Modeling for a QoS-Efficient Real-Time Video Conferencing System in the Cloud

Aklilu Daniel Tesfamicael
Science and Engineering Faculty
Queensland University of
Technology (QUT)
Brisbane, Australia
aki.tesfamicael@hdr.qut.edu.au

Vicky Liu
Science and Engineering Faculty
Queensland University of
Technology (QUT)
Brisbane, Australia
v.liu@qut.edu.au

Ernest Foo
Science and Engineering Faculty
Queensland University of
Technology (QUT)
Brisbane, Australia
e.foo@qut.edu.au

Bill Caelli
Science and Engineering Faculty
Queensland University of
Technology (QUT)
Brisbane, Australia
w.caelli@qut.edu.au

Abstract— Recently on-premises deployed real-time video conferencing (VC) systems have been facing the growing resource demand from using desktop and mobile devices. To address the cost of upgrading the existing on-premises VC infrastructure, businesses are moving towards cloud-based VC services. However, operating cloud-based VC and maintaining real-time interactive conferencing of audio and video sessions also has a number of challenges including availability, quality of video calls and performance issues. Video conferencing require significant hardware resources and stringent Quality of Service (QoS) requirements in terms of bandwidth, delay, and jitter. This research proposes a viable model for the design and implementation of a cloud-based VC system that meets time-critical requirements. We provide a numerical simulation evaluation, which shows that a relevant methodology can be used to best manage and control a cloud-based VC system to realize benefits while maintaining required level of QoS and performance.

Keywords— *queueing theory; QoS; video conferencing; cloud-based resource allocation.*

I. INTRODUCTION

Cloud computing is an evolving computing paradigm that enables access to shared pools of configurable, on-demand system resources over the internet. It generally incorporates infrastructure, platform, and software as a service [1]. Cloud-based video conferencing (VC) is a two-way communication that allows participants to interact with each other in real time. However, cloud-based video streaming is to view the streaming video file, normally is a one-way communication only between the viewer and the streaming server. In this paper, our focus is on cloud-based VC.

Recently, VC is gaining popularity, owing to the growth of using smart mobile devices as a means of video communication [14]. This has prompted the need for more resources to provide quality real-time communication services. This is the main reason that resource management for cloud-based VC has attracted attention from both academia and industry alike. Owing to financial benefits, businesses are moving on-premises VC systems towards cloud-based VC solutions. However, adopting a cloud-based VC solution has its own challenges. For real-time video conferencing, it is crucial to support QoS requirements in terms of bandwidth, delay, and jitter [2]. Therefore, using a general-purpose cloud system in the Internet to provide real-time video conferencing may have undesirable

service quality [3]. This may become more complex as the services are required to be delivered with combined data, voice, and video. An example of this is the trading turret communication system where the traders can use the system for real-time voice and video communications. A recent study [20] was conducted on the performance and security aspects for the cloud-based trading turret communication system.

A. Problem statement

Video transcoding is a resource-intensive task, and meeting QoS requirements for large video data in the cloud in a reliable and secure manner is a challenge. In a typical cloud deployment scenario, a pool of resources for the video conferencing is share. However, a modeling is needed to effectively estimate the resource demand for VC services in the cloud. This will reduce over-utilized or under-utilized resources that may have a significant impact on the overall video conferencing quality, performance and the costs of cloud services. Obviously, without an effective resource management model, this is not the most efficient solution and frequently implies waste of resources or limitations in its use. Recent research [15] suggested that efficient resource management results in better performance.

Latency in VC services can occur in any of the following situations. These are the areas where we consider applying queueing theory to formulate the delay in transcoding and the transmission of video services. Delay can be classified into:

1. The delay or waiting time in resource profiling (the time for scheduling the task)
2. The delay or waiting time in video transcoding in the cloud (the time for computing the task)
3. The transmission delay (the time for transmitting video to the recipient). Namely, from the connection is established during the video conferencing call.

B. Contribution

The contribution of this research paper is to detail and classify QoS and performance requirements to support cloud-based VC, so that delay occurring can be controlled effectively and negative performance impact can be minimized.

II. RELATED WORK

Resource allocation for a real-time video conferencing has been an active research topic, recently. Cloud-based real-time

video conferencing impose several constraints on the resource management. Numerous studies have been conducted regarding resource scheduling in the cloud. Cicalo and Tralli [13] proposed cross-layer optimization of a system for multiple scalable video delivery in orthogonal frequency division multiple access wireless networks. Their proposal for a cross-layer resource allocation scheme is to minimize the distortion difference among multiple videos. They observed differences in the efficiency and video quality fairness with different strategies. Their research is focused on the resource allocation for video transmission in a wireless network environment. Nevertheless, our study is based on meeting QoS requirement for end-to-end communications for cloud-based VC services. Chen et al. [10] studied the problem of utility maximization for peer-to-peer systems. They developed a new multi-tree routing formulation for the multicast utility maximization problem. They discovered that the lack of powerful nodes in peer-to-peer systems are unable to support the execution of high demand processing tasks. Chen *et al.* [11] also studied in a peer-to-peer environment where they designed a multi-party VC solution called Celerity. In both [10] and [11], they assumed a peer-to-peer scenario where bandwidth bottlenecks exist only at the edge of the network. For peer-to-peer video conferencing, Chen et al. proposed Celerity [10] and [11] to optimize the utility of the conferencing services, as well as to make the best use of bandwidth for peer-to-peer uplinks, which has an advantage over those in the other literature in the same area [7, 10, 11, 16]. However, our study examines the overall network latency from a node located on-premises to the server hosting the VC services in the cloud.

Ponec et al. [6] proposed a scheme to maximize video quality under uplink-downlink capacity constraints for peer-to-peer multiparty video conferencing. However, there is a lack of powerful server in the architecture to support the execution of the high processing demand tasks such as transcoding.

In the resource scheduling policies, Mell and Grance [1] presented a number of the scheduling policies for improving the performance and commercial viability of video on demand (VoD) systems.

Zhuet et al. [3] and Ghose and Kim [4] formulated the utility maximization problems. They built multi-rate multicast trees to enable multi-party VC services. Their work investigated on the physical links where the streaming rate allocation over the communication link, but their work overlooked the issue related to the transcoding flexibilities. Feng et al. [8] observed the maximization of the overall throughput of VC sessions by leveraging inter-datacentre network bandwidths, based on intra-session network coding.

[14-16] developed resource allocation schemes for on-premises system, rather than for the cloud-based resource management.

As indicated on the reviews of previous literature, there is a lack of research work on meeting end-to-end QoS requirements for cloud-based VC services.

III. VC QoS REQUIREMENT

It is important that real-time video conferencing systems meets the needs of user applications such as service reliability and QoS requirements [17]. Different types of real-time multimedia services usually require different level of QoS and performance requirements [21-22]. To evaluate the performance quality for cloud-based VC, this research is based on the Telecommunication Standardization Sector (ITU-T) recommends that a set of QoS requirements for a real-time multimedia communications system be satisfied for the video and voice services to be of a high quality. The QoS requirements are show in TABLE I.

TABLE I. VC QoS REQUIREMENTS

QoS Requirement	Recommendation
Jitter	Less than 30ms
Latency	150ms or less (one-way end-to-end)
Packet Loss	Less than 1%
Video Codec/Decoder	
H.261 and H.263	64Kbps – 1.2Mbps (Bit rate)
MPEG-2/H.262	5–25 Mbps
H.264/MPEG-4 AVC	40 kbps–10 Mbps

IV. VC SYSTEM MODELS IN THE CLOUD

A. Proposed Model

The proposed model is implemented in a public cloud environment with the backend video server is operated at the cloud service provider's data centre. The data centre, head office and branch office are connected through the Internet. A Virtual

Fig. 1 Proposed cloud-based VC system architecture

Private Network (VPN) is enabled to provide a secure tunnel between the data centre and the branch office.

B. Queueing Model Background

Queueing theory is a methodological framework for analysing network delay and systems that provides service to random demands. Due to the randomness of the arrival and service time of a VC, it is crucial to design the system so that it provides service balancing between QoS requirements of the VC service and its resource requirements. Queueing theory can provide adequate delay approximation for the design of resource allocation and management.

As illustrated in Fig.2, the overall communication latency includes:

- Processing delay, which is the delay between the time the packet is received at the head node of the link and the time the packet is assigned to an outgoing link queue for transmission.
- Queueing delay, which is the delay between the time the packet is assigned to a queue for transmission and the time it starts being transmitted.
- Transmission delay, which is the delay between the times that the first and last bits of the packet are transmitted.
- Propagation delay, which is the delay between the time the last bit is transmitted at the head node of the link and the time the last bit is received at the tail node.

Fig. 2 Types of Packet Delays

As shown in Fig. 3, the proposed model consists of three queues systems, which are the schedule queue, the computation queue, and the transmission queue. Applying queueing theory plays a key role in designing a balance between the QoS requirements of the VC service and the resource requirements of the system.

Fig. 3 Queueing System Model

V. VC MODEL USING QUEUEING THEORY

Accurate models and proper queueing rules of video traffic can provide end-to-end communication quality and provide a design foundation for video buffering, call admission control, transcoding and bandwidth allocation. The proposed model for the cloud-based VC system as an $M/G/1$ queueing model traffic with video frames arriving according to a Poisson process. In the $M/G/1$ queueing mode, M stands for a Poisson process arrival with arrival times exponentially distributed, G denotes generally distributed service times, 1 represents a single server for event processing and transcoding, and no notation after G indicates that the buffer's volume is infinite. The aim of this section is to formulate the number of video frames waiting for service in the queue $E[N_q]$, the mean service time of the system $E[S]$, the expected waiting time of the video services in the cloud-based VC system $E[W_q]$, the expected spend time of the video packets in the system $E[W]$ and the expected mean system length $E[N]$. Under the $M/G/1$ queueing model, the service times are independent and identically distributed (i.i.d.) processes with probability density function

$$p_s = P\{\text{service time equals } s \text{ slots}\} \quad s \geq 1 \quad (1)$$

Table II describes the notations we use for this model.

TABLE II. NOTATIONS USED FOR OUR MODEL

T	Transmission delay
N_q	Number of waiting video frames for service
S_n	Service time in time slot n $\{n=1, 2, \dots, n\}$
R	Residual service time (unfinished work in the server)
$E[R]$	Mean residual service time
$E[N_q]$	Mean number of video packets waiting in the queue
$E[N]$	Mean system length
$E[W]$	The mean spent time in the system
$E[W_q]$	Mean waiting time in the queue
$E[S]$	Mean service time
C^2	Squared coefficient of variation of the service process
λ	Arrival rate of requests for VC services
$\rho = \frac{\lambda}{\mu}$	VC offered load

We start with the derivation of the expectation of the waiting time T of the video frames in the queue. T is the time the video frame has to wait before it is served by the transcoding server. Each video frame has the same duration; however, the transcoding time for video frames varies based on many factors,

Fig. 4 Waiting Time Model

such as bitrate of the video type (standard frame rate or high frame rate), resolution of the video type and the overall server performance. Fig. 4 below shows the waiting time model.

We can express the expectation of the mean waiting time in the buffer, $E[Wq]$ as:

$$E[Wq] = E[S_1 + S_2 + \dots + S_n] \quad (2)$$

$$= E[R] + E[S_1 + S_2 + \dots + S_n] \quad (3)$$

From steps (2) and (3), we obtain the expected waiting time (4)

$$E[Wq] = E[R] + E[S] E[N_q] \quad (4)$$

$E[N_q]$ is the mean number of video packets waiting in the queue for service in the system and $E[S]$ is the mean service time, where both $E[N_q]$ and $E[S]$ form the mean time needed to serve the video frames ahead in the queue. $E[R]=0$ if the system is idle (i.e., the system is empty). A video frame arrives to an empty system if $E[N_q]$ and $E[R]$ are zero and in this scenario video frames enter the system for service immediately. In order to find the mean waiting time of arrival of video frames, we need to find first $E[N_q]$. So, using Little's Law, the mean queue length $E[N_q]$ of the video services can be expressed as below. Little's Law, also known as $l = \lambda w$, which asserts that the time average number of customers spend in a queueing system, l , is equal to the rate at which customers arrive and enter the system, λ , $\times$ the average sojourn time of a customer, w .

Using Little's result $E[N_q] = \lambda E[Wq]$ and from (4) we obtain (5)

$$E[Wq] = \lambda E[Wq] E[S] + E[R] \quad (5)$$

Denoting $\rho = \lambda E[S]$ and solving for $E[Wq]$:

$$E[Wq] = \rho E[Wq] + E[R] \quad (6)$$

$$E[W] = \frac{E[R]}{1-\rho} \quad (7)$$

One main property of the model which remains to be determined is the $E[R]$. To resolve this, we take a graphical approach.

$E[R]$ can be determined as a function of time using the graphical method as shown in Fig.5.

Fig.5 Residual Service $R(t)$ as a function of time

The long interval of time can be sub-divided in terms of service times $S_n \{n=1, 2, \dots, n\}$. We note the arrival of video frames separated by an idle period or an empty queue. The number of arrivals, n , is determined by the arrival rate λ with a mean number λt . R jumps by the amount of service time S when a new video frame enters the service. $R=0$ means the video frames have been served and the service is completed.

$E[R(t)] = \frac{1}{t}$ (sum and area of triangles) and time averages of $R(t)$ in $(0, t)$ is:

$$E[R[(0, t)]] = \frac{1}{t} \int_0^t R(t') dt' \quad (8)$$

Approximation of this formula becomes accurate as $t \rightarrow \infty$ and multiplying and dividing it by n we get:

$$E[r[(0, t)]] \approx \frac{n}{t} \cdot \frac{1}{n} \sum_{i=1}^n \frac{1}{2} S_i^2 \quad (9)$$

Where $\lambda = \frac{n}{t}$ and $E[S] = \frac{1}{n} \sum_{i=1}^n S_i$

As $t \rightarrow \infty$ we get $\lim_{n \rightarrow \infty} \left(\frac{n}{t}\right) \rightarrow \lambda$

$$\lim_{n \rightarrow \infty} \frac{1}{2} \left(\frac{1}{t}\right) \sum_{i=1}^n S_i^2 \rightarrow \frac{1}{2} E[S^2] \quad (10)$$

Hence, from (10) we have the residual service time $E[R]$ as

$$E[R] = \frac{1}{2} \lambda E[S^2] \quad (11)$$

Substituting the result in (7) $E[W] = \frac{E[R]}{(1-\rho)}$, we obtain the mean time spent in the queue, $E[Wq]$.

$$E[Wq] = \frac{\lambda E[S^2]}{2(1-\rho)} \quad (12)$$

This yields the desired Pollaczek–Khinchine (P-K) formula. In queueing theory, the P-K mean formula states the relationship between the queue length and service time distribution. The P-K mean value formula can be applied for the $M/G/1$ queue model in discrete time.

From the mean waiting time, we get the mean sojourn time $E[T]$.

$$E[T] = E[S] + E[W] \quad (13)$$

The squared coefficient of variation (SCV) denoted by C^2 can be rewritten as a formula in (15) below and this will help in simplifying $E[S^2]$

$$C^2 = \text{Var}[S]/E[S]^2 \quad (14)$$

$$E[S^2] = \text{Var}[S] \cdot E[S]^2 \quad (15)$$

$$= (1 + C^2) \cdot E[S]^2 \quad (16)$$

Hence, (12) can also be rewritten in the form of:

$$E[Wq] = \frac{\lambda E[S^2]}{2(1-\rho)} = \frac{1 + C^2}{2} \cdot \frac{\rho}{1-\rho} \cdot E[S] \quad (17)$$

$$E[T] = E[S] + \frac{\lambda E[S^2]}{2(1-\rho)} \quad (18)$$

$$= \left(1 + \frac{1+C^2}{2} \cdot \frac{\rho}{1-\rho}\right) \cdot E[S] \quad (19)$$

From (219), we can obtain the mean time spent in the system, $E[W]$ as:

$$\begin{aligned} E[W] &= E[Wq] + E[S] \\ &= \left(1 + \frac{1+C^2}{2} \cdot \frac{\rho}{1-\rho}\right) \cdot E[S] + E[S] \end{aligned} \quad (20)$$

By applying Little's Law's result, we obtain the mean number of waiting video traffic and video traffic in the system. using $E[N_q] = \lambda E[W]$ we write $E[N_q]$ as:

$$E[N_q] = \lambda E[W] = \frac{\lambda^2 E[S^2]}{2(1-\rho)} \quad (21)$$

From (16), we formulate $E[S^2] = (1 + C^2) \cdot E[S]^2$ and using this formula, (21) can be rewritten as (25). Also, we have from above steps $\rho = \lambda E[S]$ and applying this formula, (22) can be rewritten as (24).

$$E[N_q] = \frac{\lambda^2 (1 + C^2) \cdot E[S]^2}{2(1-\rho)} \quad (22)$$

$$= \frac{(1 + C^2) \cdot (\lambda E[S])^2}{2(1-\rho)} \quad (23)$$

$$= \frac{1 + C^2}{2} \cdot \frac{\rho^2}{1-\rho} \quad (24)$$

$$E[N] = \lambda E[T] \quad (25)$$

$$= \lambda E[S] + \frac{\lambda^2 E[S^2]}{2(1-\rho)} \quad (26)$$

$$= \rho + \frac{1 + C^2}{2} \cdot \frac{\rho^2}{1-\rho} \quad (27)$$

VI. NUMERICAL ANALYSIS

In this section, we present numerical results of the proposed model to demonstrate the performance of the expected number of video frames waiting for service in the queue $E[N_q]$, the expected mean service time of the system $E[S]$, the expected waiting time of the video services in the cloud-based VC system $E[W_q]$, the expected spend time of the video packets in the system $E[W]$ and the expected mean system length $E[N]$. In order to understand the behavior of the modeling presented in this paper, it is thus necessary to have a numerical approximation to formulate the three classified delays mentioned above. We present three numerical examples of nearly identical models, differing only in their respective coefficients of variation of the service time distribution. In these examples, we consider five service time distributions with different mean μ and with squared coefficient of variation c^2 of 0, 0.5, 1, 1.5, and 2, respectively. In all cases, video packets arrivals are generated according to an $M/G/1$ queue with a Poisson process at a rate of 0.95 ($\lambda=0.95$) and we assume that the average sojourn time of a video packet is in the value of [0.1, 0.9]. The other assumption we take in this example is that $E[S]=1$

Fig.6 Expected waiting time in the queue with different ρ and c^2

Fig.7 Expected waiting time in the system with different ρ and c^2

As can be observed, $E[W_q]$ and $E[W]$ increase with increasing values of ρ and C^2 and, by contrast a decreasing function of μ . Both means waiting times are small when the utilization of the system is low and they increase as the utilization increases. Based on the numerical examples above we observe the server stays in a stable state until ρ gets to 0.7 for all values of the c^2 . This also affects the number of video packets waiting in the queue, $E[N_q]$ as shown in Fig. 8. Obviously, the waiting time of the video packets in the queue increases when the server is busy processing existing video packets.

The mean number of requests in an $M/G/1$ queue depends on mean and mean square values $E[S]$ and $E[S^2]$ of the service time. In this example we consider $E[S]=1$ however $E[S^2]$ has impact on the waiting part, as shown in Fig.6, Fig.7 and Fig.8.

In Fig.9 we pay our attention to the behaviour of the mean system length $E[N]$ with the change of service rate μ and the squared coefficient of variation C^2 respectively. Obviously, we find that the mean system length $E[N]$ decreases with the increase of μ and increases with the increase of C^2 .

Fig.8 Expected number of video packets waiting in the queue

Fig.9 Expected number of VC services in the system $E[N]$ versus μ

VII. CONCLUSION AND FUTURE WORK

In real-time cloud-based video conferencing applications with queueing theory, there is a considerable need for modeling a QoS-efficient video traffic analysis. In this paper, we have modelled the cloud-based VC traffic considering an $M/G/1$ queue model with a heavy-tailed service time distribution. This paper demonstrates that we can model practical problems in the real-time video conferencing systems and evaluate the system performance. Numerical analysis clearly demonstrates the key performance measures for the queueing theory $E[W]$, $E[Wq]$, $E[N]$ and $E[Nq]$. Ongoing research in this direction bears the study of Markov-modulated Bernoulli process (MMBP) in the cloud-based VC, as this has been extensively used to model bursty traffics such as VC with time-correlation. For load balancing and redundancy, the future work will also consider a cloud-based VC system modeling using more than one transcoding servers. In the MMBP based system modeling, with heavy-tailed distributions of the service time, our study will focus on the queue overflow probability based on some approximations in addition to the study of the mean delay.

REFERENCES

- [1] P. Mell and T. Grance, "The NIST Definition of Cloud Computing," National Inst. of Stan and Tech., Gaithersburg, MD, Sep. 2011.
- [2] K. Kilkki, "Quality of experience in communications ecosystem," in *Journal of Universal Computer Science*, 2008, vol.14(5), pp.615-624.
- [3] W. Zhu, C. Luo, J. Wang, and S. Li, "Multimedia cloud computing," in *IEEE Signal Processing Magazine*, 2011, vol. 28(3), pp. 59-69.
- [4] D. Ghose and H. J. Kim, "Scheduling video streams in video-on-demand systems: A survey," in *Multimedia Tools Applications*, 2000, vol. 11(2), pp. 167-195.
- [5] M.-Y. Wu and W. Shu, "Optimal scheduling for parallel CBR video servers," in *Multimedia Tools Applications*, 2001, vol. 14 (1), pp. 79-99.
- [6] M. Ponc, S. Sengupta, M. Chen, J. Li, and P. A. Chou, "Optimizing multi-rate peer-to-peer video conferencing applications," in *IEEE Transactions on Multimedia*, 2011, vol.13(5), pp.856-868.
- [7] J. Li, P. A. Chou, and C. Zhang, "Mutualcast: an Efficient Mechanism for Content Distribution in a P2P Network," in *Proceedings of ACM SIGCOMM Asia Workshop*, 2005.
- [8] Y. Feng, B. Li, and B. Li. Airlift, "Video conferencing as a cloud service using inter-datacenter networks bibtex," in *Proceedings of IEEE ICNP*, 2012.
- [9] E. Kurdoglu, Y. Liu, and Y. Wang, "Dealing with user heterogeneity in P2P multi-party video conferencing: Layered distribution versus partitioned simulcast," in *IEEE Transactions on Multimedia*, 2016, vol. 18(1), pp.90-101.
- [10] M. Chen, M. Ponc, S. Sengupta, J. Li, and P. A. Chou, "Utility maximization in peer-to-peer systems with applications to video conferencing," in *IEEE/ACM Transaction on Networking*, 2012, vol.20(6), pp.1681-1694.
- [11] X. Chen, M. Chen, B. Li, Y. Zhao, Y. Wu, and J. Li, "Celerity: A lowdelay multi-party conferencing solution," in *ACM Multimedia*, 2011, pp. 493-502.
- [12] M. H. Hajiesmaili, L. T. Mak, Z. Wang, C. Wu, M. Chen, and A. Khonsari, "Cost-effective low-delay cloud video conferencing," in *IEEE 35th International Conference on Distributed Computing Systems*, 2015, pp.103-112.
- [13] S. Cicalo and V. Tralli, "Distortion-fair cross-layer resource allocation for scalable video transmission in OFDMA wireless networks," in *IEEE Transaction on Multimedia*, 2014, vol.16(3), pp.848-863.
- [14] M. Shiraz, A. Gani, R. H. Khokhar and R. Buyya, "A Review on Distributed Application Processing Frameworks in Smart Mobile Devices for Mobile Cloud Computing," in *IEEE Communications Surveys & Tutorials*, 2013, Vol.15(3), pp.1294-1313.
- [15] A. Mokhtari, M. Azizi and M. Gabli, "Optimizing management of cloud resources towards best performance for applications execution," in *International Conference on Embedded & Distributed Systems*, 2017, pp.1-5.
- [16] M. Ponc, S. Sengupta, M. Chen, J. Li, and P. Chou, "Multi-rate Peerto-Peer Video Conferencing: A Distributed Approach using Scalable Coding," in *IEEE International Conference on Multimedia and Expo*, 2009.
- [17] L. Wei, J. Cai, C. H. Foh, and B. He, "QoS-Aware Resource Allocation for Video Transcoding in Clouds," in *IEEE Transactions on Circuits and Systems for Video Technology*, 2017, vol.27(1), pp.49-61.
- [18] S. Kim and P. K. Varshney, "An Integrated Adaptive Bandwidth-Management Framework for QoS-Sensitive Multimedia Cellular Networks," in *IEEE transaction on Vehicular technology*, 2004, pp. 835-846.
- [19] W. Sugeng, J. Istiyantp, K. Mustofa and A. Ashari, "The Impact of QoS Changes towards Network Performance" in *International Journal of Computer Networks and Communications Security*, 2015, pp. 48-53.
- [20] A. D. Tesfamichael, V. Liu, E. Foo, W. Caelli, "Modeling for performance and security balanced trading communication systems in the cloud," in *IEEE International Performance Computing and Communications*, 2017, pp.1-7.
- [21] V. Liu, A. D. Tesfamichael, W. Caelli, T. Sahama, "Network security metrics and performance for healthcare systems management," in *International Conference on E-health Networking*, 2015, pp.14-17.
- [22] A. D. Tesfamichael, V. Liu, E. Foo, W. Caelli, "Design and implementation of unified communications as a service based on the OpenStack cloud environment," in *IEEE International Conference on Computational Intelligence & Communication Technology*, 2015, pp.117-122.