGP Failure includes the responses “Route de/aggregation”, “Route dampening”, “Route filtering problems” Standards compliance includes “Vendor interoperability”

	PMTUD failure: 1 * Standards compliance: 1	
Q6 - Of the OSI layer 3 issues you have mentioned, which one have you encountered the most? **Responses**: 23 **Relevant Responses** *: 13	IGP failure: 6 No QoS: 5 * PDV/Jitter: 4 * Configuration mistake: 2 DDoS attacks: 1 DF bit ignored: 1 * Duplicate IP address: 1 * Hardware overrun: 1 NAT state tables: 1 * Packet loss: 1 *	IGP Failure includes the responses “Routing loops”, “Blackholing”, “Route flapping”, “Routing protocols mess up” No QoS includes the responses “DSCP ignored”, “QoS issues”, “dscp write failure”, “dscp read failure”, “dscp/ip prec scheduling failure” Configuration mistakes includes “routing protocol typos”, “ACL misconfiguration” DF bit ignored includes the response “DF bit set” Hardware overrun includes the response “Excessive ARP or ICMP redirect traffic”
Q7 - Of the OSI layer 3 issues you mentioned, which one has been the most service degrading in your experience? **Responses**: 21 **Relevant Responses** *: 10	IGP failure: 7 No QoS: 5 * PDV/Jitter: 2 * ACL configuration issue: 1 DDoS attacks: 1 DF bit ignored: 1 * Duplicate IP address: 1 * Hardware overrun: 1 Packet loss: 1 * Software bug: 1	IGP Failure includes the responses “Routing loops”, “Blackholing”, “Route flapping”, “Routing protocols mess up” No QoS includes the responses “DSCP ignored”, “QoS issues”, “dscp write failure”, “dscp read failure”, “dscp/ip prec scheduling failure” DF bit ignored includes the response “DF bit set” Hardware overrun includes the response “Excessive ARP or ICMP redirect traffic” Software bug includes “Vendor bugs”
Q8 - Which issues have you experienced at OSI layer 4 (Such as MSS exceeded, TCP window scaling issues, bufferbloat, out of order packets, elephant/long fat networks, etc)?	Out of order packets: 11 * TCP window scaling issues: 8 * Bufferbloat: 6 * Elephant/long fat networks: 5 *	Out of order packets includes “fragmentation and re-ordering” and “Packet reordering” TCP window scaling includes “Window scaling issues”, assumed to be TCP.

Responses: 43 **Unique Answers**: 12 **Relevant Responses** *: 40 **Relevant Unique Answers** *: 9	MSS exceeded: 4 * Congestion: 3 * ICMP Filtering: 1 *Brought across from L3 issues!* NAT Time-out: 1 * Packet loss: 1 Security restrictions: 1 TCP Sync: 1 * UDP without flow control: 1 * *Brought across from L3 issues!*	Elephant networks includes "Latencies in excess of 200ms" MSS exceeded includes "MSS issues" Congestion includes the responses "congestive collapse" and "TCP Taildrop/sequencing" Security restrictions includes "Packet filtering configuration errors"
Q9 - Of the OSI layer 4 issues you have mentioned, which one have you encountered the most? **Responses**: 20 **Relevant Responses** *: 19	Out of order packets: 4 * Bufferbloat: 3 * Elephant/long fat networks: 3 * TCP window scaling: 3 * NAT issues: 2 * ICMP filtering: 1 * Incomplete TCP session: 1 * Packet loss: 1 * Security issues: 1 TCP Sync: 1 *	Elephant networks includes "High latency due to geographical distance" TCP window scaling includes "Window scaling issues", assumed to be TCP. Security issues includes "Packet filter problems"
Q10 - Of the OSI layer 4 issues you mentioned, which one has been the most service degrading in your experience? **Responses**: 17 **Relevant Responses** *: 16	TCP window scaling: 4 * Bufferbloat: 3 * Elephant/long fat networks: 3 * Out of order packets: 3 * MSS exceeded: 1 * Packet loss: 1 * Security issues: 1 UDP without flow control: 1 * *Brought across from L3 issues!*	TCP window scaling includes "Window scaling issues", assumed to be TCP. MSS exceeded includes "MSS issues" Security issues includes "Packet filter problems"

Appendix D: Laboratory Topology and Results

Figure D.1 below shows the topology of my laboratory based experiments. The below topology was created and tested with the network simulation software GNS3. The two routers R1 and R2 represent Cisco 7206VXR model routers, with NPE-400's running the IOS image c7200-adventerprisek9-mz.124-22.T.bin. The two end hosts are Qemu virtual machines running Linux Micro Core 4.0.2. All links are Ethernet with an MTU of 1500 octets (total max size is 1518 octets including Ethernet headers and frame CRC). There is an ESP+AH IPSEC tunnel between the two routers to connect the two LANs, 10.0.1.0/30 and 10.0.2.0/30. Throughout the below text the words byte and octet are used interchangeably to mean a unit of data measurement, 8 binary bits in size.

Figure D.1: Laboratory topology within GNS3 network emulation software

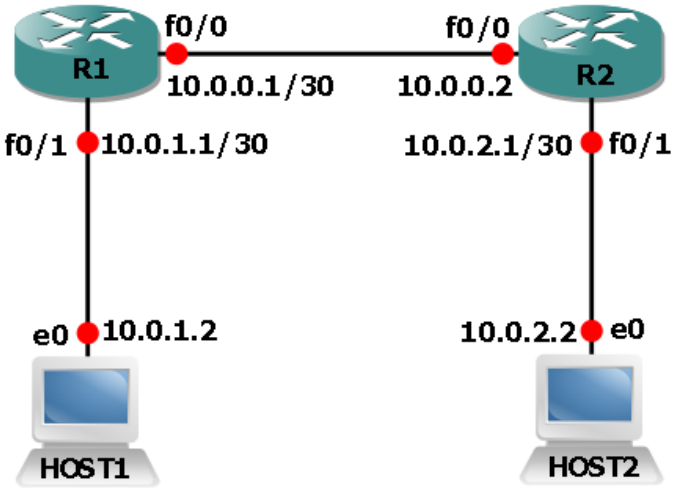


Table D.1: The configuration details of routers R1 and R2

R1	R2
crypcrypto isakmp policy 30 encr 3des hash md5 authentication pre-share lifetime 3600 crypto isakmp key 5up3rs4f3k33y address 10.0.0.2 ! crypto ipsec transform-set TS-3DES-MD5 esp-3des esp-md5-hmac ! crypto map IPSEC-TO-10-0-0-2 1 ipsec-isakmp description IPSEC L2L tunnel to 10.0.0.2 set peer 10.0.0.2 set transform-set TS-3DES-MD5 set pfs group1 match address 130 ! interface FastEthernet0/0 description Link to R2 [fa0/0] ip address 10.0.0.1 255.255.255.252	crypto isakmp policy 30 encr 3des hash md5 authentication pre-share lifetime 3600 crypto isakmp key 5up3rs4f3k33y address 10.0.0.1 ! crypto ipsec transform-set TS-3DES-MD5 esp-3des esp-md5-hmac ! crypto map IPSEC-TO-10-0-0-1 1 ipsec-isakmp description IPSEC L2L tunnel to 10.0.0.1 set peer 10.0.0.1 set transform-set TS-3DES-MD5 set pfs group1 match address 130 ! interface FastEthernet0/0 description Link to R1 [fa0/0] ip address 10.0.0.2 255.255.255.252

duplex auto speed auto crypto map IPSEC-TO-10-0-0-2 ! interface FastEthernet0/1 description Link to HOST1 [e0] ip address 10.0.1.1 255.255.255.252 duplex auto speed auto ! ip route 10.0.2.0 255.255.255.0 10.0.0.2 ! access-list 130 remark IPSEC TO 10.0.0.2 access-list 130 permit ip 10.0.1.0 0.0.0.255 10.0.2.0 0.0.0.255	duplex auto speed auto crypto map IPSEC-TO-10-0-0-1 ! interface FastEthernet0/1 description Link to HOST2 [e0] ip address 10.0.2.1 255.255.255.252 duplex auto speed auto ! ip route 10.0.1.0 255.255.255.0 10.0.0.1 ! access-list 130 remark IPSEC TO 10.0.0.1 access-list 130 permit ip 10.0.2.0 0.0.0.255 10.0.1.0 0.0.0.255

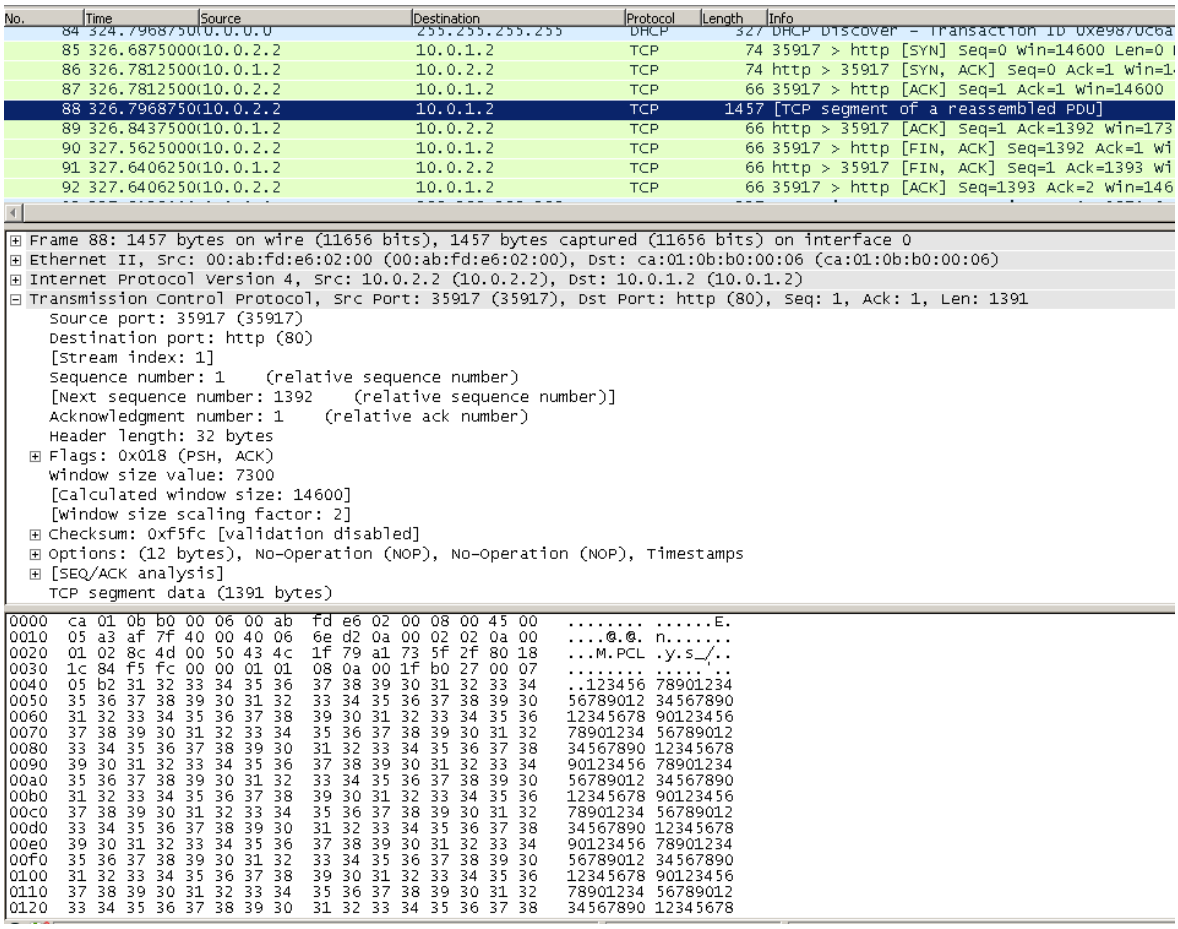
Table D.2: The commands entered to configure the test hosts HOST1 and HOST2

R1	R2
sudo ifconfig eth0 10.0.1.2/30 sudo route add default gw 10.0.1.1	sudo ifconfig eth0 10.0.2.2/30 sudo route add default gw 10.0.2.1

TCP data is sent from HOST2 to HOST1 using the Linux command line tool, 'netcat', or simply 'nc'. Figure D.2 below is the output of a packet capture, running on the Fa0/1 interface on router R2, the interface facing and receiving data from HOST2. Here we can see HOST2 sending 1391 octets of data as the TCP payload in a full TCP connection (there is a 3 way hand shake, the data is sent and acknowledge by the receiver, and then the connection is gracefully closed). The entire data frame from HOST2 to R2 is 1457 octets in size; 1391 octets of payload + 32 octets of TCP overhead + 20 octets of IPv4 overhead + 14 octets of Ethernet overhead = 1457 octets.

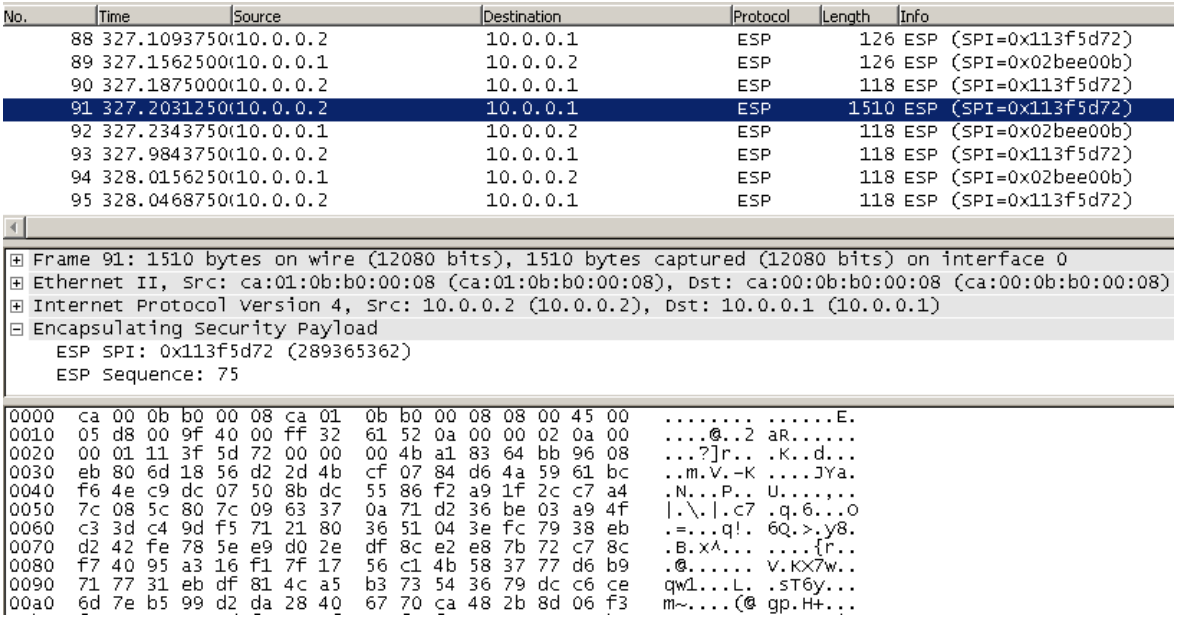
On the receiving host, HOST1, I am executing the command 'nc –v –l –p 80' to listen on TCP port 80 for incoming data, and 'nc –v –l –u –p 80' to listen on UDP port 80. On the sending host I am using files that are the desired number of bytes in length and sending them by issuing the command 'cat data1391 | nc –v 10.0.1.2 80', to send data via TCP, where 'data1391' is a file that is 1391 octets in length. For sending data over UDP I am executing the command 'cat data1391 | nc –v –u 10.0.1.2 80'

Figure D.2: Sending 1391 octets of data over TCP



The size of the frame in Figure D.2 is shown in Figure D.3 below, between R2 and R1, over the IPSEC tunnel. It is 1457 octets + 20 for the new IPv4 header + 30 for the IPSEC tunnel ESP header + 3 for ESP padding + 4 for the Ethernet CRC gives a total of 1514 octets on the wire (directly encoded onto the layer 1 medium), or 1510 as seen by the Operating System. It is worth noting that the ESP padding changes as packet size changes, and not in a clear linear fashion. '3' was calculated above for this test packet by reverse engineering the packet byte count.

Figure D.3: A 1391 octet TCP frame inside an IPSEC tunnel



In Figure D.3 the packet capture shows the data was sent un-fragmented as expected. The total frame size sent from HOST2 to R2 shown above was 1457 octets. 1457 + 53 octets for the additional overhead of the IPSEC tunnel between R1 and R2 = 1510 octets. 1518 octets is the maximum 'on the wire' MTU size. The current 1510 octet frame with a 4 octet frame CRC only leaves 4 octets. When the data payload is increased by 1 octet to 1392, we exceed this MTU size and fragmentation is required. This is because IPSEC ESP changes the header padding to align with new binary columns in the header, making the frame exceed 1518 octets on the wire. This is signalled by R2 to HOST2 and shown below.

Figure D.4: ICMP fragmentation needed packet received due to MTU exceeded with TCP

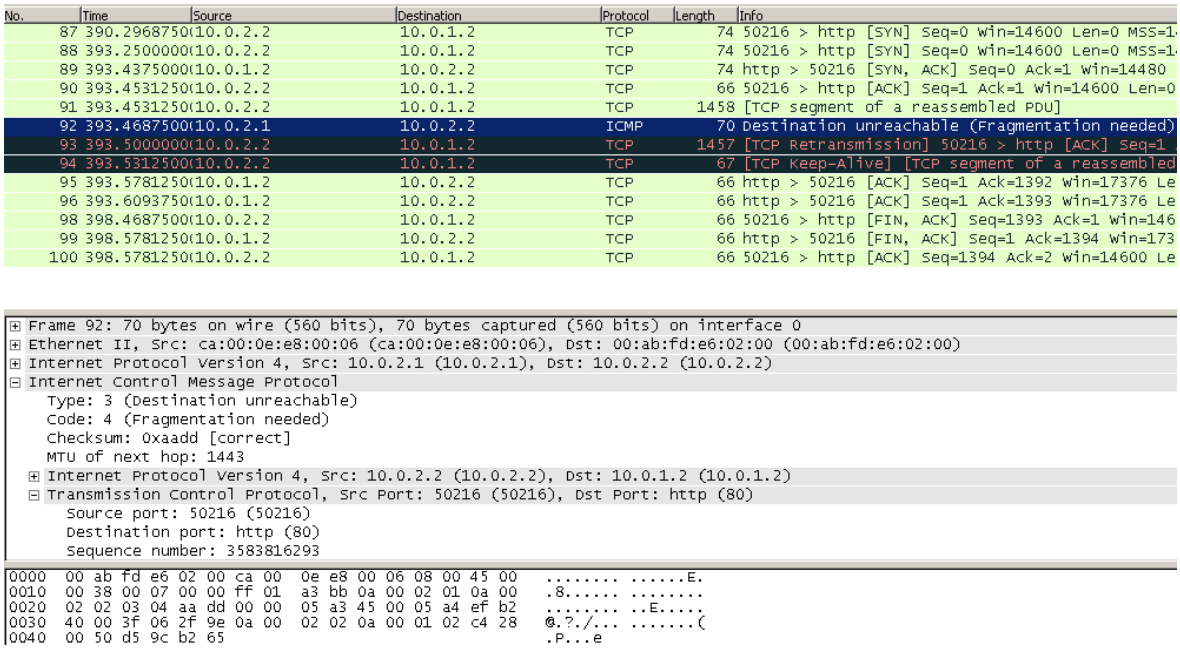


Figure D.5: ICMP fragmentation needed message received due to MTU exceeded with UDP

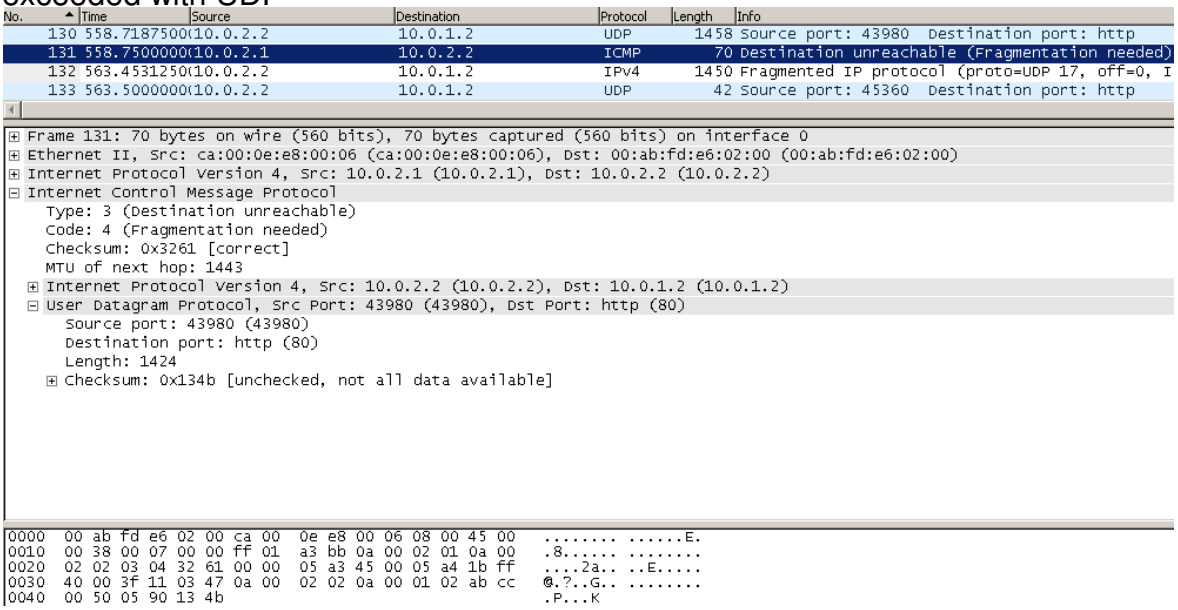
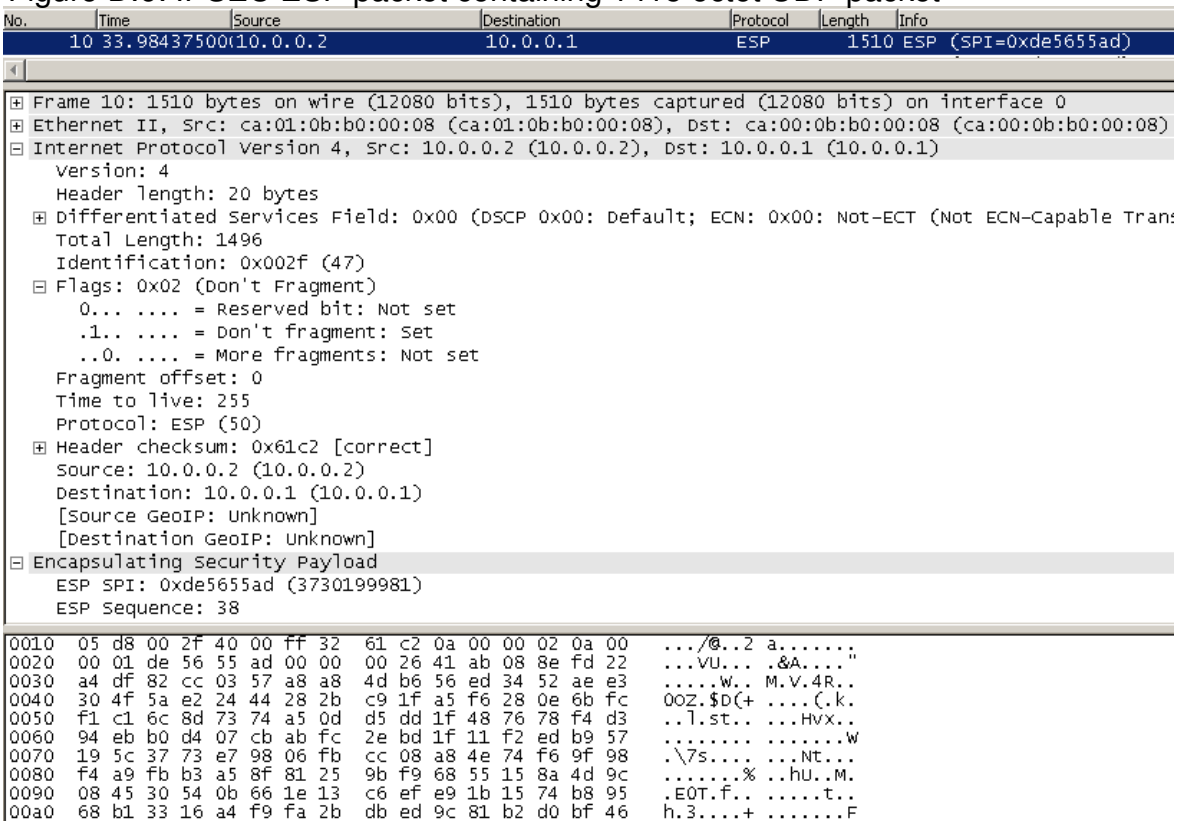


Figure D.5 above shows that with UDP the same scenario occurs as with TCP. In this figure a payload of 1416 octets is sent, which with 8 octets of UDP header, 20 octets of IPv4 header and 14 bytes of Ethernet header, that comes to 1458 octets. 1458 and 53 octets for the IPSEC tunnel is 1511 octets, 1 more than the MTU can achieve.
An ICMP fragmentation needed packet is sent back to HOST2 and the packet is fragmented and re-transmitted. These ICMP messages are only generated once and the host stores the maximum MTU value received inside the ICMP message. In order to generate this for a second time using UDP packets, I restarted GNS3 (there by, restarting all the virtual hosts and routers).

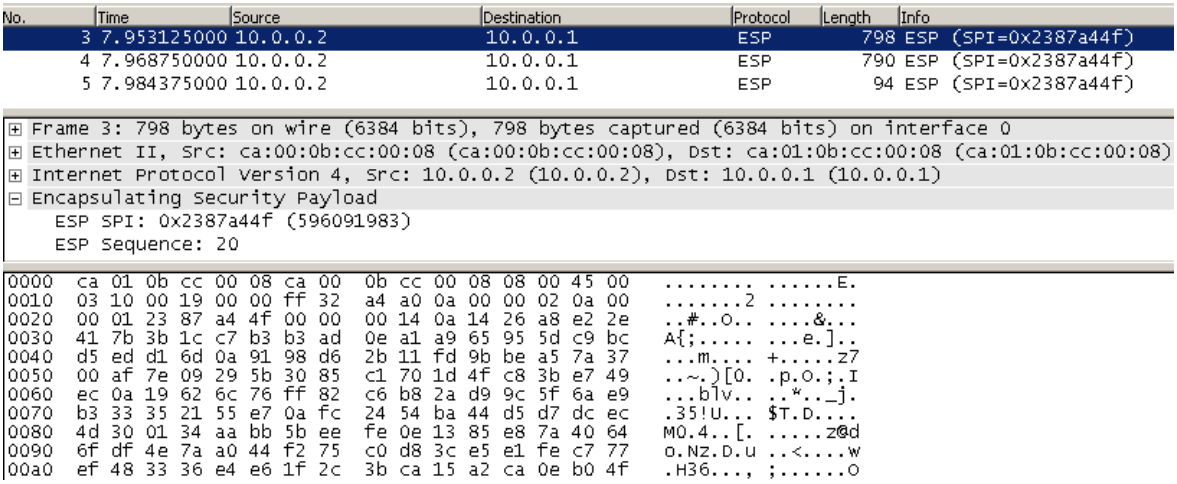
Figure D.6: IPSEC ESP packet containing 1415 octet UDP packet



Above in Figure D.6 is a packet capture running on the links between R1 and R2, showing the IPSEC tunnel traffic. When sending a 1415 octet UDP packet from HOST2 to HOST1, which is the largest possible payload size without exceeding the MTU, we can see a 1510 octet ESP frame being sent from R2 to R1.

Lastly, Figure D.7 below is a packet capture on the link between the two routers R1 and R2. Here we see that two packets of a similar size are sent, 798 and 790 octets in length. After a UDP packet has been sent from HOST2 that exceeds the IPSEC tunnel MTU, as shown in Figure D.5, HOST2 will fragment packets larger than the MTU, because it was signalled using ICMP. However, if the MTU changes on the link between R1 and R2, this isn't signalled to HOST2. I have entered the interface configuration command "ip mtu 1000" on interface Fa0/0 on router R2, and then disabled and enabled the interface. This caused the IPSEC tunnel to then re-establish at the new lower MTU of 1000 octets (down from 1500 octets). This now means that HOST2 doesn't fragments packets larger than 1000 octets because it believes 1443 is the link MTU as per Figure D.5. Now router R2 must perform fragmentation and R1 must re-assemble fragmented packets.

Figure D.7: UDP fragmentation over IPSEC performed by routers



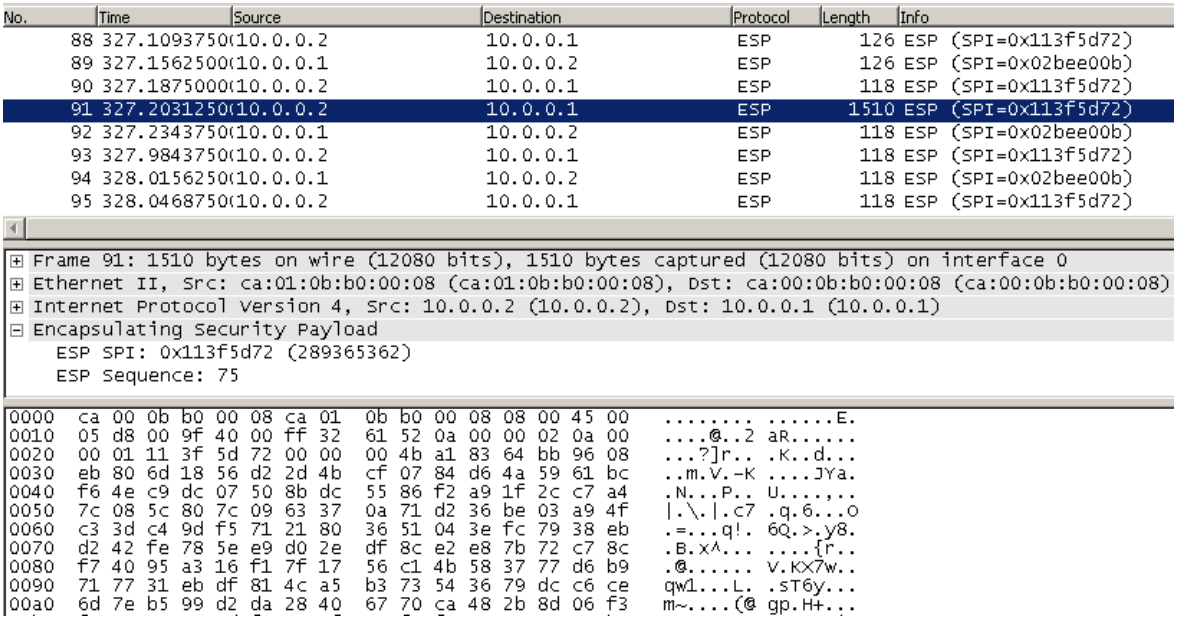
```
No.  Time          Source      Destination  Protocol  Length  Info
  3  7.953125000   10.0.0.2    10.0.0.1     ESP        798    ESP (SPI=0x2387a44f)
  4  7.968750000   10.0.0.2    10.0.0.1     ESP        790    ESP (SPI=0x2387a44f)
  5  7.984375000   10.0.0.2    10.0.0.1     ESP         94    ESP (SPI=0x2387a44f)

⊞ Frame 3: 798 bytes on wire (6384 bits), 798 bytes captured (6384 bits) on interface 0
⊞ Ethernet II, Src: ca:00:0b:cc:00:08 (ca:00:0b:cc:00:08), Dst: ca:01:0b:cc:00:08 (ca:01:0b:cc:00:08)
⊞ Internet Protocol Version 4, Src: 10.0.0.2 (10.0.0.2), Dst: 10.0.0.1 (10.0.0.1)
⊟ Encapsulating Security Payload
    ESP SPI: 0x2387a44f (596091983)
    ESP Sequence: 20

0000  ca 01 0b cc 00 08 ca 00  0b cc 00 08 08 00 45 00   ........ ......E.
0010  03 10 00 19 00 00 ff 32  a4 a0 0a 00 00 02 0a 00   .......2 ........
0020  00 01 23 87 a4 4f 00 00  00 14 0a 14 26 a8 e2 2e   ..#..O.. ....&...
0030  41 7b 3b 1c c7 b3 b3 ad  0e a1 a9 65 95 5d c9 bc   A{;..... ...e.]..
0040  d5 ed d1 6d 0a 91 98 d6  2b 11 fd 9b be a5 7a 37   ...m.... +.....z7
0050  00 af 7e 09 29 5b 30 85  c1 70 1d 4f c8 3b e7 49   ..~.)[0. .p.O.;.I
0060  ec 0a 19 62 6c 76 ff 82  c6 b8 2a d9 9c 5f 6a e9   ...blv.. ..*.._j.
0070  b3 33 35 21 55 e7 0a fc  24 54 ba 44 d5 d7 dc ec   .35!U... $T.D....
0080  4d 30 01 34 aa bb 5b ee  fe 0e 13 85 e8 7a 40 64   M0.4..[. .....z@d
0090  6f df 4e 7a a0 44 f2 75  c0 d8 3c e5 e1 fe c7 77   o.Nz.D.u ..<....w
00a0  ef 48 33 36 e4 e6 1f 2c  3b ca 15 a2 ca 0e b0 4f   .H36..., ;......O
```

For clarity of the relationship between the original packets and the ESP packets, the following two figures show firstly a 1391 octet TCP packet being sent across the IPSEC tunnel, then a 1392 octet packet. You can see how the ESP packets directly related to the TCP packets sent, with regards to the number of packets sent, in Table D.2.

Figure D.8: Maximum sized TCP packet sent across the IPSEC tunnel



```
No.  Time          Source      Destination  Protocol  Length  Info
 88  327.1093750(  10.0.0.2    10.0.0.1     ESP        126    ESP (SPI=0x113f5d72)
 89  327.1562500(  10.0.0.1    10.0.0.2     ESP        126    ESP (SPI=0x02bee00b)
 90  327.1875000(  10.0.0.2    10.0.0.1     ESP        118    ESP (SPI=0x113f5d72)
 91  327.2031250(  10.0.0.2    10.0.0.1     ESP       1510    ESP (SPI=0x113f5d72)
 92  327.2343750(  10.0.0.1    10.0.0.2     ESP        118    ESP (SPI=0x02bee00b)
 93  327.9843750(  10.0.0.2    10.0.0.1     ESP        118    ESP (SPI=0x113f5d72)
 94  328.0156250(  10.0.0.1    10.0.0.2     ESP        118    ESP (SPI=0x02bee00b)
 95  328.0468750(  10.0.0.2    10.0.0.1     ESP        118    ESP (SPI=0x113f5d72)

⊞ Frame 91: 1510 bytes on wire (12080 bits), 1510 bytes captured (12080 bits) on interface 0
⊞ Ethernet II, Src: ca:01:0b:b0:00:08 (ca:01:0b:b0:00:08), Dst: ca:00:0b:b0:00:08 (ca:00:0b:b0:00:08)
⊞ Internet Protocol Version 4, Src: 10.0.0.2 (10.0.0.2), Dst: 10.0.0.1 (10.0.0.1)
⊟ Encapsulating Security Payload
    ESP SPI: 0x113f5d72 (289365362)
    ESP Sequence: 75

0000  ca 00 0b b0 00 08 ca 01  0b b0 00 08 08 00 45 00   ........ ......E.
0010  05 d8 00 9f 40 00 ff 32  61 52 0a 00 00 02 0a 00   ....@..2 aR......
0020  00 01 11 3f 5d 72 00 00  00 4b a1 83 64 bb 96 08   ...?]r.. .K..d...
0030  eb 80 6d 18 56 d2 2d 4b  cf 07 84 d6 4a 59 61 bc   ..m.V.-K ....JYa.
0040  f6 4e c9 dc 07 50 8b dc  55 86 f2 a9 1f 2c c7 a4   .N...P.. U....,..
0050  7c 08 5c 80 7c 09 63 37  0a 71 d2 36 be 03 a9 4f   |.\.|.c7 .q.6...O
0060  c3 3d c4 9d f5 71 21 80  36 51 04 3e fc 79 38 eb   .=...q!. 6Q.>.y8.
0070  d2 42 fe 78 5e e9 d0 2e  df 8c e2 e8 7b 72 c7 8c   .B.x^... ....{r..
0080  f7 40 95 a3 16 f1 7f 17  56 c1 4b 58 37 77 d6 b9   .@...... V.KX7w..
0090  71 77 31 eb df 81 4c a5  b3 73 54 36 79 dc c6 ce   qw1...L. .sT6y...
00a0  6d 7e b5 99 d2 da 28 40  67 70 ca 48 2b 8d 06 f3   m~....(@ gp.H+...
```

Figure D.9: TCP packet 1 byte over maximum possible is sent across the IPSEC tunnel

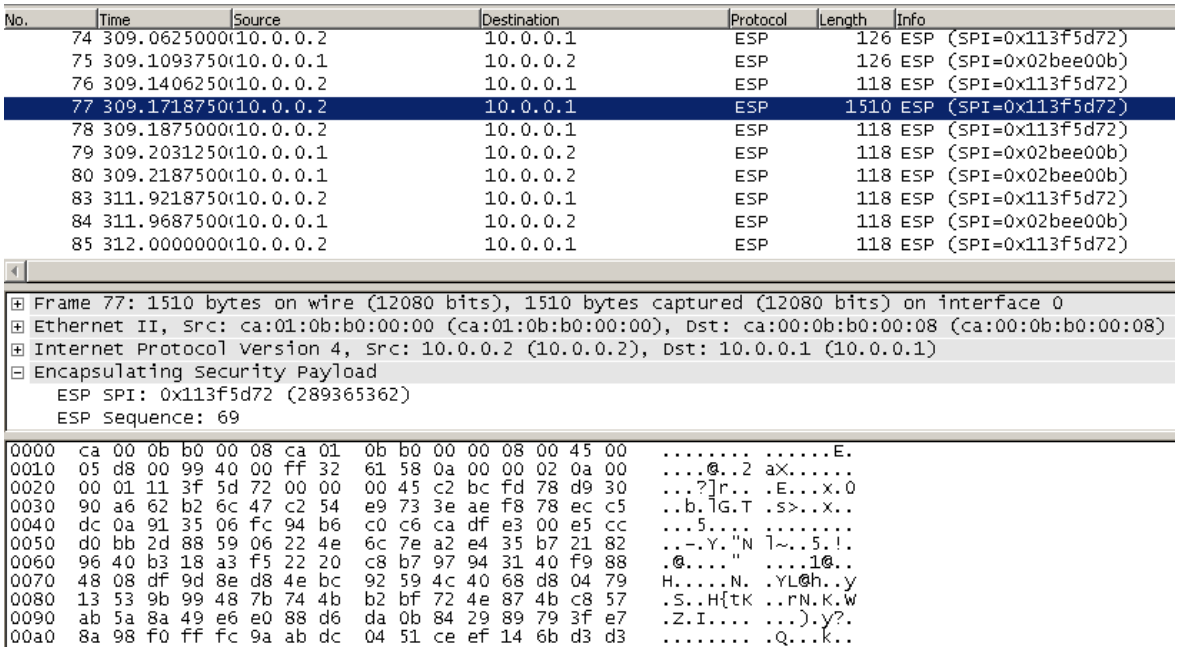


Table C.2: Comparing TCP and ESP packets

1391 Octet TCP Packet	
TCP packet flags on packets sent (>) and received (<) by HOST2	**Corresponding packet number in Figure D.8**
SYN >	88
SYN,ACK <	89
ACK >	90
PSH, ACK > (1391 octets in payload)	91
ACK <	92
FIN, ACK >	93
FIN, ACK <	94
ACK >	95
1392 Octet TCP Packet	
TCP packet flags on packets sent (>) and received (<) by HOST2	**Corresponding packet number in Figure D.9**
SYN >	74
SYN,ACK <	75
ACK >	76
ACK > (1391 bytes on payload)	77
PSH, ACK > (1 remaining byte on payload)	78
ACK <	79
ACK <	80
FIN, ACK >	83
FIN, ACK <	84
ACK >	85