# A Study of Edge Computing Enabled IoT Systems

**Thesis submitted in partial fulfillment**
**for the Award of Degree**

*Doctor of Philosophy*

**by**

*Sumit Kumar*

*DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING*
**INDIAN INSTITUTE OF TECHNOLOGY**
**(BANARAS HINDU UNIVERSITY)**
**VARANASI - 221005**

*Roll No. 17071007* *Year 2022*

# CERTIFICATE

It is certified that the work contained in the thesis titled ***"A Study of Edge Computing Enabled IoT Systems"*** by ***Sumit Kumar*** has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all requirements of Comprehensive Examination, Candidacy, and SOTA for the award of Ph.D. Degree.

**Supervisor**

**Dr. Ruchir Gupta**
Associate Professor,
Department of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi,
Uttar Pradesh, INDIA 221005.

# DECLARATION BY THE CANDIDATE

I, ***Sumit Kumar***, certify that the work embodied in this Ph.D. thesis is my own bonafide work carried out by me under the supervision of ***Dr. Ruchir Gupta*** from ***July 2017*** to ***July 2022*** at ***Department of Computer Science and Engineering***, Indian Institute of Technology (BHU) Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, *etc.* reported in journals, books, magazines, reports, dissertations, theses, *etc.*, or available at websites and have not included them in this thesis and have not cited as my own work.

**Date:**

**Place:** Varanasi **(Sumit Kumar)**

# CERTIFICATE BY THE SUPERVISOR

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

**(Dr. Ruchir Gupta)**
Associate Professor,
Dept. of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi

**Signature of Head of Department**

# COPYRIGHT TRANSFER CERTIFICATE

**Title of the Thesis:** A Study of Edge Computing Enabled IoT Systems

**Name of the Student:** Sumit Kumar

## Copyright Transfer

**The undersigned hereby assigns to the Institute of Technology (Banaras Hindu University) Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the *Doctor of Philosophy.***

**Date:**

**Place:** Varanasi **(Sumit Kumar)**

**Note:** However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and the Institute's copyright notice are indicated.

**Dedicated to my Family**

# ACKNOWLEDGEMENT

First of all, I would like to take this opportunity to express my profound gratitude and sincere regards to my supervisor, Dr. Ruchir Gupta, Associate Professor, Department of Computer Science and Engineering, Indian Institute of Technology (BHU), Varanasi, for his excellent guidance, monitoring, and constant encouragement throughout this doctorate program. Also, a special thanks for creating such a great working atmosphere in the lab.

I am grateful to the faculty and staff members of the Computer Science and Engineering Department, Indian Institute of Technology (BHU), Varanasi. I extend my special gratitude to Dr. K. Lakshmanan, Assistant Professor, Department of Computer Science and Engineering, Indian Institute of Technology (BHU), Varanasi, for his technical, moral, and emotional support throughout my thesis. Further, I am obliged to Dr. Antriksh Goswami for his technical and moral support. Also, I want to express my deepest gratitude to Mr. Saurabh Arora, Mr. Vipin Maurya, Dr. Anshul Sharma, and Mr. Vaibhav Padhye for providing moral support throughout my thesis work. I have also spent some of the craziest and most memorable moments with them.

Finally, I would like to express my deepest gratitude to my parents, Mr. Anand Prakash and Mrs. Sharda Devi, for their constant support and encouragement. I am incredibly grateful to my wife, Varuna Tyagi, and my daughter, Jahnavi Tyagi, for their tremendous understanding and encouragement over the past few years. Without their support, it would be impossible for me. I am also grateful to my other family members who have supported me along the way.

**(Sumit Kumar)**

# Contents

# List of Figures

# List of Tables

# List of Symbols

| **Symbol** | **Description** |
|---|---|
| $A$ | Accuracy |
| $U$ | the app user set |
| $V$ | the app vendor set |
| $S$ | the edge server set |
| $d$ | the resource type belongs to $\{CPU, memory, storage, \cdots\}$ |
| $W_x^d$ | initial computing resources of type $d$ on edge server $x$ |
| $\omega_i$ | app user $i$'s resource requirement of type $d$ |
| $a_i$ | denotes the user $i$'s allocation decision |
| $A_i$ | contains all allocation decisions of user $i$ |
| $a = (a_i, a_{-i})$ | the allocation profile contains one decision of each user |
| $A$ | the collection of all allocation profiles |
| $a_{-i}$ | the allocation profile contains decisions except of user $i$ |
| $R_x^a$ | the set includes the users assigned to server $x$ at $a$ |
| $\zeta_x(R_x^a)$ | the edge server $x$'s usage cost at $a$ |
| $\varrho_i(a)$ | app user $i$'s usage cost at $a$ |
| $q_x(R_x^a)$ | the QoS-degradation cost incurred by a user on server $x$ at $a$ |
| $G_\lambda(R_x^a)$ | normalization function for QoS-degradation cost |
| $c_i(a)$ | the app user $i$'s total cost (usage + QoS-degradation) at $a$ |
| $C_\zeta(a)$ | the usage cost of all users at $a$ |
| $C_q(a)$ | the sum of QoS-degradation cost incurred by all users at $a$ |
| $C(a)$ | the total cost to all app users (usage + QoS-degradation) at $a$ |
| $T_j$ | $j^{th}$ cluster of edge servers |
| $\Phi_{\zeta_x}$ | the server $x$'s potential due to usage cost |
| $\Phi_{q_x}$ | edge server $x$'s potential due to QoS-degradation cost |
| $\Phi_x(a)$ | the edge server's overall potential at $a$ |
| $\Phi(a)$ | the game's overall potential at $a$ |

| **Symbol** | **Description** |
|---|---|
| $\beta_{j,x}$ | the allocated resources of $x$ to vendor $j$ |
| $\beta^{d}_{j,x}$ | the type $d$ allocated resources of $x$ to app vendor $j$ |
| $f_x$ | the fixed cost of edge server $x$ |
| $B(.)$ | the usage cost incurred by app vendor |
| $Q_x(R^a_x)$ | the overhead cost of a server $x$ |
| $B_x(R^a_x)$ | the net benefit of a server $x$ |
| $MAX^k_x$ | the capacity in user numbers of resource type $k$ on server $x$ |
| $MIN^{MAX}_x$ | the capacity in user numbers of bottleneck resource on server $x$ |
| $e$ | the edge between any pair of nodes |
| $\varsigma(e)$ | the cost of an edge $e$ |
| $p_i$ | the path selection by destination node $i$ |
| $P_i$ | the set includes all possible paths between source and node $i$ |
| $p = (p_i, p_{-i})$ | the path selection profile |
| $p_{-i} =$ | the path selection profile except for node $i$ |
| $S_e$ | the list contains destination nodes whose paths contain edge $e$ at any $p$ |
| $c^e_i$ | the cost incurred by node $i$ |
| $c_i(p)$ | the total cost incurred by $i$ to get data via $p_i \in p$ at $p$ |
| $\zeta(p)$ | the overall cost of all the destination nodes at any $p$ |
| $\lambda_{ex}$ | the excluded edges by any destination node during the path update |
| $\lambda_{in}$ | the included edges by any destination node during the path update |

# Abbreviations

| **Abbreviation** | **Description** |
|---|---|
| IoT | Internet of Things |
| BR | Best Response |
| QoS | Quality of Service |
| NE | Nash Equilibrium |
| PNE | Pure Nash Equilibrium |
| AUA | App User Allocation |
| UAGame | User Allocation Game |
| ERA | Edge Resource Allocation |
| ERAGame | Edge Resource Allocation Game |
| PoS | Price of Stability |
| MST | Minimum Steiner Tree |
| PSGame | Path Selection Game |

# Preface

Since the past decade, the world has seen exponential growth of the Internet of Things (IoT). These things integrate the cyber world with the physical world by sensing and collecting data from the surrounding environment and transmitting it to other devices over the Internet. The rapid growth of IoT fueled the development of advanced apps. Due to limited resources, many resource-hungry IoT apps cannot process sensory data on IoT devices. To address this issue, IoT app users can offload complex computational tasks to the cloud. In recent years, apps that require low latency have emerged, e.g., interactive gaming, natural language processing, face recognition, etc. However, due to unpredictable core network latency and expensive bandwidth, the cloud often fails to meet the stringent requirements of such latency-sensitive apps. Edge computing technology is proposed to address these challenges.

Edge computing reduces latency and saves bandwidth by deploying edge servers in different geographic locations. This technology plays a crucial role in the rapidly growing app market for IoT devices. In edge computing-enabled IoT, the app vendors hire resources from edge servers and allocate them to app users to overcome the challenge of the limited computing capacities of their IoT devices. An effective allocation of edge computing and networking resources to different apps or app users is needed because the resource allocation affects the system's performance in many aspects, as follows: 1) It affects the app users' Quality of Service (QoS) measured in different parameters such as packet loss, bit rate, throughput, transmission delay, availability, jitter, etc. 2) The resource allocation affects the benefits and the cost of the app vendors. 3) The resource allocation also impacts the resource utilization of the edge servers and the benefits of the edge infrastructure providers. This thesis investigates the edge resource allocation problem from the perspectives of all the actors.

In this thesis, we divide the edge resource allocation problem into two sub-problems: the computing resource allocation problem and the networking resource allocation problem. The problem of edge computing resource allocation is viewed in terms of offloading and executing computational tasks. We allocate networking resources for group communication between the edge server and various locations in the problem of edge networking resource allocation. The contributions of this thesis to solving the problem of resource allocation to app users are as follows:

- In the first work, we investigate the problem of allocating the app user to the hired resources of multi-tenant edge servers by the app vendor while establishing a trade-off between the users' QoS and cost. This problem is studied from the perspective of individual app vendors and their app users and is referred to as the App User Allocation (AUA) problem. We proposed a game-theoretic approach to solve the AUA problem that formulates this problem as a potential User Allocation Game (AUGame).
- In the subsequent work of allocating the edge computing resources, we consider how the different app vendors (services) compete for the same edge computing resources. Here, each app vendor intends to hire optimal bundle resources and serve the most app users at the lowest cost. This problem is referred to as the Edge Resource Allocation (ERA) problem. We propose a game-theoretic approach to solve this problem, which optimizes the cost incurred by the app vendors while maximizing resource utilization. In this approach, the ERA problem is formulated as a potential Edge Resource Allocation Game (ERAGame). This game employs an ERA algorithm to reach the solution faster.

- In the following work, we investigate resource allocation from the perspective of edge infrastructure and app users to maximize resource utilization while improving the users' QoS. We propose a distributed user allocation approach that finds the bottleneck resource on each multi-tenant edge server and balances the load on them. This approach accommodates the geographically dynamic density of app users by moving the app users from the overloaded edge servers to the underutilized edge servers, which efficiently utilizes the computing resources of the multi-tenant edge servers.
- The final work allocates the edge networking resources for group communication from an edge server to app users. A cost-effective multicast tree for group communication is required to use the edge networking resources efficiently and effectively. We consider multiple objectives for building a cost-effective multicast tree, e.g., high throughput, low delay, low energy consumption, etc. These multiple objectives can be efficiently combined in the form of edge costs. In this work, a weighted cost-sharing scheme is proposed to divide the cost of network edges among their users. We then proposed a game-theoretic approach that constructs a cost-effective multicast tree using the proposed cost-sharing method. While building the optimal tree, this approach aims to maximize the throughput and reduce the delay and energy consumption for data transmission.

According to the literature, the problems mentioned above are NP-Complete. Therefore, we find the approximate solution in polynomial time and examine its convergence and optimality. Theoretically, we demonstrate that the suggested solutions are much nearer to the centralized optimum. The numerical results substantiate the theoretical analysis.