ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

Network Performance Analysis of Carrier Grade Network Address Translation with Network Traffic Prioritization

Abraham Meles

Advisor

Dr. Sosina Mengistu

A Thesis Submitted to the School of Electrical and Computer Engineering in Partial Fulfillment of the Requirements for the Degree of Masters of Science in Telecommunication Engineering

November 2021

Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

TELECOMMUNICATION ENGINEERING GRADUATE PROGRAM

Network Performance Analysis of Carrier Grade Network Address Translation with Network Traffic Prioritization

Abraham Meles

Approval by Board of Examiners

_____	_____	_____/_____/_____
Chairman/School Dean	Signature	Date
Dr. Sosina Mengistu	_____	_____/_____/_____
Advisor Name	Signature	Date
Dr. Fitsum Assamnew	_____	_____/_____/_____
Examiner	Signature	Date
Dr. Mesfin Kifle	_____	_____/_____/_____
Examiner	Signature	Date
_____	_____	_____/_____/_____
Director, Postgraduate Program	Signature	Date

Declaration

I, the undersigned, declare that this MSc thesis is my original work, has not been presented for the fulfillment of a degree in this or any other university, and all sources and materials used for the thesis have been acknowledged.

Abraham Meles _____

Name Signature

Addis Ababa, Ethiopia _____

Place Date of submission

This thesis has been submitted for examination with my approval as a university advisor.

Dr. Sosina Mengistu _____

Advisor's Name Signature

Abstract

In recent years, Internet traffic increased very rapidly due to network technology growth. Additionally, the availability of fixed broadband is continually growing; such as those based on Asymmetric Digital Subscriber Line (ADSL) and fiber technologies. This huge demand is due to technological advancement, an increase in smartphone users, and the availability of new services by Internet service providers (ISPs). With this increasing demand for fixed broadband Internet service, improvement is needed from ISP to meet customer quality of service demand.

Internet Protocol version 4 (IPv4) addresses are used to link network devices in a packet-switched network. All regional Internet registry pools of public IPv4 addresses have been depleted. With the depletion of public IPv4 addresses, ISPs are attempting to find an alternative solution to the problem. Several techniques have been proposed to address public IPv4 depletion issues. Carrier-Grade NAT (CGN) is one of the techniques utilized to solve the problem. Despite the fact that CGN solves public IPv4 address depletion by sharing a single routable IPv4 address for multiple private IPv4 addresses, it has significant flaws. CGN interferes with the smooth operation of several applications. It also breaks the Internet's end-to-end functioning premise. The CGN process is one of the causes of consumer Internet service delays, which leads to customer speed issues.

The aim of this thesis is to examine the effect of CGN on network performance and to provide a solution for enhancing network performance. Also, examines the impact of the suggested solution which is adopting CGN in conjunction with network traffic prioritization on network performance. The evaluations were carried out on two scenarios: one with CGN operation and the other CGN combined with network traffic prioritization. For implementing and evaluating network performance of the scenarios, simulation tools such as Graphical Network Simulator-3 (GNS3), Ostinato traffic

generator, and monitoring tool Paessler Router Traffic Grapher (PRTG) are employed. The performance of the two scenarios is analyzed using network metrics such as packet loss, latency, and throughput. The analysis results indicate that when CGN is used in conjunction with traffic prioritization, throughput improves by 19% on average, latency improves by 21%, and packet loss improves by 20%.

Keywords—CGN, IPv4, Performance Metrics, Traffic Prioritization, Address Depletion

Acknowledgment

First and foremost, I want to thank GOD for giving me the bravery, showing me the way, and providing me with the strength to finish the thesis work.

Second, I would like to convey my heartfelt appreciation to my adviser Dr. Sosina Mengistu for steering me in the correct way. Her outstanding assistance, information sharing, and great advice played a very crucial role in completing this study.

In addition, I am grateful for the partnership of AAiT and Ethio telecom in developing this wonderful program. My heartfelt gratitude goes to Ethio telecom for sponsoring this excellent postgraduate program and to AAiT for allowing me to participate in the program.

Finally, I would like to express my gratitude to my family, friends, coworkers, and classmates for their encouragement, assistance, sound advice, and best wishes throughout the program, particularly during this thesis work.

Table of Content

List of Figures

List of Tables

List of Acronyms

CoS	Class of Service
IETF	Internet Engineering Task Force
IS-IS	Intermediate System to Intermediate System
ToS	Type of service
TTL	Time to Live
ADSL	Asymmetric Digital Subscriber Line
ALG	Application Layer Gateway
BRAS	Broadband Remote Access Server
CGN	Carrier Grade NAT
CPU	Central Processing Unit
GNS3	Graphical Network Simulator-3
ICE	Interactive Connectivity Establishment
IEEE	Institute of Electrical and Electronics Engineers
IGP	Interior Getaway Protocol
IHL	Internet Header Length
IP	Internet Protocol
IPRAN	Internet Protocol Radio Access Network
IPSLA	Internet Protocol Service Level Agreement
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
ISP	Internet Service Provider
KPIs	Key Performance Indicators

LAN	Local Area Network
LDP	Label Distribution Protocol
MATLAB	Matrix Laboratory
MPLS	Multi-Protocol Label Switching
NAPT	Network Address and Port Translation
NAT	Network Address Translation
NGN	New Generation Network
PCP	Port Control Protocol
PRTG	Paessler Router Traffic Grapher
QoS	Quality of Service
RFC	Request for Comments
SL-NAT	Standalone Network Address Translation
TCP	Transmission Control Protocol
TP	Traffic Prioritization
UDP	User Datagram Protocol
VRF	Virtual Routing and Forwarding
WAN	Wide Area Network

1. Introduction

1.1 Background

In recent years, due to the development of network technology [1], Internet traffic has grown very rapidly. Furthermore, the availability of broadband user connections, such as those based on Asymmetric Digital Subscriber Line (ADSL) and fiber optic technologies, is increasing. With the improvement of this kind of usability and service quality, users are more willing to use various services provided by the Internet. Internet service providers (ISPs) play a crucial role in managing and monitoring this dynamic Internet traffic.

The decrease in fixed broadband data package tariffs has significantly boosted subscriber demand for Internet services in Ethio telecom. To meet the growing demand for fixed broadband service, Ethio telecom employs a variety of access network equipment. Internet access is provided via two types of Internet Protocol version 4 (IPv4) addresses. Customers that request a public IPv4 address as part of their broadband service will be assigned a block of public IP addresses based on their requirements. Other clients receive broadband Internet service through private IPv4 addresses. To send traffic to the Internet, customers with private IPv4 addresses must use the Network Address Translation (NAT) protocol to convert their private IPv4 addresses to public IPv4 addresses.

Carrier-Grade NAT (CGN) is a viable solution to the IPv4 address depletion problem. It also allowing service providers and content providers to make a smooth transition to Internet Protocol Version 6 (IPv6). CGN employs Network Address and Port Translation (NAPT) approaches to convert a large number of private IP addresses into a smaller number of public IPv4 addresses [2].

Quality of service metrics such as latency, throughput, delay, and jitter is used to assess network performance [3]. CPU usage, memory utilization, bandwidth utilization, and device temperature are all Key Performance Indicators (KPIs) for monitoring network device performance. These metrics assess the performance of a particular packet-based network.

Ethio telecom's IP New Generation Network (NGN) bearer network is built on three layers: the Backbone Layer, the Core Layer, and the Edge Layer [4]. The layer hierarchical architecture benefits service providers by making the network more controllable and scalable for future development.

The Standalone Network Address Translation (SL-NAT) router is a NAT device whose sole job is to execute IPv4 address translation. NAT devices are frequently deployed at network boundaries to translate the addresses of packets passing through the NAT [2]. The SL-NAT device allows networks on both sides of the NAT to utilize their own address space. It also saves network addresses and converts them between IPv4 address formats (i.e. private IPv4 and public IPv4).

1.2 Statement of the problem

Fixed broadband data services are in high demand. This enormous demand is the result of technological advancement, a rise in smartphone users, and the availability of new services by ISPs. There has been a surge in consumer demand after Ethio telecom reduced the monthly cost for fixed broadband. Coronavirus, in addition to pricing adjustments, contributes to the demand for fixed broadband service. With the growing demand for broadband Internet access, ISPs must upgrade to satisfy consumer quality of service demands. With the depletion of public IPv4 addresses, ISPs are attempting to find an alternative solution to the problem. The CGN technique is one method for resolving the issue. CGN may translate several clients' private IPv4 addresses into a single routable

public IPv4 address. CGN has a detrimental influence on the smooth operation of several applications. It also violates the Internet's end-to-end functioning premise. Because CGN handles millions of translations, the operation must be efficient in order to offer the needed QoS to consumers. Operators [5] expressed their dissatisfaction with the lack of well-developed best practices for configuring and dimensioning CGNs. When it comes to setting CGNs, respondents in [5] cited preserving port space and the amount of session state in CGNs as the key obstacles. The NAT procedure is one of the reasons for Internet service delays, which leads to consumer speed issues. To overcome this issue, the CGN deployment should be combined with network traffic prioritization in order to improve QoS performance.

CGN's operating strategy is to treat every traffic that arrives to be NATed identically, which makes the procedure inefficient. As a result, network performance suffers. When CGN is used with network traffic priority, traffic categorization or marking occurs. Network traffic prioritization allows administrators to treat distinct types of customer traffic based on their latency sensitivity and customer type (i.e. enterprise or residential). Clients that supply critical services to their customers need a higher priority than those who use the Internet for basic web surfing and emailing. Since the CGN operation handles all traffic equally, the CGN's limited resources must be used efficiently. In order to improve the quality of experience for consumers when using CGN, it needs to be paired with network traffic prioritization. Thus, adopting CGN along with network traffic prioritization will improve the network performance of the CGN process by making better use of the CGN's limited resources. The purpose of this thesis is to look at how CGN combined with network traffic prioritization may increase network performance. To the best of my knowledge, this study is the first attempt to examine the hypothesis.

1.3 Objective

1.3.1 General Objective

The primary goal of this thesis is to investigate the improvement of network performance by integrating CGN with network traffic prioritization.

1.3.2 Specific Objective

- To study the basic principles of CGN operations.
- To look into the effect of CGN on network performance metrics.
- To simulate and assess the CGN and CGN with network traffic prioritization architecture.
- To compare and analyze the network performance of CGN architecture with CGN paired with network traffic prioritization architecture.

1.4 Scope of the Study

This thesis is limited to network performance study on CGN and CGN combined with network traffic prioritization architecture on the IP core network, and its design is bound by Ethio telecom's IP bearer network.

1.5 Literature Review

NAT444 is a technique used for IPv4 address extension technology that allows ISPs to provide IPv4 service to their clients. This technique adds extra address translation in the service provider network [6]. CGN contributes to IPv4 continuity, even if it has an effect on specific Internet applications. The authors of [6] demonstrate the influence of CGN on various Internet applications such as peer-to-peer, online gaming, big file transfer, and so on. The paper in [7] looked into the effect of CGN on user traffic. TCP traffic between client and server is used to test the effect of CGN on web surfing. The result in [7] shows CGN has less impact on web browsing even though CGN breaks the end-to-end Internet connection principle. Both publications demonstrated the influence of CGN on specific

applications but did not address the impact on network performance; this study attempts to fill that gap by examining the impact of CGN on network performance.

The authors in [8] examined the possible impact of CGN on residential broadband customers. The pace at which entries are produced and withdrawn from the session table, as well as the overall size of the session table, are the limiting variables that define a CGN device's performance [8]. Table size and port numbers are limited resources that necessitate effective CGN operation provisioning by ISPs. One of the variables influencing CGN functioning is the number of customers who share the same CGN device without suffering substantial performance reduction [8]. For optimal usage of the session table, certain short-lived UDP sessions are required to be ended as soon as possible [8]. According to the authors of [9], CGN implementation is rapidly increasing. They also identified a lack of a consistent standard for running CGN, which leads to a lack of information on the influence of CGN on performance, which this paper attempts to investigate.

Delay is one of the most important performance metrics in IP networks [10]. Transmission delays [10] are an unavoidable aspect of communication quality. In [10], the authors demonstrate the influence of NAT operation on latency in IP networks. The NAPT procedure resulted in a 175 percent delay from the initial value. This is because, during NAT, the IPv4 and port header values will be changed with new IPv4 address and port number values. The authors of [11] investigated the influence of CGN on latency by simulating situations such as bypass, routing, and mapping modes of CGN setup. The mapping mode [11] signifies a CGN device with NAT features, routing mode denotes a CGN device operating as a regular router that routes traffic, and bypass mode denotes a CGN device that is bypassed. The result of [11] indicates that in mapping mode, there is a rise in the time it takes to establish a connection.

The authors of [5] demonstrate the breadth of CGN implementation on the Internet and its impact on end-users. According to the survey [5] they distributed to ISPs, 40 percent of the 75 ISPs adopted CGN. Following that, they attempted to determine the influence of CGN on subscribers' Internet applications, with the main issue being that some services, such as online gaming, did not function appropriately with CGN set up [5]. Furthermore, the operators were concerned about the lack of well-developed best practices for setting up and operating CGN [5]. To successfully run CGN, operators must resort to experimenting in areas such as the allocation of external IP addresses and port number ranges to consumers [5].

1.6 Methodology

The following procedures will be carried out to ensure the effective completion of this thesis. The method comprises mostly of a literature study, simulation, and interpretation of simulation findings. The approach to be used to achieve the general and particular objectives, as well as the stated problem, is:

- Works of literature will be acquired, researched, compared, and assessed in order to choose the most pertinent among them.
- The primary sources of the literature chosen will be Google Books, Institute of Electrical and Electronics Engineers (IEEE) Explore, Research Gate, and vendor websites.
- The simulation tool will be used to compare CGN base architecture with CGN combined with network traffic prioritization based architecture.
- The outcomes of the two scenarios will be compared.

- Simulators such as Graphical Network Simulator-3 (GNS3) and tools such as Paessler Router Traffic Grapher (PRTG), Ostinato, Excel, and IP-SLA were employed in this thesis.

Figure 1.1 depicts the flow chart illustrating the methodology to be used.

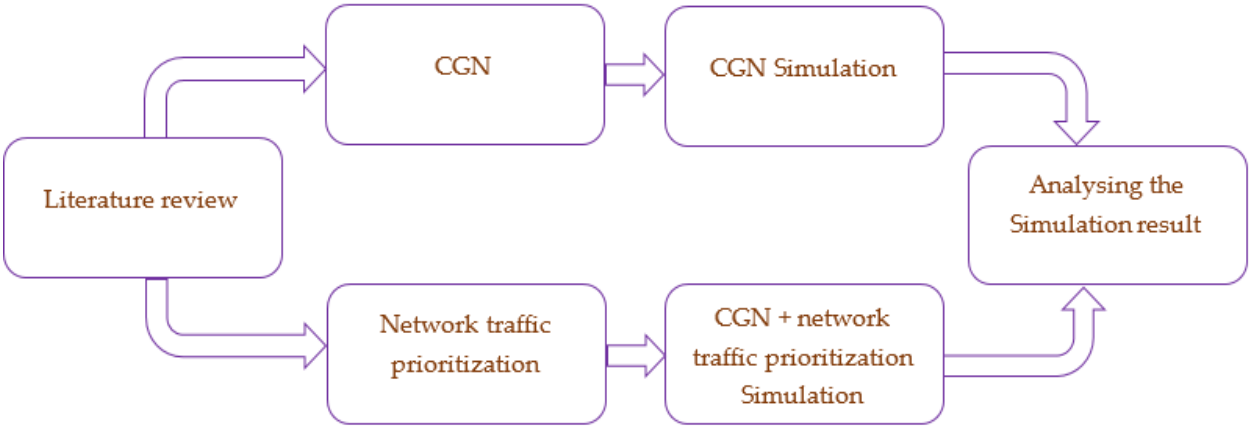


Figure 1.1: Methodology Diagram

Simulations

- There will be two scenarios.
- The simulation will be built using several network protocols.
- A GNS3 emulator will be used to run the two scenarios' topology.

1.7 Contributions

In the core network of service providers, CGN operation necessitates close monitoring. This is due to the fact that the majority of subscribers' Internet traffic must pass through the CGN. Because of its behavior, the CGN operation has the potential to cause a performance issue. Minimizing the impact of CGN operations on network performance is a top priority for service providers. The goal of this thesis is to improve the network performance of CGN operation by integrating CGN operation with traffic prioritization.

This assists the operators in making the best use of their limited resources in order to provide a high-quality service.

1.8 Thesis Layout

This thesis paper is divided into six chapters. The first chapter is devoted to the thesis's introduction. It includes the background, issues to be addressed, thesis aim, literature review, thesis scope, and the techniques this research will use to attain the specified objectives.

The second chapter explains how CGN works as well as various topics related to CGN's effect on network performance. The fundamental ideas behind CGN functioning will be discussed. It also highlighted the advantages and disadvantages of CGN implementation. The third chapter will describe several techniques for traffic categorization and prioritizing. The operation of different traffic prioritization techniques is also discussed.

The fourth chapter examines the various types of network parameters that are used to measure and assess the performance of a network. The major network performance metrics are also addressed, including latency, throughput, packet loss, and CPU and memory usage. The toolkits that were used were also explained. The fifth chapter addresses simulation scenarios, network topology, and simulation scenario outcomes. The procedure used to get the data and analysis of the data for the scenarios based on network performance metrics are also discussed.

Finally, the conclusion of the research is drawn based on the result and analysis of the simulation output. The final chapter also includes future work that may be studied for further work. Reference and appendix will also be present at the last.

2. Introduction to Carrier-Grade NAT

2.1 Overview of CGN

When the Internet was introduced, network devices could communicate seamlessly. Devices that communicate with each other on Internet use unique IPv4 addresses. An IPv4 address is an address created to uniquely identify each network device. This is enough to generate 4.3 billion unique addresses. However, the Internet's popularity and acceptability have skyrocketed, resulting in a scarcity of IPv4 addresses. To address the issue of IPv4 address depletion, a new method was created. Using the network address translation method is one of the solutions. NAT allows a globally routable address to represent a local private address. The router that performs the address translation is located at the customer's site. However, with the NAT method, IPv4 depletion is becoming more prevalent. As a result, ISPs began to use address translation in their core networks to overcome this problem. The translation process used by ISPs is known as CGN. CGN can convert millions of individual customer addresses into a single globally routable IPv4 address, which leads to reducing public IPv4 address depletion issues.

CGN is a collection of mechanisms for exchanging addresses among a large number of Internet users. Previously, each device on the network could have its own unique IPv4 address. This, however, is no longer the case. The primary function of NAT is to translate the broadband consumer's home device's private IPv4 network address to a public IPv4 address. The link between the service provider's access network and the consumer's network is comprised of Wide Area Network (WAN) IP addresses, which are then routed to the Internet through public IP address. The deployment of CGN extends the concept of consumer NAT to the ISP network. CGN enables several broadband customers to share a single globally routable address. The scarcity of IPv4 addresses is handled by sharing a globally routable IPv4 address with several users [12].

RFC 1631 proposed the concept of NAT to solve the problem of IP depletion and route scalability. The goal was to provide a short-term and long-term solution for IPv4 address storage. One of the short-term options was classless inter-domain routing, while the long-term solution was to develop a new IP format capable of generating enough addresses. While such short and long-term remedies were being implemented, the NAT was created to fill the void. Later, RFC 3022 was released with a modified version of RFC 1631, which incorporated the new notion of TCP/IP port translation. It also discusses in depth how NAT works and its limitations.

RFC6269 examines the implications of shared IPv4 addresses on public IPv4 address management. The shared IPv4 address described the process of double NAT, which entails translating two phases of a private IPv4 address. The CGN procedure involves three different IPv4 addresses. First, there is a private IPv4 to WAN IPv4 translation on the client-side, followed by a WAN IPv4 address to public IPv4 address translation on the ISP side [5][12].

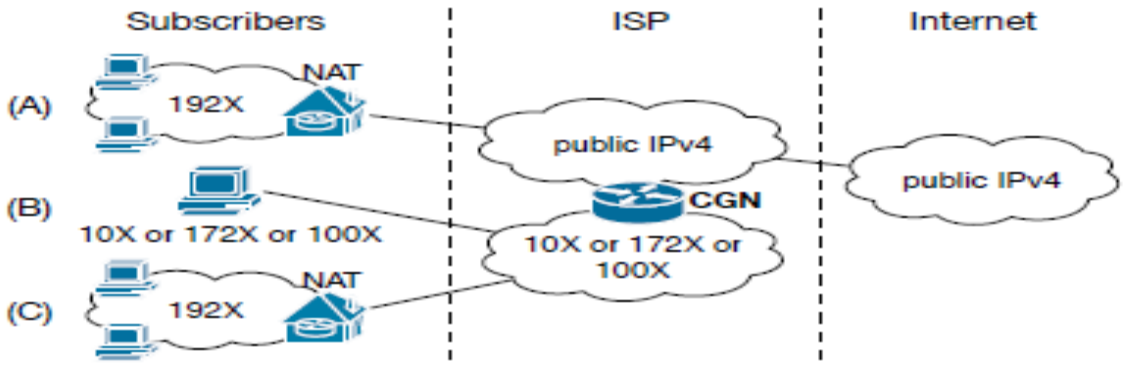


Figure 2.1: Double NAT [5]

2.2 Types of NAT

Address translation may be classified into two categories (i.e. traditional NAT and CGN). Basic address translation is a method of translating IP addresses that translates one set of

addresses to another. Network Address Port Translation (NAPT) is a type of translation mechanism that translates several network addresses to a single address together with their respective transmission protocol port numbers. NAPT and basic address translations are referred to as traditional NAT in RFC3022. The second form of address translation used at ISP core is CGN.

2.2.1 Traditional NAT

Traditional NAT involves translating IPv4 addresses as well as TCP/IP port numbers. Traditional NAT uses two variants of IPv4 addresses to execute distinct IPv4 address translations, such as one-to-one or many-to-one. RFC 1918 divided IP addresses into two groups: private IP address ranges and public IP address ranges. Private IP address ranges cannot be routed to the public Internet. They are used in the Local Area Network (LAN) by end devices for interaction internally. The requirement for IPv4 address translation comes when the internal network needs to contact the Internet. Table 2.1 shows the range of private IPv4 addresses.

Table 2.1: Private IP Address Range [5]

Range	Shorthand	RFC	comments
192.168.0.0/16	192X	1918	Commonly used in LAN
172.16.0.0/12	172X	1918	Commonly used in Enterprise network
10.0.0.0/8	10X	1918	Commonly used in Back bone network
100.64.0.0/10	100X	6598	For CGN deployments

A special network device is required for the internal network to transform or map internal IPv4 addresses to globally routable addresses. To get access to the outside network, the LAN nodes must begin a session. The ISP assigns a single or a set of public addresses to the client in order to map the internal network to the outside world. To identify each private node connection, the NAT router must create a port number [13].

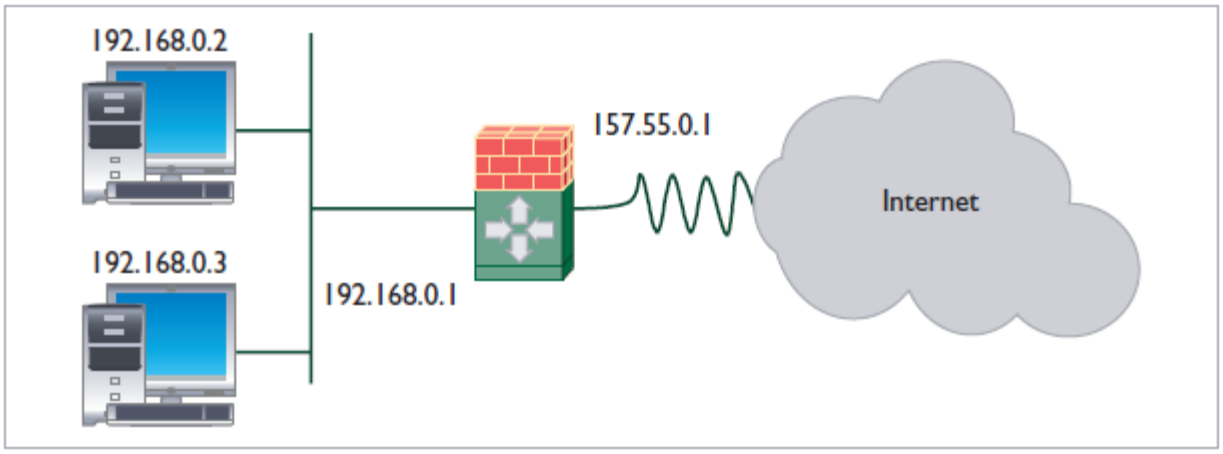


Figure 2.2: A Network Address Translation [13]

In Figure 2.2 the private network (192.168.0.0/24) is provided with one public IPv4 (157.55.0.1). For all the private network nodes to access the Internet they need to share one public IPv4 address. If only the IPv4 address is used the NAT router will face a challenge to identify and send the return traffic to each private host since each host separately is cable of generating different sessions. To solve the problem, the NAT device uses a combination of IPv4 addresses and port numbers. IPv4 address along with port number combination gives each private host a unique mapping or entry.

2.2.2 CGN

CGN is a form of IPv4 addresses and port numbers translation technique done on a huge scale. It is done by ISP's backbone network. It is capable of translating many customers' private IPv4 addresses to a single public IPv4 address for minimizing the storage of IPv4 addresses. CGN implementation at the ISP generates different problems for the ISP as

well as for the customer [14]. A CGNAT router divided the network into two sub-networks, RFC6598 private network, where the customers are assigned with the range private IPv4 addresses, and a public network, where the CGNAT router connect the customers' private network to the Internet with a public IPv4 address [15].

2.3 CGN Options

There are four possible variants of NAT. Static NAT, Static PAT, dynamic NAT, and dynamic PAT [12]. Static NAT is one to one translation between one form of IPv4 to another. It is mostly used to host a server service to be accessed from the outside network. This makes it a bidirectional operation; the internal host can be accessed from outside the network regardless of initiating a translation map by internal hosts since there is a one-to-one mapping between the private IPv4 address and public IPv4 address. It doesn't solve the shortage of IPv4 addresses since every host in the internal network will have a corresponding public IPv4 address, let's say if there are 30 hosts in the internal network those 30 hosts will require 30 public IPv4 addresses to function.

In the case of dynamic Port Address Translation (PAT), there is IPv4 address as well as port number translation. A combination of IPv4 address and port numbers are used for translation operation. This will give the internal host to use undefined port numbers for their service which needs to be accessed from the outside network. This gives extra security for the internal server from the massive scanning of well know port usage. The scanning of the well-known port is used by an attacker to get a weak point so that the service faces a vulnerability [16]. Since CGN operations are similar to NAT with a huge scale, it is equipped will all options but the main option used by ISP's is the NAPT or dynamic PAT operation. Figure 2.3 shows a sample operation of dynamic PAT.

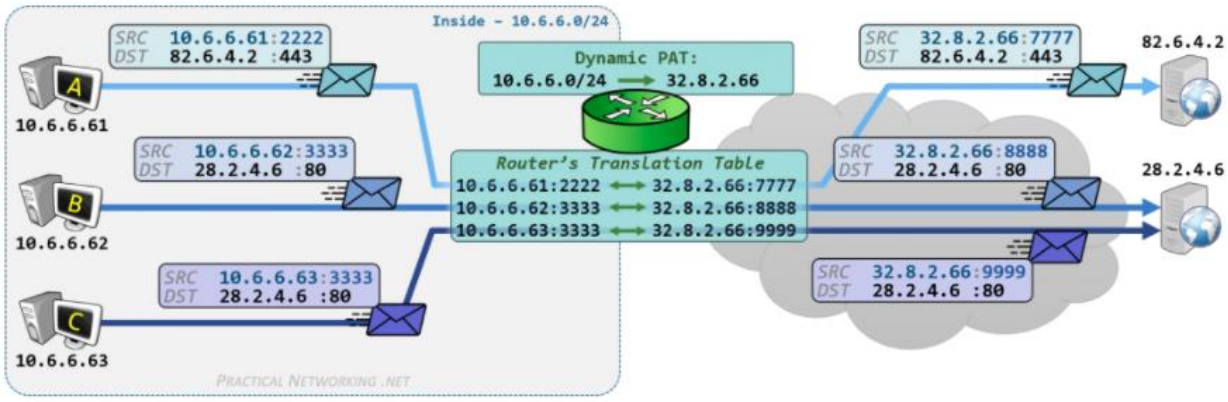


Figure 2.3: Dynamic PAT Outbound Traffic [17]

2.4 CGN Operation

The general operation of a CGNAT-enabled router is a combination of port numbers and public IPv4 addresses. When a CGN router received a packet sent to the external network from the customers, the CGNAT process will look for a matching source address and source port combination in the session mapping table. The session mapping table contains a list of internal local port numbers for conversion to specific mapped ports. If the received source port is not found on the session table, a new port is created. Then the new port is assigned to the received source port. The customer's private source IP address is replaced with the globally routable address in the CGNAT address pool. The source port is replaced with the port in the port mapping table. Finally, the packet is forwarded with the converted source IPv4 address and source port number [16].

A filtering process will be performed for the data packets that are received from the external network against the session mapping table. Then after it will be sent to the internal network after undergoing inverse conversion. The first step is looking at the destination port number in the port mapping table. If the session entry is located, then the destination port number will be replaced with the port number in the mapping table. The destination address will be replaced with the private IPv4 address in the translation

table. If no mapping exists for the destination IPv4 and destination port, the router will decide that the packet does not belong to its internal network and discard it [16].

A sample CGN packet flow for address translation and security features is shown in Figure 2.4. The flow chart shows that connection verification is the first step for the packet traveling from ingress to egress interface. In addition, if the packet does not have an already established connection, it can be dropped if any of the required conditions are not fulfilled before reaching the outgoing interface. The main process that the packet faces before it send out by the egress interface is [18]:

a. First, whether a connection already exists or not is checked, then if it needs a new connection the routing table will be checked if there is a route to the destination that exists or not.
b. Second, the access list will be checked if the network is allowed to be forwarded or not.
c. Once the packet passes the access list condition it is allowed to go through the address translation process.
d. Then it is allowed to leave via the egress interface.

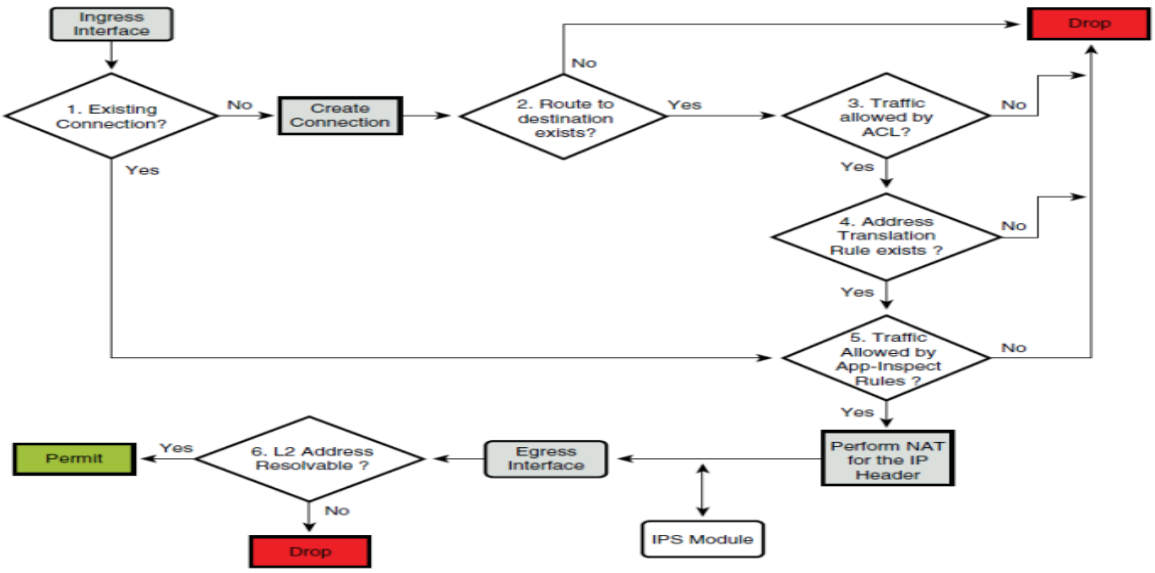


Figure 2.4: Sample Packet Flow on CGN Device [18]

2.5 Protocol Modification

For CGN operation to perform the address translation task, different protocols undergo a modification since IPv4 address and port number are changed from one type to another. RFC3022 states that to reflect the change made on IP address and port number, the protocol associated with them needs to be modified. The protocols which are associated with the address and port number are IP, Transmission Control Protocol (TCP), User Datagram Protocol (UDP), and Internet Control Message Protocol (ICMP) [12]. The modification of each protocol is done packet-by-packet basis. Each packet involved in a translation undergoes modification. IPv4 packets are consist of two parts (i.e. header and data). The process that packets undergo in IPv4 and port modification makes the CGN process resource-intensive. CPU and Memory processes are the main hardware component of the CGN device that is affected by huge address translation. During address translation, the header part of the protocols is modified for each packet that passes through address translation.

2.5.1 IPv4 Address

An IPv4 address is a numerical assignment that uniquely identifies a specific network device on the network. There are two flavors of IP address, IPv4 and Internet Protocol version 6 (IPv6). Addresses in IPv4 are 32-bits long, which allows for a maximum of 4,294,967,296 (2^{32}) unique addresses. IP belongs to a network layer in TCP/IP protocol stack. An IPv4 address provides the specification of an exact format for data, performs a routing function and it uses a set of roles.

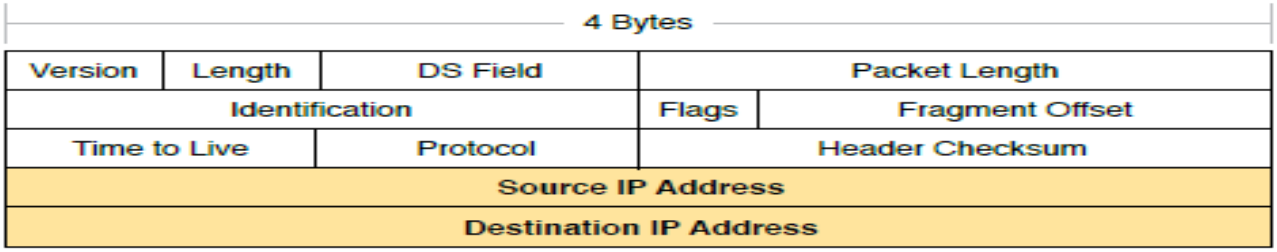


Figure 2.5: IPv4 Header Structure [19]

IPv4 address is a 32-bit numeric address that is divided into four parts. Each part is 8-bit long [19].

IPv4 address has a 20-byte header data which consists of thirteen multipurpose fields [19]. Figure 2.5 shows different parts of the IP header.

- Version: A 4-bits length field that provides the format of the Internet header. This one is version 4 format.
- Length: Internet Header Length (IHL) is 4 bits' field that indicates the length of the header.
- DS Field: Also called Type of Service (ToS) has 8-bits of length that provides network service parameters quality which used for setting different priorities. The left 3 bits are used to set the IP precedence value. There are a total of 7 precedence values for setting different traffic priorities. Those are network control, internetwork control, CRITIC/ECP, flash override, flash, immediate, priority, and routine.
- Packet Length: It is a 16-bits field that provided the length of the datagram.
- Identification: A 16-bits length field is used for identifying the fragmented datagram sent by the sender. The Data is fragmented if its length is higher than the sending rate. Based on the fragment field value the receiver reassembles the datagram at the destination.
- Flags: It is a three-bit length field used for indicating the datagram is fragmented or not during sending the packet at the sender side. It has three values to indicate either the data is fragmented or not. Based on the bit values the receiver will manage the datagram when it arrived at the destination.

- Fragmentation offset: It is a three bits' value field used to indicate a fragment identification via offset value.
- Time to Live (TTL): Has 8-bits value field. Used for loop controlling technique. When the packet traverses a router the value is decremented by 1 and when it reaches zero the packet will be dropped by the last router.
- Protocol: This 8-bit field provided header transport packet information.
- Header checksum: It is a 16-bits field, used for checking and monitoring communication errors.
- Source address: A 32-bits field that holds the source IP address.
- Destination address: A 32-bits field that stores destination IP address.

During translation, the main fields which are modified are the source and destination address but when those fields are changed their values, other related fields are also needed to be modified since they are interrelated which source and destination fields.

2.5.2 TCP, UDP, and ICMP

The transport layer protocol is used for controlling the different sessions using a port number. It has two parts, TCP and UDP. TCP is a reliable transport protocol. UDP is used for real-time traffic transmission and it is not a reliable transport protocol. If the packet is dropped for any reason there is no retransmission, unlike TCP. Both protocols can provide 65536 ports [20]. For NAPT operation port number translation is one part of the operation. As stated by RFC3022 NAT did not allow the outside traffic connection request if the session is started by the external network.

Like in the case of the IP header, the TCP and UDP headers also undergo different modifications to go along with the port number changing procedure. Different

applications use either TCP or UDP port numbers for establishing communication between the sender and receiver [12]. The main NAT behavioral requirements for TCP and UDP are disused in RFC5382 and RFC4787 respectively. The default time-out value for TCP is 24 hours whereas for UDP is 5mintus [20][21]. Session timeout value plays a very vital role in CGN operation by allowing the session to be available for new requests coming from the customers.

ICMP is the third protocol that undergoes modification during the CGN operation. Since it uses an IP address to perform its intended task, when every there is a change in IP address it will be affected too. As indicated by RFC777 ICMP is involved in two address modifications and three checksum medications. It is used to check if there is a reachability problem or not. It uses different message types to indicate the output of the intended task.

Since ICMP is used to report if there is a reachability problem, it uses an error message which undergoes the following modifications:

a. The IP header on the outer layer
b. The ICMP header of the outer layer
c. The header of the packet embedded within the ICMP-error message payload

2.5.3 Checksum Modification

The checksum field is used to check the integrity of the packet during its arrival at the destination. When a sender sends a packet it calculates a checksum value and puts it on the checksum field. At the receiver side, the receiver also calculates the checksum of the arrived packet and compares it with the checksum value in the checksum field. If the two values are the same the packet is not subjected to change during its travel and will be processed but if there is a difference the packet will be discarded by the receiver. RFC3022 states the process of checksum modification during address translation.

The checksum field uses one's complement algorithm to calculate and check the checksum value for each packet. The checksum value will be affected whenever there is a change in IPv4 address [22][16].

RFC1071 defines the steps used for calculating the checksum value.

Step one: Convert each IP header field value into a series of 16-bit binary numbers except the checksum filed

Step two: Calculate the total addition of all 16-bit binary numbers, make a carry bit to wrap around.

Step three: Use the 1's complement of the last addition.

2.6 Address and Port Mapping

The combinations address translation and port translating create a special table called the session table. The session table plays a vital role in address translation and network performance. For the customers to access the Internet the session should be initiated by themselves. The filtering process for the return traffic is entirely dependent on the existence of the session table [23].

The sessions are distinctively identified by a combination of five fields.

- Source address: The address of the customer which generates the traffic to access a piece of information from the Internet.
- Destination addresses: The address for providing the requested information from the Internet.
- Transport layer protocol: Either TCP or UDP, used based on the application type. TCP is used for reliable transmission whereas UDP is used for real-time traffic.

- Source port number: Randomly generated TCP/UDP port number used for tracking each application service the customer initiated.
- Destination port number: The port number that is used for visiting applications on the Internet.

Table 2.2: Session Table General Architecture [16].

Source-address	Source-port	Destination-address	Destination-port	IP-protocol	State	Time

There are different conditions that a subscriber node will use multiple sessions for one application. The main reasons for these conditions are [12].

- Application performance: Many sessions functioning along with in parallel can communicate with multiple servers can improve network performance.
- Application functionality: Different sections of the application use different sessions for different purposes.
- Command and control: In some cases, the sessions can be used as a controlling mechanism and data transmission mechanism.

The number of ports used by TCP and UDP has a limit. The number of ports that are generated by transmission protocol is 2^{16} or 65,536 ports. All ports are not available for random use as some are assigned for a specific purpose. These specific ports are called well-known ports. The port number pool available for a CGN operation is adequate for most applications. In some conditions, CGN needs to set a maximum number of sessions a subscriber may allow to use. Based on the number of subscribers that are sharing a single public IP address on the CGN node, this may lead to the subscriber being limited to tens, hundreds, or thousands of addresses [12][5].

The two limiting factors that affect the performance of a CGNAT node are the rate at which address and port mapping are created and canceled from the session table, and the size of the session table created. A CGNAT node generates a new session for an individual connection initiated by a subscriber. The highest session generation rate and the dimensions of the session table are limited resources. The main worry for any service provider CGN implementation is efficiently optimized CGNAT devices. Figuring out the number of subscriber's connection limit is a key for managing the limited port number resource that is generated by the CGNAT device without noticing a major degradation in network performance [8].

2.7 Applications Affected by CGN

Applications like basic web browsing work fine behind CGN operation. But it is crucial to recognize that any application has a probability of failure under CGN operation. For instance, if a customer's allocated sessions in the CGN device reached its allocated session connection, then applications will not have a new connection for handling the new request. Sometimes the applications look like working fine even if there is a performance issue. As a result, until the application stops operating, it is difficult to observe the influence of CGN operation on application performance. Internet service providers would want to eliminate service problems but as they are not able to control how their customers use the Internet, they could not able to provide a sound service behind CGN [12].

Some applications which have a port or IP address information on their payloads like file transfer protocol, IPsec, and IPv6 tunneling will fail to work in the CGN environment. For those applications to work behind a CGN operation they need the help of transversal techniques. The investigation carried out by some researchers [7][9] shows that some applications will face performance degradation under CGN. Applications like online gaming, peer-to-peer connection, real-time application, remote access service, AJAX

applications (i.e. Google Maps, torrent), and video on demand show a performance issue whereas basic Internet browsing like checking email works without a problem[7][9].

2.8 Transversal Techniques

Since NAT44 operation breaks the end-to-end Internet connection it created a problem when an outside application wants to communicate the private network. Without an established session the outside application is not able to reach nodes inside a private network. For instance, a remote desktop application cannot function from outside within a NAT deployment. For some applications to work within NAT operation, they require some transversal technique such as port forwarding.

There are several kinds of transversal techniques available for helping applications to work behind a NAT operation [12][24].

- Application Layer Gateway (ALG): ALG is applicable when the application layer has an IPv4 address and port number information. During address translating the NAT device will not check the payload of the application layer due to that when the address is translated the application will fail. So to solve this issue ALG is integrated with NAT to change the address in the payload as well. Applications like File Transfer Protocol (FTP), H.323, Session Initiation Protocol (SIP), and Domain Name System (DNS) need the help of ALG since they carry IPv4 address and port number information.
- Port Control Protocol (PCP): PCP is a protocol used for controlling IPv4 address and port number mapping between the external network and internal network work.
- Interactive Connectivity Establishment (ICE): is a protocol developed to allow UDP-based multimedia sessions to traverse NAT.

The transversal techniques developed for assisting the application to work with NAT works with NAT44 setting. But when it comes to CGN those transversal techniques will not work because ISP's does not enable those protocols within their CGN device. So, even if the NAT44 device manages to function with ALG when it comes to CGN deployment, the device's administration is out of the customer's grasp which makes the traversal technique used by the subscriber to fail.

2.9 Tractability

Tractability is one of the major issues introduce by CGN operation along with different application and network performance issues. When a globally routable address is used by the end customer it can easily be located and traced. Geolocation-based applications and services can provide their intended service with ease. Even the case of finding an abuser who disrupts the normal Internet communicating process cans be allocated by tracing its IPv4 address in the case of public IPv4 usage [12] [24].

For the lawful interception, a legal process for locating the source of Internet traffic is a necessary duty for ISPs. Identifying the source of network abuse is one key responsibility of the law enforcement bodies as well as the service providers. Logging is a technic that the service provider uses to collect every activity involved in their network and the operation is called Syslog. To keep the security of the network domain, monitoring the different parts of network devices and systems in the network domain is a necessary approach. The log files are used to analyze network vulnerabilities, and they may also be utilized to discover the attacker's network address. Logging also has a major use if the network happed to be crashed. The restoring could be done using the information gathered in the logging process [25].

Keeping a translation log between the source and destination IP addresses, as well as the port translation and timestamp, is a memory-intensive activity for CGN operation. Which

required a lot of memory space which is measured in a unit of Terabytes. Even after recording a huge amount of translation data, tracing and reading the data during an attack takes time and also impacts the performance of the CGN device. Even the working principle of CGN operation which shares a single IPv4 address to millions of customers has an impact to specifically identify the abuser. When a single public IPv4 address is blocked during a specious attack all the customer traffic involved in a translation with that public IPv4 address will also be impacted. This makes traceability a huge factor with CGN deployment [12].

2.10 CGN and IPv6

In general, CGN helps in mitigating the IPv4 depletion problem by allowing a single IPv4 address to be shared by millions of private networks. In doing that much translation, it introduced new problems by itself. It breaks the fundamental end-to-end working principle of Internet communication by translating the source IPv4 to other IPv4 addresses. For performing the translation process, it required high resources from the network device which makes it a resource-intensive process. It also affected some application performance. The CGN solution was created for a limited period while the considerably larger and more complex new address formant was being implemented. The shortage of IPv4 address problems will be solved when IPv6 implementation is successfully deployed. But the implementation of IPv6 is not progressing as expecting due to several issues [12][26].

RFC 6343 and RFC 6732 discuss the different factors CGN deployment has on IPv6 transition. The implementation of IPv6 shouldn't have been impacted by CGN operation if a dual-stack operation is deployed by the service provider. But due to the huge investment needed for implementing IPv6, most ISPs prefer working with the CGN operation alone [27]. The other transition techniques for translating IPv6 to IPv4 and vice versa will be impacted when they are operating under CGN. To avoid the impact of the

native implementation of IPv6, it is recommended to deploy it in parallel with the CGN operation. This operation has also an advantage for the customer by providing an Internet connection without NAT.

In a domain without CGNs, a customer probably uses IPv6 transition techniques like 6to4, Teredo, and Tunnel Brokers to interact with the global IPv6 Internet. During CGNs deployments, techniques 6to4, Teredo and Tunnel Brokers are not able to function when they encounter a CGN device. This indicates that the only possible methods by which customers could allow access to the IPv6 Internet are prevented from happening by CGN deployment [5][9].

3. CGN and Network Traffic Prioritization

3.1 Introduction

The goal of network traffic classification based on priority is to efficiently deliver service quality to the user. The traffic prioritization mechanism on devices seeks to optimize the routing strategy of packet forwarding across devices. Traffic prioritization refers to the ability of network components (such as a switch or router) to ensure that their service flows and service demands are met to some extent. Traffic prioritization controls and directs traffic depending on values specified based on application needs and network status. The impact of IP traffic prioritizing on performance characteristics such as packet loss, latency, throughput, availability, and jitter can be quantified. Network equipment (such as routers and switches) segregates traffic by evaluating packets as they are processed and classify them into groups of various streams depending on the class of service (CoS) traffic rule [28].

Models of traffic categorization instruct the device on how to handle traffic as it travels from the network node's input interface to the output interface. Classifying traffic will be used to differentiate one CoS from another. This may be configured to achieve the appropriate traffic classification solution in any situation. Depending on the priority settings and how the node is configured, a network device can apply various priorities to traffic using CoS values. The required behavior of the network or application that creates the traffic is required to meet a certain requirement [28]:

- Prioritizing traffic over other traffic based on protocol type, source address, destination address, source port number, or destination port number.
- Filtering traffic at input-interface or out-interface based on a certain policy.
- Controlling the bandwidth according to the traffic priority given for a certain network as it is transmitted or received on the interfaces of the device.

- Reading and writing CoS values in the packet header.
- Controlling congestion by treating high-priority traffic first.

3.2 Models of Traffic Classification

Two models are used to categorize network traffic based on the Internet Engineering Task Force (IETF) definition. The Integrated Services (IntServ) and Differentiated Services (DiffServ) models are well-known for categorizing traffic depending on a specific requirement to achieve improved latency, packet loss, and throughput. IntServ uses the per-flow technique to guarantee individual traffic, whereas DiffServ provides collective assurance for a class of networks [29].

- IntServ: It demands an explicit resource reservation per flow from the network device. Network nodes track all flows that pass through them by determining whether incoming data packets correspond to an existing flow or not. If there are adequate network resources available, the packet will be received; otherwise, the packet will be discarded.
- DiffServ: A class-based classification method is used by network nodes to separate traffic into CoS depending on the priority. DiffServ offers three possibilities for various services. Expedited Forwarding (EF), Assured Forwarding (AF), and Best-effort (BE) forwarding are DiffServ models that allocate different priorities for generated traffic based on the administrator's requirements.
- IPv4 address assigned a three-bit ToS byte in its header for classifying or prioritize traffic, as shown in Figure 3.1. RFC 791 allocates a 3-bit IP Precedence (IPP) field for traffic priority marking. IPP provides us eight separate binary values starting from 000, 001, 010, up to 111 with their equivalent decimal values 0 through 7. While IPP provided us with eight different values to label, the later RFC2474 redefined the ToS byte with the Differentiated Services Code Point (DSCP) field. DSCP provides us with an increased number of labeling bits to 6 bits, that produce

64 unique values that can be labeled for different types of traffic as well as priority allocation[19].

The major activities are performed within the network node in order to achieve the required traffic categorization based on the quality of the request in order to establish distinct criteria. Schedulers, classification, queuing, policing, and sharpening are all important functions for providing different classes. The network is set up to provide CoS for those high-priority services.

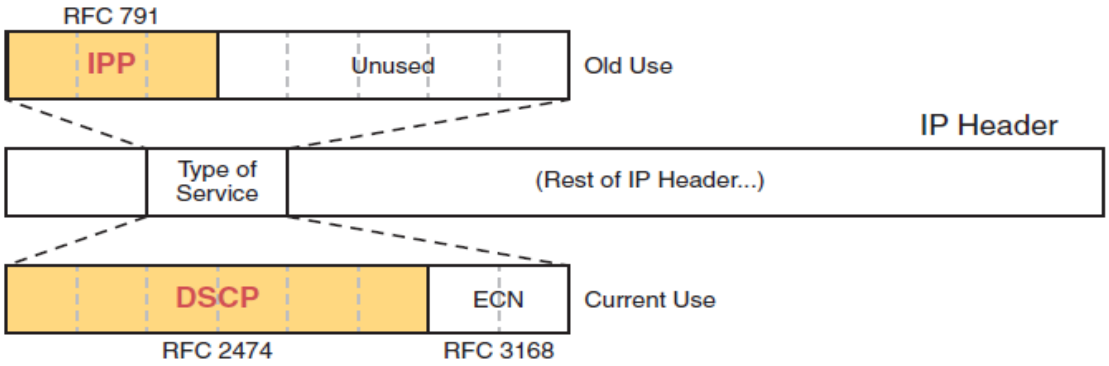


Figure 3.1: IP Precedence and DSCP Fields [19]

3.2.1 Classifiers

To accomplish a traffic classification by assigning it a different priority, the various data must first be categorized. Classified traffic will be processed differently than ordinary traffic based on the requirements. Traffic is classified based on the application, port number, or IPv4 address. Figure 3.2 depicts the traffic categorization procedure. Once the traffic has been categorized, it is transmitted for additional processing along the network's packet path. It also required to be labeled so that each network device knew what to do with the labels. Marking might take place on either layer two or layer three. Layer two markings can be applied to the ethernet header, whereas layer three markings can be applied to the IP header ToS field [30].

QoS differentiates and classifies each traffic type that enters a network node, and then

organizes that distinct traffic into classes. Traffic belonging to each class is processed based on a set of quality of service behaviors allocated to that class. Differentiating traffic and allocating it to the actual class is the duty of the differentiator. Classifiers receive incoming packets as separate inputs but provide a variety of outputs. Classifiers determine to which class a packet belongs by using a set of access lists and match rules. The IP prefix list puts the requirement whereas the match conditions determine the class to which the packet should be categorized. When the access list conditions are met, the packet is classified with the appropriate class [28][23].

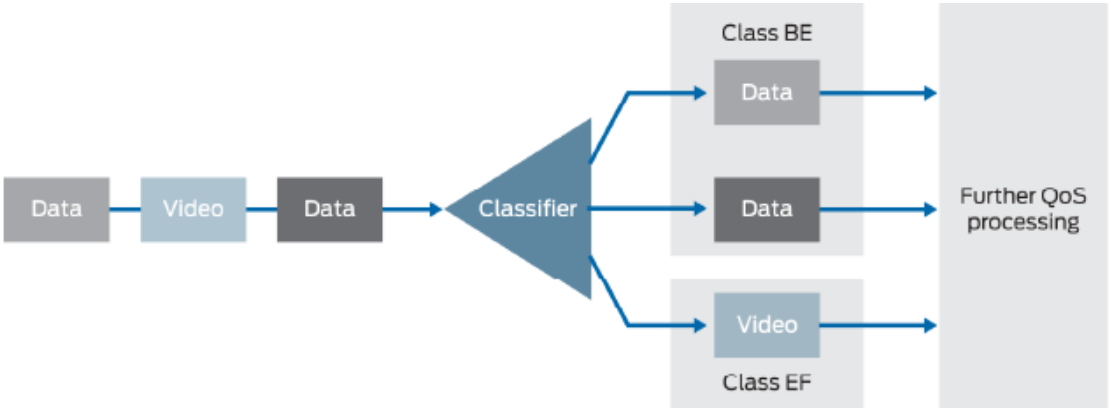


Figure 3.2: Traffic Classification Process [28].

3.2.2 Policers

Network traffic has a bursting behavior during data packet transmission. For managing the burst characteristics of network traffic, some mechanisms are developed for insuring steady traffic. One of the mechanisms used for controlling the burst behavior is applying policers.

Policers manage traffic bursts by making an assurance that forward and reverse traffic is subjected to a customized rate called the bandwidth limit. Policers give initial congestion controlling by stopping incoming traffic from heavily burden the network.

Policers manage the busty traffic by implementing hard policing or soft policing. The hard policing working principle is to discard the packet when it exceeded the set

bandwidth limit, whereas in the case of soft policing rather than discarding the exceeded packet it reassigned it to other types of congestion control mechanisms [28].

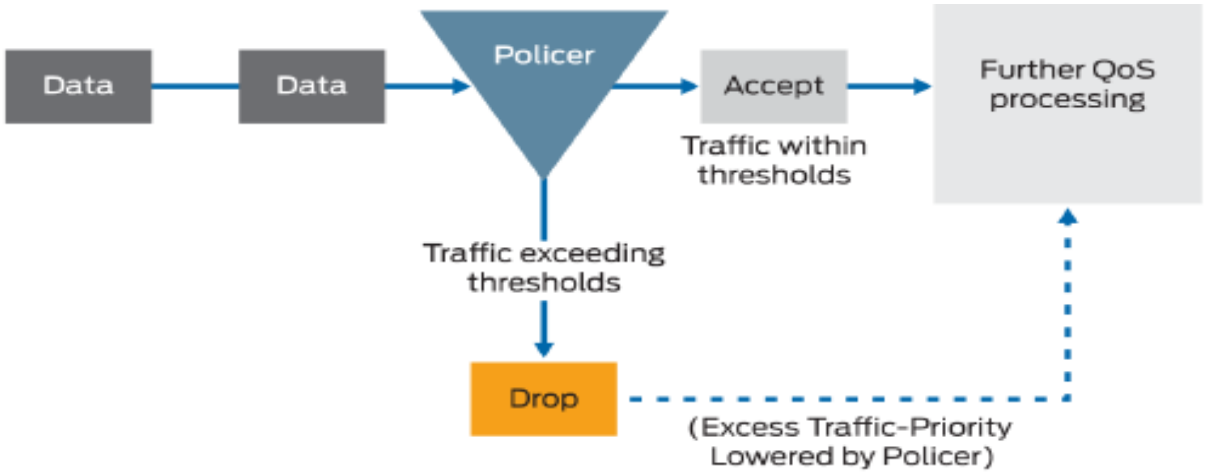


Figure 3.3: Traffic Policing Process [28]

3.2.3 Queuing

A queue is a mechanism used to keep a packet until it can be processed or wait for transmission resources. Routers and switches can provide queue service using their inbound and outbound queue. An inbound queue keeps packets until the router or switch CPU can forward the traffic to the intended interface. An outbound queue keeps packets until the packet is put on the actual physical line by the router or switch. The queues service can be provided in two forms, either in hardware form or software form by router and switch ports. A packet that waits in a queue can be processed and sent to the actual interface or discarded. Packets are discarded based on the defined fill level set on the network devices [28].

Router and switch queues are vulnerable to congestion. Congestion happens during the rate of inbound traffic arrival is greater than that of the outbound interface.

The congestion in a network device can occur due to [31]:

- The speed of an inbound interface is higher than the outbound interface.

- The aggregate traffic of different inbound interfaces transcends the amount of a single outbound interface.
- The CPU of a network device is inadequate to hold the amount of the forwarding table.

3.2.4 Shapers

Traffic shapers manage congestion by applying a threshold to the speed at which the network node can transmit traffic onto the actual wire. Shapers act on outbound traffic. Shapers apply their policy only on traffic that has already been given access to a queue. Shapers constrain outbound traffic to follow to a bandwidth speed called the shaping rate. Nevertheless, shapers do not simply drop traffic above the set shaping rate; they keep onto the data until it can be processed without crossing the threshold speed. The ability to hold a packet while it exceeds traffic rate is the key deference to that of policers.

The service provider traffic service level agreement specifies the maximum rate at which the customer may transmit packets, and any traffic that exceeds this rate is rejected by the service provider as it travels through the backbone network device. To alleviate the traffic being discarded, the subscriber can set a shaper to frame the traffic to the allowed rate which is provided by the ISP. The shaper can keep any traffic above the defined rate until it can be processed without exceeding the rate [28].

3.2.5 Schedulers

The order in which queues are managed is performed by schedulers. Schedulers control how packets are forwarded out to the next network device in the traffic path. Schedulers are also capable of providing another form of prioritizing traffic by determining which packets are sent, in what order the packet is sent in, and how the packets are processed. Figure 3.4 shows a sample process of how schedulers perform.

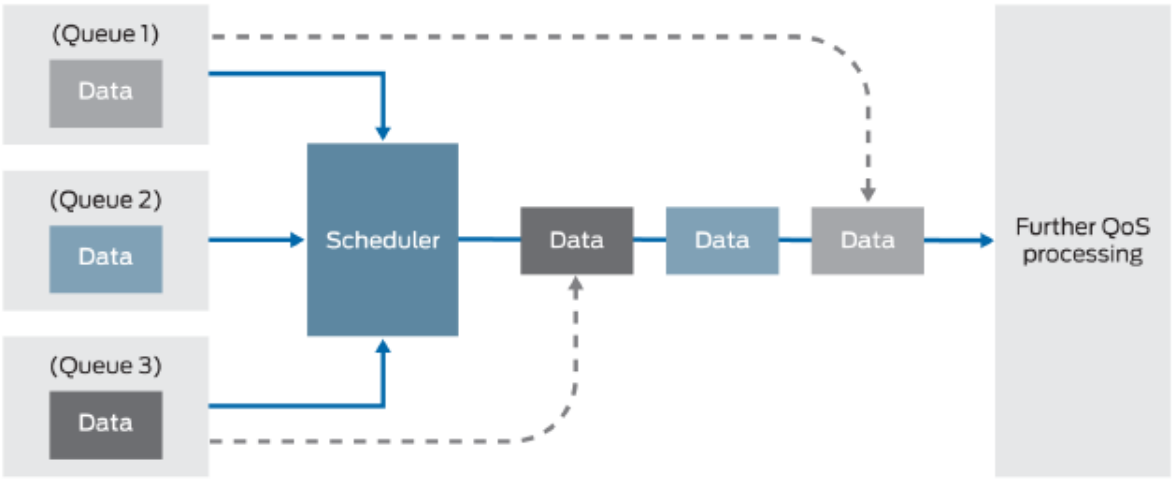


Figure 3.4: Sample Schedulers Process [28]

3.3 Traffic Class Types

DiffServ is a quality of service model that targets providing traffic classes for the IP traffic flow. There are four kinds of traffic classes in the DiffServ model. All four kinds of traffic classes provide different ways of prioritizing traffic. Since the DiffServ model uses the 6 bits in the ToS field, it provides a total of 64 different traffic classes. The 64 traffic classes are categorized under best-effort forwarding, expedite forwarding, assured forwarding, and class selector [32].

3.3.1 Best Effort Forwarding

Best effort forwarding is the default set up in the DiffServ model. All traffic has equal priority in this setting. Which affects the delay-sensitive application when there is a congestion or bandwidth limitation occurs. This implies that sensitive traffic will not be treated differently than other traffic. It applies the first come first served working principle for all traffic by setting all the DSCP values to zero. It is also easy to set up, and that makes it the default implementation in packet-based communication [29].

3.3.2 Expedite Forwarding

The expedite forwarding provides a low latency service. It is also capable of minimizing jitter and packet loss. The bandwidth allotted to expedite forwarding must be restricted

in order to avoid causing harm to other traffic classes. The queue that is assigned to expedite forwarding must be the highest priority queue for the traffic to be processed with the least delay. The successful design of expedite forwarding can give sensitive traffic the desired priority. The vital tasks of expedite forwarding are [28]:

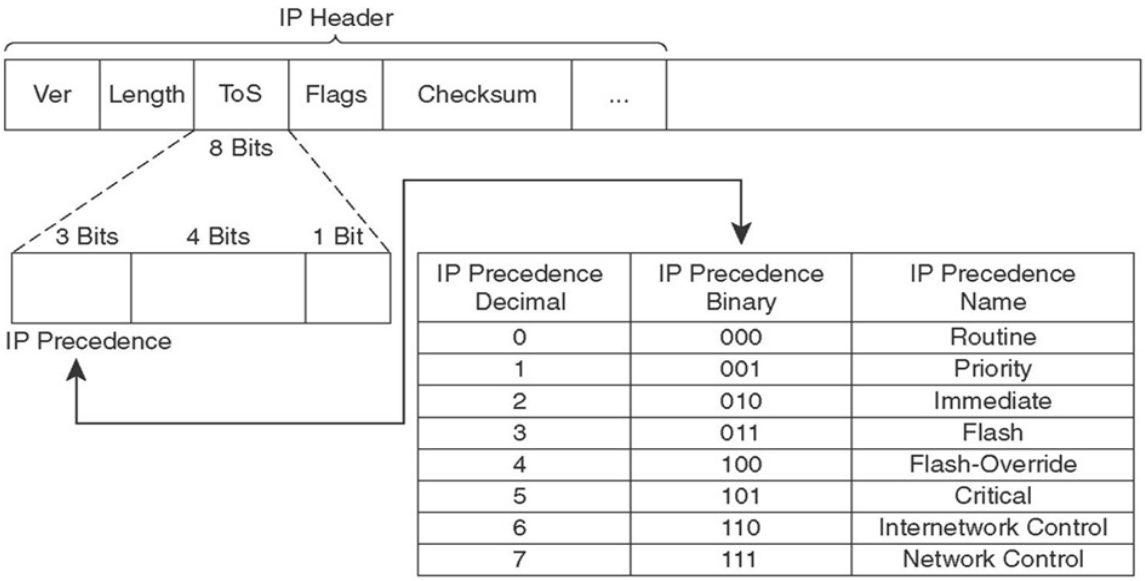
- Providing minimum delay for packets.
- Allocating the required bandwidth for the traffic set with expedite forwarding value.
- During congestion, EF policies will govern the traffic priority.

3.3.3 Assured Forwarding

The assured forwarding provides four queues for the various traffic classes. It provided a total of four classes with three drop probabilities for each class. The combination of class and drop probability gives a total of 12 values which are represented as $AF_{xy.}$ The x represents the four classes whereas y represents the drop probability of the traffic. The traffic drop is done when the traffic passes the reserved bandwidth. The drop probability has low, medium, and high categories. During traffic classification, the different values will be set according to the requested priority [29].

3.3.4 Class Selector

The class selector is one part of DiffServ that gives backward compatibility with ToS values that formulate IPv4 address precedence. IPv4 address precedence uses the three most significant bits of the ToS field on the IPv4 address header. The three IPv4 address precedence bits give eight different values. The more prioritized packet is given the highest IPv4 precedence value which gives it the ability to be forwarded first. Figure 3.6 shows the different values of the ToS field.



IP Precedence Decimal	IP Precedence Binary	IP Precedence Name
0	000	Routine
1	001	Priority
2	010	Immediate
3	011	Flash
4	100	Flash-Override
5	101	Critical
6	110	Internetwork Control
7	111	Network Control

Figure 3.5: IPv4 Precedence Values Assignment [33]

The priority of traffic can be set by enabling the IPv4 precedence. The IPv4 precedence values of 6 and 7 have a special purpose that is not used for classifying traffic. They are used for Internetwork control and network control. The rest of the six IPv4 precedence values are available for classifying different traffic classes based on the required priority set by the administrator. The least significant three bits in the DSCP field are set to 000 to relate the IPv4 precedence values [33]. Based on the IP precedence decimal value, the higher decimal value takes precedence over all those below.

Different Virtual Routing and Forwarding (VRF) are utilized in CGN operation to relate the various traffic priorities. The edge router prioritizes traffic using IP precedence techniques. Critical, flash-override, and flash IP precedence values are employed to prioritize the various classes. VRF functions as a virtual router that can segment different networks within a router. The different routing instance created by the VRF gives it the ability to improve network functionality.

4. Methodology

4.1 Introduction

For any network setup, data transmission is the main goal. Exchange of data in text, audio, or video format will have a source and destinations. This data format(text/audio) has a small unit known as packets. To investigate how the network performance of a certain network is improved and measured the provided output, different metrics are required to be implemented. The network performance metrics are used to indicate the performance of certain network scenarios. For instance, the end-to-end time taken for a packet to arrive at its destination is measured using latency. The most used key performance indicators for investigating the network performance are shown in table 4.1.

Table 4.1: Network Performance Metrics

Network Performance Metrix	Description	Measurement
Latency	It measures the time takes for a packet to reach its intended destination.	In Milliseconds
Throughput	Its measures the observed transfer rate, as related to the maximum theoretical transfer rate.	In bits per second
Jitter	It refers to variations in the latency of a connection.	In Milliseconds
Packet loss	This occurs when one or more data packets transmitted across an IP network fail to reach their destination.	Percentage of packets lost with respect to packets sent

CUP usage	used to indicate the IP packet forwarding delays and application processing delays.	Average CPU usage
Memory Usage	keeps the instructions and data needed to be executed by the CPU.	In Bytes

4.2 Latency

According to RFC 7679, latency is a measure of one-way delay experienced by data packets as the difference in the time from an ingress point to an egress point across an ISP's network. The network congestion and network device involved in the IP packet flow in a traffic path can contribute to the delay experienced by the data packet. In packet-based networks, one-way end-to-end delay is one of the key performance parameters which indicates the quality and best design practice of the ISP network. Today's Internet forwarding and reverse traffic takes a different path during packet transmission, and due to that, the round trip delay will be involved with different contributors like TCP behavior, congestion behavior; and the policy and routing setup will result in different performance. So, it is preferable to measure the one-way delay from the source to the destination due to the packet experience different performance characteristics in symmetric and asymmetric traffic flows.

In TCP/IP networks, there are four sources of delay (i.e. processing delay, queuing delay, propagation delay, and transmission delay) [3]. The International Telecommunication Union (ITU) standard for a one-way end-to-end network packet traffic delay is 150ms.

Delay can be calculated by using equation 4.1 and the standard values of latency are shown in table 4.2.

$$Total\ Delay\ = ND + QD + PD + TD[second] \ldots\ldots\ldots\ldots\ldots 4.1$$

Table 4.2: ITU-T G.114 Delay Quality Standards [7]

Delay Level	Category	Good	Medium	Poor
	Delay (msec)	0-150	150-400	>400

4.3 Jitter

Jitter which is measured in milliseconds is the variation in time delay when a data packet is transmitted and received in a packet-switched network connection. RFC 3393 defined the one-way-delay variation measure for IP packets over the network, recommending measuring one-way travel since forward and reverse-path behavior varies. The difference in time delay is known as IP packet delay variation. Undesirable glitches, like flickering monitors and untimely data transmissions between network devices, are some effects of jitter. The cause of jitter can be network device performance, network congestion, less bandwidth, and no data packet prioritization. Real-time traffic like voice-over IP is the most affected service by jitter. There are three kinds of jitter: constant jitter, transient jitter, and short-term jitter.

Constant jitter has a character of showing a constant delay variation between packet to packet transmission. Transient jitter is characterized by considerable gradual delay which is caused by a single packet. Short-term delay variation exhibits an incremental delay for a certain number of packets, followed by a greater packet to packet delay variation.

If two consecutive data packets leave the source network device with timestamps t1 and t2; and are arrived at the destination network device at times t3 and t4, then jitter can be calculated as (t4 - t3) - (t2 - t1) [34]. Equation 4.2 shows the general formula for calculating jitter.

$$Jitter = \frac{\sum variation\ delay}{\sum packet\ received} \text{[second]} \ldots\ldots\ldots\ldots\ldots 4.2$$

Table 4.3 shows the quality standards for jitter based on ITU-defined standards.

Table 4.3: ITU-T G.114 Jitter Quality Standards [3]

Jitter Level	Category	Good	Medium	Poor
	Delay (msec)	0-20	20-50	>50

4.4 Packet Loss

When some packets of data are unable to reach the intended destination in a data network is referred to as packet loss. Transport layer protocols play a different role during packet loss. For the case of TCP traffic, when there is a packet loss there will be retransmission, on the other hand, in the UDP case, there will be no retransmission even if there is a packet loss. That is the main reason real-time traffic such as voice over IP, live streaming, and video calls are the most affected services.

Some of the main causes for packet loss are network congestion, network device performance, bugs on network device software, network device failure, highly operating network, and inefficient network design [30].

- **Network congestion -** occurs when there are excessive queues in the data network.
- **Network device performance -** Since network devices have a limiting resource when comes to memory and CPU usage, the packet will be dropped when there is high traffic flow (i.e. peak hour) if the device capacity reached its threshold.

- **Network devise failure** – The network device may fail due to several reasons, and when that happens, the data packets will be dropped until an alternative path is established or the device is recovered.
- **Software bugs on network elements** – operating software's are the key elements in network devices along with the hardware, sometimes in the software bugs cause the device to behave in an undesired manner which results in packet loss.

Packet loss percentage can be calculated as shown in equation 4.3 and the quality standard for packet loss is stated in table 4.4 based on ITU standard.

$$packet\ loss = \frac{packets\ sent - oackets\ received}{packets\ sent} * 100\% \quad \text{.....4.3}$$

Table 4.4: ITU-T G.114 Jitter Quality Standards [3]

Packet loss Level	Category	Excellent	Good	Medium	Poor
	Delay (msec)	0%	3%	15%	25%

4.5 Throughput

Throughput in a data network is the actual data transfer rate from sender to receiver measured at a particular timeframe. Speed in a data network is one of the key performance indicators which tells us how fast is the data transferred from source to destination. Throughput and bandwidth are the parameters give as the information regarding network transfer speed for transmitting a certain file from source to destination. Throughput is measured in bit per second. Some factors which affect the network throughput performance are network nodes, network design, and the number of users [35].

The formula for calculating throughput value and the quality of standard set by ITU is shown in equation 4.4 and table 4.5 respectively.

$$Throughput = \frac{\sum data\ sent(bit)}{data\ delivery\ time\ (s)} [bps] \text{ .....4.4}$$

Table 4.5: ITU-T G.114 Jitter Quality Standards [3]

Throughput Level	Category	Excellent	Good	Medium	Poor
	Delay (msec)	100%	75%	50%	<25%

4.6 CPU and Memory Usage

For a data packet network, how fast is the packet forwarding in a certain network device is detected by the CPU performance of the device. With high CPU capability, the network device handles the packet forwarding task efficiently. Depending on the type of protocol or the load on the device, CPU processing time may vary. The CPU resources allocation also affects the performance of a data network. The network interface processing speed in exchanging the input traffic to the output traffic is a vital performance indicator that is determined by the CPU processing power [36].

The proper usage of CPU resources results in a good performance. Operating system instructions are executed by the CPU. Some functions of CPU are system initialization, CGNAT table lookup, routing functions, and switching functions.

Memory is the most essential component of a network device system. Network device memory determines the amount of data stored or forwarded on a device. Network node memory can shatter the performance of a network by behaving as a bottleneck. There are two types of memory, random access memory and read-only memory [37] [38].

Random-access memory:

- It is the main memory or the primary memory used to read and write data.
- It is used to store the programs and data that the CPU requires during the execution of an instruction.

- It is a volatile memory since the data will be lost when the power is switched off.
- It is used to store the operating system, running configuration file, IP routing table, and packet buffer.

Read-only memory:

- Keeps the main information essential to operate the network device system, like the program essential to boot the network device.
- It is not volatile and always keeps its data.
- It is permanent storage that keeps the startup configuration file of a network device.

CGNAT involves reading and writing to the header and payload data of every IPv4 packet to do the address and port translation which makes it a memory-intensive process. Monitoring memory usage is a very important process for providing the highest performance for network devices. Maximum memory usage results in decreased performance for the particular process in a network device which also ends up affecting the entire system [36] [37].

4.7 Overview of Relevant Toolkits

As a major study approach, simulation implementation will be employed. To demonstrate the stated hypothesis, GNS3, as well as a traffic generator and network monitoring tools, will be utilized.

It is extremely tough and challenging to do network research utilizing using real-time settings. Setting up a testbed also took a substantial amount of time and money. A network simulator is highly useful in simulating real-time scenarios, as well as saving money and time. Network simulations are increasingly being used to undertake various network academic investigations and verification methodologies for various network

assumptions. Nowadays, network simulators with diverse functionalities are available, such as GNS3, Enterprise Network Simulation Platform, Emulated Virtual Environment-Next Generation, Optimized Network Engineering Tools, and Network Simulator Version 2 and Version 3, etc. [39][40].

4.7.1 GNS3

There are several network simulation tools accessible for undertaking various academic network investigations. GNS3 is a popular simulator for testing various network scenario hypotheses. It is open-source software with a large community of users and an easy-to-use graphical user interface. It also has the ability to simulate multiple vendor network devices, giving it the flexibility to implement and test various network topologies. Because of its ability to mimic backbone network architecture, it may also be used for testing with advanced network setup. The researchers may experiment on a genuine software image thanks to the integration of GNS3 with a virtual machine environment and the ability to connect to an actual network. The GNS3 platform supports routers, switches, end devices, hubs, clouds, packet capture, and other network technologies for implementing diverse network topologies [41] [42].

4.7.2 Ostinato

When analyzing the performance of a network in a simulated environment, a network traffic generator is necessary to verify the presented hypothesis. Traffic generator tools generate the network traffic required to perform a certain task. Many network generator programs are available for producing network traffic, including WAN Killer Network Traffic Generator, Ostinato, Packet Sender, TRex, Packet Generator Tool, and others. The capabilities of different traffic generating tools differ, as does their compatibility with different operating systems [43][44].

Ostinato is a free and open-source network traffic generator with a straightforward

graphical user interface. It can give a new type of packet stream. Ostinato also supports the majority of protocols. One of the reasons for ostinato's desire to be chosen is its ability to construct and analyze packet streams. Virtual local area network QinQ, address resolution protocol, IPv4, IPv6, IP tunneling, TCP, UPP, ICMPv4 and v6, Ethernet, hypertext transfer protocol, session initiation protocol, and other protocols are supported by ostinato [45]. The different protocols that ostinato supports are depicted in Figure 4.1.

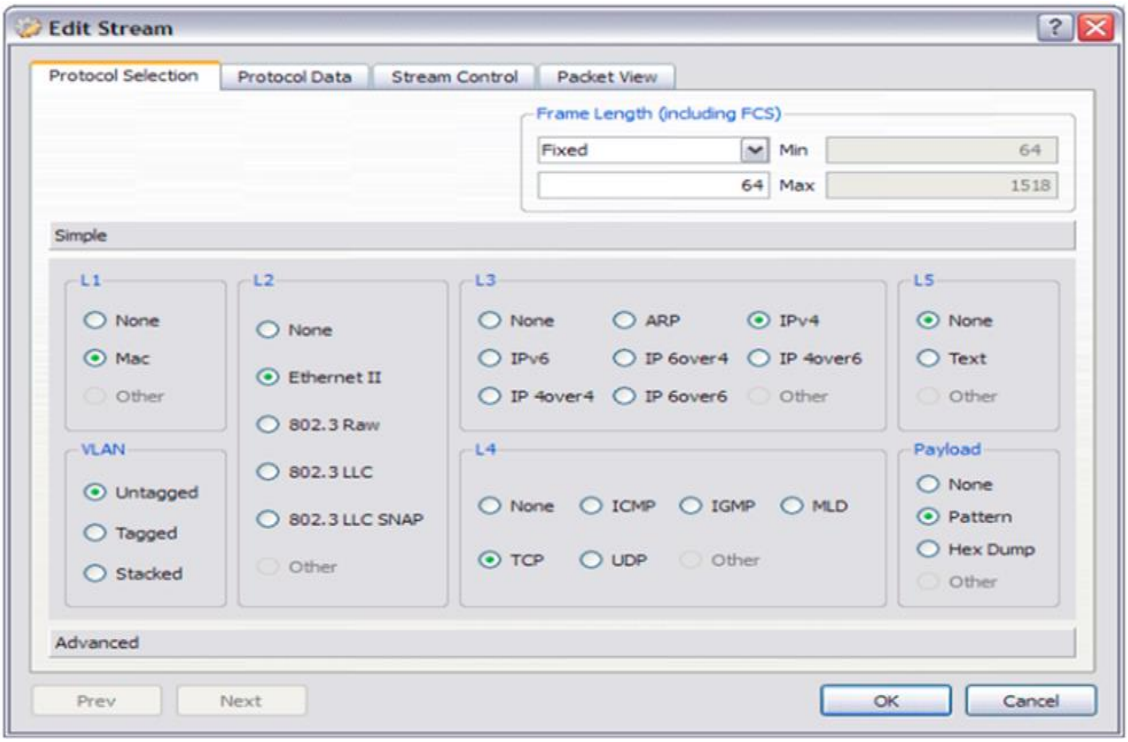


Figure 4.1: Supported Protocols by Ostinato

4.7.3 PRTG and IPSLA

Data gathering, network control, and network administration all rely on network monitoring technology. The observed data is put to use in a number of ways. Data on network performance characteristics is collected using various monitoring tools. The monitoring tools employ a packet sniffing approach in order to gather the necessary data, which may be utilized for a variety of purposes, including troubleshooting and ensuring

that network devices function within their specified threshold values. There are various types of monitoring tools available on the market, including Paessler Router Traffic Grapher (PRTG) and Internet Protocol Service Level Agreement (IP SLA) [46] [47].

IPSLA and PRTG are both active monitoring tools that inject traffic into the network and collect real-time performance parameter measurements. Cisco IPSLA collects latency, jitter, and packet loss in real-time. In the case of PRTG [48], it offers a wide range of monitoring possibilities. It also offers a plethora of monitoring capabilities that make it a highly strong monitoring tool. PRTG can also connect with other monitoring tools and provide a centralized monitoring system for gathering all of the necessary performance measurement data from its probes and other monitoring systems.

The network performance metrics throughput, latency, and packet loss will be used to measure the performance of the CGN operation in this thesis work. The GNS3 simulation tool, Ostinato traffic generator, and PRTG combined with IPSLA monitoring tool are used for simulating the network topology, generating subscriber traffic, and capturing network performance metrics output respectively. As a future endeavor, the additional performance measurements may be utilized to further analyze the impact of CGN on performance.

5. Simulation Scenarios, Result, and Analysis

The goal of this chapter is to discuss the setup process of the simulation scenarios, and how to assess the results based on the network performance indicators that were chosen. And finally, it will present the results of the network performance output graphically.

5.1 Simulation Scenario and Network Architecture

The research will be carried out utilizing the GNS3 simulation environment. Two simulation scenarios are used to achieve the research's aim. A conventional CGN operation and a CGN with Traffic Prioritization (TP) operation were simulated using the setup procedure illustrated in figure 5.1. The simulation setup is comprised of many steps that operate in tandem to fulfill the research's aim.

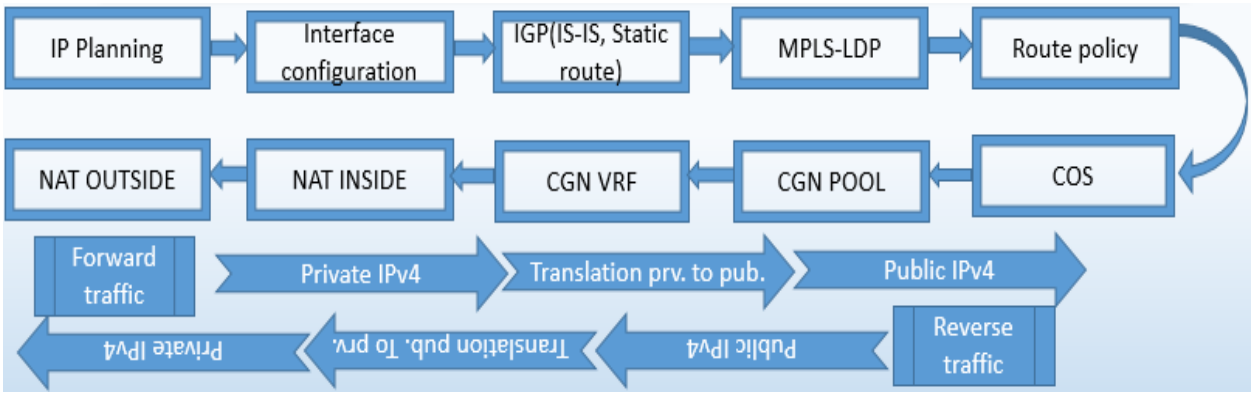


Figure 5.1: Configuration Flow

5.1.1 Network Architecture

The simulation topology is made up of routers, switches, monitoring tools, and traffic generator tools that are used to simulate client traffic and eject it into the network. The deployed topology, as seen in Figure 5.4, tries to reflect the most popular core networks among operators. The following are the key components of the simulation topology:

- IP planning: IPv4 addresses are utilized in the simulated topology. IPv4 address planning is governed by RFC1918 and RFC6598. RFC1918 established two types

of IPv4 addresses: private IPv4 addresses and public IPv4 addresses. Private IPv4 addresses are used on the customer network, whereas public IPv4 addresses are used on the ISP core network and at the subscriber gateway. The CGN device converts a customer's private address to a globally routable address. According to RFC6598, ISP core networks with CGN operation should utilize a shared IPv4 address range that is allocated for them to avoid private IPv4 address conflicts with subscriber gateway addresses.

- Interface configuration: There are two types of interfaces in the simulation topology. As a point-to-point interface, physical interfaces are used to connect various network devices. The logical or loopback interface, on the other hand, is utilized for accessing various network devices and providing network performance parameter output to monitoring tools.
- Interior getaway protocol (IGP): The simulation's IGP protocols include Intermediate System to Intermediate System (IS-IS) and static route. The IS-IS protocol is the most often used routing protocol among service providers when building a backbone network. It is a link-state protocol that uses the Dijkstra algorithm to find the best path through diverse networks. Its usage of the same default cost values for all interface types is one of its most attractive features. This allows the network administrator to steer traffic in the desired direction by manually specifying varying link costs so that traffic follows a certain path. Static routes are routing protocols that are manually set to allow a specific network to follow the desired path. Dynamic and static routing are both utilized to maximize a network device's routing decision capabilities.
- Multi-Protocol Label Switching (MPLS): Circuit switching and packet switching are the two most used switching technologies. There is a specific path for traffic to follow in the case of circuit switching. In the case of packet switching, it routes traffic using the optimal path determined by the routing protocols. As a result, it

might take different pathways to reach the destination network. When transmitting a packet with packet-switched technology, each router along the packet's path will open the packet, read it, and forward it depending on their routing table. MPLS operates between layers two and three of the TCP/IP transmission protocol stack, allowing it to forward packets quickly. It assigns a label to each network and quickly passes a packet to its destination based on the label. The labels are generated using the Label Distribution Protocol (LDP). Even though it transmits layer three IP packets, MPLS functions as a circuit-switching connection in general.

- IP routing policies are a technique that directs traffic to adhere to a set of rules. It functions as an if-else expression. It has a condition set, and if that condition is satisfied, it will act on it; otherwise, it will go on to the next condition. It includes an IP prefix list as well as a route map. It enables network administrators to grant certain traffic precedence if a predefined condition is met. The routing choice can be made based on application type, protocol type, or IP addresses to give a packet a certain priority.
- Class of Service: Using the type of service field in the IP header, distinct network traffic can be designated for different handling. It can produce seven different sorts of traffic. In the TP scenario, network traffic is classified into three types: gold, silver, and bronze. These categories are based on network addresses and are assigned various priorities. According to the Cisco IOS setup handbook, gold traffic is assigned a crucial marking, silver traffic is assigned a flash-override marking, and bronze traffic is allocated flash marking.
- CGN setup: The CGN configuration is composed of several components. It is made up of a public IP pool, an access list, Virtual Routing and Forwarding (VRF), and dedicated inside and outside interfaces for routing customer traffic to the Internet and vice versa. Session management is done per VRF for different

traffic kinds and priorities. For the TP scenario, three VRFs are constructed. The VRFs VRF-GOLD, VRF-SILVER, and VRF-BRONZE were created to correspond with the traffic categorization and priority specified at the edge router.

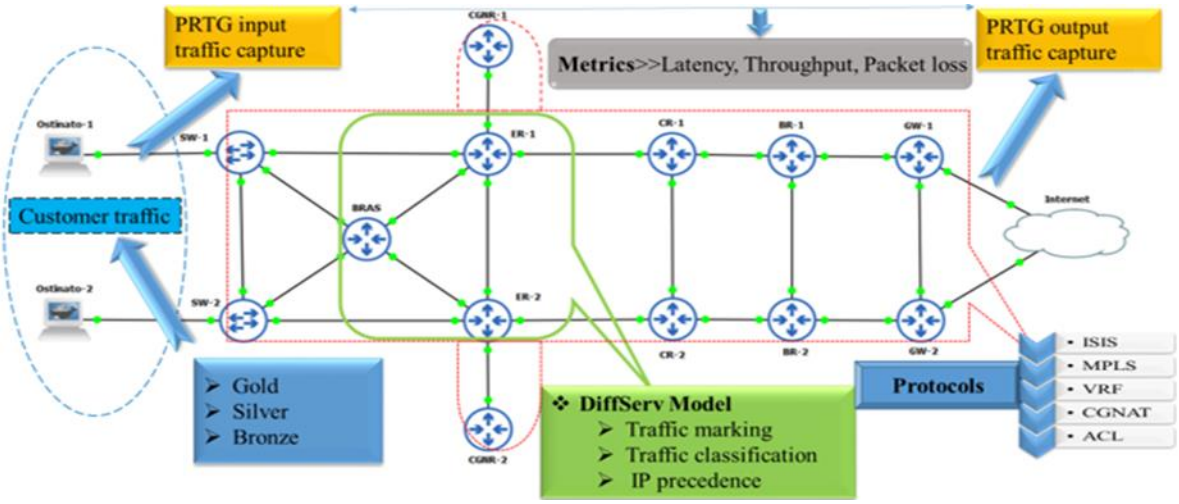


Figure 5.2: A Simulation Model

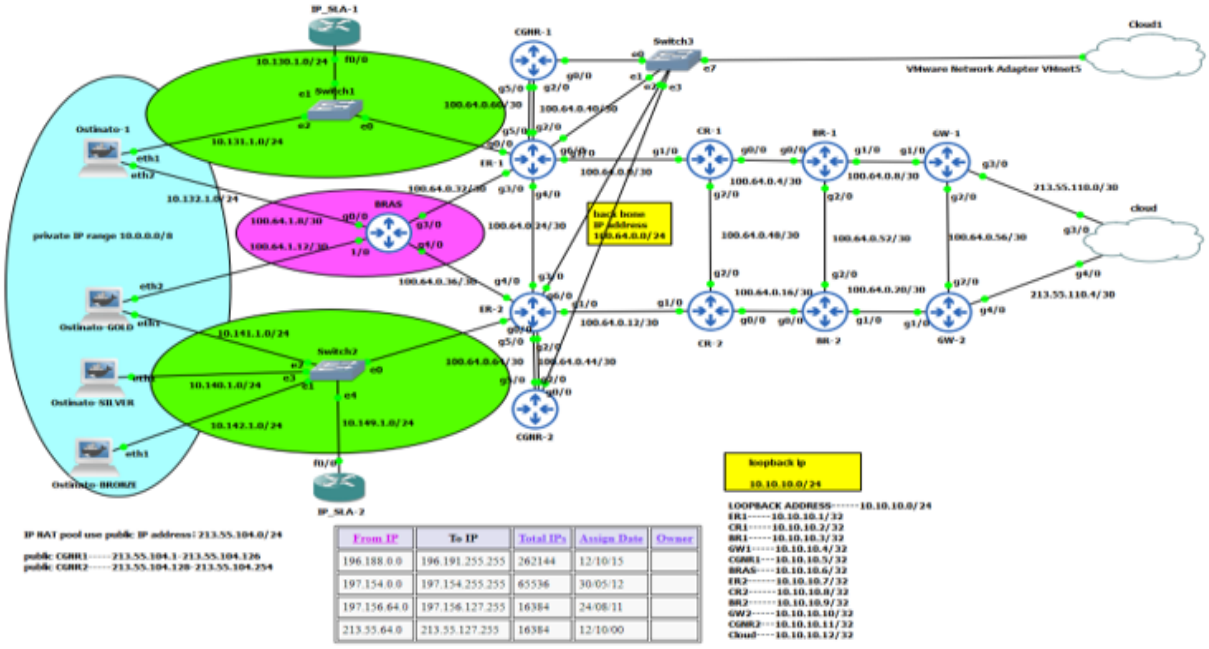


Figure 5.3: Scenarios Network Topology

5.1.2 Simulation Environment

Six steps are conducted to explore the improvement of CGN deployment using TP. The first step included a review of the literature on CGN operation, traffic prioritization, and network performance indicators. The second phase involved the creation of a simulation model. Following that, the simulation scenarios and simulation environment installation were carried out. Finally, the results and discussion portion will take place. The results of the two scenarios will be compared using the three network performance indicators (i.e. latency, packet loss, and throughput).

Several simulations were run to investigate the impact of CGN with TP and regular CGN operation on network performance. As mentioned in Chapter 4, latency, throughput, and packet loss are employed as network performance parameters to compare the performance of the two scenarios. Ostinato generates two traffic figures in order to obtain a reliable result from the simulations. Various protocols are chosen for traffic generation using Ostinato. TCP, UDP, and ICMP are the traffic streams involved in creating the required traffics. Figure 5.4 is an example of traffic created by ostinato.

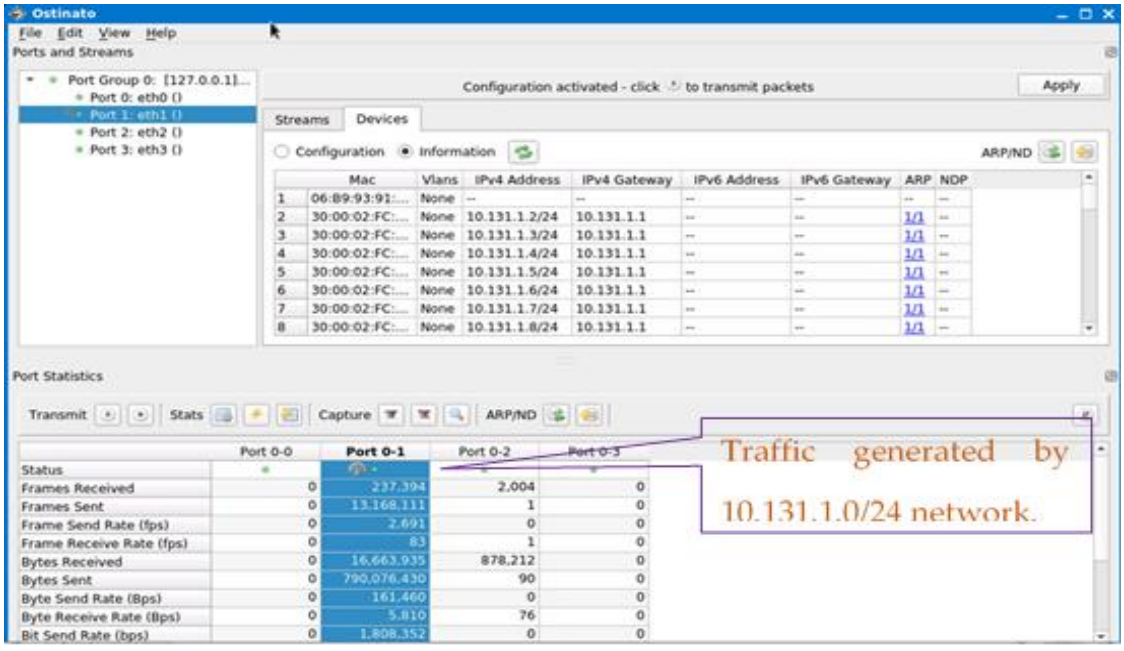


Figure 5.4: Traffic Generated Using Ostinato.

Three class C networks are used to replicate the customer's traffic. In the first scenario, conventional deployment CGN with no traffic distinction is utilized. In this case, each host has equal access to the CGN's restricted resources. The customer's traffic is prioritized into three groups in the second scenario. Depending on the traffic classification, CGN deployment takes a different approach. Three IP precedence values are used to identify customer traffic priority. Gold traffic is given critical priority, whereas silver and bronze traffic got flash-override and flash IP precedence values, respectively. To correspond with this traffic marking, the CGN was equipped with three VRF. The edge router labels traffic and routes it to CGN through three sub-interfaces.

In CGN plus TP operation, the VRFs are configured with a distinct number of session entries for each traffic priority. The gold traffic configuration uses 15000 maximum-session entries, the silver traffic configuration uses 35000 maximum-session entries, and the bronze traffic configuration uses 40000 maximum-session entries. 150 gold customers, 225 silver customers, and 375 bronze customers are used to generate traffic. In order to efficiently employ translation session entries, several timeout configurations were used for the protocols. For dynamic NAT translation entries, the default timeout settings are 24 hours for TCP connections and 5 minutes for UDP connections. To make the most of port numbers and session generation, idle connection waiting time must be controlled. TCP connections are set to timeouts of 1500, 1000, and 500 seconds for gold, silver, and bronze traffic, respectively. For gold, silver, and bronze traffic, the NAT translation timeout for the UDP protocol was set to 90 seconds, 60 seconds, and 30 seconds, respectively.

Based on the network performance indicators specified, the simulation results are utilized to evaluate the hypothesis. The simulation's implementation strategy is depicted in Table 5.1.

Table:5.1: Simulation Plan

Simulation Scenarios Scheme			
No. of Simulation	**Traffic Generated**	**Scenario**	**Duration**
Simulation one	550Kbps	CGN	60 Minutes
Simulation two	550Kbps	CGN+TP	60 Minutes
Simulation three	850Kbps	CGN	60 Minutes
Simulation four	850Kbps	CGN+TP	60 Minutes

5.2 Analysis of Simulation Results

In this section, the results gained from the simulation will be presented followed by analysis and discussion.

5.2.1 Analysis of Throughput

As mentioned in Chapter 4, throughput tells us how much data was transported from a source to a destination at any given time through a medium. A packet generator is used to deliver a packet at a specific rate to test throughput for specific traffic that goes through the CGNAT operation. The produced packets are then collected using a monitoring tool after passing through the CGN process. Finally, a comparison will be performed to determine throughput performance.

Table 5.2 shows the output of throughput for 550Kbps traffic. The data is extracted from the simulation using the PRTG monitoring tool.

Table 5.2: Throughput for CGN and CGN + TP of 550Kbps Traffic

Sampling Time(min)	5	10	15	20	25	30	35	40	45	50	55	60
CGN(Kbps)	440.8	442.6	445.4	432.5	465.6	456.8	436.2	466.8	442	434.4	460.7	460.2
CGN+TP(Kbps)	512.8	546.4	517.9	502.9	538.2	521.4	539	541.4	540	522.9	540.6	536

The graphical depiction of the simulation result for 550Kbps traffic is shown in Figure 5.5. The CGN operation has an average throughput of 448.63 Kbps, whereas the CGN With TP operation has an average throughput of 529.98. According to the findings, CGN with TP operation performs better. CGN coupled with TP increased throughput by 15.2 % on average when compared to regular CGN operation.

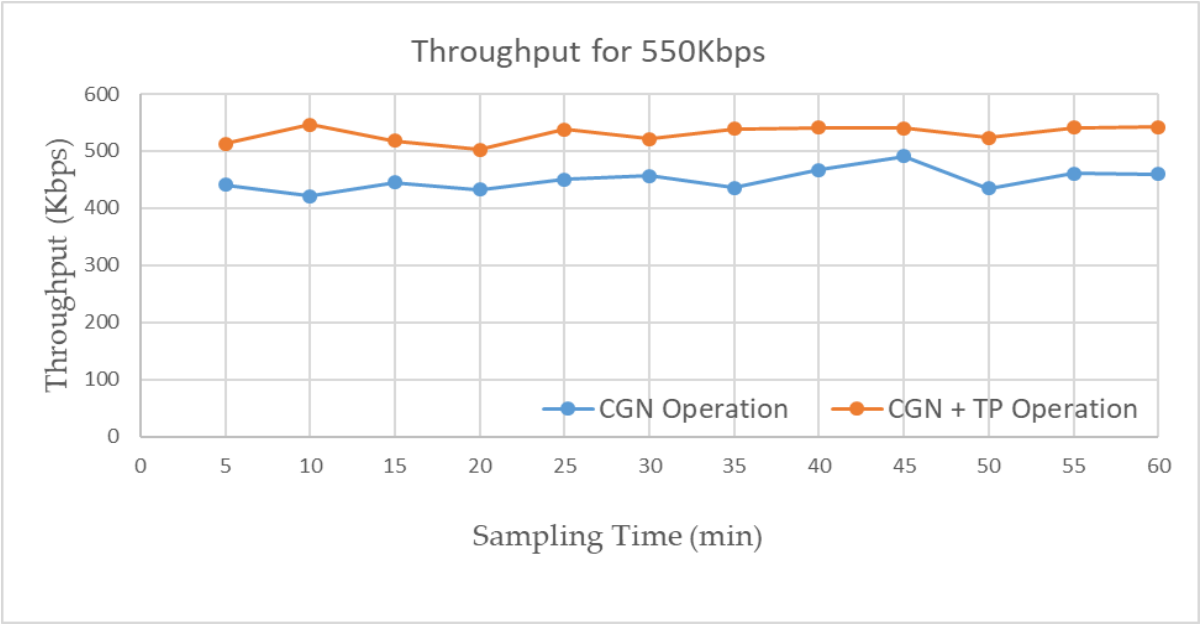


Figure 5.5: Graph of Throughput for Scenarios 1 and 2 of 550Kbps Traffic

Table 5.3 displays the throughput output for CGN and CGN with TP operating with 850Kbps traffic. The simulation is set to run for 60 minutes. The output is collected over a 5-minute sample interval.

Table 5.3: Throughput for CGN and CGN + TP of 850Kbps Traffic

Sampling Time(min)	5	10	15	20	25	30	35	40	45	50	55	60
CGN(Kbps)	678.4	680.5	650.5	645.9	665.7	562.4	623.7	620.8	630.2	590.6	623.8	588.9
CGN+TP(Kbps)	771.8	790.5	784.1	750.8	788.6	750.8	802.4	750.7	805.6	780.6	788.1	778.2

Figure 5.6 displays the result of throughput for scenarios 1 and 2. The average throughput for CGN + TP operation is 778.52Kbps. For standard CGN operation, the average throughput is 630.12Kbps. When compared to CGN operation, CGN with TP operation boosts throughput by 19% on average.

CGN throughput is determined by the rate at which the CGN router delivers a packet from the inbound interface to the outgoing interface. The CGN device translates each packet's IP address and port number and maintains the session for filtering the returned traffic. Changes to IPv4 addresses result in changes to the checksum header, which increases processing time. Due to that, the throughput will be affected.

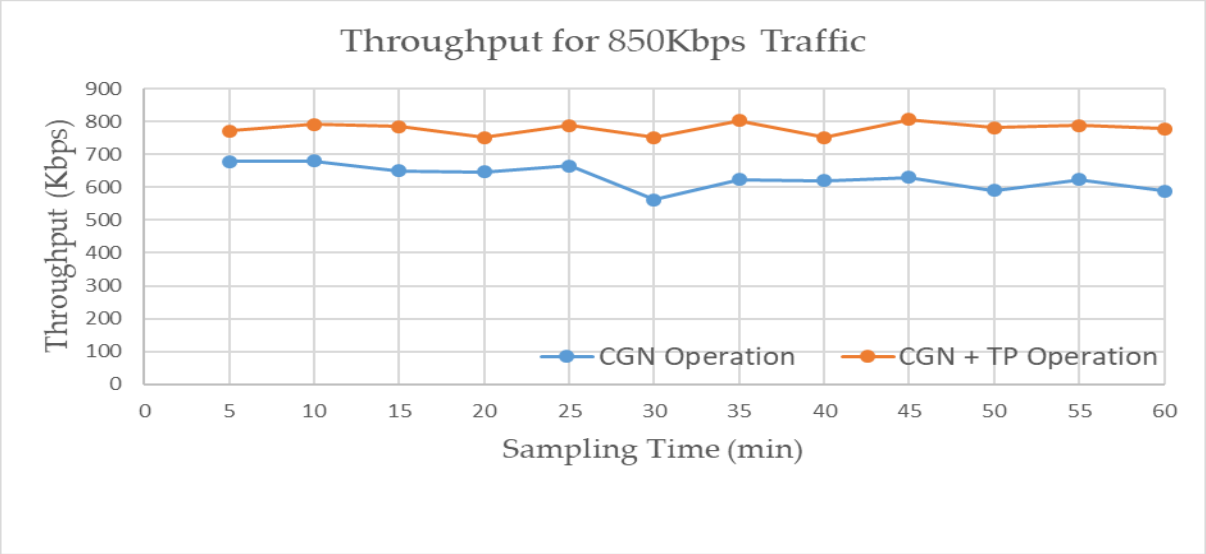


Figure 5.6: Graph of Throughput for Scenarios 1 and 2 of 850Kbps Traffic

5.2.2 Analysis of Latency

One of the network performance factors is the time it takes for a packet to arrive at its destination. The time it takes for a packet to be transmitted is referred to as delay or latency. The ITU recommends a maximum one-way end-to-end delay of 150ms. The latency of the two scenarios is measured using the identical simulation setup. Smooth CGN operation is essential for stable and high-quality packet transmission. PRTG was used to continuously monitor network performance. PRTG utilizes a sensor as a fundamental monitoring element, which gives a measured output for the needed network performance statistic. IP SLA uses the Internet Control Message Protocol (ICMP) ping test to assess how long it takes a packet to reach its destination. For assessing network latency, PRTG monitoring software uses a Cisco IP SLA sensor.

Table 5.4 displays the latency output for scenarios 1 and 2 of 550kbps traffic.

Table 5.4: Latency for CGN and CGN + TP of 550Kbps Traffic

Sampling Time (min)	5	10	15	20	25	30	35	40	45	50	55	60
CGN (msec)	77.1	80.8	86.4	89	91.2	111.4	120.5	122.6	127.4	137.8	149.2	165.4
CGN+TP (msec)	69.3	70.4	65.4	76.5	82.5	97.0	100.6	96.9	111.8	113.7	131.1	138.6

The CGN with TP combination performs better, as seen in Figure 5.7. CGN operation has an average latency of 113.8msec. The average latency for CGN with TP operation is 96.2msec. The CGN paired with TP implementation reduced latency by 15% on average.

The time is measured using a time-stamp in the packets. The monitoring tool computes the time elapsed between the original packet time-stamp and the destination arrival time. The CGN device and the subscriber's relative distance from the CGN device both contributed to the delay in CGN operation. The latency of the CGN device is taken into account here.

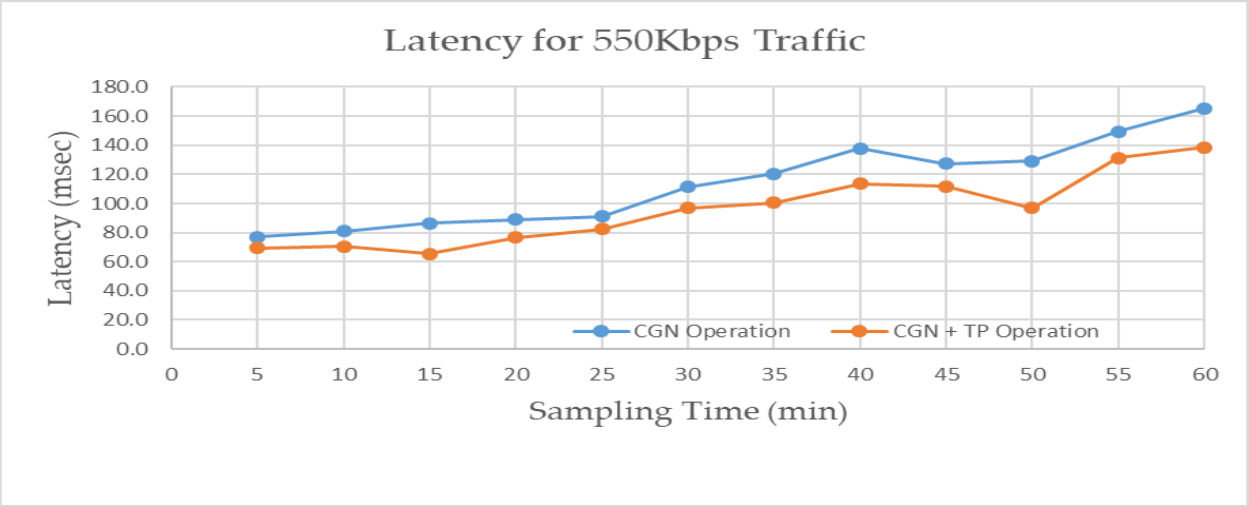


Figure 5.7: Graph of Latency for Scenarios 1 and 2 of 550Kbps Traffic

Table 5.5 shows the output of latency for 850Kbps traffic.

Table 5.5: Latency for CGN and CGN + TP of 850Kbps Traffic

Sampling Time (min)	5	10	15	20	25	30	35	40	45	50	55	60
CGN (msec)	85.3	89.2	97.4	98.6	106.4	121.2	127.5	131.2	134.6	140.7	159.7	175.5
CGN+TP (msec)	70.3	73.4	74.4	76.54	81.5	94.4	98.6	93.9	107.8	115.7	127.1	142.6

Figure 5.8 shows the result of latency for the two scenarios for 850Kbps traffic. As observed from the graph the CGN operation has higher latency on average when compared with CGN with TP setup. The average latency for CGN operation is 122.3msec.

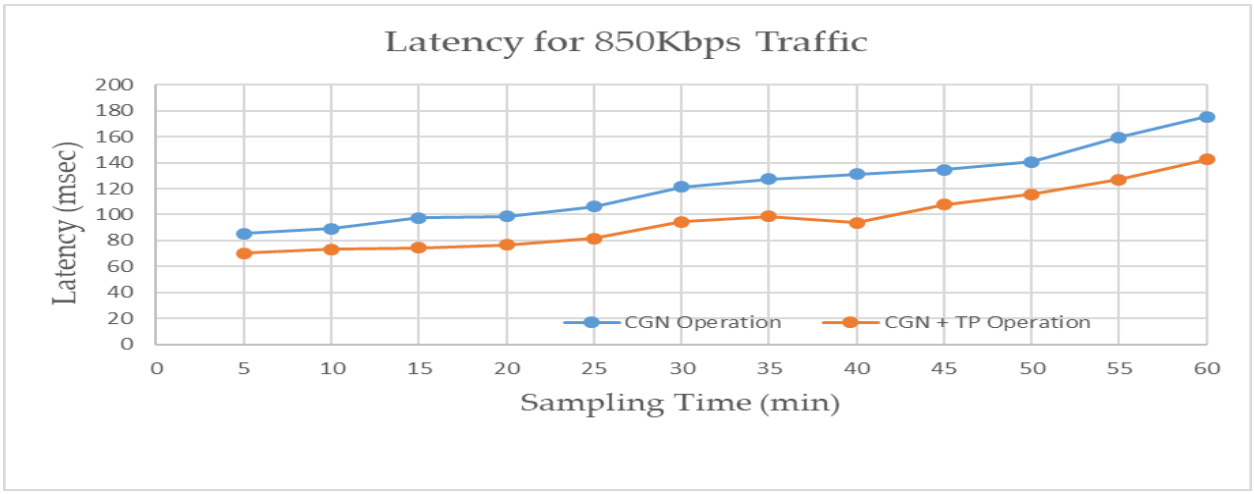


Figure 5.8: Graph of Latency for Scenarios 1 and 2 of 850Kbps Traffic

CGN with TP configuration reduced latency by 21% when compared to conventional CGN operation.

5.2.3 Analysis of Packet loss

IP-SLA creates packet loss data by utilizing a variety of payloads. IP-SLA is linked to PRTG in order to obtain the packet loss result. To get an acceptable result, two traffics are generated. Packet loss was used to determine whether or not packets arrived safely at their destination. Using PRTG and an IP-SLA sensor, the ICMP test is performed to measure packet loss in the two scenarios. Table 5.6 displays packet loss results for 550Kbps traffic.

Table 5.6: Packet Loss for CGN and CGN + TP of 550Kbps Traffic

Sampling Time (min)	5	10	15	20	25	30	35	40	45	50	55	60
CGN (%)	3	9	14	16	19	22	25	28	31	36	37	40
CGN+TP (%)	3	8	11	14	17	19	20	22	23	30	31	34

Figure 5.9 depicts the graphical representations of the outcomes for Scenarios 1 and 2. CGN operation with TP deployment has lower packet loss than CGN operation without TP deployment.

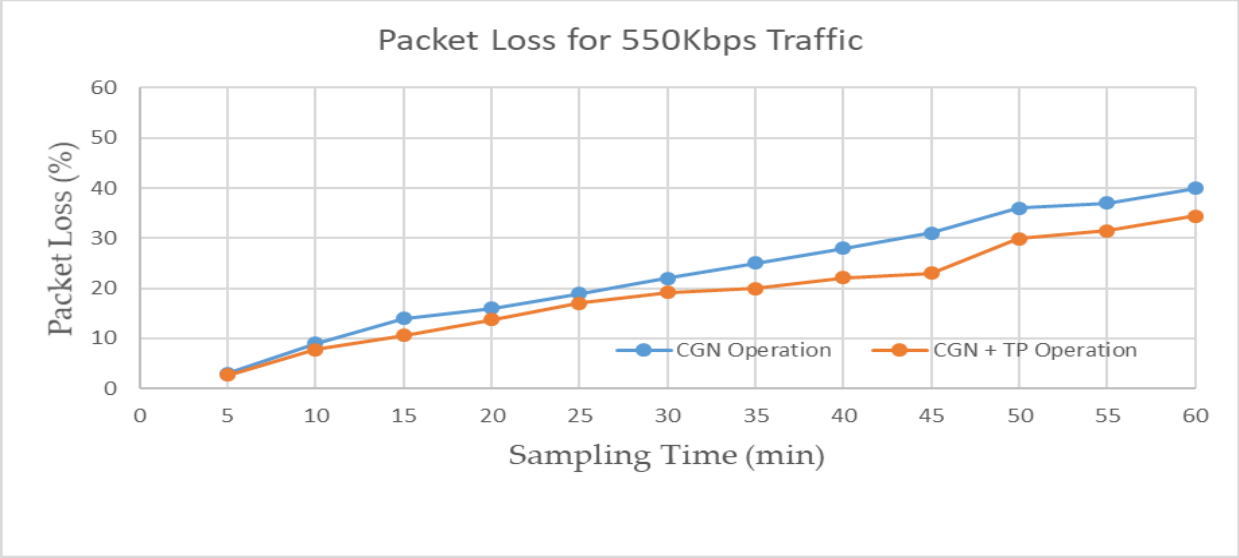


Figure 5.9: Packet Loss Graph for Scenarios 1 and 2 with 550Kbps Traffic

The average packet loss for the CGN operation is 23%. For CGN with TP operation the average packet loss is 19%. On average CGN with TP improved the pack loss by 16%.

Table 5.7 shows the output of pack loss for 850Kpbs traffic for both scenarios.

Table 5.7: Packet Loss for CGN and CGN + TP of 850Kbps Traffic

Sampling Time (min)	5	10	15	20	25	30	35	40	45	50	55	60
CGN (%)	14	19	23	25	29	31	35	41	40	45	47	50
CGN+TP (%)	11	16	17	23	20	27	26	27	30	37	40	43

As shown in figure 5.10 the CGN with TP operation has less packet loss. On average the CGN Operation with TP has improved the packet loss by 20%.

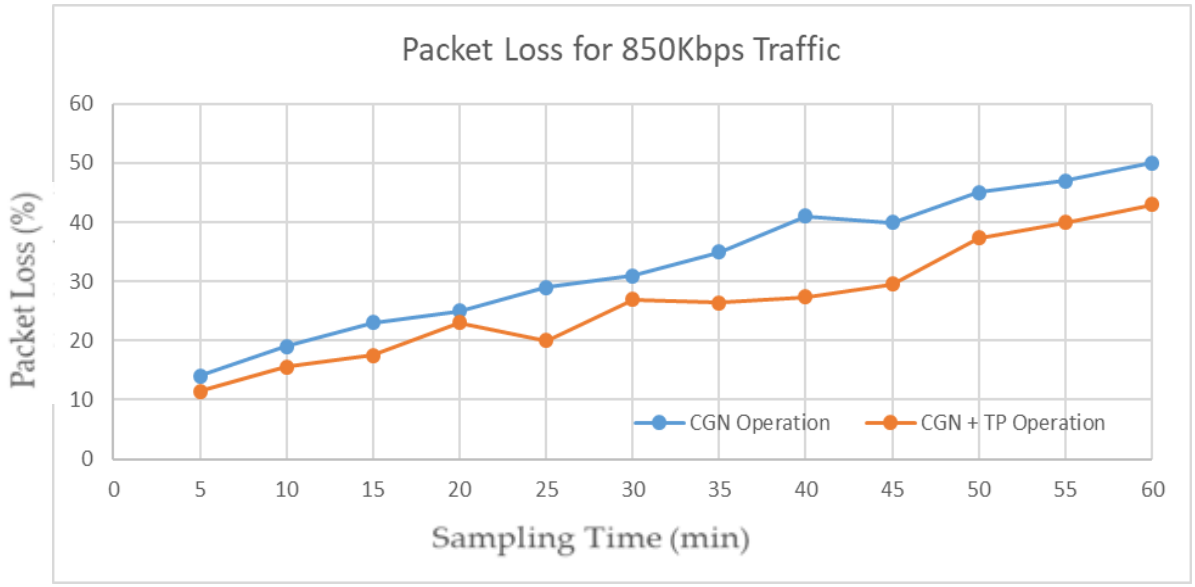


Figure 5.10: Packet Loss Graph for Scenarios 1 and 2 with 850Kbps Traffic

In general, standard CGN keep the translation state in the session table for a longer length of time for each active traffic to keep the inbound to outbound and outbound to inbound traffic flows steadily. As a result, additional sessions are created for new subscriber traffic. Because no NAT entry limitation is specified, the growing number of translation entries consumes a lot of memory and CPU resources. As a result, CGN's performance suffers, resulting in high latency, high packet loss, and lower throughput. Whereas in the

case of CGN paired with TP operation there is session management put in place that controls the session limit and expires the idle connection quickly in order to avail the NAT entry for the new session request.

The operation power of the CGN device determines the maximum translating rate at which new session entries are created and removed. With a limited number of port numbers, the creation and removal of translation entries have an impact on network performance. The impact of conventional CGN operation on latency, throughput, and packet loss is higher than that of CGN with TP operation, as demonstrated in the findings since the formation of a new session is limitless and each NAT entry state is retained for a long period, resulting in memory and CPU overload. The fact that CGN combined with TP operation effectivity optimized the session management resulting in improved performance. CPU usage, memory utilization, traffic load, traffic type, number of session entries, and CGN operation configuration are all variables that influence the network performance of CGN deployment. As demonstrated in the results, with more session formation and higher traffic load, the impact of conventional CGN operation on network performance is greater than that of CGN with TP operation.

The performance of the personal computer was a challenge during the simulation. In order to get the best possible output, certain redundant routers were disabled, which had no effect on the output. The fact that the simulation done on a personal computer could be considered as a challenge during the analysis and the performance issue could be avoided if the analysis is performed in a real environment or on a testbed.

6. Conclusion and Future work

6.1 Conclusion

The influence of installing CGN combined with TP architecture operation on network performance is analyzed in this thesis work in contrast to conventional CGN operation. This research has looked at the impact the various configuration setup operation has on CGN along with TP and standard CGN operation. The translation creates-rate, translations delate-rate, inside to outside forwarding rate, and outside to inside forwarding rate all played important roles in assessing the performance characteristic of CGN integration with TP and conventional CGN operation.

Network performance metrics such as latency, pack loss, and throughput are used to analyze and compare the deployment of standard CGN operation and CGN paired with TP operation. GNS3 has been used to set up the two network scenarios along with all required protocol configurations. After the simulation is set up, an ostinato traffic generator is used to generate customer traffic. Then the result of the network performance metrics is gathered from the two scenarios using PRTG. Finally, the result is presented in graphical form using excel. From the simulation results, the following conclusions can be drawn:

- CGN paired with TP can improve network performance by optimally use the limited resource of CGN operation.
- CGN operation has an impact on network performance.
- TP integration with CGN improves the latency by 21% on average when compared with standard CGN operation.
- When comparing CGN paired with TP operation to conventional CGN operation, throughput improves by 19% on average.

- The scenarios demonstrate that CGN deployment has an influence on network performance however the impact is mitigated in TP with CGN integration.
- Standard CGN operation shows greater packet loss than CGN integration with TP setup. CGN with TP integration shows reduced packet loss with an average value of 20%.

6.2 Future Work

This research conducted an analysis to achieve the objectives stated in Chapter One. Certain concerns, however, will need to be addressed in the future. Among these concerns are the following:

- Examine the network performance impact of activating application layer gateway and port control protocol support on CGN devices in order to allow applications to run smoothly.
- Analyze the impact of installing a dual-stack backbone on CGN performance. Where IPv6 and IPv4 services are coexisting.
- Implement and analyze software-defined networking along with the CGN operation to further improve the network performance.

Reference

[1] Cisco, "Cisco Annual Internet Report (2018–2023)," *white paper*, pp. 1–41, 2020

[2] N. Skoberne, O. Maennel, I. Phillips, R. Bush, J. Zorz, and M. Ciglaric, "IPv4 address sharing mechanism classification and tradeoff analysis," *IEEE/ACM Transactions on Networking*, vol. 22, no. 2, pp. 391–404, 2016.

[3] W. Sugeng, J. Istiyanto, K. Mustofa, and A. Ashari, "The Impact of QoS Changes towards Network Performance," *International Journal of Computer Networks and Communications Security*, vol. 3, no. 2, pp. 48–53, 2015.

[4] ETC IP NGN Project, "ETC IP NGN Project Implementation Plan-V3.0," 2008.

[5] P. Richter, F. Wohlfart, N. Vallina-Rodriguez, et al., "A Multi-perspective Analysis of Carrier-Grade NAT deployment," Proceedings of the 2016 Internet Measurement Conference, pp. 215–229, 2016.

[6] C. Donley, L. Howard, V. Kuarsingh, J. Berg, and J. Doshi, "Assessing the Impact of Carrier-Grade NAT on Network Applications," *RFC 7021*, 2013.

[7] E. Bocchi, A. Ktahouni, S. Traverso, *et al.*, "Impact of Carrier-Grade NAT on web browsing," *IWCMC 2015 - 11th International Wireless Communications and Mobile Computing Conference*, pp. 532–537, 2015.

[8] S. Alcock, R. Nelson, and D. Miles, "Investigating the Impact of Service Provider NAT on Residential Broadband Users," Computer Communications Networks, pp. 4–9, 2014.

[9] I. Livadariu, K. Benson, A. Elmokashfi, A. Dhamdhere, and A. Dainotti, "Inferring Carrier-Grade NAT Deployment in the Wild," *Proceedings - IEEE INFOCOM*, vol. 2018-April, pp. 2249–2257, 2018.

[10] D. Samociuk, B. Adamczyk, and A. Chydzinski, “Impact of Router Security and Address Translation Mechanisms on the Transmission Delay,” *INTERNET 2015: The Seventh International Conference on Evolving Internet Figure*, pp. 1–5, 2015.

[11] Y. Ohara, K. Nishizyka, K. Chinen, *et al.*, “On the impact of mobile network delays on connection establishment performance of a carrier-grade NAT device,” *Asian Internet Engineering Conference, AINTEC 2014*, pp. 1–8, 2014.

[12] B. Aitken, “Report on the Implications of Carrier Grade Network Address Translators,” *InterConnect Communication*, 2013.

[13] D. Wing, “Network address translation: Extending the Internet address space,” *IEEE Internet Computing*, vol. 14, no. 4, pp. 66–70, 2010.

[14] O. Maennel, R. Bush, L. Cittadini, and S. Bellovin, “A Better Approach than Carrier-Grade-NAT,” *Technical Report CUCS*, 2008.

[15] S. Mahmoud, H. Sqalli and N. Sakthi, “A Scalable NAT-based Solution to Internet Access Denial by Higher-tier ISPs,” *Security and Communication Networks*, vol. 5, no. June, pp. 422–437, 2012.

[16] Cisco, “How Can Service Providers Face IPv4 Address Exhaustion? - Cisco,” no. June, pp. 1–28, 2012.

[17] “Dynamic PAT – NAT Series," *Practical Networking.net,* 26-Mar-2019. [Online]. Available: https://www.practicalnetworking.net/series/nat/dynamic-pat/ (accessed July. 22, 2021).

[18] A. Moraes, "Cisco Firewalls” Cisco Press, 2011.

[19] O. Wendell, "CCNA 200-301 Official Cert Guide”, Cisco Press, 2020.

[20] Z. Wang, Z. Qian, Q. Xu, Z. Mao, and M. Zhang, "An Untold Story of Middleboxes in Cellular networks," *Computer Communication Review*, vol. 41, no. 4, pp. 374–385, 2011.

[21] A. Müller, F. Wohlfart, and G. Carle, "Analysis and topology-based traversal of cascaded large scale NATs," *HotMiddlebox 2013 - Proceedings of the 2013 Workshop on Hot Topics in Middleboxes and Network Function Virtualization*, pp. 43–48, 2013.

[22] K. Akpinar, M. Akpinar, I. Ozcelik, and N. Yumusak, "Carrier-grade NAT-Is it really secure for customers? A test on a Turkish service provider," *Application of Information and Communication Technologies, AICT 2017 Conference Proceedings*, 2017.

[23] M. Mostafa, A. Abou, E. Kalam, and C. Fraboul, "Extending Firewall Session Table to Accelerate NAT, QoS Classification and Routing," *Toulouse Computer Science Research Institute,* 2017.

[24] F5 Global Services, "Manage IPv4 Address Depletion and Seamlessly Migrate to IPv6," *White paper, BIG-IP Carrier-Grade NAT,*2016.

[25] H. Ma, Y. Wu, Y. Ma, and Z. Wang, "Optimization Scheme Of CGN Logs," *Proceedings of IEEE CCIS2016*, pp. 1–4, 2016.

[26] J. Amutha, S. Albert, and R. Meenakshi, "An integrated secure architecture for IPv4/IPv6 address translation between IPv4 and IPv6 networks," *Advances in Intelligent Systems and Computing,* vol. 380, pp. 669–679, 2015.

[27] M. Haroon and H. Ali, "Hurdles in Implementing IPv6 in Pakistan," *IJCSNS International Journal of Computer Science and Network Security*, vol. 17, no. 12, pp. 151–157, 2017.

[28] C. Feerick, "Learn About Quality of Service (QoS) Why QoS is Essential in Today's

Networks," *Juniper Networks*, vol. 1, 2015.

[29] A. Toguyéni and O. Korbaa, "Quality of service of Internet service provider networks: State of the art and new trends," *ICTON-MW'07 - International Conference on Transparent Optical Networks "Mediterranean Winter" 2007- Conference Proceedings*, pp. 1–8, 2007.

[30] I. Octavio, J. Parra, I. Angela, P. Rios, and G. Rubio, "Quality of Service over IPV6 and IPV4," *7th International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 4–7, 2011.

[31] Z. Li, Y. Liu, and Y. Jing, "Design of adaptive backstepping congestion controller for TCP networks with UDP flows based on minimax," *ISA Transactions*, vol. 95, pp. 27–34, 2019.

[32] M. Barry, J. Tamgno, C. Lishou, and M. Cisse, "QoS impact on multimedia traffic load (IPTV, RoIP, VoIP) in best-effort mode," *International Conference on Advanced Communication Technology, ICACT*, vol. 93, pp. 694–700, 2018.

[33] "The DiffServ Model,"[Online].Available: https://www.scribd.com/document/324954154/The-DiffServ-Model/ (accessed July. 09, 2021).

[34] C. Williams, "The Impact of Advanced QoS Mechanisms on the Performance of Converged Networks" *A Dissertation*, Glasgow Caledonian, August 2019.

[35] B. Vellambi, N. Torabkhani, and F. Fekri, "Throughput and latency in finite-buffer line networks," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3622–3643, 2011.

[36] J. Li, D. Li, Y. Huang, Y. Cheng, and R. Ling, "Quick NAT: High-performance NAT

system on commodity platforms," *IEEE Workshop on Local and Metropolitan Area Networks*, vol. 93, no. 1, pp. 1–2, 2017.

[37] R. Ling, Y. Cheng, Y. Huang, and J. Li, "High Performance and Scalable NAT system on commodity platforms," *Attaining High-Performance Communications*, pp. 57–92, 2016.

[38] D. Tutsch, "Performance analysis of network architectures" *Springer*, 2006.

[39] V. Thapar, "A Comparative Study of Various Network Simulation Tools," *International Journal on Computer Science and Engineering (IJCSE)*, vol. 9, no. 06, pp. 385–392, 2017.

[40] S. Gupta, M. Ghonge, P. Thakare, and P. Jawandhiya, "Open-Source Network Simulation Tools: An Overview," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 2, no. 4, pp. 1629–1635, 2015.

[41] C. Neumann , "The Book of GNS3: Build Virtual Network Labs Using Cisco, Juniper, and More" [Online]. Available: https://books.google.com.et/books?hl=en&lr=&id=9wcvDwAAQBAJ&oi=fnd&pg=PR5&dq=gns3&ots=aEKh8oMZPa&sig=gziGE8kbnPKO_84z38taE5VWBYc&redir_esc=y#v=onepage&q=gns3&f=false (accessed Aug. 20, 2021).

[42] J. Castillo, "GNS3 Limitations when Emulating Connectivity and Management for Backbone Networks: A Case Study of CANARIE", 2020.

[43] A. Soumya, B. Amulya, and M. Moharir, "A study of tools to develop a traffic generator for L4 - L7 layers," *Proceedings of the 2016 IEEE International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2016*, pp. 114–118, 2016.

[44] M. Swann, J. Rose, G. Bendiab, S. Shiaeles, and N. Savage, "A Comparative Study of Traffic Generators: Applicability for Malware Detection Testbeds," *Journal of Internet Technology and Secured Transactions*, vol. 8, no. 1, pp. 705–713, 2020.

[45] B. Rahuldhev, M. Moharir, and P. Kumar, "Ostinato-A powerful traffic generator," *2nd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2017*, pp. 1–5, 2017.

[46] P. Allavi and H. Patel, "Analysis of Various Packet Sniffing Tools for Network Monitoring and Analysis," *International Journal of Electrical, Electronics and Computer Engineering*, vol. 1, no. 1, pp. 55–58, 2012.

[47] S. Lee, K. Levanti, and H. Kim, "Network monitoring: Present and future," *Computer Networks*, vol. 65, pp. 84–98, 2014.

[48] D. Zobel, "Monitoring Applications and Services with Network Monitoring," *PAESSLER*, pp. 1–8, 2017.

Appendix

Network Performance Analysis of Carrier-Grade Network Address Translation with Network Traffic Prioritization

Abraham Meles

Electrical and Computer Engineering department

Addis Ababa Institute of Technology

Addis Ababa, Ethiopia

abm26t11@gmail.com

Sosina Mengstu

Electrical and Computer Engineering department

Addis Ababa Institute of Technology

Addis Ababa, Ethiopia

lesosinamg@gmail.com

Abstract*—Because of advancements in network technology, Internet traffic has increased rapidly in recent years. Furthermore, the availability of broadband, such as those based on Asymmetric Digital Subscriber Line (ADSL) and Fiber technologies, is constantly increasing. This enormous demand is the result of technological advancement, a rise in smartphone users, and the availability of new services by Internet service providers (ISPs). With the growing demand for broadband Internet access, ISPs must improve to satisfy consumer quality of service demands. In a packet-switched network Internet Protocol version 4 (IPv4) addresses are used for connecting network devices. All regional Internet registers have exhausted their IPv4 address pools. For solving IPv4 depletion problems several mechanisms were developed. One of the mechanisms used for solving the problem is Carrier-Grade NAT (CGN). Even though CGN provides a solution for IPv4 address depletion by sharing a single routable IPv4 address for many private IPv4 addresses, it has some problems. CGN does not allow some applications to work seamlessly. It also breaks the end-to-end working principle of Internet traffic. CGN process is one of the causes for customers Internet service delay which result in customers speed problem. The purpose of this paper is to examine the influence of CGN on network performance and offer a technique for enhancing network performance, as well as to assess the effects of adopting CGN in conjunction with network traffic prioritization on network performance. The evaluations are performed on two scenarios: one with regular CGN and the other is CGN combined with network traffic prioritization. For implementing and evaluating network performance of the scenarios, simulation tools such as Graphical Network Simulator-3 (GNS3), Ostinato traffic generator, and monitoring tool Paessler Router Traffic Grapher (PRTG) are employed. Network performance metrics such as packet loss, latency, and throughput are used for analyzing the performance of the two scenarios. The analysis result indicates that in CGN combined with traffic prioritization throughput is improved on average by 19%, latency is improved by 21%, and packet loss is improved by 20%.

Keywords—CGN, IPv4, Performance Metrics, Traffic Prioritization, Address Depletion

I. INTRODUCTION

In recent years, Internet traffic increased very rapidly due to network technology growth [1]. Additionally, the availability of broadband user connections is continually growing; such as those based on Asymmetric Digital Subscriber Line (ADSL) and Fiber technologies. With such availability and increased quality of service, users are more attracted to the use of different services provided by the Internet. Internet Service Provider (ISP) plays a huge role in managing and supervision this dynamic Internet traffic. the Internet traffic uses an IPv4 address for connecting devices to the web. IPv4 addresses are classified into two types. Private IPv4 addresses and public IPv4 addresses are the two types of IPv4 addresses. Private IPv4 addresses are used for local area networks, whereas public IPv4 addresses are used to connect local area networks to the Internet. Customers with private IPs are required to translate their private IPv4 address to a public IPv4 address using Network address translation (NAT) protocol to send their traffic to the Internet.

Carrier-Grade NAT (CGN) is a feasible solution to the IPv4 address consummation problem and offers a way for service provider subscribers and content providers to implement a seamless transition to Internet Protocol Version 6 (IPv6). CGN engages Network Address and Port Translation (NAPT) methods to aggregate many private IP addresses into fewer public IPv4 addresses [2].

Quality of service metrics such as latency, throughput, delay, and jitter is used to assess network performance [3]. Key Performance Indicators (KPIs) for monitoring network device performance comprise CPU utilization, memory utilization, bandwidth utilization, and device temperature. These metrics assess the performance of a particular packet-based network. The NAT device whose sole responsibility is performing IPv4 address translation is referred to as the Standalone Network Address Translation (SL-NAT) router. NAT devices are commonly installed at network edges to modify the addresses of packets crossing the NAT [2]. SL-NAT devices let the networks on either side of the NAT use their own address space, conserve network addresses, and translate between IPv4 address formats (i.e. private IPv4 and public IPv4).

II. RELATED WORK

One of the methods used for IPv4 address extension technology is NAT444, which helps the ISPs to give IPv4 service to their customers. This technology adds extra CGN in the service provider network [4]. CGN helps IPv4 continuity even if it has some impact on some Internet applications. The authors in [4] show the CGN impact on some Internet applications like peer-to-peer, online gaming, large file transfer, etc. The paper in [5] investigated the impact of CGN on user traffic. TCP flow between client and server is used for checking the impact of CGN on web browsing. The result in [5] shows CGN has less impact on web browsing even though CGN breaks the end-to-end Internet connection principle. Both papers showed the impact CGN has on some applications but they didn't consider the impact it has on network performance, this paper tries to fill the gap by analyzing the impact CGN has on network performance.

The authors in [6] evaluated the potential impact of CGN on residential broadband users. The limiting factors which

define the performance of a CGN device are the rate at which entries are created and removed from the session table and the total size of the session table [6]. Since table size and port numbers are finite, efficient provisioning of CGN operation is the main concern of ISPs. The number of subscribers who share the same CGN device without experiencing a significant degradation in performance is one of the factors with CGN operation [6]. Some short-lived UDP sessions needed to be expired as fast as possible for efficient use of the session table [6]. The authors in [7] show CGN deployment becoming rapidly increased. They also discovered there is a lack of a common standard on operating CGN which results in a lack of information on the impact of CGN on performance which this paper tries to explore.

Delay is one of the most performance characteristic indicators in IP networks [8]. Transmission delays [8] are an inherent problem of communication quality. The authors in [8] show the impact of NAT operation delay in IP networks. NAPT operation induced a 175% delay of the original value. This is due to IPv4 and port header values will be replaced with new IPv4 address and port number values during NAT operation. The authors in [9] tested the impact of CGN by creating scenarios like bypass, routing, and mapping mode of CGN configuration to analyze the effect CGN has on delay. The mapping mode in [9] indicates the CGN device with the NAT functions and routing mode is the CGN device acting like a normal router that routes traffic whereas bypass mode indicated the CGN device is bypassed. The result in [9] shows on mapping mode there is an increase in a delay to establishes a connection.

The authors in [10] illustrate the scope of CGN deployment on the Internet and its effect on end-users. Based on the survey [10] they circulated to ISPs, they find out that out of 75 ISPs 40% of them deployed CGN. Next, they tried to find out the impact of CGN on subscribers' Internet applications and the primary concern was some applications such as online gaming did not work seamlessly with CGN set up [10]. Besides, the operators' concern was the lack of well-developed best practices when setting up and operating CGN [10]. Operators need to resort to experimentation on aspects such as the distribution of external IP addresses and port ranges to customers to efficiently operate CGN [10].

I. METHODOLOGY

A. Problem Statement

There is a huge demand for broadband data services. This huge demand is due to technological advancement, an increase in smartphone users, and the availability of new services by ISPs. Since Ethio telecom decreased the monthly payment for fixed broadband, there is huge customer demand. Besides tariff adjustment coronavirus also contributes to the demand for fixed broadband service. With this increasing demand for broadband Internet service, improvement is needed from ISP to meet customer quality of service demand. With the depletion of public IPv4 addresses, ISPs are trying to find a different solution for addressing the issue. One of the ways to solve the problems is using the CGN approach. CGN is capable of translating many customers' private IPv4 addresses to a single routable public IPv4 address. CGN has a negative impact on some applications to work seamlessly. It also breaks the end-to-end working principle of Internet traffic. Since CGN handles millions of translations the operation needs to be efficient to provide a required QoS for customers. Operators in [10] voiced concerns about a lack of well-developed best practices for configuring and dimensioning CGN. When it comes to configuring CGN, respondents in [10] named maintaining the port space, as well as the amount of session state in CGNs, are the primary challenges when configuring CGNs. The NAT process is one of the causes for customers' Internet service delay which result in customers speed problem. To address this problem, the CGN process needs to be coupled with network traffic prioritization to improve QoS performance.

The working principle of CGN is handling every traffic that comes to be NATed equally, which makes the operation to be inefficient. This leads to poor network performance. When CGN is coupled with network traffic prioritization, there will be traffic classification or marking. Network traffic prioritization helps to treat different customer traffic based on their delay sensitivity and based on customer type (enterprise or residential). The enterprise customers which provided sensitive service to their customer needs a higher priority than those customers using the Internet for basic browsing and emailing. Since the CGN treats all traffic equally, the limited resource for CGN operation needs to be managed effectively so that customers' quality of experience will be improved when CGN is coupled with network traffic. So implementing CGN plus network traffic prioritization will increase the network performance of the CGN process by efficiently use the CGN limited resources. This paper aims to investigate the improvement of network performance by using CGN coupled with network traffic prioritization.

B. CGN Operation

When the Internet was introduced, network devices could communicate seamlessly. Devices that communicate with each other on Internet use unique IP addresses. The address developed for communication between network devices is an IPv4 address. That is capable of providing 4.3 billion unique addresses. But the popularity and acceptance of the Internet have become huge, which leads to the depletion of public IPv4 addresses. A different solution was devised to resolve the exhaustion problem of public IPv4 addresses. One of the solutions is to use the network address translation technique. NAT allows the use of a globally routable address to represent local private addresses. The router that performs the address translation is located at the customer's site. However, with the NAT technique, IPv4 exhaustion continues to evolve. So, to solve this problem, ISPs started to apply address translation in their core network. The translation mechanism adopted by ISPs is called CGN. CGN can translate millions of individual customer addresses into a single globally routable IPv4 address, which helps the depletion of the IPv4 addresses.

CGN is a form of IPv4 addresses and port numbers translation technique done on a huge scale. It is done by ISP's backbone network. It is capable of translating many customers' private IPv4s to a single public IPv4 for minimizing the storage IP addresses. CGN implementation at the ISP generates different problems for the ISP as well as for the customer [11]. A CGNAT router divided the network into two sub-networks, RFC6598 private network, where the customers are assigned with the range private IPv4 addresses, and a public network, where the CGNAT router

connect the customers' private network to the Internet with a public IPv4 address [12].

The general operation of a CGNAT-enabled router is a combination of port numbers and public IPv4 addresses. When a CGN router received a packet sent to the external network from the customers, the CGNAT process will look for a matching source address and source port combination in the session mapping table. The session mapping table contains a list of internal local port numbers for conversion to specific mapped ports. If the received source port is not found on the session table, a new port is created. Then the new port is assigned to the received source port. The customer's private source IP address is replaced with the globally routable address in the CGNAT address pool. The source port is replaced with the port in the port mapping table. Finally, the packet is forwarded with the converted source IPv4 address and source port number [13].

A filtering process will be performed for the data packets that are received from the external network against the session mapping table. Then after it will be sent to the internal network after undergoing inverse conversion. The first step is looking at the destination port number in the port mapping table. If the session entry is located, then the destination port number will be replaced with the port number in the mapping table. The destination address will be replaced with the private IPv4 address in the translation table. If there is no mapping existed for destination IPv4 and destination port, then the router will decide that the packet does not belong to its internal network and proceed to dropping the packet [13].

For CGN operation to perform the address translation task, different protocols undergo a modification since IPv4 address and port number are changed from one type to another. RFC3022 states that, to reflect the change made on IP address and port number protocol associated with them need to be modified. The protocols which are associated with the address and port number are IP, Transmission Control Protocol (TCP), User Datagram Protocol (UDP), and Internet Control Message Protocol (ICMP) [12]. The modification of each protocol is done packet-by-packet basis. Each packet involved in a translation undergoes modification. IPv4 packets are consist of two parts (i.e. header and data). The process that packets undergo in IPv4 and port modification makes the CGN process resource-intensive. CPU and Memory processes are the main hardware component of the CGN device that is affected by huge address translation. During address translation, the header part of the protocols is modified for each packet that passes through address translation.

Some applications which have a port or IP address information on their payloads like file transfer protocol, IPsec, and IPv6 tunneling will fail to work in the CGN environment. For those applications to work behind a CGN operation they need the help of transversal techniques. The investigation carried out by some researchers [5][7] shows that some applications will face performance degradation under CGN. Applications like online gaming, peer-to-peer connection, real-time application, remote access service, AJAX applications (i.e. Google Maps, torrent), and video on demand show a performance issue were us basic Internet browsing like checking email works without a problem [5][7].

A. Traffic Prioritization

The aim of network traffic classification based on its priority is to efficiently provide the user with good service quality. The traffic prioritizing mechanism on devices aims to optimize the routing policy of packet forwarding among devices. Traffic prioritizing means network elements (such as switch or router) can ensure to some degree that their service flows and service requirements can be met. Traffic prioritization manages performance and directs traffic based on values set according to demands of applications and network status. IP traffic prioritization effect on performance parameters can be measured using packet loss, latency, throughput, availability, and jitter. Network devices (such as routers and switches) separate the traffic by analyzing the packets as they process them and classify them based on the class of service (CoS) traffic role into groups of different streams [14].

Based on the Internet Engineering Task Force (IETF) definition, two models are used for classifying network traffic. The Integrated Services (IntServ) and the Differentiated Services (DiffServ) are the two well-known models used for classifying traffic based on a certain requirement to achieve a better latency, packet loss, and throughput. IntServ follows the per-flow method for providing a guarantee to individual traffic, whereas DiffServ gives a collective assurance for a class of networks [14]. IntServ requests the network device for an explicit resource reservation per flow. Network nodes follow all the flows passing the nodes by checking whether new data packets belong to an existing flow or not. The packet will be accepted if there are enough network resources available otherwise the packet will be dropped. In the case of DiffServ, network nodes use a class-based classification process, which differentiates traffic into CoS based on the required priority. DiffServ has three kinds of options for providing different services. Expedited Forwarding (EF), Assured Forwarding (AF) and Best-effort (BE) forwarding are the types of DiffServ models that give the generated traffic different priority based on the requirement set by the administrator. DSCP provides us with an increased number of labeling bits to 6 bits, that produce 64 unique values that can be labeled for different types of traffic as well as priority allocation [15].

The IPv4 address assigned a three-bit ToS byte in its header for classifying or prioritize traffic, as shown in Figure.1. RFC 791 allocates a 3-bit IP Precedence (IPP) field for traffic priority marking. IPP provides us eight separate binary values starting from 000, 001, 010, up to 111 with their equivalent decimal values 0 through 7. While IPP provided us with eight different values to label, the later RFC2474 redefined the ToS byte with the Differentiated Services Code Point (DSCP) field [15].

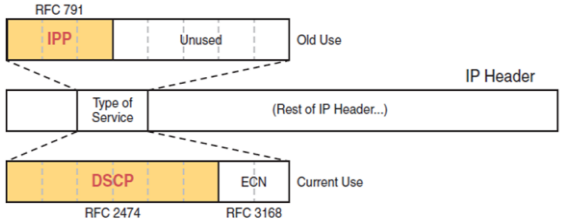


Fig. 1. IP Precedence and DSCP Fields

A. *Network Performance Metrics*

For any network setup, data transmission is the main goal. Exchange of data in text, audio, or video format will have a source and destinations. This data format(text/audio) has a small unit known as packets. To investigate how the network performance of a certain network is improved different metrics are required. The network performance metrics are used to indicate the performance of certain network scenarios. For instance, the end-to-end time taken for a packet to arrive at its destination is measured using latency.

According to RFC 7679, latency is a measure of one-way delay experienced by data packets as the difference in the time from an ingress point to an egress point across an ISP's network. The network congestion and network device involved in the IP packet flow in a traffic path can contribute to the delay experienced by the data packet. In packet-based networks, one-way end-to-end delay is one of the key performance parameters which indicates the quality and best design practice of the ISP network. Today's Internet forwarding and reverse traffic takes a different path during packet transmission, and due to that, the round trip delay will be involved with different contributors like TCP behavior, congestion behavior; and the police and routing setup will result in different performance. So, it is preferable to measure the one-way delay from the source to the destination due to the packet experience different performance characteristics in symmetric and asymmetric traffic flows. In TCP/IP networks, there are four sources of delay. (i.e., processing delay, queuing delay, propagation delay, and transmission delay). Equation 1 shows the formula for calculating latency.

$$\text{Total Delay} = ND + QD + PD + TD \text{ [second]} \quad (1)$$

Jitter which is measured in milliseconds is the variation in time delay when a data packet is transmitted and received in a packet-switched network connection. RFC 3393 described the one-way-delay variation metric for IP packets across the network which recommended measuring the one-way travel since the behavior of forward and reverse-path are different. The difference in time delay is known as IP packet delay variation. Undesirable glitches, like flickering monitors and untimely data transmissions between network devices, are some effects of jitter. The cause of jitter can be network device performance, network congestion, less bandwidth, and no data packet prioritization. Real-time traffic like voice-over IP is the most affected service by jitter. The different types of jitter are constant jitter, transient jitter, and short-term jitter. If two consecutive data packets leave the source network device with timestamps t1 and t2; and are arrived at the destination network device at times t3 and t4, then jitter can be calculated as (t4 - t3) - (t2 - t1) [16].

When some packets of data are unable to reach the intended destination in a data network is referred to as packet loss. Transport layer protocols play a different role during packet loss. For the case of TCP traffic, when there is a packet loss there will be retransmission, on the other hand, in the UDP case, there will be no retransmission even if there is a packet loss. That is the main reason real-time traffic such as voice over IP, live streaming, and video calls are the most affected services.

Throughput in a data network is the actual data transfer rate from sender to receiver measured at a particular timeframe. Speed in a data network is one of the key performance indicators which tells us how fast is the data transferred from source to destination. Throughput and bandwidth are the parameters give as the information regarding network transfer speed for transmitting a certain file from source to destination. Throughput is measured in bit per second. Some factors which affect the network throughput performance are network nodes, network design, and the number of users [17].

B. *System Model*

Simulation implementation will be used as a primary research method. Graphical Network Simulator (GNS3) along with traffic generator and network monitoring tools will be used to show the stated hypothesis. The desired network topology with all the required network protocols will be configured to achieve the objective. Conducting researches on networks using real-time scenarios is very challenged and difficult. It also required a significant amount of time and cost to set up a testbed. A Network simulator plays a very significant role in emulating the real-time scenario; and also saves money and time. Network simulations are becoming the main source of performing different network academic researches and verification methodology for different network hypotheses. There are various types of network simulators available nowadays with different features like GNS3, Enterprise Network Simulation Platform, Emulated Virtual Environment-Next Generation, Optimized Network Engineering Tools, Network Simulator Version 2 and Version 3, etc.

To investigate the improvement of CGN deployment with Traffic Prioritization (TP) tools such as GNS3 for topology setup, Ostinato for traffic generation, and PRTG combined with IP-SLA for performance data retrieval are used. Figure 2 depicts the network model used to perform the two scenarios (standard CGN operation and CGN paired with TP operation). Several simulations were run to investigate the impact of CGN with TP on network performance. To compare the performance of the two scenarios, network performance parameters such as latency, throughput, and packet loss are used. Ostinato generates two traffic figures to obtain a reliable result from the simulations. Various protocols are chosen for traffic generation using Ostinato. TCP, UDP, and ICMP are the traffic streams involved in generating the requested traffics.

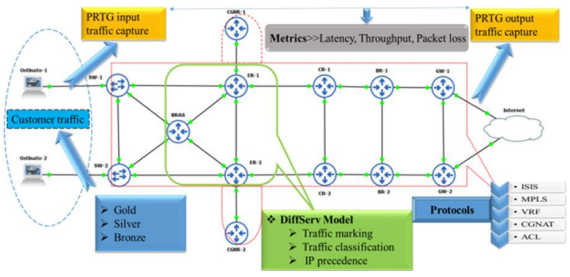


Fig. 2. Network Model

Three class C networks are used to simulate the customer's traffic. In the first scenario, standard deployment CGN with no traffic differentiation is used. In this case, each host has equal access to the CGN's limited resources. The customer's traffic is prioritized into three categories in the second scenario. Depending on the traffic classification, CGN deployment takes a different approach. Three IP

precedence values are used to distinguish customer traffic priority for scenario two. Gold traffic assigned critical priority, while silver and bronze traffic assigned flash-override and flash IP precedence values, respectively. To relate to this traffic marking, the CGN was set up with three VRF. The edge router is used to label traffic and deliver it to CGN through three sub-interfaces.

VRFs in CGN combined with TP operation are configured with a different number of session entries for each traffic priority. The gold traffic configuration uses 15000 maximum-session entries, the silver traffic configuration uses 35000 maximum-session entries, and the bronze traffic configuration uses 40000 maximum-session entries. 150 gold customers, 225 silver customers, and 375 bronze customers are used to generate traffic. To effectively use translation session entries, various timeout configurations were used for the protocols. For dynamic NAT translation entries, the default timeout values are 24 hours for TCP connections and 5 minutes for UDP connections. To make the most of port numbers and session creation, idle connection waiting time must be managed. TCP connections are set to timeouts of 1500 seconds, 1000 seconds, and 500 seconds for gold, silver, and bronze traffic, respectively. For gold, silver, and bronze traffic, the NAT translation timeout for the UDP protocol was set to 90 seconds, 60 seconds, and 30 seconds, accordingly. The simulation results are used to test the hypothesis based on the network performance metrics chosen. Table 5.1 depicts the simulation's implementation strategy.

TABLE 1 SIMULATION PLAN

Simulation Scenarios Scheme			
No. of Simulation	**Traffic Generated**	**Scenario**	**Duration**
Simulation one	550Kbps	CGN	60 Minutes
Simulation two	550Kbps	CGN+TP	60 Minutes
Simulation three	850Kbps	CGN	60 Minutes
Simulation four	850Kbps	CGN+TP	60 Minutes

I. Result and Discussion

Throughput is used for measuring how a certain network performs by providing the actual traffic rate passing through a medium at a certain time. A method used to measure throughput for certain traffic that passes through CGNAT operation is done by sending a packet at a certain rate from a packet generator. Then the generated packets are captured after they pass through CGN operation using a monitoring tool. Finally, the comparison will be done to see the throughput.

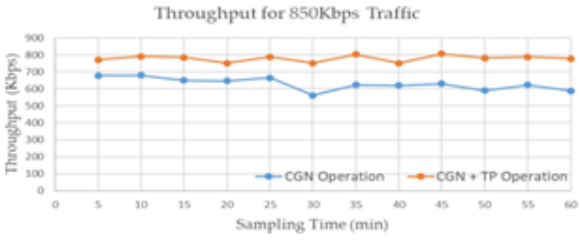


Fig. 3. Graph of throughput for scenarios 1 and 2 of 850kbps traffic

Figure 3 displays the throughput results for Scenarios 1 and 2. For CGN integration with TP operation, the average throughput is 778.52Kbps. The average throughput for standard CGN operation is 630.12Kbps. CGN with TP operation improves throughput by 19% on average when compared to CGN operation. The rate at which the CGN router transmits a packet from the inbound interface to the outgoing interface determines CGN throughput. The CGN device translates the IP address and port number of each packet and maintains the session to filter returning traffic. Changes to IPv4 addresses cause changes to the checksum header, which adds to processing time.

The time taken for the packet to reach its destination is used as one of the network performance parameters. The time taken for packet transmission is referred to as delay or latency. The maximum one-way end-to-end latency which is recommended by ITU is 150ms. The same simulation setup is used for measuring the latency of the two scenarios. For a stable and quality packet transmission, smooth CGN functioning is required. PRTG is used to monitor network performance in real-time. As a fundamental monitoring element, PRTG employs a sensor, which provides a measured output for the required network performance statistic. The Internet Control Message Protocol (ICMP) ping test is used by IP SLA to determine how long it takes a packet to reach its destination. PRTG monitoring software has a Cisco IP SLA sensor for measuring network latency.

Figure 4 shows the result of latency for the two scenarios for 850Kbps traffic. As observed from the graph the CGN operation has higher latency on average when compared with CGN with TP setup. The average latency for CGN operation is 122.3msec. CGN with TP configuration reduced latency by 21% when compared to conventional CGN operation.

Fig. 4. Graph of latency for scenarios 1 and 2 of 850kbps traffic

Using a variety of payloads, IP-SLA generates packet loss data. IP-SLA is linked with PRTG to obtain the packet loss result. To achieve a good result, two traffics are generated. Packet loss was used to determine whether packets arrived at their destination successfully. Using PRTG and an IP-SLA sensor, the ICMP test is used to determine packet loss for the two scenarios.

Fig. 5. Graph of packet loss for scenarios 1 and 2 of 850kbps traffic

As shown in figure 5 the CGN with TP operation has less packet loss. On average the CGN Operation with TP has improved the packet loss by 20%.

In general, the operation of standard CGN needs to maintain the translation state in the session table for a longer time for every active traffic to make the inbound to outbound and outbound to inbound traffic flow stable. This leads to more session creation for the new subscriber traffic. Since there is no NAT entry limitation is configured, the increasing number of translation entries leads to high memory and CPU usage. Due to that, the performance of CGN is affected which leads to high latency, high packet loss, and less throughput. Whereas in the case of CGN paired with TP operation there is session management put in place that controls the session limit and expires the idle connection quickly to avail the NAT entry for the new session request.

I. Conclusion and Future Work

The influence of deploying CGN combined with TP architecture operation on network performance is analyzed in this documentation in contrast to conventional CGN operation. This study looked at the influence of several configuration setups on CGN with TP operation and regular CGN operation. The translation entry creates-rate, translations entry delate-rate, inside to outside forwarding rate, and outside to inside forwarding rate all contribute to evaluating the performance CGN coupled with TP and standard CGN operation.

To assess and compare the deployment of conventional CGN operation and CGN combined with TP operation, network assessment methods such as latency, pack loss, and throughput are employed. GNS3 was used to build the two network scenarios along with all the required protocol configurations. An ostinato traffic generator is used to create client traffic after the simulation has been set up. The results of the network performance measurements from the two scenarios are then obtained using PRTG. Finally, the outcome is shown graphically. The following conclusions are inferred from the simulation results:

- CGN paired with TP can improve network performance by optimally use the limited resource of CGN operation.
- CGN operation has an impact on network performance.
- TP integration with CGN improves the latency by 21% on average when compared with standard CGN operation.
- When comparing CGN paired with TP operation to conventional CGN operation, throughput improves by 19% on average.
- Both scenarios show there is an impact on network performance by CGN deployment where the impact is minimized in TP with CGN integration.
- Standard CGN operation shows greater packet loss than CGN integration with TP setup. CGN with TP integration shows reduced packet loss with an average value of 20%.

This paper has conducted an investigation to achieve the suggested solution for improving network performance with CGN deployment. However, there are some issues needed to be addressed in the future. Study the impact on network performance by enabling application layer gateway and port control protocol support on CGN devices to allow application function seamlessly. Analyze the improvement of CGN performance by deploying a dual-stack backbone. Where IPv6 and IPv4 services are operating together. Implement and analyze software-defined networking along with the CGN operation to further enhance the network performance.

References

[1] Cisco, “Cisco Annual Internet Report (2018–2023),” white paper, pp. 1–41, 2020.

[2] N. Skoberne, O. Maennel, I. Phillips, R. Bush, J. Zorz, and M. Ciglaric, “IPv4 address sharing mechanism classification and tradeoff analysis,” IEEE/ACM Transactions on Networking, vol. 22, no. 2, pp. 391–404, 2016.

[3] W. Sugeng, J. Istiyanto, K. Mustofa, and A. Ashari, “The Impact of QoS Changes towards Network Performance,” International Journal of Computer Networks and Communications Security, vol. 3, no. 2, pp. 48–53, 2015.

[4] A. Greca, Y. Li, and S. Cho, “Network survivability management system design for broadband networks,” 2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), vol. 1, 2017.

[5] E. Bocchi, A. Ktahouni, S. Traverso, et al., “Impact of Carrier-Grade NAT on web browsing,” IWCMC 2015 - 11th International Wireless Communications and Mobile Computing Conference, pp. 532–537, 2015.

[6] S. Alcock, R. Nelson, and D. Miles, “Investigating the Impact of Service Provider NAT on Residential Broadband Users,” Computer Communications Networks, pp. 4–9, 2014.

[7] I. Livadariu, K. Benson, A. Elmokashfi, A. Dhamdhere, and A. Dainotti, “Inferring Carrier-Grade NAT Deployment in the Wild,” Proceedings - IEEE INFOCOM, vol. 2018-April, pp. 2249–2257, 2018.

[8] D. Samociuk, B. Adamczyk, and A. Chydzinski, “Impact of Router Security and Address Translation Mechanisms on the Transmission Delay,” INTERNET 2015: The Seventh International Conference on Evolving Internet Figure, pp. 1–5, 2015.

[9] Y. Ohara, K. Nishizyka, K. Chinen, et al., “On the impact of mobile network delays on connection establishment performance of a carrier-grade NAT device,” Asian Internet Engineering Conference, AINTEC 2014, pp. 1–8, 2014.

[10] P. Richter, F. Wohlfart, N. Vallina-Rodriguez, M. Allman, et al., “A Multi-perspective Analysis of Carrier-Grade NAT deployment,” Proceedings of the 2016 Internet Measurement Conference, pp. 215–229, 2016.

[11] O. Maennel, R. Bush, L. Cittadini, and S. Bellovin, “A Better Approach than Carrier-Grade-NAT,” Technical Report CUCS, 2008.

[12] S. Mahmoud, H. Sqalli and N. Sakthi Priya, “A scalable NAT-based solution to Internet access denial by higher-tier ISPs,” Security and Communication Networks, vol. 5, no. June, pp. 422–437, 2012.

[13] Cisco, “How Can Service Providers Face IPv4 Address Exhaustion? - Cisco,” no. June, pp. 1–28, 2012.

[14] C. Feerick, “Learn About Quality of Service (QoS) Why QoS is Essential in Today’s Networks,” Juniper Networks, vol. 1, 2015.

[15] O. Wendell, "CCNA 200-301 Official Cert Guide", Cisco Press, 2020.

[16] C. Williams “The Impact of Advanced QoS Mechanisms on the Performance of Converged Networks" A Dissertation, Glasgow Caledonian, August 2019.

[17] B. Vellambi, N. Torabkhani, and F. Fekri, “Throughput and latency in finite-buffer line networks,” IEEE Transactions on Information Theory, vol. 57, no. 6, pp. 3622–3643, 2011.