Designing and Building a Virtual Cyber Security Range

(Technical Paper)

Case Study Analysis of IPv4's Ongoing Usage Despite Technical Limitations

(STS Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree

Bachelor of Science, School of Engineering

Emil Baggs

Fall, 2022

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signature _____ Date _____

Emil Baggs

Approved _____ Date _____

Yonghwi Kwon, Department of Computer Science

Approved _____ Date _____

Joshua Earle, Department of Engineering and Society

Introduction

In computer science and specifically computer networking, protocols are technical definitions for how two processes talk to each other. Protocols ensure that two processes or devices know exactly what to send and expect to receive in order to communicate efficiently and effectively for whatever purpose the protocol serves (*What is the internet protocol?*). For example, the Hyper Text Transfer Protocol (HTTP) defines how a web browser should communicate with a web server so that end users can request specific pages, provide inputs to forms, upload files, etc (*What is HTTP?*). In most cases, these protocols are designed to rely on each other, often encapsulating one protocol within another.

For my STS Project, I plan to perform a deeper dive into some of the complex factors that go into determining how widely used a protocol becomes, aside from the protocol's explicit technical advantages. I am going to perform a case study looking at the Internet Protocol version 4 (or IPv4), which is one of the most widely used protocols used today, and what factors have led to its continual usage, despite its well known technical flaws. Internet Protocol version 6 (IPv6) has existed for over 25 years and was designed specifically to address flaws in IPv4. Despite the flaws that it addresses, only around 25% of the world has transitioned from IPv4 to IPv6 as of 2018 (*State of IPv6 deployment 2018,* 2021). So why is IPv4 still so widely used? What factors have led to IPv4's continued success?

My computer science technical project is an independent research project working on building a Cyber Range utilizing various virtualization and automation technologies. This Cyber Range will be designed to automatically create virtual lab networks to test Cyber Security methods on. The goal of the project is to have a cluster of servers that has tools to automatically build brand new networks according to a single configuration file, greatly reducing the

administrative time spent manually building the networks. A few permanent configurations will exist on the cluster, such as a centralized firewall to provide internet access, but otherwise the cluster will dynamically build and destroy networks as needed. This makes it possible to create lab scenarios that are easy to deploy and practice with, without any risk of disrupting anything beyond the lab network. This project relates to my STS Research as they both involve diving into networking protocols in order to increase my understanding. My STS Research will primarily focus on the social and nontechnical factors related to the protocol, while my technical project will look specifically at the technical aspects involved with actually using the protocol.

For the remainder of the prospectus, I'll have a three sections talking about different aspects of my project. First I'm going to go more in depth about my STS project and the research question I'll be focusing on. I'll then talk about my technical project and the goal that I hope to accomplish with it. Finally, I'll talk about some key texts that I have found in my early research that have guided my questions surrounding my project.

STS Project

The main focus of my STS research will be focusing on how computer networking has rapidly developed over the last several decades and how that history showcases the way that social factors influence the progression and adoption of new technologies. Prior to the mass adoption of computers and then the internet, there were very few technological innovations that had such a massive global impact. While there are still major issues surrounding unequal access to the internet, the internet has also helped give people a voice who otherwise would struggle to be heard (Dyer, 2017). Given the immense user base for the internet, I think it is important to look more closely at what technology actually drives the connectivity between every single

online device. Compared with some other technologies, network protocols are very publicly designed, tested, and implemented. The Internet Engineering Task Force (IETF) is a large group of individuals from across the globe who collect, design, and publish standards for many protocols that the internet relies on. The IETF, when trying to standardize a new protocol or protocol version, publish Request For Comments (RFCs) that serve to outline the current design of a protocol or internet standard, then collect any feedback it possibly can from those that would use the protocol. This helps ensure that protocols are thoroughly thought out and not designed in a way to inadvertently disadvantage a specific social group. This also helps alleviate unhealthy competition within computer networking that could fracture access between social groups from proprietary protocols (*About the internet engineering task force (IETF)*). While there is nothing stopping a company from designing a proprietary protocol or not following the standards published by the IETF, the fact that there is a published standard means it is in everyone's best interest to either abide by those standards or help improve them.

While the IETF has helped govern the creation and standardization of protocols since their founding in the early 1990's, IPv4 was a protocol originally designed in 1981 by DARPA, a United States Defense Agency. The protocol had a much more limited scope at the time and DARPA had no idea how ingrained the protocol would become in everyday life. Likely the biggest issue that the IPv4 protocol has run up against is the extremely limited number of IP addresses, which are the unique addresses for computers connected to the internet. An IPv4 address is made up of four bytes, meaning that there are only 2^{32} or a little less than 4.3 billion unique IP addresses and is used to represent a unique device connected to the internet. While this number seemed large at the conception of IPv4, even in the 1990s it became apparent that we pretty quickly would have well more than 4.3 billion devices needing to connect to the internet

(*Internet protocol version 4(IPv4) history & evolution: Prefixx News,* 2020). And, on top of that, as the United States created the internet, it controlled the distribution of these unique IP addresses, ensuring that the US had a step up from other countries in how quickly it could develop network infrastructure. The US currently has over a third of all IPv4 addresses globally, roughly five times more than China who has the second most IPv4 addresses (*IP address by country 2022*).

The IETF is an obvious social group that has had a lot of influence in both versions of IP, but perhaps the most influential social group are the many Internet Service Providers all across the world. The Internet Service Providers directly influence which consumers are using IPv4 vs IPv6, so researching the role that ISPs played, especially around the turn of the century, will be core to my research. Other major social groups include the United States government, large digital content providers, and several major Universities, all of which played (and still play) a major role in deciding the direction that Internet technology developed.

Despite flaws that have been known for more than half of IPv4's life, it is still fundamentally ingrained into nearly every interaction we have online. Because of this, I will be using IPv4 as a subject for a case study analysis on computer networking protocols and why they stay relevant. Most of research will come from historical references, looking at the fairly young and well documented history of the internet. I'll also look for examples of public policy which have impacted how people use the internet and consequentially IPv4. I will be predominantly focusing on the early stages of IPv4 and IPv6, as the initial competition between these protocols established many of the trends that are reflected today. I except to perform the majority of this research in the next three to four months so I can talk about the answers I've found in my final thesis.

Technical Project

In order to get my technical project to a fully working state, I am going to use a number of different preexisting tools and software that are designed for creating compute clusters and automating their usage. The server cluster is built off of the software Proxmox, which is an open source operating system for running several virtual machines on a single physical server (*What is Proxmox Ve?,* 2021) . Virtual machines are a widely used concept in computing, where a single physical server runs software that allows it to carve up its own resources (like CPU cores, RAM, disk space, etc.) into different virtual machines, which can act independently as if they were their own physical machines. This makes it possible to effectively run hundreds of virtual computers off of a single sufficiently powerful physical server (*What is a virtual machine (VM)?,* 2022). Proxmox also supports clustering, which means that we can connect several physical servers together and manage them all through a single interface. This makes it trivial to move a virtual machine onto a physically different server, allowing you to distribute the computational load that running many virtual machines can have.

Aside from Proxmox, another big tool that I am using is Terraform, which interacts with Proxmox to create, destroy, or modify virtual machines according to the state defined in a configuration file. Terraform is also very modular, making it possible to write complex logic for building small networks and fitting them together. For example, you could implement logic in a module that would make it so that your configuration file just contains a small lab network, which then gets cloned several times to have multiple copies that several people can work with. Terraform will automatically create whatever machines you specified, which can be a massive time save when deploying hundreds of machines. Terraform also does a small amount of

provisioning to the virtual machines, which makes sure that when the virtual machine gets created, it has working network access and a user we can manage it with. This enables us to remotely login to the machine and perform more complex automated setup (Austin, *How to deploy VMS in Proxmox with Terraform,* 2021).

Finally, the third major tool I will use in my technical project is Ansible. Ansible runs after Terraform and is the tool that uses remote login to automate whatever tasks I want. Ansible runs with a configuration file and playbook file to determine what machines to connect to and what to do once connected. A significant portion of the project is writing the playbook, which implements the logic to determine which tasks to run on each remote system. For example, the configuration file might specify three web servers that I want Ansible to provision. After Terraform has created the blank machines on the Proxmox cluster, Ansible will run, connect to each virtual machine, determine that they should be web servers from the configuration, then install and setup a simple website. Ansible is especially powerful as it can perform these tasks in parallel, saving a lot of time by configuring dozens of machines at the same time (*How Ansible Works*). Most of the work I need to do with Ansible will be writing different roles, which correspond to different services I want the cluster to build. These roles will tell Ansible to install software, update configurations, create users, etc. depending on what the configuration says a virtual machine needs.

Together I can use these three tools to accomplish the overall goal of the technical project. Most of the work involved is writing the specific modules that I want to have at my disposal in order to make the system as robust as possible. This involves creating several template images with different operating systems at different versions, such as Windows 7, Windows Server, Ubuntu, CentOS, etc. I then need to make sure that I have Terraform modules

that work with each of these templates. Finally, I'll need to write Ansible playbooks to support the wide range of services that I want to be able to automatically setup. As this project is very highly modular, most of the work will be expanding the configuration options available, rather than setting up an initial "working" state that is very limited in what it can do.

Key Texts

The most useful text that I found so far on my STS topic is the article "State of IPv6 Deployment 2018", which was posted on the Internet Society website. It gives a great overall explanation of the issue the world is facing with IPv4 and why IPv6 is better, then goes on to explain how slowly it has been phasing out IPv4. It also talks about who is switching to IPv6, which helps give some insight onto the social factors involved with the protocols. It even includes a few small case studies on large network entities who have made the transition. The article also does a great job of showing actual data on the matter, with a combination of graphs and statistics.

I also found the article on World Populate Review, "IP Address by Country 2022", which gives a good overview of how IPv4 addresses are distributed and some of the ramifications of that. It also gives a good explanation of how IP addresses are given out and where those IPs end up. Similar to the State of IPv6 article, this web page also has specific statistics that I can use for backing up claims about the issues with IPv4.

The page on the Internet Society website that explains who the IETF are was also very helpful in providing an explanation of who actually helps develop these protocols and computer standards. The IETF have played a huge role in making the internet as accessible as it is, but don't tend to get a lot of credit. The Internet Society partners with the IETF and beyond the

single page outlining who they are, there are also several other pages on their site that provide additional background into specific aspects of the IETF.

Finally, although it isn't technical, I found the article "The internet is giving a voice to those on the margins – losing net neutrality will take it away" on The Conversation, which talks about the way that internet access has made it possible for anyone's voice to be amplified, and what the real impact of that is. The article discusses how this has helped bridge knowledge gaps as information is available to everyone, but also how not all information online is correct or a complete picture. It also talks about how there is still a significant majority that doesn't have their voice heard among the masses online, which can lead to even more issues.

Works Cited

About the internet engineering task force (IETF). Internet Society. (n.d.). Retrieved from https://www.internetsociety.org/about-the-ietf/

Austin. (2021, September 1). *How to deploy VMS in Proxmox with Terraform*. Austin's Nerdy Things. Retrieved from https://austinsnerdythings.com/2021/09/01/how-to-deploy-vms-in-proxmox-with-terraform/

Cloudflare. (n.d.). *What is the internet protocol?* Cloudflare. Retrieved from https://www.cloudflare.com/learning/network-layer/internet-protocol/

Cloudflare. (n.d.). *What is HTTP?* Cloudflare. Retrieved from https://www.cloudflare.com/learning/ddos/glossary/hypertext-transfer-protocol-http/

Dyer, H. T. (2017, December 19). *The internet is giving a voice to those on the margins – losing net neutrality will take it away*. The Conversation. Retrieved from https://theconversation.com/the-internet-is-giving-a-voice-to-those-on-the-margins-losing-net-neutrality-will-take-it-away-89259

Internet protocol version 4(IPv4) history & evolution: Prefixx News. Prefixx. (2020, January 16). Retrieved from https://prefixx.net/news/ipv4-history

IP address by country 2022. World Population Review. (n.d.). Retrieved from https://worldpopulationreview.com/country-rankings/ip-address-by-country

Redhat. (n.d.). *How Ansible Works*. Ansible.com. Retrieved from https://www.ansible.com/overview/how-ansible-works

State of IPv6 deployment 2018. Internet Society. (2021, July 8). Retrieved from https://www.internetsociety.org/resources/2018/state-of-ipv6-deployment-2018/

What is a virtual machine (VM)? Red Hat - We make open source technologies for the enterprise. (2022, May 11). Retrieved from https://www.redhat.com/en/topics/virtualization/what-is-a-virtual-machine

What is Proxmox Ve? Hivelocity Hosting. (2021, October 22). Retrieved from https://www.hivelocity.net/kb/what-is-proxmox/