COLLABORATIVE-BASED APPROACH UTILIZING ENSEMBLE FEATURE SELECTION FOR DETECTING HTTP-GET DDOS ATTACKS IN CLOUD COMPUTING ENVIRONMENTS

ZIYAD REEFAT HAMZEH AL ASHHAB

UNIVERSITI SAINS MALAYSIA

2025

COLLABORATIVE-BASED APPROACH UTILIZING ENSEMBLE FEATURE SELECTION FOR DETECTING HTTP-GET DDOS ATTACKS IN CLOUD COMPUTING ENVIRONMENTS

by

ZIYAD REEFAT HAMZEH AL ASHHAB

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosphy

May 2025

ACKNOWLEDGEMENT

بِسمِ ٱللَّهِ ٱلرَّحمَٰنِ ٱلرَّحِيمِ

With the name of Allah, Most Gracious, Most Merciful

﴿ نَرْفَعُ دَرَجَاتٍ مَّن نَّشَاء وَفَوْقَ كُلِّ ذِي عِلْمٍ عَلِيمٌ ﴾ [٦٧ ، سورة يوسف]

First and foremost, gratitude is expressed to Allah, the Creator, for bestowing blessings and providing the opportunity to achieve a PhD. Acknowledgment is also extended to the numerous individuals who have contributed, either directly or indirectly, to the academic journey over the years.

Special thanks are offered to Professor Dr. Selvakumar A/L Manickam (main supervisor), Associate Professor Dr. Manmeet Mahinderjit Singh (co-supervisor), and Dr. Mohammed F. Anbar (field supervisor) for their unwavering support, innovative ideas, and invaluable guidance throughout the PhD programme at the Cybersecurity Research Centre (CYRES), USM, a prestigious institution. The research committee members are also acknowledged for their constructive and insightful feedback.

Furthermore, Profound gratitude is extended to Reefat and Hayat (may Allah rest her soul) for their unwavering belief and encouragement in pursuing academic aspirations. Their guidance has fostered the determination and resilience needed to overcome challenges. Sincere thanks are also offered to the siblings for their support. Lastly, it is hoped that this study will serve as a meaningful contribution dedicated to future generations, particularly the children.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

CCE	Cloud Computing Environment
ICT	Information and Communication Technology
DDoS	Distributed Denial of Service
TCP	Transmission Control Protocol
ICMP	Internet Control Message Protocol
WWW	World Wide Web
IT	Information Technology
FBI	Federal Bureau of Investigation
ISP	Internet Service Provider
CSP	Cloud Service Provider
DoS	Denial of Service
SLA	Service Level Agreements
OSI	Open Systems Interconnection
VM	Virtual Machine
VMM	Virtual Machine Monitor
SaaS	Software as a Service
PaaS	Application as a Service
IaaS	Infrastructure as a Service
OS	Operating Systems
AWS	Amazon Web Services
NIST	National Institute of Standards and Technology
CIA	Confidentiality, Integrity, and Availability

ML	Machine Learning
AI	Artificial Intelligence
DL	Deep Learning
IP	Internet Protocol
MIME	Multipurpose Internet Mail Extensions
RNN	Recurrent Neural Networks
CPU	Central Processing Unit
UDP	User Datagram Protocol
ICMP	Internet Control Message Protocol
EPDoS	Economic Permanently Denial of Service
PDoS	Permanent Denial of Service
CSP	Cloud Service Provider
ACK	Acknowledgement
PoD	Ping of Death
CSP	Cloud Service Provider
FRC	Fraudulent Resource Consumption
HTTPD	Apache Hypertext Transfer Protocol Server
HTTP	Hypertext Transfer Protocol
WWW	World Wide Web
RfCs	Requests for Comments
HTML	Hypertext Markup Language
URL	Uniform Resource Locator
URI	Uniform Resource Identifier
CLF	Common Log Format
MS	Microsoft

RAM	Random Access Memory
HDFS	Hadoop Distributed File System
JVM	Java Virtual Machine
HA	High Availability
FT	Fault Tolerant
DFS	Distributed File System
IG	Information Gain
GainR	Gain Ratio
PCA	Principal Component Analysis
Chi	Chi-Squared
GRU	Gated Recurrent Units
LSTM	Long Short-Term Memory
FRC	Fraudulent Resource Consumption
DNNs	Deep Neural Networks
ANN	Artificial Neural Network
SVM	Support Victor Machine
RNN	Recurrent Neural Networks
EDoS	Economic Denial of Sustainability
HDFS	Hadoop Distributed File System
EFS	Ensemble Features Selection
Enfs	Enrichment Features Set
BFS	Basic Features Set
IG	Information Gain
GainR	Gain Ratio
PCA	Principal Component Analysis

AI	Artificial Intelligence
ANNs	Artificial Neural Networks
VM	Virtual Machine
AVWS	Apache Virtual Web Server
V-CPU	Virtual Central Processing Unit
V-RAM	Virtual Random Access Memory
GRPS	GET Request Per Second
FPR	False-Positive Rate
DA	Detection Accuracy
DV	Dataset Validation
C C	Command and Control
ECLTAS	Extract, Cleansing, Labeling, Transformation, Aggregation, and Store Data

PENDEKATAN BERASASKAN KOLABORATIF YANG MENGGUNAKAN PEMILIHAN CIRI GABUNGAN UNTUK MENGESAN SERANGAN DDOS HTTP-GET BANJIRAN DALAM PERSEKITARAN KOMPUTASI AWAN

ABSTRAK

Perkhidmatan berasaskan Persekitaran Pengkomputeran Awan (Cloud Computing Environment, CCE) merupakan suatu pendekatan baharu dan alternatif untuk pengurusan perniagaan secara jarak jauh. Salah satu kelebihan utama menggunakan CCE ialah ketersediaan perkhidmatan atas permintaan, yang memungkinkan wujudnya model pembayaran setiap penggunaan. Ini menjadikan teknologi CCE sebagai suatu kaedah yang memudahkan perkhidmatan melalui Internet. Walau bagaimanapun, kerentanan keselamatan, seperti serangan nafi khidmat teragih (DDoS), terutamanya serangan HTTP-GET DDoS pada lapisan aplikasi, mengancam ketersediaan perkhidmatan dalam CCE. Tesis ini mencadangkan pendekatan kolaboratif menggunakan pemilihan ciri ensemble untuk mengesan serangan HTTP-GET DDoS dalam CCE. Pendekatan ini melibatkan enam fasa. Fasa pertama ialah pengumpulan dan pra-pemprosesan data, yang bertujuan mengumpul dan memproses data dari pelbagai sumber. Fasa kedua melibatkan penjanaan set data, iaitu penciptaan set data sintetik khusus CCE. Fasa ketiga menfokuskan pengayaan ciri, bertujuan untuk menambah ciri baharu ke dalam log akses AVWS yang diekstrak daripada log aktiviti VM dan log sumber untuk meningkatkan pengesanan serangan HTTP-GET DDoS. Fasa keempat melibatkan pengesahan set data, bertujuan untuk memastikan kesahihan dan kesediaan set data serta mengesahkan bahawa ia memenuhi keperluan set data penanda aras. Fasa kelima melibatkan pemilihan ciri ensemble, bertujuan untuk memilih set ciri yang paling penting dan minima yang menyumbang kepada pengesanan serangan HTTP-GET DDoS. Fasa keenam bertujuan untuk membina model pengesanan pembelajaran mendalam berdasarkan Memori Jangka-Pendek Panjang (LSTM) untuk mengesan serangan HTTP-GET DDoS pada CCE dengan tepat. Pendekatan ini dinilai menggunakan set data CCE yang dihasilkan sendiri, mencapai ketepatan pengesanan yang tinggi (96.26%–99.83%), kadar positif palsu yang rendah (0.003%–0.042%), serta skor kejituan, recall, dan F1 yang amat baik

(0.963–0.998). Analisis kualitatif mengesahkan kesesuaian set data sebagai penanda aras, mengatasi set-set data sedia ada.

COLLABORATIVE-BASED APPROACH UTILIZING ENSEMBLE FEATURE SELECTION FOR DETECTING HTTP-GET DDOS ATTACKS IN CLOUD COMPUTING ENVIRONMENTS

ABSTRACT

Cloud Computing Environment (CCE)-based services present a novel paradigm for remote business management. One of the primary advantages of utilizing CCE is the availability of on-demand services, thereby facilitating a pay-per-use model. This makes CCE technology a convenient means of facilitating services over the Internet. However, security vulnerabilities, such as Distributed Denial of Service (DDoS) attacks, particularly HTTP-GET DDoS attacks at the application layer, pose a significant threat to service availability in CCEs. This thesis proposes a collaborative approach utilizing ensemble feature selection to detect HTTP-GET DDoS attacks in CCEs. The proposed approach comprises six phases. The first phase entails data gathering and pre-processing, responsible for collecting and processing data from multiple sources. The second phase involves dataset generation, comprising the creation of a synthetic CCE-specific dataset. The third phase focuses on feature enrichment, aiming to augment the AVWS access log extracted from VM activity and resource logs to enhance the detection of HTTP-GET DDoS attacks. The fourth phase entails dataset validation, aimed at validating the dataset to ensure its validity and readiness, and confirming that it meets the requirements of a benchmark dataset. The fifth phase involves ensemble feature selection, aimed at selecting the most crucial and minimal feature set that contributes to detecting HTTP-GET DDoS attacks. The sixth phase aims to develop a deep learning detection model based on Long short-term memory (LSTM) to detect HTTP-GET DDoS attacks on CCE accurately. Evaluated on a self-generated CCE dataset, the proposed approach achieved high detection accuracy (96.26%–99.83%), low false positive rates (0.003%–0.042%), and strong precision, recall, and F1 scores (0.963–0.998). Qualitative analysis confirmed the dataset's suitability as a benchmark, outperforming existing datasets.

CHAPTER 1

INTRODUCTION

1.1 Overview

Cloud Computing Environments (CCEs) have significantly impacted the Information and Communication Technology (ICT) sector, offering a flexible and cost-effective delivery of computing resources and services to end-users. In contrast to traditional computing models, CCEs provide a scalable and on-demand infrastructure, enabling organisations to rapidly deploy and adapt to changing business needs. The increasing reliance of online services on CCEs is evident across various sectors, including commerce, economy, education, and healthcare. The architecture of CCEs is characterised by the provision of dynamically scalable internet services, enabled by remotely managed computing resources (Furht & Escalante, 2010). This approach encompasses both hardware and software components of data centres, facilitating the operation of web-based applications and system software (Fox et al., 2009; Winans & Brown, 2009).

The growing adoption of CCEs by businesses is driven by their numerous benefits, including rapid development and deployment, improved customer service, scalability, and cost control. However, this shift also raises significant concerns regarding security, availability, and reliability (Sarna, 2010). To fully realise the potential of CCEs, these challenges must be addressed. While CCEs offer numerous advantages, they also present certain challenges and potential risks, including the exploitation of key characteristics such as resource sharing, elasticity, and multi-tenancy by cyber threats (SentinelOne, 2024). Distributed Denial of Service (DDoS) attacks are a particularly significant threat, as they can severely compromise the availability of online services and networks, eroding user confidence and increasing their risk exposure (Cloudflare, 2023).

The increasing prevalence and frequency of Distributed Denial of Service (DDoS) attacks pose a significant economic threat globally, with far-reaching consequences for

businesses and individuals alike. For instance, a cyber attack on a financial institution's website in California resulted in a loss exceeding $900,000, highlighting the devastating impact of such attacks on financial institutions (Harris et al., 2014). Furthermore, a separate incident led to losses exceeding $100 million in online banking transactions (Silva & Karine, 2017). These examples underscore the critical need for robust security measures to protect against such cyber threats.

According to data from AO Kaspersky Lab (Figure 1.1), HTTP was the third most common type of DDoS attack on application layer services in the second and third quarters of 2018 (Kupreev et al., 2018). This finding highlights the significance of HTTP as a major security concern, emphasising the need for an efficient method to accurately detect HTTP-GET DDoS attacks on CCEs. This thesis proposes a novel approach to detect HTTP-GET DDoS attacks on CCEs, aiming to enhance the security of CCEs and protect them from these prevalent and disruptive attacks.

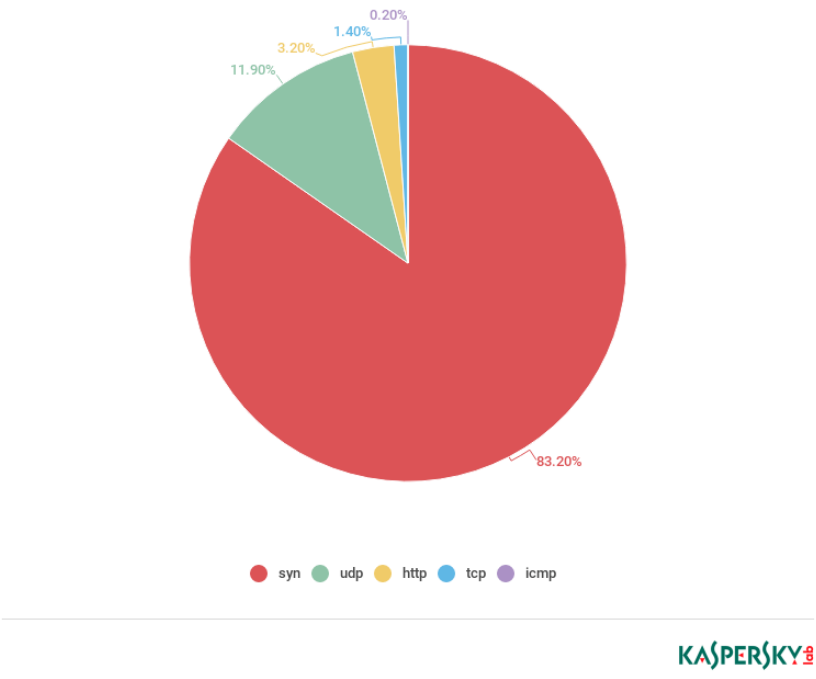

Figure 1.1: DDoS Attacks by Type, Q2 and Q3 2018

1.2 Background

CCEs offer several benefits, including scalability, flexibility, and cost-efficiency. However, these advantages are accompanied by inherent security challenges. The key characteristics of CCEs, such as resource sharing, elasticity, and multi-tenancy, can be exploited by cyber threats, thereby compromising cloud services. Notably, DDoS attacks pose a significant threat to the availability of online services, highlighting the need for robust security measures to mitigate these risks.

1.2.1 Cloud Computing Environment

The CCE is a concept rooted in virtualisation technology. It is a framework in which the resources of a data centre are shared and allocated using advanced virtualisation technology. This framework enables multiple users to share the computational power of various hardware devices. A single physical host may host multiple Virtual Machines (VMs), providing customers with elastic, instant, and on-demand services, with billing structured similarly to utility bills (Catteddu, 2010; Linthicum, 2009). This concept can be visualised graphically, as illustrated in Figure 1.2 (Haltdos, 2018).

Virtualisation technology, which provides a software abstraction layer between the applications running on it, the operating system (OS), and the hardware, is primarily responsible for the widespread deployment of CCEs (Medina & García, 2014). This technology plays a crucial role in the efficient and flexible operation of CCEs.

A hypervisor, or Virtual Machine Monitor (VMM), serves as an abstraction layer and a hardware resource manager. It enables multi-tenancy, allowing multiple OSs to coexist on the same hardware and share resources. This scalability and instant access functionality are key features of CCEs.

However, despite these advantages, valid concerns regarding security gaps, threats, and risks persist. While CCEs represent a significant advancement in IT infrastructure, addressing these security issues remains a critical area of focus. Ongoing research

and development are necessary to enhance the security of CCEs and ensure their safe and efficient operation. This is a crucial aspect of maximising the potential benefits of CCEs.

Figure 1.2: Schematic Definition of CCE

The CCE model involves two key players: the Cloud Service Providers (CSPs), which deliver CCE services, and the CCE users, which include individuals or businesses that utilise the services provided by CSPs. Additionally, the CCE broker and the CCE auditor are two other potential participants in this model (Huang & Nicol, 2013). The development of a CCE model to offer software, infrastructure as a service, hardware, platform, and other services, rather than standalone products, is driven by the stringent requirements of modern applications (Mell & Grance, 2011). These service models are commonly referred to as XaaS, where X denotes the type of service provided (Xu, 2012). As illustrated in Figure 1.3 and defined by the National Institute of Standards and Technology (NIST) Rajkumar Buyya and Venugopal (2008), the three primary types of CCE services are Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) (Jansen, 2011).

The CCE service models are typically categorised into three fundamental classes, as summarised in Table 1.1 and illustrated in Figure 1.3. Additionally, Table 1.2 provides an overview of the four types of CCE deployment models.

Despite offering diverse service and deployment models, CCE still faces numerous security challenges.

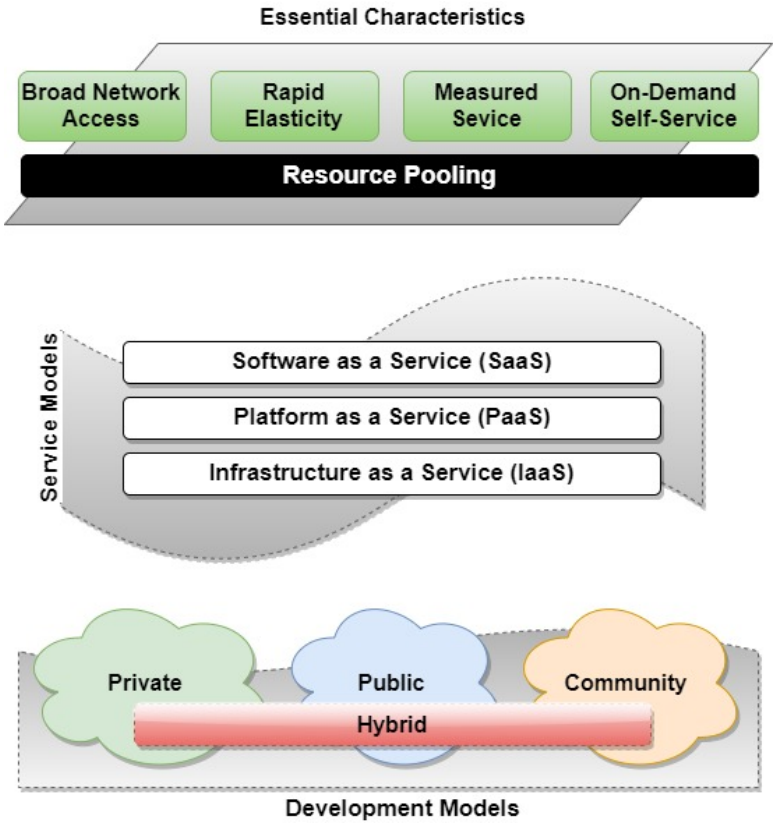

Figure 1.3: CCE Service and Development Models

Table 1.1: Summary of CCE Service Models

Service Models	Ownership Level	Control Level	Example
SaaS	End Clients	Users can use the app but cannot control it or the server, storage, OS, Network Interface Card (NIC), antivirus.	Google and Microsoft applications
PaaS	Software Developer	Users can create and control own app but cannot control the OS, antivirus.	GAE, and Microsoft's Azure
IaaS	Infrastructure Architects	Users can create and control own app, OS, NIC, antivirus.	GoGrid, Flexiscale, Layered Technologies, and Rackspace

Table 1.2: Summary of CCE Deployment Models

Deployment Models	Ownership	Services
Private	Owned by Single Company	Serve Internal Clients
Public	Owned by Single Company	Serve General Public
Community	Owned by Specific Community	Serve to Specific Community
Hybrid	Owned by a private company or/and owner of CSP	Serve to the general public and internal clients

1.2.2 Security Challenges in Cloud Computing Environment

Despite the numerous benefits provided by CCE, certain distinctive features render it susceptible to reliability and security concerns. The provision of high-availability CCE services highlights the challenges faced by CCE providers, which ultimately hampers widespread adoption (Machida et al., 2013). CCE inherits the security vulnerabilities associated with availability, integrity, and confidentiality, as previously identified by Gonzalez et al. (2012). Furthermore, the multi-tenancy features and resource sharing inherent to CCE may compromise data confidentiality, thereby jeopardising the security of sensitive information (Khorshed et al., 2012).

In terms of integrity, CCE encompasses the protection of systems from unauthorised data tampering, deletion, and modification. Additionally, the concept of availability in CCE services ensures that authorised users can access and utilise CCE resources in accordance with their requirements (Zissis & Lekkas, 2012). Notably, the consequences of resource scarcity and service limitations in CCE can be severe, resulting in partial or complete service disruptions, which may be either temporary or permanent. Common challenges to CCE services include service denial or degradation, often precipitated by factors such as competition among CSPs, proficiency evaluation, punitive measures, protests from special interest groups, political pressures, and bribery.

A DDoS attack constitutes a significant threat to CCE security. The denial or degradation of service for web servers in CCE occurs when DDoS attacks compromise users' access to targeted online services. Notably, HTTP DDoS attacks pose a particularly formidable challenge, as they target the application layer and are often difficult to detect (Dhanapal & Nithyanandam, 2021; Kanber et al., 2022). These attacks exploit system resources, resulting in service unavailability. The typical methodology employed by attackers involves a two-stage approach, commencing with an HTTP-GET DDoS attack, which is generally more harmful and harder to detect, followed by an HTTP-POST DDoS attack (Chovanec et al., 2023; Dhanapal & Nithyanandam, 2021).

1.3 Research Motivation

The threats to information integrity and confidentiality in CCEs have significantly increased, overshadowing the benefits of resource and service availability. Notably, DDoS attacks remain a persistent security risk, making it increasingly challenging to maintain online service availability. These attacks have become the most destructive and prevalent threats, targeting authorised cloud users and causing substantial economic losses in the services they utilise. Furthermore, the detection of DDoS attacks is crucial in preventing Economic Denial of Sustainability (EDoS). Many organisations, particularly those engaged in online businesses or operating business-critical applications such as banking and financial services, are deeply concerned about service availability in CCEs. Prominent organisations, including the Sony PlayStation Network, the Stock Exchange of Hong Kong, Estonia, Visa, MasterCard, PayPal, Google, and GitHub, have been severely impacted by these attacks in the past (Security, 2011).

Figure 1.4 demonstrates an upward trend in DDoS attacks over time, as reported by Jackson (2022). Additionally, Figure 1.5 illustrates a significant increase in DDoS attacks targeting application layers in 2022, as documented by Yoachimik (2022).

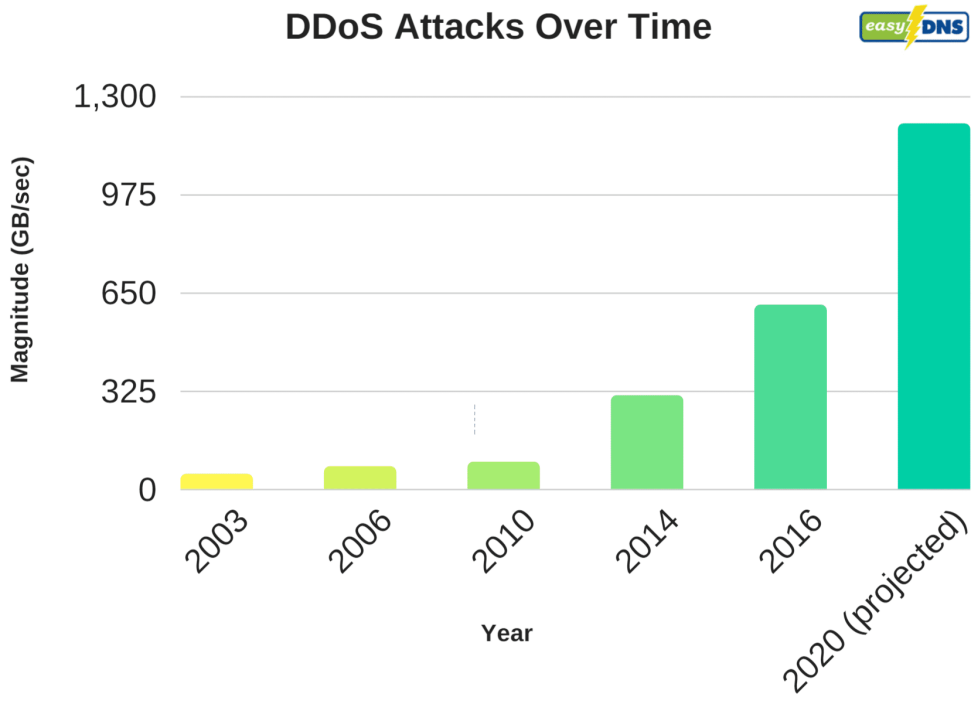

Figure 1.4: DDoS Attack Trends in Last Years

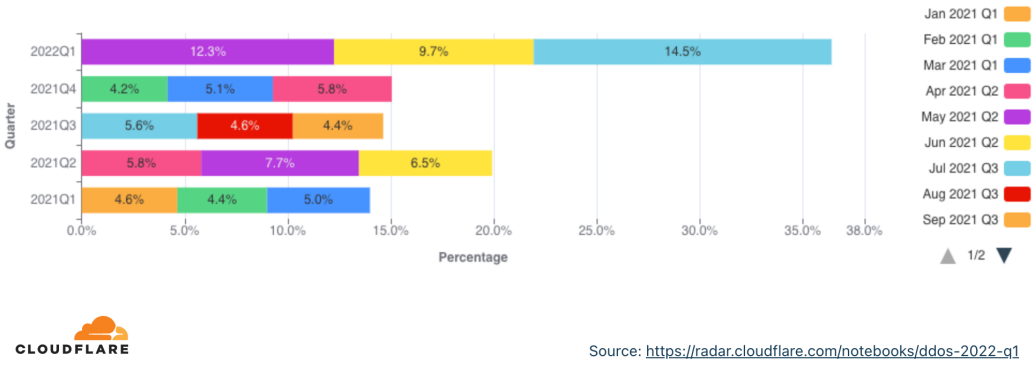

Figure 1.5: Application Layer DDoS Attacks - Yearly Distribution by Month

The detection of HTTP DDoS attacks is a particularly challenging task within CCE and has garnered increased attention in recent research. Reports indicate a rising trend in DDoS attacks, establishing them as the primary security concern in cloud environments. As reported by CNN on October 21, 2016 Brien (2016), flooding DDoS attacks have disrupted high-profile websites such as Twitter, Netflix, and GitHub. Furthermore, Amazon Web Services (AWS), a leading public CSP, has revealed repeated attacks on its infrastructure. The website Tripwire David (2016) identified the top five DDoS attacks of 2016, including incidents targeting Russian banks (Sberbank and Alfabank), the Rio Olympics websites, the blogs of American journalist Brian Krebs, services provided by Dyn Inc, and the Clinton and Trump campaigns. Additionally, Google and GitHub have also faced significant DDoS attacks. In particular, DDoS attacks at the application layer, particularly HTTP flooding attacks, have increased in frequency, affecting numerous sites and services and resulting in substantial disruptions.

The prevalence of cloud-based cyberattacks necessitates a comprehensive understanding of the current threat landscape. A 45% increase in application layer attacks in the cloud, as reported by Alert Logic, Inc. based on data from over 3,000 corporate clients, underscores the significance of this issue (Korolov, 2016). Furthermore, a survey conducted by Vanson Bourne and commissioned by Sophos revealed that India was the most targeted country for cyberattacks among 25 surveyed nations in 2019 Ghosh (2020). Notably, a Forrester research report published in 2022 highlights the prevalence of web application attacks, while Palo Alto emphasises the importance of

securing web applications in CCE from attacks such as DDoS (Hill, 2022). Moreover, recent statistics indicate that 45% of breaches are cloud-based (Jones, 2023), and a survey titled The 2022 State of Cloud Security Report found that 80% of companies have experienced at least one cloud-based security incident in the last year, with 27% of organisations experiencing a public cloud security incident, representing a 10% increase from the previous year Snyk (2022). As 72% of organisations rely on cloud-based services when upgrading or purchasing new technologies, it is evident that most businesses have experienced a breach.

A comprehensive review of DDoS attacks in CCE is provided by Lemos (2010), highlighting the significance of detecting HTTP DDoS attacks in this context. The authors emphasise the importance of addressing the complexities of HTTP DDoS detection in CCE, which are exacerbated by the inherent characteristics of the environment, including multi-tenancy, elasticity, virtualisation, and the computing model Angamuthu and Pandian (2020), Bohn et al. (2020), and Dhanapal and Nithyanandam (2020). Current approaches to detecting and mitigating DDoS attacks, particularly HTTP-GET DDoS attacks, are deemed inadequate, necessitating further research to develop more effective strategies. Notably, these attacks pose a significant threat to the resources of both customers and service providers in CCE.

CCE offers scalable access to IT resources over the internet, but this feature is exploited by adversaries to launch attacks. Consequently, safeguarding data in the cloud is a critical concern for all stakeholders, and defending against various attacks is essential. Specifically, HTTP DDoS attacks target customer and CSP resources, underscoring the need for effective countermeasures. While restricting resource allocation for web service servers can serve as a preventive measure, it compromises the essential features of CCE. Given the current limitations in preventing DDoS attacks, detection emerges as a vital stage in protecting CCE from this pervasive threat (Bochman & Freeman, 2021).

Research on detecting and mitigating DDoS attacks in CCE is still in its infancy (Valdovinos et al., 2021). Current algorithms predominantly rely on network traffic

data, yet distinguishing between legitimate and malicious packets remains a significant challenge (Azab et al., 2024). Notably, the majority of research in this field has been conducted within simulated environments (Gray, 2002). A central tenet of this thesis is that requests in HTTP-GET DDoS attacks exhibit distinct characteristics that differentiate them from normal requests. Consequently, a comprehensive investigation is necessary to address the complexities that impede the delivery of high-availability CCE services.

1.4 Problem Statement

The widespread adoption of CCE has made web servers a high-priority target for attackers seeking to exploit the inherent characteristics of elasticity and service availability. Among these attacks, HTTP-GET DDoS attacks on CCE are particularly significant, as their undetected execution can lead to EDoS, a unique and economically devastating form of DDoS attack that poses substantial risks to both providers and customers. The combination of these attacks presents a considerable threat to service availability. Notably, the detection of HTTP-GET DDoS attacks in CCE is complicated by the fact that traditional approaches used in non-virtualised environments may not be effective due to the impact of virtualisation technology on server overload (Jalili et al., 2005; Kaizaki et al., 2002; Shea & Liu, 2012b). Furthermore, even a small-scale attack in CCE can render the server inoperable, highlighting the necessity for adapting traditional detection methods to suit the CCE.

Various approaches have been proposed to detect HTTP-GET DDoS attacks in CCE, which are commonly categorised into (i) shallow Machine Learning (ML), (ii) statistical, (iii) data mining, and (iv) Deep Learning (DL) approaches (Osanaiye et al., 2016). DL-based approaches demonstrate superior Detection Accuracy (DA) compared to other methods. Furthermore, DL-based approaches, unlike shallow ML-based approaches, can effectively handle and process the large volumes of data generated by CCEs. However, DL-based approaches are associated with low to medium DA and high false positives due to the following reasons (Asiri et al., 2019; Kim, 2016).

First, an inadequate dataset containing non-specific CCE data is utilised. This data is typically collected from a single source of non-virtualised networks, leading to improper evaluation of existing approaches. Therefore, collecting relevant CCE data from multiple sources is essential for accurately detecting HTTP-GET DDoS attacks in CCE, while employing data collection techniques capable of handling massive volumes of data.

Second, existing approaches use inappropriate features selected through simple heuristics, which do not reflect the characteristics of CCE. This limitation degrades the effectiveness of HTTP-GET DDoS attack detection.

Third, the generated DL detection models are trained using irrelevant CCE data and features, which compromises their performance in detecting HTTP-GET DDoS attacks in CCE.

The problem statement can be summarised as follows:

- Inadequate Dataset: The use of non-specific CCE data collected from a single source of a non-virtualised network results in the improper evaluation of existing approaches. This underscores the necessity of collecting relevant CCE data from multiple sources to ensure accurate detection of HTTP-GET DDoS attacks in CCE.
- Inappropriate Feature Set: Existing approaches frequently utilise an inappropriate set of features, which adversely affects the detection of HTTP-GET DDoS attacks. These features are often selected based on simple heuristics, thereby limiting the effectiveness of the detection process.
- Irrelevant Features in DL Models: Deep Learning (DL) models employed for detecting HTTP-GET DDoS attacks are trained using irrelevant data and features, which diminishes their performance. To enhance accuracy, DL models should be trained with relevant data and features specific to the attack scenario.

1.5 Research Objectives

The primary objective of this thesis is to propose a collaborative-based approach for the accurate detection of HTTP-GET DDoS attacks in CCE. To achieve this objective, the following research objectives have been outlined:

- RO-1: To generate a CCE-specific dataset that entails CCE features from multiple sources of CCE such as Apache Virtual Web Server (AVWS) access log, VM activity log, and VM resources log.
- RO-2: To propose a new CCE-specific feature set that contributes to detecting HTTP-GET DDoS attacks on CCE.
- RO-3: To design an ensemble feature selection mechanism to select the best subset of features. This mechanism is applied to the features that were proposed in RO-2.
- RO-4: To adapt a DL-based model to detect HTTP-GET DDoS attacks on CCE.

1.6 Scope and Limitation

The proposed approach involves analysing the HTTP protocol requests and monitoring the activity, resource, and access logs of the AVWS. However, it is important to note that developing a comprehensive approach to address all types of DDoS attacks is beyond the scope of this research, as it would require significant time and resources. Consequently, the proposed approach is limited to detecting HTTP-GET DDoS attacks on web servers using Internet Protocol version 4 (IPv4) addresses in CCE. The focus on Apache Web servers is justified by their widespread use, as more than 60% of web servers targeted by such attacks utilise Apache W3Techs (2025) (as of 14th April 2021). Table 1.3 summarises the scope of this research.

Table 1.3: Scope of Research

No	Items	Scope of Research
1	Environment	Real-Private CCE, IPv4, VMware, Apache web server, assumes all resource allocations are elastic
2	Attack type	HTTP-GET DDoS attack
3	Target layer	Application layer (HTTP)
4	Detection type	Detection the HTTP-GET DDoS attacks
5	Dataset	Realistic Self-generated (log-based) benchmark dataset
6	Attacking Tool	Goldeneye
7	Performance Metrics	False-Positive Rate (FPR), Precision, Recall, F1 score (F-Measure), and Detection Accuracy (DA)

1.7 Research Contributions

This thesis contributes to the field of internet infrastructure security for CCE services at the application layer. Its primary contribution is an approach for the accurate detection of HTTP-GET DDoS attacks targeting web servers in CCE, based on the AVWS access log, VM activity log, and VM resource logs. The performance of the approach is further enhanced through the integration of an additional set of informative features extracted from an enrichment method also proposed in this research.

The contributions of this thesis are outlined as follows:

- A CCE-specific log-based dataset collected from multiple sources within CCE, comprising two components: first, a Basic Feature Set (BFS) obtained from a single source, specifically the AVWS access log; and second, a New Basic Feature Set (NBFS) collected from two sources, namely the VM activity log and VM resource logs, for detecting HTTP-GET DDoS attacks in CCE. The proposed dataset has been published on a dedicated website.

- A novel CCE-specific feature set designed to enhance the detection of HTTP-GET DDoS attacks in CCE. These features represent the first identified characteristics of such attacks, establishing their novelty. An enrichment method incorporating informative features from logs introduces additional relevant features from CCE, thereby improving detection capabilities. The enriched dataset, augmented with

these features, is processed before applying the attack detection system. Furthermore, these features reduce the likelihood of false alarms generated by the detection system. When combined with the previously identified basic features, they contribute to improved detection accuracy.

- An Ensemble Feature Selection (EFS) method to identify the optimal subset of features characterising HTTP-GET DDoS attacks in CCE. The EFS employs feature ranking algorithms combined with a majority voting system to select the most effective set of features based on all features from previous contributions (BFS and NBFS). Compared to earlier contributions, these selected features enhance detection performance while reducing modelling time.

- A Deep Learning (DL)-based model for detecting HTTP-GET DDoS attacks in CCE. The DL model is selected following a comparative analysis of various classifiers to determine the most suitable approach.

1.8 Thesis Organization

This thesis is structured into the following chapters:

Chapter 2 examines the current literature on DDoS attacks in CCE, their taxonomy, and the research context. This chapter reviews existing approaches for detecting HTTP-GET DDoS attacks that target web servers in CCE. Additionally, it identifies gaps in previous research and highlights the necessity of detecting and classifying HTTP-GET DDoS attacks based on their anomalous characteristics.

Chapter 3 outlines the design and architecture of the proposed approach (DDoSCCE), detailing the integrated phases and the methodology for detecting HTTP-GET DDoS attacks targeting web servers in CCE.

Chapter 4 (Implementation of the Proposed Research): This chapter describes the software and tools utilised to develop and implement the proposed approach (DDoSCCE) on a realistic VMware platform testbed (private cloud). Furthermore,

it presents the design and implementation phases of the proposed approach.

Chapter 5 presents the experiments conducted and their corresponding results, analysed in detail using the proposed DDoSCCE approach. The performance of the proposed approach is evaluated in comparison to existing methods. Finally, the proposed dataset is compared with publicly available HTTP datasets.

Chapter 6 concludes the research, summarises the findings, and suggests potential directions for future research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Chapter 1 highlights the significance of HTTP-GET DDoS attacks. This chapter provides the necessary background information and reviews relevant research to establish a comprehensive understanding of the thesis. The literature review is divided into seven sections. Section 2.2 presents the background, while Section 2.3 covers data acquisition and processing for attack detection. Related work is discussed in Section 2.4, followed by the research gap and discussion in Section 2.5. Dataset enrichment is addressed in Section 2.6. Finally, the chapter concludes with a summary in Section 2.7.

2.2 Background

Cloud Computing (CC) has significantly revolutionised the Information and Communication Technology (ICT) industry. It enables the flexible delivery of new services and computing resources at a fraction of the costs for end-users compared to traditional computing. However, numerous potential cyber threats impact CC-deployed services due to the exploitation of CC's characteristics, such as resource sharing, elasticity, and multi-tenancy, by cybercriminals (Jones, 2015; Oliveira et al., 2019).

Over the past decade, advances in ICT have dramatically transformed the consumption and dissemination of information, particularly through online services (Konsbruck, 2000). CCEs are now widely used to deploy client services. However, the delivery and provision of these online services are at risk of disruption due to certain user actions, whether intentional or unintentional (Jones, 2023). Such situations occur when a large number of users simultaneously access a web server or when malicious actors launch a DDoS attack against the web server in CCE, rendering it inaccessible (Cloudflare, 2023).

2.2.1 Overview Threats Cloud Computing Environment

CCE represent the realisation of a long-held concept known as computing-as-utility, which has emerged with significant potential. CCE, as a utility, enables online services to be accessible anytime, anywhere, and to anyone through the internet, while also offering cost-effective and straightforward deployment. As a result of adopting this technology, the popularity of web applications is anticipated to experience substantial growth in the near future. The fundamental characteristic of CCE is its transformative impact on the IT field, positioning it as the next major innovation following the internet. CCE consolidates IT infrastructure, network services, and applications with the resources of a Data Centre (DC) through extensible virtualisation technology, resulting in flexible, scalable Singh et al. (2014), load-balanced, and on-demand services for CCC. This allows CSP to charge CCCs based on usage (Bhardwaj et al., 2016).

The popularity of modern web-based applications, which have stringent requirements, necessitates continuous improvements to the CCE model to ensure effective service delivery. Many businesses adopt various forms of CCE models to optimise user operations and reduce costs. However, numerous potential cyber threats impact services deployed in CCE by exploiting its characteristics, such as resource sharing, elasticity, and multi-tenancy (Jones, 2023).

2.2.2 Underlying Concept of DoS and DDoS Attacks on CCE

A DoS attack is a malicious attempt by an adversary using a single attacking host to prevent the targeted victim from accessing required services or a node providing services to its consumers. In contrast, a DDoS attack involves multiple attacking hosts flooding the victim's network or host with attack packets, resulting in a distributed multi-point attack (Waller et al., 2011). DDoS attacks are considered one of the most serious threats to CCE.

DoS attacks can be categorised into two main types. ***(i) Flood attacks*** involve an adversary targeting the victim's network bandwidth or connectivity by sending a large

volume of continuous packet streams. This depletes the victim's resources and exhausts the bandwidth, causing delays or preventing users' requests from being fulfilled. ***(ii) Vulnerability attacks*** involve an adversary exploiting a vulnerability in the victim's system and sending crafted messages to cause a DoS. DDoS attacks are frequently executed by flooding the targeted system or network with a significant volume of traffic from multiple sources (Ghaben et al., 2021).

2.2.2(a) DoS, DDoS Attacks on CCE

A DDoS attack is a malicious act perpetrated by adversaries involving the transmission of a large volume of seemingly normal packets to a target, making them difficult to detect. These packets are sent to the victim machine to disrupt applications or protocol execution on the victim (Ghaben et al., 2021; Jaber et al., 2021b). DDoS attacks that disrupt users' connectivity by exhausting network bandwidth and reducing the router's processing capacity fall under the network-layer attack category. Conversely, attacks that deny legitimate user access to services by exhausting VM resources, such as input/output bandwidth, sockets, Central Processing Unit (CPU), disk/database bandwidth, and memory, are classified as application-layer (Layer 7) attacks (Jaber et al., 2021a). Application-layer attacks target server applications or services by attempting to fully exhaust their resources through the creation of numerous transactions and processes. These attacks are particularly challenging to detect and mitigate because the generated attack traffic, such as HTTP requests, is often indistinguishable from traffic generated by legitimate users. DDoS attacks are executed via a remotely controlled network, which is distributed and well-organised, enabling compromised machines, known as zombies, to transmit a high volume of simultaneous requests to continuously attack the target system.

The widespread adoption of CCE has led to a significant increase in the number of attacks. The commonality of CCE structures and components makes them more susceptible to attacks, particularly DoS and DDoS attacks. Adversaries typically target CSPs with numerous connected devices, as the scalability and dependability of CCE

make it accessible from anywhere. DDoS attacks often involve the transmission of a large volume of malicious packets to the target to overwhelm it. While traditional networks employ various strategies to defend against DDoS attacks, the unique properties of CCE make defence mechanisms more challenging to implement. Abusaimeh (2020) explores the increasing threat of DDoS attacks in CCEs. It highlights how the scalability, resource sharing, and multi-tenancy features of CCEs make them vulnerable to such attacks. Various types of DDoS attacks, such as flooding, fragmentation, and amplification attacks, are categorised, and their impact on cloud services is discussed. The motivations behind these attacks, including financial gain, revenge, ideological beliefs, and cyber warfare, are also examined. Existing detection, prevention, and mitigation techniques, such as Moving Target Defence, CAPTCHA, and DNS-based methods, are reviewed. Additionally, defence mechanisms like source-based, network-based, and hybrid approaches are evaluated, with their strengths and limitations addressed. The study concludes by emphasising the need for robust security policies and advanced defence strategies to effectively protect cloud services from DDoS attacks.

Organisations can leverage CCE to access on-demand, elastic, and fully managed computer system resources and services. However, attacks on CCE, particularly DoS and DDoS, can result in substantial losses for both CSPs and users. Successful DDoS attacks have serious consequences, including poor user experience, service failures, and, in the worst-case scenario, complete shutdowns and financial repercussions. The volume, frequency, and intensity of DDoS attacks have increased significantly, driven by the growing popularity of the Internet of Things (IoT) and widespread network connectivity, which have inadvertently facilitated their proliferation (Ghazaleh & Ahmad, 2017).

Fast-growing web-based applications are typically developed and deployed within the CCE ecosystem. CCE significantly reduces the cost of monitoring and maintaining IT infrastructure. The control and management of CCE resources often rely on standard networking protocols, enabling administrators to centrally manage and con-

trol distributed IT infrastructure. However, the use of standard networking protocols also allows adversaries to gain unauthorised access if security measures are inadequate. Attacks such as DDoS are among the most common threats in private CCEs, often resulting in service degradation or complete denial of service (Virupakshar et al., 2020).

DDoS attacks targeting application-layer services are not new. The first well-documented DDoS attack occurred in August 1999 against a higher educational institution in the United States. Since then, such attacks have targeted various sectors, including corporate institutions such as CNN, eBay, Yahoo, and Amazon (Hovav & D'Arcy, 2003). In 2009, a DDoS attack disrupted numerous popular online services, including Facebook, Live Journal, Twitter, and Amazon (Yevsieieva & Helalat, 2017). By 2014, more than 7,000 DDoS attacks were being launched daily (Mousavi & St-Hilaire, 2018). The first quarter of 2013 saw the average attack volume reach approximately 50 Gbps, representing a 718% increase from the last quarter of 2012 (Chaudhari & Talmale, 2019; Mousavi & St-Hilaire, 2018). The largest DDoS attack in history occurred in late 2016, involving a botnet called Mirai Woolf (2017), which infected the servers of a Domain Name System (DNS) provider. Additionally, the duration of DDoS attacks has been increasing. For example, until the last quarter of 2020, no attack lasted more than 302 hours. However, the longest attack in the first quarter of 2021 lasted 746 hours (over 31 days), which was later surpassed by a 776-hour onslaught (over 32 days) in the second quarter. Figure 2.1 Somani et al. (2016) illustrates the distribution of DDoS attacks over time, spanning the first and second quarters of 2022 (Kaspersky, 2022). A projection for the global market of DDoS protection systems SoftActivity Team (2022) estimates that its value will expand from $3.3 billion in 2021 to $6.7 billion by 2026.

2.2.2(b) Impact of HTTP-GET DDoS Attacks on CCE

Security in CCE is critical for maintaining service availability for end users. Typically, CCCs obtain CCE services from CSPs via the HTTP protocol, which is susceptible to misconfigurations and vulnerable to attacks if not adequately secured. Clients ac-

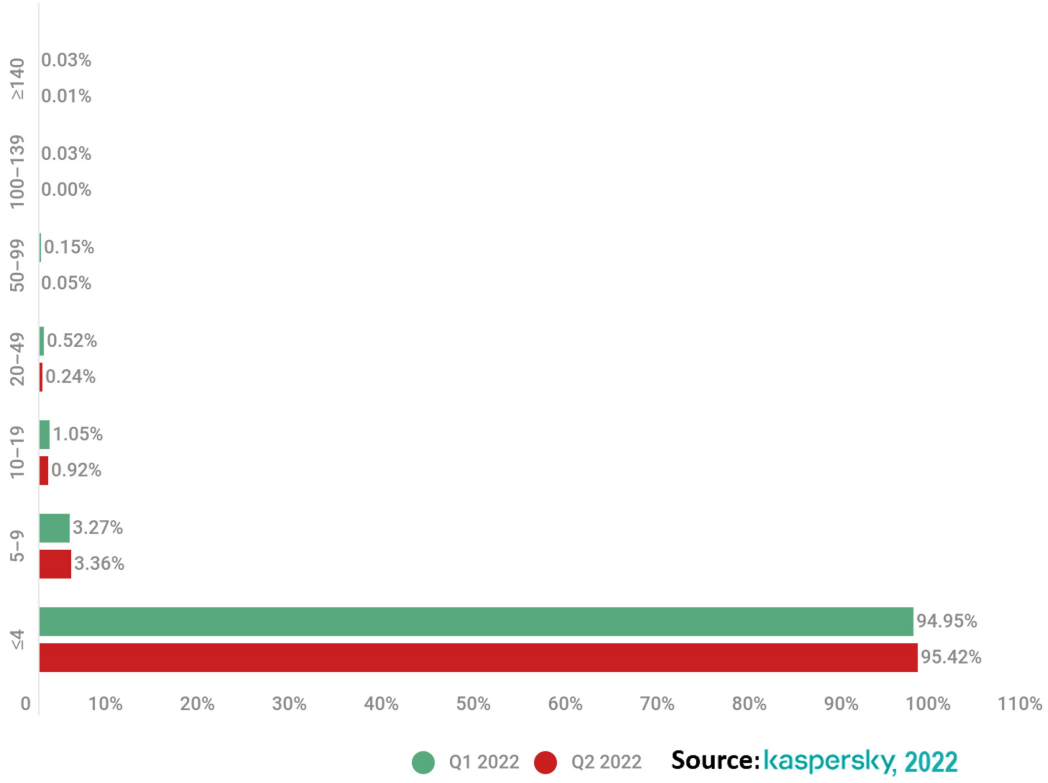

Figure 2.1: Distribution of DDoS Attacks by Duration, Q1-Q2-2022

cessing online services from CSPs via HTTP are often targeted by HTTP-GET DDoS attacks. The HTTP-GET attack is similar to traditional DDoS attacks, except it overloads the server with a high volume of HTTP-GET requests instead of data packets (Alanazi et al., 2019). During an adversary-initiated HTTP attack, attackers may rent or use their own servers to bombard the targeted victim's VM with HTTP GET/POST requests, leading to a significant increase in resource consumption and financial losses for the victim. Ultimately, the targeted host becomes overburdened and overwhelmed by HTTP attacks, compromising the entire CC infrastructure.

HTTP attacks are challenging to detect because they utilise valid HTTP requests to the web server Beitollahi et al. (2022), overloading VM resources. Furthermore, HTTP-GET attack traffic is difficult to distinguish from legitimate traffic, as it employs regular Uniform Resource Locator (URL) queries at a normal rate. Since the traffic volume in HTTP-GET attacks often falls below the thresholds of most rate-based detection approaches, the attack frequently goes undetected (Gupta et al., 2013).

Additionally, attackers use valid HTTP packets without any anomalies in the packet's structure or flags, whereas IDS are primarily designed to detect malformed or anomalous combinations of flags. The consequence of such attacks is an EDoS attack, as the auto-scaling feature of CC exacerbates the damage by enabling CSPs to allocate excessive resources to CCCs to fulfil Service-Level Agreements (SLAs). Although limits on the CCC's pay-per-use invoicing system could prevent excessive user charges, they could also result in the shutdown of multiple CCE services during a single attack. Detecting HTTP-GET DDoS attacks in CCE requires a comprehensive understanding of both the attacks and user behaviour.

2.2.3 Security Issues and Challenges

As the popularity of CCE services continues to grow and their usage becomes more widespread, the associated security issues and challenges are also expected to increase. This section examines the security issues and challenges inherent to CC.

2.2.3(a) Security Issues

CCE face several security issues, including the misuse and excessive utilisation of resources, malicious insiders, insecure and unreliable APIs, data corruption or leakage, account takeovers, and vulnerabilities in shared technologies (Alzahrani & Hong, 2018; Maghrabi, 2014). Among the most serious risks to CCE security are DoS and DDoS attacks Maghrabi (2014), as highlighted by ongoing research conducted by the Cloud Security Alliance (CSA) (Ren, 2015). A well-secured CCE ensures the security of the websites it hosts, and conversely, the security of hosted websites contributes to the overall security of the CCE (Hezavehi & Rahmani, 2020). The demand for robust CCE security is becoming increasingly prominent and critical. Although CCE security is still in its nascent stages and continues to evolve, a diverse range of CCE clients already demand various security requirements, which may differ across CSPs. Even the same client may have varying security needs depending on the context. CCE clients can include individuals, academic institutions, corporations, and software developers, each

of whom may have specific security prerequisites. According to several researchers Kumar et al. (2012) and Somani et al. (2015), the following security issues are associated with CCCs and CSPs:

- ***Data-Related Issues:*** CSPs must ensure the security and availability of data transferred by CCCs and prevent potential security breaches. CSPs should implement robust security measures to guarantee data availability and security. For instance, various authorisation approaches can be employed to control data access, such as isolating clients' data within the CCE storage unit to ensure that only authorised clients can access or modify the data.

- ***Data Integrity:*** In an independent system, data integrity can be ensured by adhering to Atomicity, Consistency, Isolation, and Durability (ACID) properties. This can be achieved through a decentralised approach using a central manager. However, ensuring data integrity in CCE poses significant challenges. The primary challenge with web services at the application layer is transaction management, as the HTTP protocol cannot guarantee a reliable and available service. Consequently, the responsibility often falls on the API level. However, existing standards, such as web service transactions and reliability, remain underdeveloped. The lack of control at the data level can result in substantial integrity issues.

- ***Availability:*** Maintaining the availability of CCE applications is crucial for ensuring user accessibility. Therefore, CSPs must guarantee uninterrupted services to their CCCs, including scalability to support business continuity for their users.

- ***Privileged User Access:*** As CCCs often store sensitive data in the cloud, users require varying levels of privileges to access this information. A robust control system must be implemented to manage access permissions for different users.

- ***Data Recovery and Backup:*** CSPs should ensure data recoverability by providing backup services in the event of disasters or accidental damage. Backup

data should be replicated across multiple distributed servers while maintaining integrity and privacy, enabling data restoration whenever necessary.

- ***Investigative Support:*** Investigating incidents in CCE becomes challenging when logging data is constantly distributed across multiple sites. CCCs must secure contractual obligations from CSPs to investigate any inappropriate or criminal activities.

- ***Long-Term Viability:*** CCCs should obtain guarantees that their data will remain accessible even after changes in CSP policies or terms.

- ***Network Security:*** CSPs are required to ensure data security during transit or network transfers to prevent the leakage of critical or sensitive information.

- ***Web Application Security:*** Services on CCE are typically managed via web browsers. Any security flaws in the service application affect all CCCs using the same application. Although numerous conventional security approaches exist, they often fail to address security issues effectively.

- ***Virtualisation:*** VMs that lack proper isolation procedures may inadvertently allow guest clients to execute code on the hypervisor. The multi-tenancy property of VMs provides opportunities for adversaries to exploit vulnerabilities, potentially gaining unauthorised access to data from other VMs within the cloud.

- ***Identity Management (ID):*** Identity management enables systems to identify and manage clients, including controlling their permission-based access. Identity management systems should be configurable to comply with corporate policies (Lopez & Rubio, 2018).

- ***PaaS-Related Issues:*** PaaS allows CSPs to enable CCCs to develop applications through platform services. CSPs are accountable if illicit data is transferred between applications. The hypervisor remains the responsibility of CSPs. While PaaS offers greater elasticity and adaptability compared to SaaS, it is more vulnerable to cyber-attacks due to security trade-offs.