International Journal of advancement in electronics and computer engineering (IJAECE)
Volume 11, Issue1, Jan.. 2022, pp.141-150, ISSN 2278 -1412
Copyright © 2012: IJAECE (www.ijaece.com)

HierTrain: Fast Hierarchical Edge AI Learning with Hybrid Parallelism in Mobile-Edge-Cloud Computing

[1] Reeta Kushwaha , [2]Chhatrapani Gautam
[1]Research Scholar, CSE, VITS Satna (M.P.), India
[2]Assi. Prof. , CSE, VITS Satna (M.P.), India

Abstract

Mobile-edge cloud computing is a new paradigm to provide cloud computing capabilities at the edge of pervasive radio access networks in close proximity to mobile users. In this paper, we first study the multi-user computation offloading problem for mobile-edge cloud computing in a multi-channel wireless interference environment. We show that it is NP-hard to compute a centralized optimal solution, and hence adopt a game theoretic approach for achieving efficient computation offloading in a distributed manner. We formulate the distributed computation offloading decision making problem among Mobil e device users as a multi-user computation offloading game. We analyze the structural property of the game and show that the game admits a Nash equilibrium and possesses the finite improvement property. Then design a distributed computation offloading algorithm that can achieve a Nash equilibrium, derive the upper bound of the convergence time, and quantify its efficiency ratio over the centralized optimal solutions in terms of two important performance metrics. We further extend our study to the scenario of multi-user computation offloading in the multi-channel wireless contention environment. Numerical results corroborate that the proposed algorithm can achieve superior computation offloading performance and scale well as the use r size increases

Keywords: Edge AI, Deep Learning, Fast Model Training, Mobile-Edge-Cloud Computing.

I. Introduction

Mobile Edge Computing (MEC) is a new technology which is currently being standardized in an ETSI Industry Specification Group (ISG) of the same name. Mobile Edge Computing provides an IT service environment and cloud-computing capabilities at the edge of the mobile network, within the Radio Access Network (RAN) and in close proximity to mobile subscribers. The aim is to reduce latency, ensure highly efficient network operation and service delivery, and offer an improved user experience. Mobile Edge Computing is a natural development in the evolution of mobile base stations and the convergence of IT and telecommunications networking. Based on a virtualized platform, MEC is recognized by the European 5G PPP (5G Infrastructure Public Private Partnership) research body as one of the key emerging technologies for 5G networks (together with Network Functions Virtualization (NFV) and Software-Defined Networking (SDN)). In addition to defining more advanced air interface technologies, 5G networks will leverage more programmable approaches to software networking and use IT virtualization technology extensively within the telecommunications infrastructure, functions, and applications. MEC thus represents a key technology and architectural concept to enable the evolution to 5G, since it helps advance the transformation of the mobile broadband network into a programmable world and contributes to satisfying the demanding requirements of 5G in terms of expected throughout, latency, scalability and automation. MEC is based on a virtualized platform, with an approach complementary to NFV: in fact, while NVF is focused on network functions, the MEC framework enables applications running at the edge of the network. The infrastructure that hosts MEC and NFV or network functions is quite similar; thus, in order to allow operators to benefit as much as possible from their investment, it will be beneficial to reuse the infrastructure and infrastructure management of NFV to the largest extent possible, by hosting both VNFs (Virtual Network Functions) and MEC applications on the same platform.

Copyright © 2012: IJAECE (www.ijaece.com)

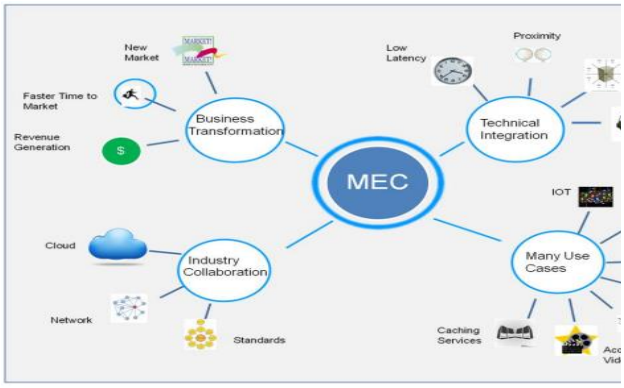

Figure 1: MEC

The environment of Mobile Edge Computing is characterized by low latency, proximity, high bandwidth, and real-time insight into radio network information and location awareness. All of this can be translated into value and can create opportunities for mobile operators, application and content providers enabling them to play complementary and profitable roles within their respective business models and allowing them to better monetize the mobile broadband experience. Mobile Edge Computing opens up services to consumers and enterprise customers as well as to adjacent industries that can now deliver their mission-critical applications over the mobile network. It enables a new value chain, fresh business opportunities and a myriad of new use cases across multiple sectors. The intention is to develop favourable market conditions which will create sustainable business for all players in the value chain, and to facilitate global market growth. To this end, a standardized, open environment needs to be created to allow the efficient and seamless integration of such applications across multi-vendor Mobile Edge Computing platforms. This will also ensure that the vast majority of the customers of a mobile operator can be served.

I.2 Cloud Computing

Cloud computing is becoming one of the next IT industry buzz words: users move out their data and applications to the remote “Cloud” and then access them in a simple and pervasive way. This is again a central processing use case. Similar scenario occurred around 50 years ago: a time-sharing computing server served multiple users. Until 20 years ago when personal computers came to us, data and programs were mostly located in local resources. Certainly currently the Cloud computing paradigm is not a recurrence of the history. 50 years ago we had to adopt the time-sharing servers due to limited computing resources. Nowadays the Cloud computing comes into fashion due to the need to build complex IT infrastructures. Users have to manage various software installations, configuration and updates. Computing resources and other hardware are prone to be outdated very soon. Therefore outsourcing computing platforms is a smart solution for users to handle complex IT infrastructures. At the current stage, the Cloud computing is still evolving and there exists no widely accepted definition. Based on our experience, we propose an early definition of Cloud computing as follows: A computing Cloud is a set of network enabled services, providing scalable, QoS guaranteed, normally personalized, inexpensive computing infrastructures on demand, which could be accessed in a simple and pervasive way.

Figure 2: Cloud Computing

II. Method

In the below flow, we have explained each block in the following way and on the basis

Copyright © 2012: IJAECE (www.ijaece.com)

of this method, we have worked in this paper.

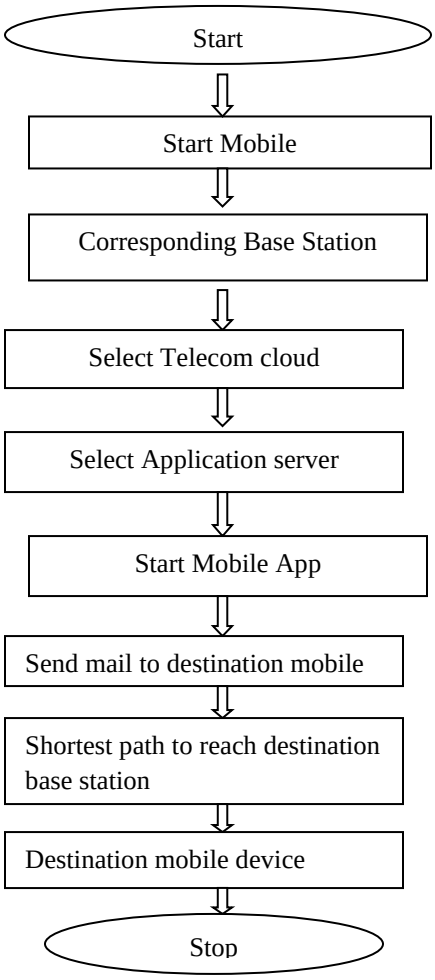

Figure 3: Data Flow Diagram

Start Mobile

The history of mobile phones covers mobile communication devices that connect wirelessly to the public switched telephone network.

While the transmission of speech by signal has a long history, the first devices that were wireless, mobile, and also capable of connecting to the standard telephone network are much more recent. The first such devices were barely portable compared to today's compact hand-held devices, and their use was clumsy.

Drastic changes have taken place in both the networking of wireless communication and the prevalence of its use, with smartphones becoming common globally and a growing proportion of Internet access now done via mobile broadband.

The conduct businesses has transformed from what it used to be 10 years ago. With the ever-changing customer needs and user behavior along with advancements in the landscape of communications driven by trends such as AI, ML, and NLP, 5G, and IoT, the need to continue innovating with cloud technologies to meet customer needs is just getting started and businesses are finding ways to keep their customers engaged in order to stay ahead of the curve.

Corresponding Base Station

Base station (or base radio station) is – according to the International Telecommunication Union's (ITU) Radio Regulations (RR) – a "land station in the land mobile service."

The term is used in the context of mobile telephony, wireless computer networking and other wireless communications and in land surveying. In surveying, it is a GPS receiver at a known position, while in wireless communications it is a transceiver connecting a number of other devices to one another and/or to a wider area. In mobile telephony, it provides the connection between mobile phones and the wider telephone network. In a computer network, it is a transceiver acting as a switch for computers in the network, possibly connecting them to a/another local area network and/or the Internet. In traditional wireless communications, it can refer to the hub of a dispatch fleet such as a taxi or delivery fleet, the base of a TETRA network as used by government and emergency services or a CB shack.

Select Telecom Cloud

A telecom cloud represents the data center resources that are required to deploy and manage a mobile phone network with data transfer capabilities by carrier companies in production operations at scale. Telecom clouds have traditionally been based in private data center facilities which are used to manage the telecommunication requirements of 3G/4G and LTE networks. With the current roll-out of 5G equipment across the mobile service provider community internationally, vendors have adopted strategies related to network function virtualization

Copyright © 2012: IJAECE (www.ijaece.com)

(NFV) and software-defined data center (SDDC) management. This makes the deployment of required operating software to carriers more efficient.

The inherent advantages of VMware's cloud orchestration software platforms which include vSphere, the ESXi hypervisor, NSX distributed firewalls, and other products are all combined into the vCloud NFV suite. These products streamline 5G mobile data network installation for telecom cloud requirements. Network function virtualization (NFV) greatly accelerates the speed at which 5G networks carriers can launch, as well as reducing the number of trained staff needed for deployment. Accordingly, many experts now consider NFV and SDN to be key characteristics of the telecom cloud by definition.

Select Application server

An application server is a mixed framework of software that allows both the creation of web applications and a server environment to run them.

It can often be a complex stack of different computational elements running specific tasks that need to work as one to power multiple clouds and web-based software and application.

Sitting between the primary web-based server tier and the back-end tier of a database server, the application server is essentially a go-between for the database server and the users of the business or consumer apps it supports through putting various protocols and application programming interfaces (APIs) to use.

An application server is designed to install, operate and host applications and associated services for end users, IT services and organizations and facilitates the hosting and delivery of high-end consumer or business applications.

Depending on what is installed, an application server can be classified in a number of ways, such as a web server, database application server, general purpose application server or enterprise application server.

It's commonly paired with a web server or contains a web server, which means the two can be converged and named a web application server. It is also versatile enough to be used with other application servers simultaneously.

Application servers can also contain their own graphical user interfaces for management through PCs, but they can also take care of their own resources, as well as transaction processing, messaging, resource and connection pooling, and performing security tasks.

Start Mobile App

A mobile app is a software application developed specifically for use on small, wireless computing devices, such as smartphones and tablets, rather than desktop or laptop computers. Mobile apps are designed with consideration for the demands and constraints of the devices and also to take advantage of any specialized capabilities they have. A gaming app, for example, might take advantage of the iPhone's accelerometer. Mobile apps are sometimes categorized according to whether they are web-based or native apps, which are created specifically for a given platform. A third category, hybrid apps, combines elements of both native and Web apps. As the technologies mature, it's expected that mobile application development efforts will focus on the creation of browser-based, device-agnostic Web applications.

Send Mail To Destination Mobile

You're at your computer, and your friend is out with their phone. You want to send them a message, and your phone is dead.You could send an email, fire off a Facebook message, or hit them up on Twitter. They use iMessage, right? All of these methods are available on a desktop. But if you're talking to someone who isn't carrying around a smartphone, these options don't work. What then? Simple---send an email to their phone number. This works with virtually ***any*** SMS-capable phone, whether it runs apps or not, thanks to SMS gateways.

Shortage Path To Reach Destination Base Station

The shortest path is the problem of finding a path between two vertices (or nodes) in a graph such that the sum of the weights of its constituent edges is minimized. The problem of finding the shortest path between two intersections on a road map may be modeled as a special case of the shortest path problem in graphs, where the vertices correspond to intersections and the edges correspond to road segments, each weighted by the length of the segment.

The shortest path problem can be defined for graphs whether undirected, directed, or mixed. It is defined here for undirected graphs; for directed graphs the definition of path requires that consecutive vertices be connected by an appropriate directed edge.

Copyright © 2012: IJAECE (www.ijaece.com)

Two vertices are adjacent when they are both incident to a common edge. A path in an undirected graph is a sequence of vertices $P = (\upsilon_1, \upsilon_2 ....., \upsilon_n) EV \times V \times .... \times V$ such that υ_i is adjacent to υ_{i+1} for $1 \leq i \prec n.$ Such a path P is called a path of length $n-1$ from υ_1 to υ_n (The υ_i are variables; their numbering here relates to their position in the sequence and needs not to relate to any canonical labeling of the vertices.)

Let $e_{i,j}$ be the edge incident to both υ_i and υ_j Given a real-valued weight function $f : E \to \Re$ and an undirected (simple) graph G, the shortest path from υ to υ' is the path $P = (\upsilon_1, \upsilon_2 ....., \upsilon_n)$ (Where $\upsilon_1 = \upsilon$ and $\upsilon_n = \upsilon'$) that over all possible n minimizes the sum $\sum_{i=1}^{n-1} f(e_{i,i+1})$. When each edge in the graph has unit weight or $f : E \to \{1\}$ this is equivalent to finding the path with fewest edges.

Destination mobile device

A mobile device (or handheld computer) is a computer small enough to hold and operate in the hand. Typically, any handheld computer device will have an LCD or OLED flat screen interface, providing a touchscreen interface with digital buttons and keyboard or physical buttons along with a physical keyboard. Many such devices can connect to the Internet and interconnect with other devices such as car entertainment systems or headsets via Wi-Fi, Bluetooth, cellular networks or near field communication (NFC). Integrated cameras, the ability to place and receive voice and video telephone calls, video games, and Global Positioning System (GPS) capabilities are common. Power is typically provided by a lithium-ion battery. Mobile devices may run mobile operating systems that allow third-party applications to be installed and run.

Early smartphones were joined in the late 2000s by larger, but otherwise essentially the same, tablets. Input and output is now usually via a touch-screen interface. Phones/tablets and personal digital assistants may provide much of the functionality of a laptop/desktop computer but more conveniently, in addition to exclusive features. Enterprise digital assistants can provide additional business functionality such as integrated data capture via barcode, RFID and smart card readers. By 2010, mobile devices often contained sensors such as accelerometers, magnetometers and gyroscopes, allowing detection of orientation and motion. Mobile devices may provide biometric user authentication such as face recognition or fingerprint recognition.

For Nonhomogeneous Image Dehazing, we employ a Multi-patch and Multi-scale network. We go over these two architectures in depth in this section.

III.1. Multi-patch Architecture:

Deep Multi-patch Hierarchical Networks are used (DMPHN). DMPHN was designed with Single Image Deblurring in mind. In this research, we use a DMPHN version. The architecture will be discussed in the following sections for completeness' sake. The DMPHN architecture is multi-level. Each level has a pair of encoders and decoders. Each level has a varied quantity of patches to work with. The number of patches used in DMPHN(1-2-4) is 1,2 and 4, correspondingly, from top to bottom. Only one patch per image is considered at the highest level (level-1). The image is divided into two sections vertically in the next level (level-2). The patches from the previous level are further divided horizontally in the bottom-most level (level 3), resulting in a total of four patches. Consider the hazy image I H as an example. I H i,j denotes the j-th patch in the i-th level. I H is not separated into any patches in level one. I H is separated vertically into I H 2,1 and I H 2,2 in level 2. I H 2,1 and I H 2,2 are split horizontally in level 3.

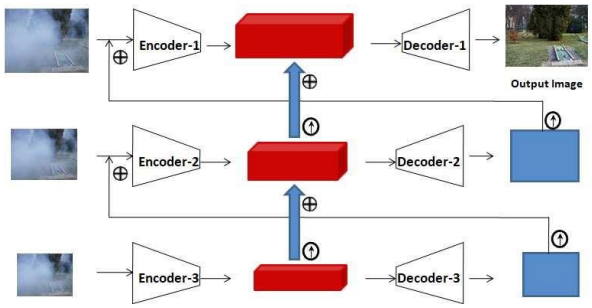

Figure 4: Architecture diagram of Deep Multi-Patch Hierarchical Network

To create 4 patches, I_3^H ,1 I_3^H ,2 I_3^H ,3 and I_3^H ,4 Encoders and Decoders at i-th level is denoted as Enci and Deci respectively. In DMPHN, information flows from the bottom up. Patches at the lowest level are passed into Enc3's encoder, which generates feature maps.

$$F3,j = Enci(I_3^H, j), \forall j \in [1, 4] \quad (2)$$

Copyright © 2012: IJAECE (www.ijaece.com)

To create a new feature representation, we concatenate geographically neighbouring feature maps.

$$P_{3,j} = [F_{3,2j-1}, F_{3,2j}], \forall j \in [1, 2] \quad (3)$$

Concatenation is represented by [.,.]. Dec3 is used to decode the new concatenated features

$$Q_{3,j} = Dec_3(P_{3,j}), \forall j \in [1, 2] \quad (4)$$

The decoder output is added with patches in the next level and fed to encoder. F2,j = Enc2(I H 2,j + Q3,j), ∀j ∈ [1, 2] (5)

The encoder outputs are combined with the preceding level's decoder inputs. The feature maps that result are then spatially concatenated.

$$F_2^*,j = F_{2,j} + P_{3,j}, \forall j \in [1, 2] \quad (6)$$

$$P2 = [F_2^*,1, F_2^*,2] \quad (7)$$

P2 is then loaded into Dec2 to build level-2 residual feature maps.

$$Q_2 = Dec_2(P_2) \quad (8)$$

The output of the level-2 decoder is combined with the input image and routed via Enc1. At level 2, Q2, encoder output F1 is combined with decoder output Q2.

$$F1 = Enc_1(I^H + Q_2) \quad (9)$$

To generate the final dehazed output I, F1 is combined with P2 and sent to Dec1.

$$P_1 = F_1 + P_2 \quad (10)$$

$$\hat{I} = Dec_1(P1) \quad (11)$$

3.2. Multi-scale Architecture:

We also experiment with a multi-scale architecture. We name this architecture Deep Multi-scale Hierarchical Network(DMSHN). The details of the architecture are described as follows.

Input hazy image I^H is downsampled by factor of 2 and 4 to create an image pyramid. We call these downsampled images F_1^* and $I_{0.25}{}^H$ respectively. The architecture consists of 3 levels where each level has a pair of encoder and decoder. Encoder and decoder at level i is denoted as Enci and Deci respectively. At the lowest level $I_{0.25}{}^H$ is fed to encoder Enc_3 to obtain feature map F_3 and is further passed through decoder Dec_3 to feature representation P_3.

$$F_3 = Enc_3(I_{0.25}{}^H) \quad (12)$$

$$P_3 = Dec_3(F_3) \quad (13)$$

P_3 is upscaled by factor of 2 and added to $I_{0.25}{}^H$ and passed through encoder Enc_2 to generate F_2^*. Encoder output from

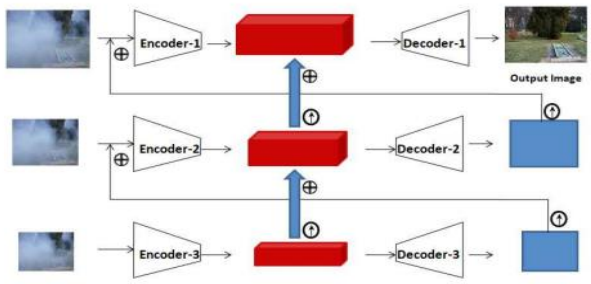

Figure 5: Architecture diagram of Deep Multi-Scale Hierarchical Network. ⊕ denotes Upsampling by factor of 2 and L denotes residual addition.

previous level is upscaled and added to intermediate feature map F_2^* and fed to the decoder Dec2.

$$F_2^* = Enc2(I_{0.5}{}^H + up(P_3)) \quad (14)$$

$$F_2 = F_2^* + up(F3) \quad (15)$$

$$P_2 = Dec_2(F_2) \quad (16)$$

where up(.) denotes Upsampling operation by a factor of 2. Residual feature map P2 from level-2 is added to the input hazy image and fed to encoder Enc1. Encoder output is added with upscaled F2 and passed through decoder to synthesize the dehazed output

$$\hat{I}. F_2^* = Enc_2(I H + up(P_2)) \quad (17)$$

$$F_1 = F_1^* + up(F2) \quad (18)$$

$$I = Dec_1(F_1) \quad (19)$$

For No homogeneous Image Dehazing, we employ a Multi-patch and Multi-scale network. We go over these two architectures in depth in this section.

3 Results

Here we show proposed method working output.

Copyright © 2012: IJAECE (www.ijaece.com)

Figure-6 : Initial Architecture

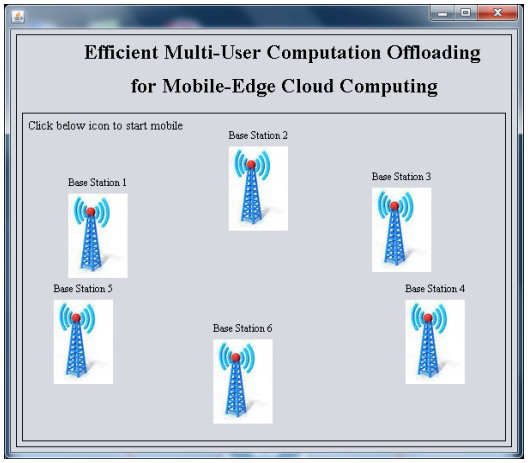

Figure-7: Different Section

In Efficient Multi-User Computation Offloading for Mobile-Edge Cloud Computing, first of all we will press the click button, and then our mobile will be started. After that, all the icons of the given mobile will be visible on the screen.

All the icons can be seen in the above Fig- 7, 5 icons are visible in the given picture, this icon is divided into section-1 section-2 section-3 section-4 and section-5. When we press on these icons our mobile will start.

Figure-8: Cloud Server 1

Figure-9: Cloud server 2

When we will start Telecom cloud service, first we will select the given cloud then after that click on check availability.

Copyright © 2012: IJAECE (www.ijaece.com)

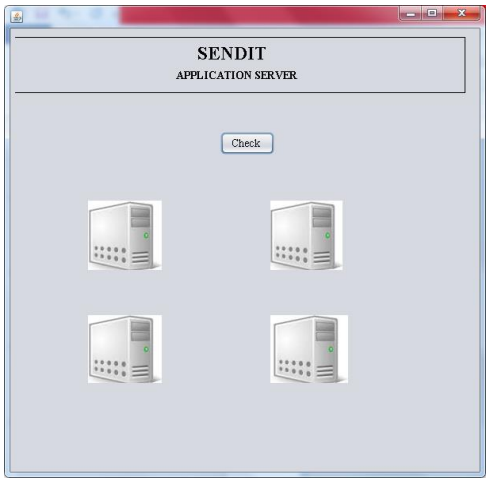

Figure-10 Application server 1

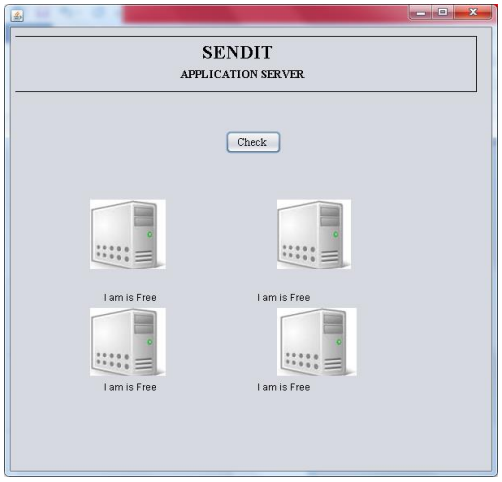

Figure-11 Application server 2

Figure -12 Share with Friend

Figures -13 Share with Friend 2

After downloading all the servers that are visible in Figure 10 and Figure-11 in the application server, after that, we will check the server by clicking on the check button.

In Sendit We will check the sever; the application form has to be filled as given in Figure-12 and Figure-13, after that we can check the server.

After the set transmitter and receiver server we initialize our proposed architecture and send data.

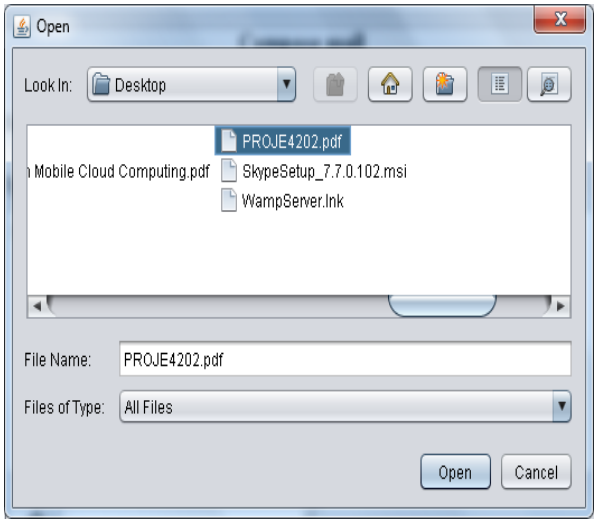

Figure-14: File open

Copyright © 2012: IJAECE (www.ijaece.com)

Figure- 15 Send File

4 Conclusions

A game theoretic approach for the computation offloading decision making problem among multiple mobile device users for mobile-edge cloud computing.

We formulate the problem as a multi-user computation offloading game and show that the game always admits a Nash equilibrium. We also design a distributed computation offloading algorithm that can achieve a Nash equilibrium, derive the upper bound of convergence time, and quantify its price of anarchy.

Numerical results demonstrate that the proposed algorithm achieves superior computation offloading performance and scales well as the user size increases.

REFERENCES

[1] Deyin Liu; Xu Chen; Zhi Zhou; Qing Ling "HierTrain: Fast Hierarchical Edge AI Learning With Hybrid Parallelism in Mobile-Edge-Cloud Computing" 2021.

[2] Silvana Trindade, Luiz F. Bittencourt, Nelson L. S. da Fonseca "Management Of Resource At The Network Edge For Federated Learning" 2021.

[3] Zahra Makki Nayeri Toktam Ghafarian Bahman Javadi "Application placement in Fog computing with AI approach: Taxonomy and a state of the art survey" 2021.

[4] Jun Zhu and Yushen Wang "Parallel Implementation of Swarm Intelligent Algorithms in a Spark-Based Cloud Computing Environment" 2021.

[5] Gayashan "Distributed Data Stream Processing and Task Placement on Edge-Cloud Infrastructure"2021.

[6] Fatsuma Jauro, Haruna Chiroma, Abdulsalam Y. Gital , Mubarak Almutairi , Shafi'i M. Abdulhamid, Jemal H. Abawajy "Deep Learning Architectures in Emerging Cloud Computing Architectures: Recent Development, Challenges and Next Research Trend" 2020.

[7] Shuai Yu, Xu Chen, Zhi Zhou, Xiaowen Gong, and Di Wu "When Deep Reinforcement Learning Meets Federated Learning: Intelligent Multi-Timescale Resource Management for Multi-access Edge Computing in 5G Ultra Dense Network" 2020.

[8] Wazir Zada Khana, Ejaz Ahmedb , Saqib Hakakb , Ibrar Yaqoobc , Arif Ahmedd "Edge Computing: A Survey" 2020.

[9] Zicong Hong, Wuhui Chen, Huawei Huang, Song Guo, and Zibin Zheng "Multi-hop Cooperative Computation Offloading for Industrial IoT–Edge–Cloud Computing Environments"2019.

[10] Abdulrahman Alreshidi, Aakash Ahmad, Ahmed B. Altamimi, Khalid Sultan and Rashid Mehmood "Software Architecture for Mobile Cloud Computing Systems" 2019.

[11] Li Lin, Xiaofei Liao, Hai Jin, And Peng Li "Computation Offloading Toward Edge Computing" 2019.

[12] DadmehrRahbari, Mohsen Nickray "Computation Offloading and Scheduling in Edge-Fog Cloud Computing" 2019.

[13] AbdulrahmanAlreshidi, AakashAhmad , Ahmed. Altamimi , Khalid Sultan and Rashid Mehmood "Software Architecture for Mobile Cloud Computing Systems" 2019.

[14] Sa´ul Alonso-Monsalve, F´elixGarc´ıa-Carballeira, Alejandro Calder´on "A Heterogeneous Mobile Cloud Computing Model for Hybrid Clouds" 2018.

Copyright © 2012: IJAECE (www.ijaece.com)

[15] Jiale Zhang , Bing Chen, Yanchao Zhao , Xiang Cheng , And Feng Hu "Data Security and Privacy-Preserving in Edge Computing Paradigm: Survey and Open Issues" 2018.

[16] AhmetCihatBaktir, AtayOzgovde, and CemErsoy "How Can Edge Computing Benefit from Software-Defined Networking: A Survey, Use Cases & Future Directions" 2017.

[17] Yun Liu1 Ming-Ming Cheng1 Xiaowei Hu1 Kai Wang1 Xiang Bai "Richer Convolutional Features for Edge Detection" 2017.

[18] Míriam Bellver Bueno, Xavier Giró-i-Nieto, Ferran Marqués, Jordi Torres "Hierarchical Object Detection with Deep Reinforcement Learning" 2017.

[19] OleksandrLemeshko ,OleksandraYeremenko, OlenaNevzorova "Hierarchical Method Of Inter-Area Fast Rerouting" 2017.