Economic Incentive-based Schemes for Improving Data Availability in Mobile-P2P Environments

A dissertation submitted for the degree of

Doctor of Philosophy

New Delhi, INDIA, 2015–2016

Submitted by

Nilesh Padhariya

Department of Computer Science

Indraprastha Institute of Information Technology

Delhi, INDIA

February 2015

Dedication

I dedicate this dissertation to
my guru, H.D.H. P.P. Hariprasad Swamiji Maharaj,
my mentors, P. Tyagvallabh Swamiji, and P. Sarvatit Swamiji, and
my ATMIYA family for their constant support and unconditional love.
I love you all dearly.

Dissertation Certificate

This is to certify that the dissertation entitled **"Economic Incentive-based Schemes for Improving Data Availability in Mobile-P2P Environments"** submitted by **Nilesh Padhariya** to the Indraprastha Institute of Information Technology, Delhi (IIITD) for the award of the degree of Doctor of Philosophy is a bonafide record of research work carried out by him under our supervision. The contents of this dissertation, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Guide: Dr. Anirban Mondal
Senior Research Scientist
Xerox Research Centre, Bangalore, INDIA

Co-Guide: Prof. Sanjay Kumar Madria
Professor, Department of Computer Science
Missouri University of Science & Technology - Rolla, USA

Director: Prof. Pankaj Jalote
Professor, Department of Computer Science
Indraprastha Institute of Information Technology - Delhi, INDIA

February 2015

Declaration

I hereby declare that with effect from the date on which the dissertation is deposited in the Library of the Indraprastha Institute of Information Technology - Delhi, I permit the Librarian of the Institute to allow the dissertation to be copied in whole or in part with reference to me on the understanding that such authority applies to the provision of single copies made for study purposes or for inclusion within the stock of another library. IT IS A CONDITION OF USE OF THIS DISSERTATION THAT ANYONE WHO CONSULTS IT MUST RECOGNISE THAT THE COPYRIGHT RESTS WITH THE AUTHOR AND THAT NO QUOTATION FROM THE DISSERTATION AND NO INFORMATION DERIVED FROM IT MAY BE PUBLISHED UNLESS THE SOURCE IS PROPERLY ACKNOWLEDGED.

Nilesh Padhariya
PhD Scholar, Department of Computer Science
Indraprastha Institute of Information Technology - Delhi, INDIA

February 2015

Abstract

In a Mobile ad hoc Peer-to-Peer (M-P2P) network, mobile peers (MPs) interact with each other in a peer-to-peer (P2P) fashion. Proliferation of mobile devices (e.g., laptops, PDAs, mobile phones) coupled with the ever-increasing popularity of the P2P paradigm (e.g., KaZaa, Gnutella) strongly motivate M-P2P network applications. However, challenges such as free-riding, data accessibility and mobile resource constraints (e.g., energy) need to be addressed for realizing M-P2P applications. In particular, economic incentive schemes become a necessity to entice mobile peers to share their data, given the generally limited resources of mobile devices. Furthermore, in M-P2P networks, data availability is typically low due to rampant free-riding, frequent network partitioning and mobile resource constraints. Hence, this dissertation focuses proposes economic incentive-based approaches for effective data management in M-P2P networks.

In this regard, this dissertation makes the following key research contributions. First, we propose the economic incentive-based top-k query processing system in M-P2P networks. The system assigns rewards/penalties (payoffs) to MPs for incentivizing their participation and for enabling them to re-evaluate their data item scores for top-k query processing. Furthermore, we extend the system to incorporate the notion of a peer group-based economic incentive scheme. Second, we propose the system for improving data availability in M-P2P networks by incentivizing broker MPs to provide *value-added routing service*, which includes pro-active search for the query results

by maintaining an index of the data items (and replicas) stored at other MPs (as opposed to just forwarding queries). Moreover, the system also incentivizes relay peers to act as information brokers for improving data availability and efficient load sharing. Third, we propose the system for efficiently managing the vehicular traffic in road networks using economy-based reward/penalty framework with traffic congestion control. In particular, a user is rewarded for following system-assigned paths, while it is penalized for any deviations from the system-assigned paths. Finally, we present an economic incentive system for improving rare data availability by means of licensing (with group-based) replication in M-P2P networks.

Our performance evaluation demonstrates significant improvements in the processing of top-k queries in terms of query response times and accuracy at reasonable communication traffic cost, as compared to existing schemes. We also determine the number of brokers, beyond which the mobile peers are better off without a broker-based architecture i.e., they can directly access data from the data-providing peers. Furthermore, our performance study for E-VeT shows that it is indeed effective in managing vehicular traffic in road networks by reducing the average time of arrival and fuel consumption. Finally, our performance study indicates considerable improvements in query response times and availability of rare data items in M-P2P networks.

Keywords: Mobile-P2P networks, mobile computing, data management, data replication, top-k query processing, economic schemes and incentives

We are pleased to note that the contributions of this dissertation work have been published at following reputed conferences and journals.

- Nilesh Padhariya, Anirban Mondal, Vikram Goyal, Roshan Shankar and Sanjay Kumar Madria. "EcoTop: An Economic Model for Dynamic Processing of Top-k Queries in Mobile-P2P Networks." *Database Systems for Advanced Applications*, DASFAA(2) 2011:251-265

- Nilesh Padhariya, Anirban Mondal, Sanjay Kumar Madria, and Masaru Kitsuregawa. "Economic incentive-based brokerage schemes for improving data availability in mobile-P2P networks." *Computer Communications*, (2013) 36(8):861-874

- Nilesh Padhariya, Ouri Wolfson, Anirban Mondal, Varun Gandhi and Sanjay Kumar Madria. "E-VeT: Economic Reward/Penalty-based System for Vehicular Traffic Management." *Mobile Data Management*, MDM(1) 2014: 99-102

Moreover, the following papers are currently under review in journals.

- Nilesh Padhariya, Anirban Mondal, Sanjay Kumar Madria and Masaru Kitsuregawa. "Economic Incentive Schemes for Improving Availability of Rare Data in Mobile-P2P Networks"

- Nilesh Padhariya, Anirban Mondal and Sanjay Kumar Madria. "Top-k Query Processing in Mobile-P2P Networks using Economic Incentive Schemes"

Acknowledgments

I would like to thank to my PhD advisors, Professors Dr. Anirban Mondal and Dr. Sanjay Kumar Madria, for supporting me during these past four years. Anirban is someone you will instantly get inspired and never forget once you meet him. He's the funniest advisor and one of the smartest people I know. I hope that I could be as lively, enthusiastic, and energetic as Anirban and to someday be able to command an audience as well as he can. Sanjay has been supportive and has given me the freedom to pursue various projects without objection. He has also provided insightful discussions about the research. I am also very grateful to Professors Dr. Pankaj Jalote, Director, IIITD; Dr. Masaru Kitsuregawa, Professor, University of Tokyo, and Dr. Ouri Wolfson, Professor, University of Illinois at Chicago, for their scientific advice and knowledge and many insightful discussions and suggestions. I also have to thank the members of my PhD committee, Professors Dr. Pushpendra Singh, Assistant Professor, IIITD; Dr. Anupam Joshi, Professor, University of Maryland; and Dr. Mukesh Mohania, Director, IBM Research, India for their helpful career advice and suggestions in general.

I will forever be thankful to my former college research advisors, Professors Dr. Dheeraj Bhardwaj, Professor, IIT Delhi; and Dr. Kolin Paul, Professor, IIT Delhi. Dheeraj has been helpful in providing advice many times during my graduate school career. He was and remains my best role model for a scientist, mentor, and teacher. I still think fondly of my time as an graduate student in his lab. Dheeraj was the reason why I decided to go to pursue a career in research. His enthusiasm and love for teaching is contagious.

I'm also glad to have worked with a wonderful undergraduate student, Varun Gandhi for a couple of quarters last year. I value his friendship and support as well. He has helped me to formulate many experiments and to obtain insightful results for our research. I also thank Vivek Tiwari, Sweta Sachen, for also making this an interesting group. Each person listed has played some kind of role in my time here.

From the bottom of my heart, I especially thank my gurus and my friends. My hard-working gurus, P. Tyagvallabh Swamiji and P. Sarvatit Swamiji have sacrificed their lives for myself and provided unconditional love and care. I love them so much, and I would not have made it this far without them. My both the brothers, Parag and Prabhumay, and my dearest friends, Stavan and Donga have been my best friends all my life and I love them dearly and thank them for all their advice, support, and care. I know I always have my family to count on when times are rough.

Contents

List of Algorithms

List of Figures

List of Tables

List of Notations

B	Broker
L	Location
Q	Query
QI	Query-issuer
T_G	Top-k Global Ranking List
ECR	EConomic incentive-based rare data Replication scheme
ECR+	group-based EConomic incentive-based rare data Replication scheme
EIB	Economic Incentive-based Brokerage scheme
EIB+	enhanced Economic Incentive-based Brokerage scheme
ETG	Economic Top-k peer Group-based query processing scheme
ETK	Economic Top-K query processing scheme
M-P2P	Mobile Peer-to-Peer
MANET	Mobile Ad hoc NETwork
MP	Mobile Peer
P2P	Peer-to-Peer

1

Introduction

In a Mobile ad hoc Peer-to-Peer (M-P2P) network, mobile peers (MPs) interact with each other in a peer-to-peer (P2P) fashion [SF04, MMK09]. Proliferation of mobile devices (e.g., laptops, PDAs, mobile phones) coupled with the ever-increasing popularity of the P2P paradigm (e.g., KaZaa [Kaz06], Gnutella [Gnu]) strongly motivate M-P2P network applications. Mobile devices wirelessly communicating in a P2P fashion (e.g., Microsoft's Zune [Zun06]) facilitate M-P2P applications by enabling information sharing *on-the-fly*. Moreover, the proliferation of mobile devices with embedded GPS sensors coupled with the growth in the popularity of infotainment services for vehicles have created new avenues for improving vehicular traffic management in road networks. Thus, schemes for improving transportation system efficiency are becoming increasingly popular [AWX+11, SWYX11].

In M-P2P networks, an MP obtains the required information from the neighbouring MPs via short-range communications such as Bluetooth, 802.11g. For example, Alice wants to find the top-k restaurants with "lunch specials" (or "manager's special hours") within 1 km of her current location. Here, top-k can be determined based on the parameters (e.g., star rating, price and distance from the point of query reference) selected by the user. In a similar vein, another application could involve a parking lot, where MPs can collect information about available parking slots and charges based on their preferences such as *nearby* floor. The parking slot availability information has to be current and therefore the such information is *temporal* in nature. Similarly, an MP may want to find top-k stores selling Levis jeans in a shopping mall with criteria such as (low) price during a specific time duration. In a similar vein, people want to find others with similar interests (e.g., tennis, music) at a party. Figure 1.1 illustrates some of these application scenarios.

(a) Finding top-k restaurants with "lunch specials"

(b) Finding nearby car parking slots

Figure 1.1: Application scenarios for Mobile-P2P environments

On the other hand, in cases of the sudden events like an earthquake, users often get disconnected from the respective centralized base stations [NWT12]. Hence, a user in that area can either ask nearby people for information concerning shelters or distribute such information to other MPs in its vicinity. Such kNN queries can be processed in a mobile environment in a P2P fashion, where collaboration among the peers provides possibilities for them to interact in the absence of functioning communication infrastructure such as base stations. In a similar vein, M-P2P networks can also be used to do effective distribution of *rare* data items, which get sudden bursts in accesses based on *events*. For example, suppose a group of archaeologists is in the course of an expedition in a remote Amazonian forest, where communication infrastructures (e.g., base stations) do not exist. When there is a sudden unexpected decrease in temperature and gusty winds, they need to look for information about protective clothing such as shops selling sweaters and wind-cheaters, photos of such clothing and so on.

Additionally, suppose a group of adventure tourists *unexpectedly* encounters a cave during their journey. They would like to find information about where to buy gas-masks and associated safety equipment for added safety, video tutorials on how to use this equipment and so on. Similarly, when a tourist moving in a forest notices a rare bird, she may wish to find additional information about the bird and videos about its living habits. In a similar vein, due to the sudden and unexpected onset of a heat wave, a group of botanists on an expedition in a forest may want to find information such as non-drinking water sources and pictures/videos of the locations of such water sources. In these application scenarios, M-P2P interactions can facilitate the MPs in finding the required information.

Observe that such ad hoc queries are spatio-temporal in nature (e.g., parking slot availability information), hence they cannot be answered without obtaining information from other MPs. Incidentally, such P2P interactions, which facilitate *spatio-temporal* queries among MPs, are generally not freely supported by existing wireless communication infrastructures. Notably, this research will also contribute towards CrowdDB [FKK$^+$11], which uses human input via crowdsourcing to process queries that cannot be answered by database systems or search engines. The inherently ephemeral nature of M-P2P environments suggests that *timeliness* of data delivery is of paramount importance in these applications, thereby necessitating query deadlines. For example, an MP looking for top-k restaurants with "happy hours" would generally prefer to receive the answer within a specified deadline.

Incidentally, Amazon.com has developed Mechanical Turk [Ama05], which is an online marketplace for match-making between the requirements of businesses and the skill sets of developers. Developers can select from a large pool of tasks based on their skill sets. The Mechanical Turk system also provides economic incentives. Observe that technologies, such as WiFi and

Bluetooth networks, are nowadays adequately capable of providing a platform for incentive-based mobile P2P collaborations.

Data availability in M-P2P networks is typically lower than in fixed networks due to frequent network partitioning arising from peer movement and also due to mobile devices being autonomously switched 'off'. Rampant free-riding further reduces data availability i.e., most peers do not provide any data [HA05, KSGM03a]. (Nearly 90% of the peers in Gnutella were free-riders [AH00].) Observe that increased MP participation in providing service to the network would likely lead to better data availability, better data quality, higher available bandwidth and multiple paths to answer a given query. Incidentally, data availability is less than 20% even in a wired environment [SGG01].

Given the generally limited resources (e.g., bandwidth, energy, memory space) of MPs and the fact that relaying messages requires energy, the relay MPs may not always be willing to forward queries in the absence of any incentives, let alone search *pro-actively* for query results in order to ensure timeliness of data delivery. Thus, providing incentives for relay MPs to pro-actively search for query results becomes a necessity to improve data availability in M-P2P networks. Notably, many schemes such as incentive-based schemes, replication-based schemes and caching-based schemes can be used for improving data availability. Hence, we will provide an overview of these schemes in this dissertation.

1.1 Trends in mobile applications

Mobile devices have nowadays become handy computing tools for effective analysis and rapid sharing of information on-the-fly. This section examines the current *business* and *technological* trends across some of the mobile application areas.

Business trends

Ecommerce

Electronic Commerce is the sale and procurement of supplies and services using information systems technology. There are three steps of Ecommerce: (i) put marketing information on web (ii) allow online order taking (iii) construct electronic exchanges. Online shopping or online retailing are the examples of Ecommerce. Moreover, mobile e-commerce is exactly the same as e-commerce except that the access mechanism is via a wireless phone or terminal rather than the fixed telephone network. The research work in [LW08] shows how mobile e-commerce can be potentially useful in this new era of mobile technology. Consider mobile-P2P dissemination of merchant's sale and inventory information, which enables a customer with a smart-phone to locate a desired product at the time of entry into the mall. The system proposed in [LW08] motivates merchants to provide inventory/sale/coupons information electronically to nearby potential customers.

Medical healthcare

The healthcare environment comes with major constraints and requirements such as confidentiality of the medical data, privacy of the doctor-patient relationship and genuineness of the source of medical information. Such vital issues must be satisfied by the mobile environment [AMM03] in order to deploy the mobile applications for medical healthcare systems. Mobile applications related to biomedical information have been increasingly becoming attractive to the medical community as the small-screen devices (e.g., PDAs and smart-phones) permit the healthcare practitioners to access online biomedical resources anywhere at any time. Thus, the recent application proposals on mobile devices have been explored in [AMM03, PSS06].

Due to mobile resource constraints such as the limited display screens, the various information to be displayed on the same screen also tends to be limited. To overcome this barrier, [PSS06] proposed the system called multi-modal transcoder. This technique transcodes full-text biomedical information resource, which can be supported by various types of mobile devices. Furthermore, [PSS06] proposed a novel algorithm, which uses visual template matching and piglet detection process to understand the structure of biomedical resources. In this work, the usability study showed that the system's usability is improved by the simplification and summarization technique, and it is also useful to deliver the compressed information to the mobile user.

Moreover, the usage of wireless technologies play a significant role in telemedicine, which is also known as mobile-health. The work proposed in [XTL+11] shows that telemedicine is used to calculate, to communicate and to deliver high-quality medical care.

In developing regions such as Africa, where the rural areas still may not be benefited by the basic civilized facilities, especially, the health care and medications. Due to inappropriate healthcare facilities, sometimes people may lose their life in the rural areas. One of the solutions could be the wearable sensors, which can continuously monitors the patients and issues the warnings to the doctors or care takers by sending messages on their mobile devices. [RAT12] represents the decision support system, which collects data from various wearable sensors and analyzed this data for the variety of diseases. Furthermore, this result will be stored and sent to the required person as an Short Message Service (SMS). The proposed system in [RAP12] integrates the wearable sensors with mobile device and developed a platform, which continuously monitors the patient. In case of emergency, this system is capable of sending SMS on the doctor's mobile phone. This system collects data from the various wearable sensors in the mobile, analyze that data and send it to the centralized server.

Public services

Mobile devices became a part of the every user in the society. Hence, mobile applications, which are related to public services, are rapidly being developed to provide facilities to the society. Many public services such as transportation, weather etc. can be facilitated by mobile applications. For example, an extremely important public service is transportation. To facilitate the users in transportation application scenarios, mobile devices are capable of providing the following services to the user: (i) navigation (ii) provide current traffic situation such as congestion or accident (iii) suggest alternate pathways for user's traveling (iv) explore the spots of interest during a given user's itinerary. The works in [GK03, GK02] discuss a wide gamut of possibilities for deploying such mobile-based services for the benefit of the users.

[GK03] proposed the development of a travel guide, i.e., a mobile passenger guide, which helps the passengers to purchase electronic tickets using mobile terminals via Internet. Moreover, the travel guide works as an electronic ticket during the travel as well as guides the passengers via short messages. The system uses one personal database to work as an electronic ticket based on the user requirements. Similarly, the work related to the passenger support system had also been discussed previously in [GK02]. In this work, system allows user to make their travel plans and purchase necessary tickets by accessing the booking system via mobile computing devices. Such a system can be really helpful for public transport to provide easiness and quick access to the end-users. In addition to this, for visually handicapped passengers, the proposed system has been implemented and tested in a railway station.

In a similar vein, the research work in [LW08] shows the numerous uses of mobile-P2P databases. In particular, the work focuses on transportation problems. According to the research work of [LW08], Mobile P2P database software enables travelers to cooperate intelligently, which improves safety and mobility. When a vehicle encounters an accident, a congestion or a dangerous road surface, it will able to send “slow-speed message” to trailing vehicles, and this helps other drivers to make decisions such as finding alternative roads and also may help in avoiding pile-ups in some situations.

Moreover, [LW08] described a car-sharing application, which can potentially improve the efficiency of transportation systems. Notably, a car-sharing system can address issues such as environmental pollution, fuel consumption, public safety and congestion. Such car-sharing requires matchmaking and provision of information that is simultaneously relevant in time, location and interest.

Integration of mobile P2P databases with navigational devices and PDA’s can be used to disseminate information about relevant resources, like car-sharing partners, free parking slots and available taxicabs. The commercial purpose of Mobile P2P databases has also been shown in [LW08] e.g. airports have stores, kiosks and malls, where there is significant potential for information dissemination among Mobile-P2P users. Here, merchants can provide their location information and help users to search for products. Mobile-P2P interactions can facilitate in disseminating real-time information related to flight changes, delays, queue length, parking information, special security alerts and procedures and baggage information, which benefits both consumers and airport operators.

Technological trends

Near Field Communication (NFC)

More recently, e-Transaction widely adopted to simplify the process of transaction across various platforms such as e-ticket, e-commerce, etc. Moreover, the virtual money is involved in every virtual transaction. For example, shopping with virtual money becomes a common practice using NFC, especially, when we have smart phones.

A novice prototype for Train Ticket Application using NFC is shown in

[NHW12]. In this system, with the help of NFC enabled mobile device, the passenger gets the information about the vacant seats in the train and passenger feels like he/she purchasing this ticket because, a passenger can get the scanned copy of the e-ticket on the mobile device, when he/she completes the payment using the voucher system. When conductor approaches the passenger in the train, the passenger sends ticket data to the conductor and once the conductor receives the data from passenger, the ticket data will be destroyed from the passengers device.

Similarly, the work in [TZF12] demonstrates Thumb, a system to share information instantly on smart phones, especially for resources with extreme short life time. Using Thumb, a third-party can share the information with others, while the users get spontaneously notification about that. For example: any passenger wants to cancel the railway ticket just before one day of the train departure, and wants to sale ticket, then that user is able to advertise this information over the internet.

Data Dissemination

Data dissemination is a passive mode of communication in which the usage rate of information is much higher than the rate of information production [RR09] . In other words, it is an asymmetric communication, where downloading rate is much higher than uploading rate. Data dissemination related applications become very crucial due to the rapid growth of information generation and their distribution over communities. The recent social networking-based applications became a vital platform for the data dissemination. For managing such a wide variety of information and retrieving the specific information in real time in mobile environment, we need to have some semantics to be associated with the every piece of information. This also helps to distribute the information across the boxes of choices. The works on semantics-based information management on resource-constrained mobile environment have been proposed in [PCR+10, RHTA10, EHTA11].

The work in [EHTA11] proposed OntoWiki framework, which is the novel approach for semantics-based information management through mobile semantic collaboration. OntoWiki allows users to browse data in offline mode. This enables users to retrieve information even though they are in no-network area such as forest, villages etc. Such platform is very handy to collect semantically rich information like biodiversity expeditions to remote areas, where network connectivity is very low and discontinued or sometime totally unavailable.

1.2 Research challenges in M-P2P networks

Two major challenges associated with data management in M-P2P networks are **free-riding** and **data availability**. Other challenges include mobility and resource constraints (e.g., energy, bandwidth, processing power and memory) of mobile devices. As a consequence of peer mobility, the underlying physical ad hoc network keeps changing dynamically, thereby making it challenging to maintain a P2P overlay network for an optimal or reasonable topology [BCFN03]. Furthermore, peer mobility also causes frequent network partitioning, thereby leading to reduced data availability as well as decreased connectivity. Additionally, issues concerning privacy, security and trust also arise in M-P2P environments. These issues are discussed in [MK06].

Free-riding

Free-riding is defined as "A fundamental tension between *individual rationality* and *collective welfare*". In P2P networks, majority of the users generally choose to do free-riding i.e., they want to consume resources, but not contribute any resources of their own. For example, in Gnutella [AH00], 70% of the users share no data at all. Thus, free-riding is a rampant problem in P2P networks [RL03]. It causes limited growth of data in the system, which causes reduction in users' interest in accessing and providing services in the

P2P network. Consequently, over a period of time, it leads to system collapse. Notably, the problem of free-riding is further exacerbated in M-P2P environments due to mobile resource constraints such as energy, bandwidth and memory.

One solution to free-riding is to rely on altruism e.g., assume that all peers are generous and would contribute resources unselfishly. If social generosity is sufficiently high, there is no need for intervention [FC05]. However, in this case, users do not have any *incentive* to perform such altruistic acts. Another solution could be to provide incentives to the users for encouraging them to contribute to the M-P2P network. Here, we are considering the realistic assumption that users are selfish and respond to reward/punishment.

Data availability

Data availability in M-P2P networks is typically lower than in fixed networks due to frequent network partitioning [HM06] arising from peer movement, mobile resource constraints (e.g., bandwidth, energy, memory space) and mobile devices being autonomously switched 'off'. (Incidentally, data availability is less than 20% even in a wired environment [SGG01].) Rampant free-riding further reduces data availability since a large percentage of MPs are typically free-riders [HA05, KSGM03a, GBM01, LDHS05] i.e., they do not provide any data. Thus, economic incentive schemes become a necessity to entice resource-constrained MPs with incentives to provide data for answering queries.

1.3 Contributions of the Dissertation

The following are the contributions of this dissertation.

E-Top: Top-k Query Processing in Mobile-P2P Networks using Economic Incentive Schemes

A peer in M-P2P networks is used to have top-k queries. For example, someone wants to find the top-k restaurants with "happy hours" (or "manager's special hours") within 1 km of her current location. Here, top-k is determined based on the parameters (e.g., star rating, price and distance from the point of query reference) selected by the user. Similarly, another application could involve a parking lot, where MPs can collect information about available parking slots and charges, and then they can inform the brokers. The parking slot availability information has to be current and therefore, the broker can compare such current information with its current list of parking slots. The broker can then provide the top-k available slots to the query-issuing MP in terms of price or distance (from the MP's current location). Similarly, an MP may want to find the top-k stores selling Levis jeans in a shopping mall with criteria such as (low) price during a specific time duration.

Observe that such ad hoc queries are temporal in nature (e.g., parking slot availability information), hence they cannot be answered by the broker without obtaining information from other MPs. Notably, this research will also contribute towards CrowdDB [FKK$^+$11], which uses human input via crowdsourcing to process queries that cannot be answered by database systems or search engines. Additionally, such M-P2P interactions among peers are generally not freely supported by existing wireless communication infrastructures. The inherently ephemeral nature of M-P2P environments suggests that *timeliness* of data delivery is of paramount importance in these applications, thereby necessitating query deadlines. For example, an MP looking for top-k restaurants with "happy hours" would generally prefer to receive the answer within a specified deadline.

E-Top provides the efficient processing of such top-k queries in M-P2P networks using economic incentive schemes. E-Top issues economic rewards to the mobile peers, which send relevant data items (i.e., those that contribute to the top-k query result), and penalizes peers otherwise, thereby optimizing the communication traffic. Peers use the payoffs (rewards/penalties) as a means of feedback to re-evaluate the scores of their items for re-ranking purposes. In E-Top, brokers facilitate top-k query processing in lieu of a commission.

The main contributions of E-Top are three-fold:

1. It proposes two economic incentive schemes, namely ETK and ETK+, in which MPs act individually towards top-k query processing. These schemes assign payoffs to MPs for incentivizing participation and for enabling them to re-evaluate their data item scores.

2. It extends ETK and ETK+ to propose a peer group-based economic incentive scheme ETG, which defines three payoff allocation approaches.

3. It is indeed effective in improving the performance of top-k queries in terms of query response times and accuracy at reasonable communication traffic cost, as demonstrated by our performance evaluation.

E-Top also discourages free-riding due to its economic nature. ETK and ETK+ differ in that while ETK performs equal distribution of payoffs to the rankers, ETK+ uses a weighted distribution. In ETG, ad hoc groups of MPs are formed in the vicinity of the query location. Each group has a leader for coordinating the top-k query processing. In contrast with ETK and ETK+, where individual MPs directly send their top-k items to the broker, query processing in ETG proceeds by means of group members sending their individual top-k items to the group leader. The group leader selects (i.e., 'filters') the top-k items to be sent to the broker based on the relative frequencies of the items in the individual top-k lists. In our application scenarios, some of the restaurant managers in the vicinity of the query location can be the group leaders.

In our performance evaluation, as a baseline reference, we adapt a non-economic non-incentive-based existing top-k query processing scheme for

MANETs, proposed in [HHS+10], which is designated as **NETK (Non-Economic Top-K)** scheme. Furthermore, NETK does not incorporate the notion of item re-ranking as no feedback has been sent back to the MPs, who participated into top-k query processing. NETK is the closest to our top-k query processing schemes since it addresses dynamic top-k query processing in mobile networks. None of the existing proposals on economic issues addresses top-k query processing in M-P2P networks.

E-Broker: Economic Incentive-based Brokerage Schemes for Improving Data Availability in Mobile-P2P Networks

This work proposes the E-Broker system for improving data availability in M-P2P networks by incentivizing MPs to provide *value-added routing service.* Here, the term "value-added routing service" refers to the broker MPs enabling pro-active search for the query results by maintaining an index of the data items (and replicas) stored at other MPs (as opposed to just forwarding queries).

The main contributions of E-Broker are three-fold:

1. It proposes the EIB (Economic Incentive-based Brokerage) scheme, which incentivizes relay peers to act as information brokers for performing value-added routing and replication in M-P2P networks, thereby effectively improving data availability.

2. It proposes the EIB+ (enhanced Economic Incentive-based Brokerage) scheme, which extends the EIB scheme by incorporating three different broker scoring strategies for providing additional incentives to brokers towards providing better service. Brokers with higher scores become *preferred brokers* and they earn higher commissions than *common brokers.* EIB+ also facilitates load-sharing among the peers.

3. It experimentally determines the number of brokers, beyond which the

mobile peers are better off without a broker-based architecture i.e., they can directly access data from the data-providing peers.

Similar to E-Top, E-Broker also discourages free-riding in M-P2P networks. Both EIB and EIB+ use economic incentives in that every data item is associated with a *price* (in *virtual currency*). Data item price depends upon several factors such as access frequency, data quality and estimated response time of access. The query-issuer pays the price of the queried item to the data-provider, and a commission to the broker and the relay MPs in the successful query path.

We have evaluated the performance of EIB and EIB+ w.r.t. the non-economic **E-DCG+** replication scheme [HM06]. Notably, E-DCG+ is the closest to our schemes since it aims at improving data availability in MANETs. As a baseline, we also do performance comparison w.r.t. a non-incentive and non-broker-based **NIB** (Non-Incentive without Brokerage) scheme to show the performance gain due to brokerage.

E-VeT: Economic Reward/Penalty-based System for Vehicular Traffic Management

This work proposes the E-VeT system for efficiently managing the vehicular traffic in road networks using economy-based reward/penalty schemes. In this work, the cost of traversing a path in the road network corresponds to the time required for the path traversal, unless otherwise specified. Hence, we shall use the terms "**path cost**" and "**path time-cost**" interchangeably. Observe that defining path cost in terms of time encompasses factors such as path distance, the speed limit relevant to the path and the path's traffic congestion.

In E-VeT, base stations collaboratively facilitate dynamic vehicular route assignments for mitigating the traffic congestion, thus reducing the average time of arrivals and fuel consumption. However, vehicles may not follow the paths assigned by the base stations e.g., when they can find lesser-cost paths. To incentivize vehicles towards following the system-assigned paths, E-VeT uses **rewards/penalties (payoffs)**, which are in terms of *real currency.* Hence, these payoffs can be used towards paying road taxes, car registration, and license/toll fees.

This work assumes that all the vehicles fall under the purview of the E-VeT reward/penalty framework, which could be implemented as part of a government-mandated program for facilitating traffic management. Note that the proposed scheme is a government-mandated system, it is always operational, but only the price changes dynamically based on the congestion. Thus users, will not know the pricing scheme and congestion information well ahead of time. Though the system suggests and offers options to users, they still have a choice of paying the penalty and taking the higher-priced paths; the objective is not to force users for explicit load-balancing. Since we have a reward/penalty system, it is using incentives for load-balancing. It is also different and better than randomization where users can get one of the options, which they have to follow, and they have no choice to alter the option they received. Thus, in our scheme, we preserve the notion that a user is the final entity to decide the path taken.

This work can be seen as a further extension to the initial proposal for routing of the VS-scheme for parking introduced in [AWX+12]. In the VS-scheme, a central authority (CA) makes an optimal assignment, and penalizes vehicles severely for deviating from it. Furthermore, in the VS-scheme, the CA guarantees that each vehicle v will pay a travel-cost to slots that is not higher than v's cost in equilibrium. Since in an optimal assignment some vehicles may travel longer than in equilibrium, the CA compensates them in dollars so that the total cost that v pays is not higher than v's travel-time in equilibrium. The CA also charges vehicles that travel less in the optimum assignment than in the equilibrium assignment. This dollar-charge is equivalent to the saving in travel-time.

Our work here is mainly focused on routing in V2V different from parking of vehicles in [AWX+12] in terms of policies for route allocation of vehicles based on revenues, modeling the pricing problem for revenue generation and finding a suitable reward/penalty scheme that adapts to changing behavior of drivers over period of time. In addition, the performance metrics directly focus on the impact of different revenue allocation schemes on the average fuel saving, average time of arrival and the number of messages exchanged among others.

In summary, our proposed schemes differ from existing proposals [Bra96, Mor10, Xu06, Yan12, Iss11] in mainly two ways. First, we introduce a reward/penalty framework for controlling the traffic congestion. Second, users' good behavior (i.e., following the system advice) is considered in the congestion control decision-making in the sense that the system remembers past behavior and rewards/penalty earned in the past. Thus, our scheme is user-centric and it inspires users to earn rewards so that they can get preferred assignment of paths when needed by redeeming rewards.

The contributions of E-VeT are three-fold:

1. It proposes an R^2A (Revenue-based Route Allocation) scheme, which rewards vehicles for following system-assigned longer-time paths, and charges a fee for following system-assigned shorter-time paths. Furthermore, it penalizes (charges much higher fee) vehicles for any deviations from the system-assigned paths.

2. It presents the R^2A^+ (extended R^2A) scheme by incorporating the notion of *revenue-scales* for further incentivizing vehicles based on their past system usage.

3. It discusses a route allocation algorithm, which gives lesser-time paths as a preference to vehicles that have earned higher revenue based on the scheme used i.e., either R^2A or R^2A^+.

Note that both R^2A and R^2A^+ schemes are designed to ensure fairness in the sense that vehicles pay when they travel faster, and they earn currency when they travel slower. Both schemes penalize vehicles, which deviate from system-assigned paths, thereby incentivizing them to adhere to the system-assigned paths. Furthermore, when vehicles follow the system-assigned paths, they are rewarded either in terms of time-savings (i.e., lower time-cost routes being allocated) or in terms of real currency (i.e., payments for following longer time-cost routes).

R^2A and R^2A^+ differ in that while R^2A assigns payoffs to vehicles based on every individual journey, R^2A^+ performs the payoff assignment based on the *consistency* of a given vehicle in following the system-assigned paths across *multiple* journeys. To achieve this, R^2A^+ uses a set of pre-defined *revenue-scales* and provides better payoffs to the vehicles that are associated with higher revenue-scales. This entices vehicles to consistently follow the system-assigned routes. Our performance study shows that the proposed schemes are indeed effective in managing vehicular traffic in road networks by reducing the average time of arrival and fuel consumption.

E-Rare: Economic Incentive Schemes for Improving Availability of Rare Data in Mobile-P2P Networks

E-Rare focuses on handling *rare* data items in an M-P2P environment. *Rare* data items are those, which get sudden bursts in accesses based on *events* as they are only hosted by only a few peers in comparison to the network size. Thus, they may not be available within few hops of query-issuing peers. The sudden burst in accesses to rare items generally occurs within a given time-frame (associated with the event), before and after which such items are rarely accessed.

Some application scenarios are as follows. Suppose a group of college students in the course of an expedition in a remote forest, where communication infrastructures (e.g., base stations) do not exist. When there is a sudden unexpected decrease in temperature and gusty winds, they need to look for information about protective clothing such as shops selling sweaters and wind-cheaters, photos of such clothing and so on. In a similar vein, suppose a group of tourists *unexpectedly* encounters a cave during their journey. They would like to find information about where to buy gas-masks and associated safety equipment for added safety, instructional tutorials on how to use this equipment and so on. Similarly, when a motorist driving in a mountainous region, sees a rare animal, she may wish to find additional information about living habits. Additionally, due to the sudden and unexpected onset of a heat wave, a group of botanists on an expedition in a forest may want to find information such as non-drinking water sources and pictures of the locations of such water sources. In these application scenarios, M-P2P interactions can facilitate the MPs in finding the required information.

Such M-P2P interactions for effective sharing of rare data are currently not freely supported by existing wireless communication infrastructures. Observe how the sudden urgent demand of several MPs for information concerning rare items (e.g., protective clothing or gas-masks) arises due to the occurrence of *events* such as the sudden onset of harsh weather conditions or the users unexpectedly encountering a cave.

E-Rare is a *novel* economic incentive model for improving rare data availability by means of licensing-based replication in M-P2P networks. E-Rare comprises two replication schemes, namely ECR and ECR+, both of which use its incentive model for improving rare data availability. The key difference between these schemes is that in ECR, the MPs act individually towards replication, while for ECR+, the MPs perform replication in groups. In both these schemes, a given MP issues queries specifying its desired data item, its location and query deadline. In E-Rare, each data item is associated with four types of prices (in *virtual currency*), which provide different rights to the query-issuer concerning the usage of the item. E-Rare requires a query-issuer to pay one of these prices for its queried data item to the query-serving peer, thereby effectively increasing data availability and combating free-riders.

The main contributions of E-Rare are three-fold:

1. It provides incentives for replication of rare data items by means of a novel licensing mechanism, thereby improving rare data availability.

2. It provides additional incentives for MPs to collaborate in groups, thereby further improving rare data availability.

3. A detailed performance evaluation has been done to show the improvement in query response times and availability of rare data items in M-P2P networks.

Incidentally, virtual currency incentives are suitable for P2P environments due to the high transaction costs of real-currency micro-payments [TR04]. The works in [DPGB03,ET04,ZCY03] discuss how to ensure secure payments using a virtual currency. Notably, these secure payment schemes are complementary to our proposal, but they can be used in conjunction with our proposal.

We have performed a detailed performance evaluation of both ECR and ECR+. As a baseline reference, we have also compared against an existing non-incentive and non-economic replication E-DCG+ scheme for MANETs, proposed in [HM06], which is closure to our scenario. We have used average response times of queries, query success rates, query hop-counts and the number of messages as performance metrics. ECR+ outperforms ECR due to its group-based incentives (such as discounts), which facilitate collaborative replication among MPs. ECR outperforms E-DCG+ essentially due to its economic licensing scheme, which incentivizes MP participation in the creation of multiple copies of rare items. Both ECR and ECR+ incur more messages than E-DCG+ because in case of E-DCG+, a large percentage of unsuccessful queries result in decreased amount of data transfer, albeit at the cost of reduced query success rates.

1.4 Organization of the Dissertation

The dissertation is organized as follows:

Chapter 2 provides a general introduction and discusses about related work in this field.

Chapter 3 discusses our proposed system E-Top for top-k query processing in Mobile-P2P Networks using economic incentive schemes, when operating on resource constrained mobile devices. E-Top incorporates the economic incentive-based schemes to perform the effective query processing and to improve the data availability by means of increasing peer participation in M-P2P networks, which relies not just on the available peers but also on the query's answer-rate as well as its answer-quality at the query-issuing peer.

Chapter 4 provides a description of our second research component E-Broker, which looks at effective broker participation in M-P2P network towards serving better quality to the MPs by means of improving query response-time. The proposed brokerage schemes are based on various types of incentives to the broker-MPs towards serving M-P2P networks. We justify in this work that in a mobile environment, the nearly optimal broker's participation would effectively improves data availability thereby resulting faster response and better quality.

Chapter 5 discusses our third research component of this dissertation E-VeT, which focuses on efficient vehicular traffic management in road networks using economy-based reward/penalty schemes. In E-VeT, our goal is to identify that how effectively system can assign the paths to the vehicles to manage vehicular traffic by reducing the average time of arrival and fuel consumption.

Chapter 6 discusses our forth research component of this dissertation E-Rare, which focuses on leveraging network heterogeneity of rare data availability in M-P2P networks. In E-Rare, our goal is to identify that how effectively a rare data is adequately replicated on mobile peers to improve rare data availability while keeping its rarity (i.e., value of a data for a given time) up.

Chapter 7 concludes this dissertation with a summary of our contributions. We have also provided the directions for future work.

We are pleased to note that the contributions of this dissertation work have been published at following reputed conferences and journals.

- Nilesh Padhariya, Anirban Mondal, Vikram Goyal, Roshan Shankar and Sanjay Kumar Madria. "EcoTop: An Economic Model for Dynamic Processing of Top-k Queries in Mobile-P2P Networks." *Database Systems for Advanced Applications*, DASFAA(2) 2011:251-265

- Nilesh Padhariya, Anirban Mondal, Sanjay Kumar Madria, and Masaru Kitsuregawa. "Economic incentive-based brokerage schemes for improving data availability in mobile-P2P networks." *Computer Communications*, (2013) 36(8):861-874

- Nilesh Padhariya, Ouri Wolfson, Anirban Mondal, Varun Gandhi and Sanjay Kumar Madria. "E-VeT: Economic Reward/Penalty-based System for Vehicular Traffic Management." *Mobile Data Management*, MDM(1) 2014: 99-102

Moreover, the following papers are currently under review in journals.

- Nilesh Padhariya, Anirban Mondal, Sanjay Kumar Madria and Masaru Kitsuregawa. "Economic Incentive Schemes for Improving Availability of Rare Data in Mobile-P2P Networks"

- Nilesh Padhariya, Anirban Mondal and Sanjay Kumar Madria. "Top-k Query Processing in Mobile-P2P Networks using Economic Incentive Schemes"

2

Related Work

This chapter provides an overview of existing works related to economic schemes for data management in M-P2P environments. Notably, the combination of issues such as node mobility, free-riding, network partitioning and resource constraints (e.g., energy, memory space) are more relevant to M-P2P environments, although some of these issues may also arise in other environments. As a single instance, in static P2P environments, the issue of node mobility does not arise and resource constraints are not as severe as in M-P2P environments. The free-riding issue in traditional static P2P environments may be handled by blocking the free-riders. However, in M-P2P environments, in order to have connectivity in the network, we need to attract free-riders to provide services.

2.1 Economic incentive schemes

This section provides an overview of economic incentive schemes.

2.1.1 Economic schemes for resource allocation in P2P networks

Economic schemes have been discussed for resource allocation in distributed systems [FNY93, FYN88, KS89]. The proposed model in [FNY93] manages distributed data objects by means of revenue-based resource allocation. In this economic model, each job/transaction pays to the processor for performing data object-related read/write operations, while the processor uses revenues to lease copies of data objects to other processors. Here, the price of the data objects are decided by themselves. Thus, the economic model facilitates efficient management of data objects in the system.

The work in [FYN88] discusses microeconomics load-balancing in distributed systems. In this economic model, the prices of the resources in the system are decided based on their demands. The system assumes that the demand generates the competition among the non-cooperative devices in the distributed system. In particular, each job is assigned to a host using an auction mechanism i.e., the winning host serves the job. This economic approach has been shown to provide better performance due to its limited complexity and intrinsic decentralization.

In a similar vein, [KS89] also examined the effective distribution of divisible resources in a distributed system. For example, the file allocation based on communication cost and its processing time are optimized by dividing a file into several parts distributed fashion. Thus, a file is processed by executing each of its individual parts on different processors. [Gro03] has described the economic aspects in a pure peer-to-peer (P2P) networks. They considered GNUnet file-sharing, which is a distributed framework that uses trust-based mechanisms for performing economic resource allocation among equal peers, while minimizing free-riding. Thus, it defined a game of cooperative players to maximize individual outcomes, while distinguishing between friendly and malicious players. However, this work does not incorporate any pricing model.

The proposals in [LI04, XLN06b, XLN06a] discuss economic schemes for resource allocation in wireless ad hoc networks. The goal of [LI04] is to perform service provisioning in a mobile ad hoc network (MANET) environment, hence it proposes a distributed algorithm for effective service provisioning. Moreover, it uses the Vickrey auction mechanism for allocating services among selfish peers in the MANET. The proposal [XLN06b] assumes that the peers would prefer to be selfish rather than cooperative in wireless ad hoc networks. Hence, incentive mechanisms would inspire selfish and greedy peers to participate and share their resources, thereby reducing free-riding. This work presented the price-pair mechanism to allocate resources across peers, while incentivizing them for their services to the network, thereby further increasing the extent of cooperative behavior in the system. By means of cooperation, the system is converged to the desired global optimal operating point, even though the peers are independent and the system is decentralized. Similarly, the proposal [XLN06a] provides the price-based resource allocation framework for fair and optimal resource utilization among the peers. The price of each resource is computed based on the notion of 'maximum cliques' of that resource in the wireless ad hoc network. They also showed that the proposed distributed algorithm converges to a global network optimum w.r.t. resource allocations.

Observe that the schemes in [FYN88,KS89,FNY93,Gro03] do not address M-P2P issues such as node mobility, free-riding, frequent network partitioning and mobile resource constraints. Furthermore, the schemes in [LI04,XLN06b, XLN06a] do not incorporate replication or data rarity issues.

2.1.2 Incentive schemes for encouraging peer participation in static P2P networks

Incentive-based schemes for encouraging peer participation in static P2P networks involve formal game-theoretic approaches such as the proposal in [GBM01]. The work in [GBM01] proposed an incentive-based model for

P2P file-sharing systems. They provided various payment schemes under a game-theoretic model in which each selfish user is trying to increase his rewards. The work also analyzed the equillibria of users' strategies with respect to micro-payment and quantized micro-payment mechanisms.

The works in [HA05, RL03] encourage peer participation using utility functions. [HA05] aimed at encouraging resource sharing by providing incentives to the peers and they also address the issue of malicious peers. [HA05] defined utility functions based on credit i.e., contributed bytes - consumed bytes, for measuring users' behavior. However, they do not address data quality i.e., a given user may share fake files to increase his credit. [RL03] also uses utility functions to measure the user's value to the system. It described three different utility functions based on (a) the total number of files shared (b) the total size of the data shared and (c) the popularity of the shared files. These utility functions consider reward (i.e., user's sharing) and penalty (i.e., user's consumption).

Furthermore, [KSGM03a] has also addressed free-riding in P2P networks and proposed a solution based on a peer participation metric called the EigenTrust score [KSGM03b]. The parameters in the EigenTrust metric are used for defining the incentives that are to be awarded to the peers. EigenTrust ensures that participatory peers obtain rewards, but less active peers do not get excluded from the system, thereby providing opportunity to the free-riders to be active participants in the network. Moreover, [LDHS05] considers the asymmetric angle of the P2P system, where the uploaded bytes have less value then the downloaded bytes because selfish and rational peers would prefer downloads as opposed to uploads. Hence, [LDHS05] provides more incentives for the uploaded bytes w.r.t. the downloaded bytes.

Observe that the approaches [HA05,RL03,KSGM03a] are too static to be deployed in M-P2P networks because they assume peers' availability and fixed topology. Furthermore, they do not address few mobile resource constraints (e.g., energy) and data rarity issues.

The works in [Azz10a, Azz10b] focus on addressing the problem of free-riding in decentralized collaborative environments. In particular, these works propose a taxonomy for classifying and tracking free-riders in multimedia systems based on trustworthiness. The proposal in [LGS12] addresses free-riding in the popular eMule/eDonkey P2P file-sharing network by evaluating and improving the fairness policy, which rewards contributors. Notably, the works in [LGS12, Azz10a, Azz10b, HA05, KSGM03a] do not address replication and mobile resource constraints.

2.1.3 Incentive schemes for combating free-riding in MANETs

The proposals in [BH01, BH03, ZCY03, CN04, CGKO03, SNCR03] address free-riding in MANETs. The work in [BH01] introduces a virtual currency to stimulate node cooperation for packet forwarding services. They provide two different models, namely the Packet Purse Model (PPM) and the Packet Trade Model (PTM), for defining the price of the packet forwarding service. In PPM, nuglets (a form of virtual currency) are loaded into a packet by the source, and intermediate nodes takes off these nuglets from the given packet according to their respective prices for the forwarding service. On the other hand, in PTM, each node in the message path 'buys' a packet from the predecessor node and 'sells' it to the successor node, thereby implying that the destination node has to pay the price for the packet forwarding service. Observe that PTM is vulnerable to network overload, since senders do not have to pay. Thus, PPM is more promising. The work also proposed the hybrid and extended PPM (with Fixed Per Hop Charges and Auctions) approaches for packet forwarding service.

Similarly, the works in [BH03, ZCY03] also use virtual currency to stimulate the cooperation of mobile nodes in forwarding messages. In [BH03], virtual currency is defined as a 'nuglet counter' on each peer, which is decreased by one in case of peer has generated a packet (i.e., peer wants to obtain service from the network) and is increased by one in case a peer forwards a packet (i.e., the peer serves the network). In order to use the message-forwarding service, the value of nuglet counter on a peer must be positive, thereby facilitating the avoidance of free-riding. The work also suggested mechanisms for protecting the nuglet counter. [ZCY03] proposed Sprite, a simple cheat-proof, credit-based mechanism to provide message-forwarding service among selfish nodes in a MANET. This work considered incentivizing peers for their cooperation as follows: a node keeps a receipt of its receiving/forwarding message; later it clears with Credit Clearance Service (CCS) provider by uploading receipts; and obtains currency for its services. This mechanism does not require any tamper-proof hardware at any node.

The auction-based iPass [CN04] incentive scheme and the works in [CGKO03, SNCR03] also provide incentives for relaying messages. iPass pays to each flow i.e., the message path from source to destination with relay peers as the intermediate nodes, for the message-forwarding service in a non-cooperative MANET environment. The resource allocation is performed by bidding, and a flow is chosen by a generalized Vickrey auction with reverse pricing mechanism. It shows that truthful bidding of utility is a dominant strategy and incentivization leads to higher social welfare for the whole network. The work in [CGKO03] explores an incentive-based model for a MANET. A mobile node can earn as much as it is capable of transmitting messages, but its capacity is constrained by its remaining energy, hence the system will be balanced. [SNCR03] considers a market-based approach, in which each node independently charges prices for relaying the data packets. The method effectively converges to the equilibria of the resulting market due to its iterative price and rate allocation algorithm. In particular, the proposals in [CGKO03, SNCR03] concentrate on compensating forwarding cost in terms of battery power, memory and CPU cycles.

However, these works do not consider M-P2P architecture. Furthermore, they do not consider data rarity issues, data item prices and incentives for data replication i.e., they do not entice peers to host data. Also, they do not address the issue of creating pro-active mobile peers for providing value-added routing services.

2.1.4 Incentive schemes for M-P2P networks

The work in [MMK07c] discussed ABIDE, which is an auction-based economic model for M-P2P networks. In ABIDE, the relay mobile peers are encouraged to provide value-added routing services in lieu of a commission. ABIDE also considers load-balancing issues. Furthermore, the work in [MYM10] proposed EcoBroker, which is a novel economic incentive-based brokerage model for improving data availability via replication for multiple-item queries in Mobile-P2P networks. In EcoBroker, data requestors need to pay the price (in virtual currency) of their requested data items to data-providers. The economic incentive model of EcoBroker effectively combats free-riding by incentivizing MPs to become brokers and to host replicated data, thereby improving data availability. Moreover, its brokerage model facilitates efficient processing of queries involving multiple data items.

The proposals in [XWR06, WXS04] discuss incentive schemes for combating free-riding in M-P2P networks. The work in [XWR06] provides incentives to mobile peers for participation in the dissemination of reports about resources in M-P2P networks. Each disseminated report contains information concerning a spatial-temporal resource e.g., availability of a parking slot at a given time and location. The work in [WXS04] considers opportunistic resource information dissemination in transportation application scenarios. A mobile peer transmits its resources to the mobile peers that it encounters, and obtains resources from them in exchange. The works in [WXS04, XWR06] primarily address data dissemination with the aim of reaching as many peers as possible i.e., they focus on how every peer can get the data. However, they do not incentivize relay peers to perform value-added routing and to host data. Furthermore, they do not consider licensing-based data replication and data rarity issues.

The work in [MMK09] proposes an economic incentive model for the efficient processing of constraint queries in M-P2P networks, given that M-P2P users may issue queries with varying constraints on query response time, data quality of results and trustworthiness of the data source. The focus in [MMK09] is on how to index the constraints in user queries by using the CR*-tree. Furthermore, the work in [MMK09] provides incentives for peers to form collaborative peer groups for maximizing data availability and revenues by mutually allocating and deallocating data items using royalty-based revenue-sharing. Thus, the focus in [MMK09] is completely different from the focus of E-Broker in that E-Broker focuses on brokerage schemes for performing value-added routing and replication (and load-sharing) in M-P2P networks.

2.1.5 Economic schemes for top-k query processing

The proposal in [SIC08] addresses top-k queries and aggregate queries for probabilistic databases with focus on data uncertainty and semantics. Uncertainty imposes probability as a new ranking dimension that does not exist in the traditional settings. This work has novel formulations based on traditional top-k semantics, which are combined with real-world semantics. The proposed framework supports query processing and indexing by encapsulating a state space model and an efficient search algorithm for computing query answers. The state space model divides the search space into small sub-spaces, thereby minimizing the number of accessed tuples and the size of the materialized search space.

The work in [HC07] examines the optimization of top-k queries in middleware by means of a cost-based optimization approach. The work incorporates various search and optimization algorithms. In particular, each top-k request is treated differently in the sense that their access costs vary. In contrast to relational queries, where "focused" search is possible through relational definitions, top-k queries are handled by using logical tasks as building blocks for identifying a comprehensive and focused search space. The work has defined several search schemes over a spectrum of possible algorithms to identify an optimal algorithm for a given top-k query.

The proposals in [LXL10,WXTL07] discuss top-k query processing in wireless sensor networks. In [LXL10], each top-k query retrieves k number of data objects, where the top-k evaluation is done by a scoring function on the queried features from sensor data. For conserving the energy of the sensor nodes, the work minimizes redundant data transmissions by means of both a cluster-tree routing structure for locally aggregating objects as well as a cross-pruning technique for filtering purposes. The work in [WXTL07] exploits semantics and facilitates energy-efficiency by installing a filter at each sensor node to avoid unnecessary updates.

The work in [LCLC04] uses a probabilistic approach towards cost-effectively selecting sensor nodes for processing continuous probabilistic queries in wireless sensor networks by reducing sensor data aggregation. The tutorial in [ZYV08] provides a comprehensive overview of top-k query processing in wireless sensor networks. The proposal in [HSHN09] discusses a message processing method for top-k queries in MANETs for reducing the communication traffic. The work in [JCCL10] discusses location-based top-k query processing for wireless broadcast environments using two R-tree variants, namely the broadcast aggregate R-tree and bit-vector R-tree. The work in [LLKL09] presents a search engine geared towards providing mobile users with top-k web search results. Notably, the proposals in [SIC08,HC07,LXL10,WXTL07, LCLC04,HSHN09,JCCL10,LLKL09] do not incorporate M-P2P architecture and economic schemes for incentivizing top-k query processing.

Incidentally, P2P replication suitable for mobile environments has been incorporated in systems such as ROAM [RRPK01], Clique [RNC03] and Rumor [GRR$^+$98]. However, these systems do not incorporate economic incentive schemes and top-k queries.

The work in [PMG$^+$11] addresses the processing of top-k queries in M-P2P networks by using economic incentive schemes. In the proposed economic model (designated as EcoTop), which is based on a super-peer architecture, brokers facilitate top-k query processing in lieu of a commission. EcoTop issues economic rewards to the mobile peers, which send *relevant* data items (i.e., those that contribute to the top-k query result), and penalizes peers for sending irrelevant items, thereby optimizing the communication traffic. Peers use the rewards/penalties as a means of feedback to re-evaluate the scores of their respective items for item re-ranking purposes. EcoTop also incorporates commissions for relay peers to incentivize them in forwarding messages quickly. A performance study demonstrates that EcoTop is indeed effective in improving the performance of top-k queries, while minimizing the communication traffic. Notably, this novel economic incentive model also discourages free-riding in M-P2P networks.

2.1.6 Payment schemes

A small study [MdRK04], which was conducted on users' motivation and decision to share resources in P2P networks, revealed that 50% of the questioned users would share more, if some materialistic incentives (e.g., money) are dispensed by the application. Herein lies the motivation for coupon-based systems like adPASS [SH04]. The works in [DPGB03, ET04, ZCY03] discuss how to ensure secure payments using a virtual currency. Another way proposed in [GA04] describes Coupons, an incentive scheme that is inspired by the eNcentive framework [RFJY03], which allows mobile agents to spread digital advertisements with embedded coupons among mobile users in a P2P manner.

Several non-repudiation [KMZ02, SS05] systems, which can be incorporated to control the deceiving behaviour of peers, have been developed. In many applications such as content distribution, the price can also be controlled by the service-providers [FST04].

MoB [CABP05] is an open market collaborative wide-area wireless data services architecture, which can be used by mobile users for opportunistically trading services with each other. MoB also handles incentive management, user reputation management and accounting services.

A bootstrap kind of mechanism can also be used in many applications [DHA03a]. Symella is a Gnutella file-sharing client for Symbian smartphones. It expects that illegal acts occur, such as interpolation or destruction of the distribution history to get incentives. Hence, the distribution history attached to the e-coupon [CN04] is enciphered with a public-key cryptographic system so that users cannot peruse the distribution history. Moreover, a message digest (MD) of the distribution history is embedded by digital-watermarking technology to check the validity of the history.

2.1.7 Trust-based schemes

The work in [QMK10] analyzes various existing decentralized and distributed trust management schemes. Based on this analysis, it proposes the M-trust scheme for mobile-P2P networks. M-Trust is a robust and scalable lightweight trust ratings aggregation scheme. Notably, the M-trust scheme also considers issues such as system mobility and dynamic network topology. In a similar vein, the work in [RSB11] proposes a generalized distributed trust management scheme to estimate peer trust based on their encounter history in different environmental contexts. Moreover, the work also discusses how to prioritize contexts depending upon the level of association with them. Furthermore, the proposal in [SL03] presents the TrustMe protocol for managing trust and anonymity in P2P environments. The work also demonstrates that the TrustMe protocol is reasonably secure against a wide variety of potential attacks.

The proposal in [AR10] examines the role of recommenders in P2P systems with the objective of managing trust. In particular, it provides an in-depth treatment of the feedback behavior of the recommenders as well as their role in trust assessment for P2P systems. Non-repudiation systems [SS05] can also be incorporated to control the deceiving behaviour of peers. The work in [BBS10] discusses an experimental model for trust and cooperation for partner selection in social networks.

The work in [VLdOC+10] proposes a human-based model for building a trust relationship between nodes in an ad hoc network. In particular, it proposes the Recommendation Exchange Protocol (REP), which enables nodes to exchange recommendations about their neighbors. Trust is based not only on previous individual experiences, but also on the recommendations of other nodes. Nodes maintain and exchange trust information about nodes within their respective radio ranges.

Notably, the trust-based schemes discussed above can be used in conjunction with our proposal as countermeasures to the selfish and deceiving behaviors of the peers.

2.2 Data Caching in mobile environments

To improve the response time of data retrieval in mobile environments, Data caching plays an important role. This section provides an overview of data caching schemes for mobile environments.

2.2.1 Cooperative caching

In a mobile environment, the mobile client can access data items from the cache of its neighbouring client. This concept is known as "cooperative caching". Notably, cooperative caching can also be used in conjunction with the P2P paradigm. [CLC04] proposes a cooperative caching scheme, designated as COCA, for mobile systems. COCA categorizes the mobile clients into two categories: Low Activity Mobile clients (LAM) and High Activity Mobile Clients (HAM). Notably, mobile clients from both of these categories share their respective caches. COCA reduces the server workload because the server replicates data items on the LAMs, while the HAMs take advantage of these replicas. Thus, COCA improves the overall system performance, reduces the number of requests as well as the access miss ratio when the mobile hosts are outside of the service region.

Wireless Sensor Networks support several applications such as environment control, intelligent buildings, and target tracking in battlefields. Over the past few years, Wireless Sensor networks have been growing in importance. To serve data in short latency and with minimal energy consumption, these applications require optimization in communication among the sensors. Hence, cooperative data caching protocols has been proposed. The selection of sensor nodes is at the heart of these protocols, and it plays an important role in making the caching and request forwarding decisions. The [DKTM11] introduces two new metrics to aid in the selection of such nodes. On the basis of these metrics, the work proposed two new cooperative caching protocols.

2.2.2 Techniques for maintaining cache consistency

In mobile database systems, if data is cached on a mobile host, it will reduce the query response time and also conserve the generally limited bandwidth. However, there is a need for *cache consistency*. A basic cache consistency scheme works as follows. The server broadcasts the invalidation report, which identifies the updated data objects so that the mobile hosts may remove the old data from their cache. Due to this reporting, the reconnecting process of a given mobile host may be slow as the mobile host requests the server for validating a cache as it receives an invalidation report. [KL01] proposes a set of new cache validation schemes, which are capable of conserving the bandwidth for cache validation as well as for query processing.

Caching is also useful for reducing the server load as it facilitates data access at clients, thereby improving the overall performance of the system. In mobile computing environments, there are chances of frequent disconnections. In such situations, coherence between servers and clients becomes a necessity. [ZCY06] proposes a category of cache invalidation strategy and mathematical model, and develops a high-performance caching technique.

Moreover, the work evaluates the performance for practical wireless mobile computing scenarios.

Furthermore, the cache invalidation methods are record ID-based, hence they are not adequate to manage the cache consistency of the mobile clients efficiently. [Chu08] proposes a cache invalidation scheme for continuous partial query in mobile computing environment, which is predicate-based. Here, the cache state of the mobile client is the predicate. The server broadcasts the cache invalidation report (CIR) and the predicate to the client for cache management. This method is useful for reducing the requirement of data for cache management. There are a number of methods to generate the CIR in the server and to identify the invalid data in the client.

Additionally, in dynamic environments, users may not always be able to stay in permanent contact with the network, but message delivery should be guaranteed for all active users of the network. [SPFT09] introduces two caching policies: basic caching and leaf caching for providing guaranteed message delivery.

2.2.3 Cache replacement strategies

While caching frequently accessed data items on the mobile clients improve the system performance, the *cache size* is generally limited. Hence, effective cache replacement techniques become a necessity to determine the set of data items that should be evicted from the cache. [KMS07] proposes a cache replacement policy called the Weighted Predicted Region-based Cache Replacement Policy (WPRRP) for location-dependent data. WPRRP works on the basis of client's movement by selecting the predicted region to calculate the weighted distance of a given item.

In a mobile computing environment, the mobile user uses cache to access the data easily, thereby enhancing the data availability as well as improving the data access time. Information is transferred from the server to the query-issuer depending on its current location. This is known as Location

Dependent Information Services (LDISs). [KSM10] proposes a cache replacement policy named Prioritized Predicted Region based Cache Replacement Policy (PPRRP), which uses a cost function for the data eviction based on the client's movement pattern.

[HXW$^+$05] proposes a proactive caching model, which caches the result objects along with the index that supports these objects as the results. This is helpful for object reusability for all common types of queries. To optimize the query response time, [HXW$^+$05] also proposes an adaptive scheme to cache an index. In mobile environments, proactive caching achieves significant performance gains as compared to page caching and semantic caching.

As the cache size is limited on mobile devices, there are number of cache replacement policies, used to discover a proper subset of items for eviction. The Euclidean distance and Euclidean space are important parameters for eviction in existing policies. In spatial networks, position and movement of the objects are constraints and network distance is an important measure. By considering the network density, network distance and the probability of access, [JPNS08] proposes a cache replacement policy which uses Progressive incremental network expansion (PINE) technique to calculate the network distance.

Moreover, [JYLK02] proposes a caching policy and broadcast scheme in which the geographical adjacency and characteristics of target area in Location Dependent Queries (LDQ) are reflected. By applying the moving distance of mobile host, [JYLK02] develop the caching policy suitable for urban area. The broadcast scheme uses the space-filling curve to cluster data based on adjacency of data in LDQ. The expectation is: when executing LDQ in local cache, the caching policy offers more accurate answers and significantly improves the workload of mobile hosts. Also, the broadcast scheme improves the battery life of the mobile host.

Mobile environment is dynamic, in which the mobile users are moving around a number of service areas. Notably, as the mobile user goes from one service area to another, the new server takes responsibility of that user. This process is known as handoff. In the process of handoff, the new server will not get benefit to access the cache. As a solution to this, [PC05] discovers numerous cache retrieval schemes to improve the cache retrieval efficiency. The use of 'coordinator buffer' shows the improvement in the cache retrieval. Moreover, Dynamic and Adaptive cache Retrieval scheme (DAR) is developed, which can deal with the service of handoff by utilizing proper cache methods according to specific criteria.

An adaptive per-user per-object cache consistency management (APPCCM) scheme is proposed in [LC11]. The scheme supports strong data consistency semantics through integrated cache consistency and mobility management in wireless mesh networks. Minimization of overall network cost is the main objective of APPCCM. In APPCCM, caching of data objects is done dynamically, depending on mesh client's mobility and data query/update characteristics and network conditions.

2.2.4 Semantic caching

The bandwidth of the mobile devices is also a challenge in developing large spatial database application on mobile environment. Here, the spatial data is used to process the query in mobile environments. [SZS05] attempted to combine multi resolution spatial data structure and semantic caching techniques for efficient processing of spatial queries. [SZS05] also proposed a new semantic caching model named Multi-resolution Semantic Caching (MSC) by considering the characteristics of multi-resolution spatial data and multi-resolution spatial query (MSQ) in mobile environments. MSC improves the performance in three ways: (a) a reduction in the amount and complexity of the remainder queries; (b) the redundant transmission of spatial data already residing in a cache is avoided; (c) a provision for satisfactory answers before 100% query results have been transmitted to the client side.

Furthermore, the two features of semantic cache, namely less network traffic and improved response time, make it efficient for mobile environments. The [LHC12] extends the traditional semantic cache management in three ways: (a) extension of quadtree-based index structures to semantic caches, (b) availability of a query processing strategy and (c) discussion on object-oriented implementation of the semantic cache.

2.3 Data Replication

Data replication means that the same data is stored at multiple nodes. Data replication is generally used for improving data availability, system reliability and performance. This section provides an overview of data replication schemes.

2.3.1 Data replication in P2P networks

Replication schemes for static P2P networks [BMSV03, DHA03b] and traditional replication strategies [KA00] for distributed systems do not consider peer mobility issues. The proposal in [BMSV03] tries to overcome the various failures that may potentially occur in P2P systems. The work suggested an analytical method based on reasoning about the efficiency of replication with redundancy to handle failure tolerance and its recovery albeit at a small scale. They also proposed a bulk replication scheme, in which the groups of files or file systems are replicated across the network for high data availability, hence the system has persistent storage failures and fast access.

The study in [DHA03b] addressed data consistency in P2P systems, where the data has been replicated at several peers. For maintaining data consistency, the update strategy needs to be capable of providing the same result throughout the network. The work has proposed a hybrid push/pull algorithm based on rumor spreading/gossiping mechanism. The algorithm provides an efficient and robust communication scheme for replication with high probability of consistency in a distributed environment. Moreover, [KA00] also addressed database replication rather than file replication with data consistency as a key objective. Several replication techniques for addressing data consistency in a distributed environment have been discussed in [DGMS85].

The work in [MLK04] shows how data replication impacts the performance of a static P2P system, where the data dependability issue is critical. In particular, the work in [MLK04] proposes a dynamic data replication strategy for effective load balancing among the peers and dynamic query redirection to reduce the query response time. The data to be replicated is decided by an individual peer based on access frequency i.e., a data with high access frequency is considered as 'hot' item and therefore a suitable candidate for replication. Moreover, the replicas of data items with relatively low access frequencies are periodically deallocated for optimizing the disk space of the peers. The work also considered the *distance* as a replication parameter. A given query is redirected based on the index available at each peer. Here, the index comprises the list of peer ids, which host replicas of a given data item. The performance shows that dynamic data replication scheme indeed outperforms the traditional replication schemes due to its effective load-balancing mechanism. Furthermore, the work in [MK05] discusses replication schemes for efficient data management in a wide area network (WAN) environment, where the major challenge is node heterogeneity in terms of processing capacity and storage. The work also addresses issues such as bandwidth variations, decentralized control, incomplete knowledge about the network, distributed ownership and scalability of WANs.

The work in [SYHN10] considers the reduction of delays i.e., interruption time in replica downloading in P2P streaming environments, but it is focused at a lower level networking layer. In this work, load is distributed across the clients by storing partial (pieces) of streaming data on them. Interruption time occurs due to client's disconnectivity or its low bandwidth. Hence, the work proposed a method to reduce this interruption time by considering the importance of the pieces of data based on their *immediacy* and *scarcity*. Peers receive more important pieces faster, hence more replication is done for the more important pieces. Increased replication of a data piece reduces its importance, hence eventually other data pieces also get opportunities to be replicated.

2.3.2 Data replication in MANETs

A network, where content exchange or delivery is done by autonomous peers, it becomes challenging to construct efficient distributed algorithms for content replication. This is due to the autonomy of the peers and their freedom to decide which objects they want to replicate. Additionally, churn (i.e., peers leaving the network autonomously) poses significant challenges to data availability.

The proposals in [HM06, HM05] discuss replication in MANETs. **E-DCG+** [HM06] creates groups of mobile peers (MPs) that are biconnected components in a MANET, and shares replicas in larger groups of MPs to provide high stability. An RWR (read-write ratio) value in the group of each data item is calculated as a summation of RWR of those data items at each MP in that group. In the order of the RWR values of the group, replicas of items are allocated until memory space of all MPs in the group becomes full. Each replica is allocated at an MP, whose RWR value to the item is the highest among MPs that have free memory space to create it.

The work in [HM05] aims at classifying different replica consistency levels in a MANET based on application requirements, and proposes protocols to realize them. In this work, each replica is valid till its original owner updates it. Hence, applying strict consistency updates may potentially degrade the system performance, given the inherently dynamic nature of the environment. Thus, the work assumes that all applications do not necessarily require such strict consistency, and it defines consistency based on group-level information consistency. For example, in case of a disaster management group, the information must be consistent within the group, but not strictly consistent w.r.t. to the other groups. Here, the local consistency maintenance within a given group is performed via quorums and it is based on local conditions such as location and time. Notably, the proposals in [HM06, HM05] do not consider any economic model, M-P2P architecture and data rarity issues.

Incidentally, P2P replication suitable for mobile environments has been incorporated in systems such as ROAM [RRPK01], Clique [RNC03] and Rumor [GRR+98]. ROAM, which is a system designed based on the Ward model [RPR96], satisfies a replication solution redesigned specifically for mobile environments. ROAM further considers replication factors such as local replication, appliance compatibility for replication and consistent updates throughout the network.

Clique, a server-less file system model, uses optimistic replication algorithms to store replicas in users' native file systems. It provides mechanisms for ensuring consistent updates (i..e., the replicas are consistent), periodic update management and conflict management. Moreover, it guarantees replica convergence, thereby ensuring data consistency at the group level. In essence, updates are propagated to all nodes within the group to provide reliable and robust data management in the distributed environment.

The Rumor file system is also based on an optimistic replication algorithm, where updates are propagated based on opportunistic cost model among the sites replicating the files. It is built at the application level of the users' mobile devices to provide higher portability, while limiting replication costs. The files are updated through a periodic reconciliation mechanism, which ensures the maintenance of consistency when communication can be restored. However, these systems do not incorporate economic models and data rarity issues.

Various data replication techniques have been proposed for MANET databases. By considering the issues of MANET data replication, [PGVA08] tries to attempt the classification of existing MANET data replication techniques, and proposes various criteria for selecting the appropriate replication technique for a given application scenario. The work also considers several data replication issues relevant to MANET databases such as energy, mobility, real-time data availability and frequent network partitioning, based on which the replication schemes have been classified.

Moreover, in a MANET, the mobile peers move freely and disconnections take place frequently, thereby reducing the data accessibility due to the dynamically changing network topology. [HC06] proposes a group mobility model and a replica allocation scheme to address the problem of data accessibility by replicating data items and using the concept of group mobility, where a group of mobile nodes move together.

2.3.3 Data replication in M-P2P networks

The work in [MMK06c] has proposed CLEAR, a context and location-based approach for replica allocation in M-P2P networks. It exploits user mobility patterns, and considers load and different levels of replica consistency.

The works in [MMK06a, KKMM10] propose CADRE (Collaborative Allocation and De-allocation of Replicas with Efficiency), which is a dynamic replication scheme for improving the typically low data availability in dedicated and cooperative mobile ad-hoc peer-to-peer (M-P2P) networks. In particular, replica allocation and de-allocation are collaboratively performed in tandem to facilitate effective replication. Such collaboration is facilitated by a hybrid super-peer architecture in which some of the mobile hosts act as the 'gateway nodes' (GNs) in a given region. GNs facilitate both search and replication.

The main contributions of CADRE are as follows. First, it facilitates the prevention of 'thrashing' conditions due to its collaborative replica allocation and de-allocation mechanism. Second, it considers the replication of images at different resolutions to optimize the usage of the generally limited memory space of the mobile hosts (MHs). Third, it addresses fair replica allocation across the MHs. Fourth, it facilitates the optimization of the limited energy resources of MHs during replication.

The proposals in [MMK06b, MK10] discuss E-ARL, which is a novel Economic scheme for Adaptive Revenue-Load-based dynamic replication of data in dedicated M-P2P networks with the aim of improving data availability. Thus, E-ARL considers a mobile cooperative environment, where the MPs are working towards the same goal, and the network performance is facilitated by the economic scheme. E-ARL essentially allocates replicas based on its economic scheme. Each data item has a price in virtual currency. E-ARL requires a query issuing peer to pay the price of its queried data item to the query-serving peer and a commission to relay peers in the successful query path.

The main contributions of E-ARL follow. First, it uses an economic scheme for efficiently managing M-P2P resources in a context-aware manner by facilitating effective replica hosting and message relaying by peers. Second, it collaboratively performs bid-based replica allocation to facilitate better quality of service. Third, it incorporates both revenue-balancing and load-balancing to improve peer participation and performance. Fourth, it conserves the energy of low-energy MPs to facilitate network connectivity.

The work in [MMK07a] considers that M-P2P users may issue queries with varying constraints on query response time, data quality of results and trustworthiness of the data source. Thus, this work proposes ConQuer, which addresses constraint queries in economy based M-P2P networks. ConQuer proposes a broker-based incentive M-P2P model for handling user-defined constraint queries. It also provides incentives for MPs to form collaborative peer groups for maximizing data availability and revenues by mutually allocating and deallocating data items using a royalty-based revenue-sharing method. Such reallocations facilitate MPs in providing better data quality, thereby allowing them to further increase their revenues.

The work in [MMK06b] presented the economic model for efficient replica management in M-P2P networks, in which mobile peer has been incentivized to host replica. Here, mobile peers choose which data should be replicated based on its importance. In this manner, mobile peers earn revenues from their hosted queried data items. Hence, it encourages peer participation to improve data availability and discourages free-riding. Progressively, [MMK07a] proposed ConQuer: a group-based replication method with incentivization in M-P2P networks. This work assume the super-peer architecture for M-P2P network, in which a broker i.e., super-peer has been incentivized for serving constrained query processing by query-issuing peer. Moreover, collaborative peer groups further improves data availability and revenues by mutually allocating and deallocating data items based on royalty-based model. In a similar vein, a collaborative replication approach for M-P2P networks is also proposed by [KKMM10].

The proposal in [MMK07b] discussed an economic model LEASE, in which data-providers lease data items to the free-riders in lieu of a lease payment. Hence, it provides free-riders the opportunity to earn revenue by hosting data, thereby incentivizing them towards data hosting. [MM08] also discussed incentive-based services for a dynamic data management in M-P2P networks.

2.4 Vehicular Network (VANET) Management

Economic models for resource allocation in distributed systems [KS89] implicitly assume that every node in the system would follow the system-assigned policies. In contrast, our environment considers autonomous vehicles that may not necessarily adhere to system-assigned routes in the absence of incentives. Incentives have been proposed for stimulating data sharing and combating free-riding in mobile-P2P networks [PMG+11, WXS04, XWR06] as well as for encouraging nodes to forward messages in mobile ad hoc networks [BH03]. However, these works do not incorporate economic rewards/penalties for vehicular routing.

The proposal in [Bra96] proposes the use of an incentive-compatible pricing and routing scheme, which also compensates users for sharing their vehicular movement information. Moreover, the work in [Mor10] proposed a dynamic pricing model, which is based on values of time (VOT). The work assumed that the travel-time is a function of vehicles' types e.g., VOT may differ across public transportation vehicles and personal vehicles. Furthermore, the work in [Xu06] proposes a dynamic congestion pricing model, which captures users' personal choices by means of a discrete choice framework.

A predictive model for dynamic pricing in traffic management has been proposed in [Yan12]. In particular, it formulates a mathematical model for addressing distance-based dynamic congestion pricing. Based on the types of measurements volume, speed and occupancy of vehicles, three types of toll collections are proposed; pass-based, per use-based and distance-based associated with different types of rate patterns.

The feasibility of applying dynamic congestion pricing to traffic management has been studied in [Iss11]. The study focuses on spatial or temporal variations in pricing of road-usage, thereby discouraging overuse during rush hours by incentivizing users to travel using alternate routes or at alternate times. Thus, tolls can be adjusted continually based on road conditions e.g., prices increase when the tolled lane(s) are busy and decrease when the tolled lane(s) are relatively less busy.

The work in [RSKM09] proposes a P2P traffic information system for purposes of dynamic route guidance. Cellular Internet access is used for establishing a P2P overlay over the Internet. However, it does not use economic payoffs to encourage vehicles towards following system-assigned paths. The emphasis in [BGJL06] is on routing messages in vehicle-based disruption-tolerant networks, and this is orthogonal to our focus.

As a first differentiating factor from others, our scheme provides users some rewards for following system suggested routes at different times and assigns penalty for any deviations. Our second differentiating contribution is that our scheme considers users' history and incorporates that to suggest different paths and pricing to different users. This is similar to frequent flyer (traveller) schemes where users are given certain privileges and rewards, which they can use for availing better service. Note that our system also introduces penalty so that users are discouraged from deviating system-assigned paths.

2.5 Crowdsourcing

Crowdsourcing refers to the process of outsourcing activities from a firm to an online community or crowd in the form of an 'open call'. Any member of the crowd can then complete an assigned task and be paid for their efforts [Whi09]. In this concept the company pays only for products or services that meet its prospects; beyond the cost and benefits of the company. Characterization of crowdsourcing from management science perspective is given in [Sch09].

The task that requires human intelligence for the crowdsourcing has become an acceptable medium for creation of resources. For the purpose of system building and evaluation, the information retrieval and related fields regularly use it. In this case there are chances of fraudulent attempts by malicious workers and it is also challenging and time consuming process to identify these persons for both crowdsourcing providers and requesters. The work in [EdV11] explains that how to reduce such fraud attempts.

The usage of mobile devices is going to increase day by day because of the availability of number of facilities other than just a communication device. Mcleark, AMT, txteagle, mCrowd and SMSAssassin are the various applications and platforms developed to take the best advantage of the crowdsourcing. Moreover, few mobile companies are planning to provide crowdsourcing enabled mobile device to support such applications.

Moreover, [Eag09] presents a system to provide open-access working platform to people, who can earn small amount by completing the tasks e.g., translations, transcriptions, surveys etc. Here, people are paid either in airtime or MPESA (mobile money) by the corporations. Such services have been recently launched in few countries like Kenya. When Crowdsourcing is extended to sensor-rich mobile devices like smart phones, it has potential that can be truly set free. The work in [YMH+09] proposes a new iPhone-based mobile crowdsourcing platform called mCrowd; which facilitates users to work on sensor related crowdsourcing tasks at fingertips e.g., geolocation aware image collection, image tagging, road traffic monitoring etc. through the rich sensor equipped with iPhone.

We can relate or combine the concept of crowdsourcing with mobile for the effective usage of mobile devices and for the beneficial usage of the Crowd-sourcing; it is not just enough we can also relate the crowdsourcing with database systems. Sometimes the queries processing requires human input as they cannot be answered by machines only. For ex: the queries like matching, ranking or aggregating results based on fuzzy criteria. To process the queries, which is sufficiently answered neither by database system nor does search engine require human input via crowdsourcing which is used by CrowdDB. The CrowdDB uses the SQL for two purposes: as a language for posing complex queries and as a way to model data. There are some differences between CrowdDB and traditional database system like CrowdDB waits for the human inputs and performance and cost of the query depend on a number of new factors. The [FKK+11] describes the design of CrowdDB.

2.6 Summary

In this chapter, we have provided the detailed discussion of various economic schemes for effective resource allocation, encouraging peer participation and combating free-riding in static P2P networks. Then, we have discussed the economic schemes using incentivization strategies for mobile networks: MANETs and M-P2P networks. Interesting related work in mobile environment inspired us to carry our work in this direction. Moreover, a survey of some existing incentive-based economic systems have been presented to demonstrate the importance of such systems in today's technological world. Since free-riding among the MPs is a major drawback of shared and distributed systems, we have discussed both static and dynamic economic approaches. However, none of these works together addresses the issues such as free-riding, data availability and node mobility concerning the effective data dissemination in mobile environment. We have also surveyed existing works on economic schemes for pure M-P2P networks and noted that none of these works consider issues concerning incentivization. Furthermore, we are glad to introduce the concept of top-k query processing using economic incentive approach in M-P2P networks.

Additionally, we have surveyed background information concerning data caching and data replication in mobile networks. However, most of these works focus on memory management on a peer, while our work is distinguished from these existing works since our proposed techniques are aimed at improving the managing, controlling and distributing data across the network, irrespective of the initial data placement across the nodes. In this regard, we have also studied the recent data dissemination approach through crowdsourcing, which are aligned to our M-P2P scenario. In future, we would like to focus on emerging concept of M-P2P with crowdsourcing using economic incentive and game theoretic approaches together.

3

E-Top: Top-k Query Processing in Mobile-P2P Networks using Economic Incentive Schemes

3.1 Overview

In a Mobile ad hoc Peer-to-Peer (M-P2P) network, mobile peers (MPs) interact with each other in a peer-to-peer (P2P) fashion. Proliferation of mobile devices (e.g., laptops, PDAs, mobile phones) coupled with the ever-increasing popularity of the P2P paradigm (e.g., Kazaa) strongly motivate M-P2P applications.

Suppose Alice wants to find the top-k restaurants with “happy hours” (or “manager’s special hours”) within 1 km of her current location. Top-k is determined based on the parameters (e.g., star rating, price and distance from the point of query reference) selected by the user. A broker can facilitate such range-constrained top-k queries by soliciting information from the MPs in its vicinity, and it can then compare this information with its current top-k list of restaurants to generate the top-k result to be provided to the query-issuing MP. The broker compiles its current top-k list by periodically collecting information from various sources such as the Web and social networking sites. Notably, a broker is a trusted entity, which manages the peers in its vicinity and provides value-added services. As we shall see later, brokers also distribute rewards/penalties across the MPs. Moreover, brokers are those nodes that do not make wide-area movements.

In a similar vein, another application could involve a parking lot, where MPs can collect information about available parking slots and charges, and then they can inform the brokers. The parking slot availability information has to be current and therefore, the broker can compare such current information with its current list of parking slots. The broker can then provide the top-k available slots to the query-issuing MP in terms of price or distance (from the MP's current location). Similarly, an MP may want to find the top-k stores selling Levis jeans in a shopping mall with criteria such as (low) price during a specific time duration.

Observe that such ad hoc queries are temporal in nature (e.g., parking slot availability information), hence they cannot be answered by the broker without obtaining information from other MPs. Notably, this research will also contribute towards CrowdDB [FKK+11], which uses human input via crowdsourcing to process queries that cannot be answered by database systems or search engines. Additionally, such M-P2P interactions among peers are generally not freely supported by existing wireless communication infrastructures. The inherently ephemeral nature of M-P2P environments suggests that *timeliness* of data delivery is of paramount importance in these applications, thereby necessitating query deadlines. For example, an MP looking for top-k restaurants with "happy hours" would generally prefer to receive the answer within a specified deadline.

Incidentally, Amazon.com has developed Mechanical Turk [Ama05], which is an online marketplace for match-making between the requirements of businesses and the skill sets of developers. Developers can select from a large pool of tasks based on their skill sets. Similar to our work, the Mechanical

Turk system also provides economic incentives. Observe that technologies, such as WiFi and Bluetooth networks, are nowadays adequately capable of providing a platform for incentive-based mobile P2P collaborations.

Existing economic schemes for distributed systems [Gro03, KS89] and static P2P networks [GBM01, KSGM03a, LDHS05] do not address top-k queries and M-P2P issues such as frequent network partitioning and mobile resource constraints. Economic incentive schemes for mobile ad-hoc networks (MANETs) [BH03] and M-P2P networks [WXS04, XWR06] do not address top-k query processing. Furthermore, the top-k query processing approaches [SIC08, HC07, JCCL10, HSHN09, LCLC04, LXL10, WXTL07] do not consider economic incentive schemes and M-P2P architecture.

Incidentally, data availability in M-P2P networks is typically lower than in fixed networks due to frequent network partitioning [HM06] arising from peer movement and/or peers autonomously switching 'off' their mobile devices. Data availability is further exacerbated due to rampant free-riding [GBM01, KSGM03a, LDHS05], which is characteristic of P2P environments. Furthermore, MPs generally have limited resources (e.g., bandwidth, energy, memory space). Since sending/receiving messages expend the limited energy resources of MPs, minimizing the communication traffic becomes a necessity to address energy constraints. Thus, economic incentive schemes become a necessity to entice resource-constrained MPs with incentives to provide data for answering queries.

This work proposes the **E-Top** system for addressing efficient top-k query processing in M-P2P networks. In E-Top, we have considered that a given query-issuer sends location-based *range-constrained* top-k queries to the M-P2P network. Brokers facilitate top-k query processing in lieu of a commission. E-Top requires a query-issuing MP to pay a *price* (in *virtual currency*) for obtaining its queried top-k result. This price is used for making payments for incentivizing *rankers* (i.e., peers that send data items to answer the query), brokers and relay peers. Thus, an MP has to earn adequate currency by providing *service* (as a broker, ranker or relay peer) before it can issue its own top-k queries, thereby discouraging free-riding.

E-Top issues economic rewards to the rankers, which send relevant data items (i.e., those that contribute to the top-k query result), and penalizes peers for sending irrelevant items. This incentivizes MPs to send only those data items (to the broker), which have a higher probability of being in the top-k results, thereby optimizing the communication traffic. MPs use the rewards/penalties as feedback to re-evaluate their items' scores. We shall henceforth use the term **payoffs** to refer to rewards/penalties.

The main contributions of E-Top are three-fold:

1. It proposes two economic incentive schemes, namely ETK and ETK+, in which MPs act individually towards top-k query processing. These schemes assign payoffs to MPs for incentivizing participation and for enabling them to re-evaluate their data item scores.

2. It extends ETK and ETK+ to propose a peer group-based economic incentive scheme ETG, which defines three payoff allocation approaches.

3. It is indeed effective in improving the performance of top-k queries in terms of query response times and accuracy at reasonable communication traffic cost, as demonstrated by our performance evaluation.

E-Top also discourages free-riding due to its economic nature. ETK and ETK+ differ in that while ETK performs equal distribution of payoffs to the rankers, ETK+ uses a weighted distribution. In ETG, ad hoc groups of MPs are formed in the vicinity of the query location. Each group has a leader for coordinating the top-k query processing. In contrast with ETK and ETK+, where individual MPs directly send their top-k items to the broker, query processing in ETG proceeds by means of group members sending their individual top-k items to the group leader. The group leader selects (i.e., 'filters') the top-k items to be sent to the broker based on the relative frequencies of the items in the individual top-k lists. In our application scenarios, some of the restaurant managers in the vicinity of the query location can be the group leaders.

For simplicity we have considered the uniformly distributed grid regions to show closed-group system. The other suggested approach based on dynamic density-based grid can also be considered to form groups in M-P2P networks. In that case, system may require more brokers to provide better services to the mobile peers, but at the cost of communication overhead in terms of energy and bandwidth. Furthermore, we considered to choose top-1 broker towards serving top-k query into a given query path. There is no restriction over considering multiple brokers (i.e., top-2, top-3 etc.), but in that case the inter-broker communication further increases communication traffic, thereby degrading overall performance of the system. This is due to that the nearby brokers periodically exchange the information (such as global ranking list (T_G), number of unique MPs that interacted with brokers, etc.) with each other, to maintain consistency into such dynamic environment.

In the three approaches deployed by ETG for payoff allocation among group members for any given top-k query, group penalties are equally distributed, thus the schemes differ in their allocation of group rewards. Group rewards are allocated in the following three ways i.e., equally, based on the number of relevant items sent and based on the revenue earned from those items. The group leader receives a percentage of the group rewards as a commission, thereby incentivizing it to participate. Group-based collaboration provides better incentivization since it is likely to lead to higher rewards and lower penalties due to the following reasons. First, MPs risk a lower amount of individual penalties due to the sharing of penalties among group members. Second, MPs have a higher probability of obtaining rewards because the 'filtering' performed by the group leader ensures that collective top-k answers from group members are likely to be of higher quality (i.e., more relevant and accurate) than individual answers.

To the best of our knowledge, none of the existing top-k query processing schemes in M-P2P environment uses incentives. Hence, as reference, we adapt an existing non-incentive-based top-k processing scheme for MANETs. We designate this scheme as **NETK (Non-Economic Top-K)**, proposed in [HHS+10]. Although NETK does not provide incentives to the MPs. it is closest to our top-k query processing scheme. Notably, NETK does not incorporate the notion of item re-ranking as no feedback has been sent back to the MPs, who participated in the top-k query processing.

The results of our performance evaluation indicate that ETG outperforms both ETK and ETK+ due to its group-based scheme, which better incentivizes MP collaboration in top-k query processing due to effective sharing of rewards and penalties among group members. Moreover, ETK+ outperforms ETK due to its weighted distribution (of rewards and penalties to ranker MPs), which provides better incentives to ranker MPs than ETK's equal distribution. ETK, ETK+ and ETG outperform NETK essentially due to the effectiveness of economic payoffs and item re-ranking.

The results also indicate that at higher values of k, query response times increase for all the schemes due to longer query paths. This is because fewer nearby rankers are able to provide enough relevant data items pertaining to the top-k query. Our schemes exhibit good scalability with increasing number of MPs because larger network implies the presence of more rankers. Our schemes exhibit improvement in performance as the communication range of MPs increases. This is because increase in communication range has the effect of bringing the MPs 'nearer' to each other, thereby improving data accessibility.

As the percentage of MP failures increases, our schemes degrade in performance partly due to decreased overall MP participation and partly because of failure of MPs that host data relevant to the top-k queries. ETG performs best when the group sizes are neither too small nor too large. This is because medium-sized groups are better able to leverage the benefits of group-based collaboration.

The remainder of this chapter is organized as follows. Section 3.2 details the architecture of E-Top. Section 3.3 discusses the ETK and ETK+ economic incentive schemes in E-Top. Section 3.4 presents the peer group-based ETG economic incentive scheme of E-Top. Section 3.5 reports our performance study. We summarize E-Top in Section 3.6 with directions for future work.

3.2 Architecture of E-Top

The architecture of E-Top consists of MPs that can assume one of the four following roles: *query-issuer*, *broker*, *ranker* and *relay*. Notably, these roles are interchangeable e.g., a given MP can be a broker for a top-k query Q_1, but a ranker for another top-k query Q_2.

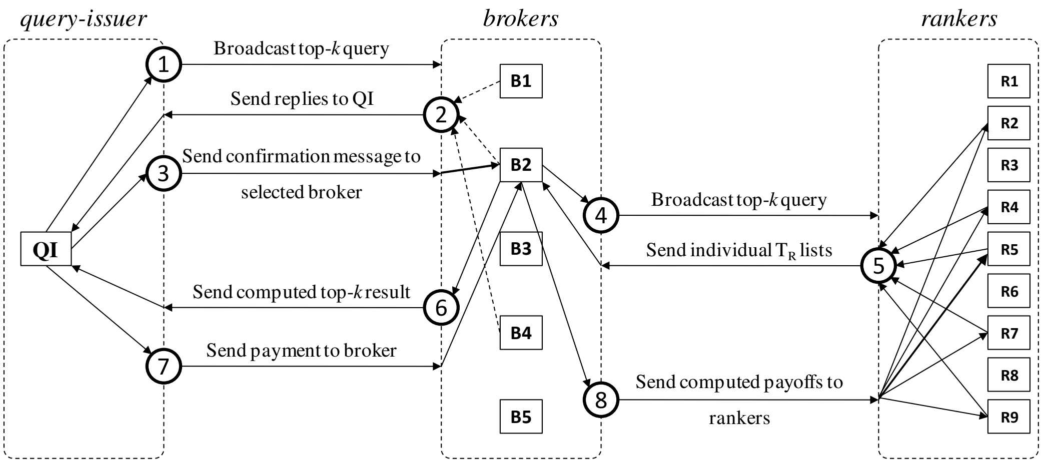


Figure 3.1: Illustrative example of query processing in E-Top

Query-issuer QI issues queries of the form (k, L, τ_Q, ρ), where k is the number of data items that are requested in the location-based *range-constrained* top-k query. L represents the query location, and is of the form of $\{(x, y), rad\}$. Here, (x, y) represents the spatial coordinates associated with a given query Q, while rad represents the radius. For example, QI may want to find restaurants within 1 km of its current location L. τ_Q is the deadline time of Q. ρ is the query price that QI will pay to obtain the top-k query result[1]. An MP decides the query price based on his/her information. Moreover, brokers

[1]Query results received by QI after the deadline entail no payments.

periodically broadcast price ranges for data from different domains such as restaurants, travel and so on. MPs can also subscribe for such information, and brokers can inform them from time to time. *Broker* B acts as a mediator, which facilitates efficient top-k query processing in lieu of a commission. As we shall see in Section 3.3, B also performs economic incentive functions i.e., distribution of payoffs.

Rankers are MPs, which provide data items for answering the top-k query. Rankers are rewarded if their items contribute to the top-k result, otherwise they are penalized. Relay MPs forward messages in multi-hop M-P2P networks in lieu of a small constant commission. Notably, payments to rankers are typically higher than that of broker commissions in order to better incentivize MPs to provide data. This is because MPs providing data generally contribute significantly more to data availability than brokers. Furthermore, relay commission is lower than that of broker commission to better incentivize brokerage functions as compared to relay functions.

During the network configuration phase in the beginning, the broker will be pre-defined, but can also be elected based on the resources. We assume that an MP with relatively more resources may want to become a broker for a given query Q, as it can provide better services, while in case of low resources, an MP should play a role of relay peer, as it requires very few resources for relay service. Moreover, an MP, which has an answer to a given query Q, may be more interested to be a ranker for that query to earn rewards. Hence, our system does not assign specific roles to the MPs, thereby providing them with the flexibility to decide their respective roles for a given query. However, the role assignment for a broker is done in a pre-defined manner.

Notably, we divide the region of interest into square cells of equal area in a grid. Since MPs may not be uniformly distributed across the cells, the density can vary across cells. Observe that the density d of a broker's cell is an important consideration for E-Top because a broker connected with more MPs (or groups in case of ETG) is preferable over one connected with less MPs (or groups). In E-Top, a broker estimates d for its cell by examining the average number of unique MPs, which had connected to it, during the past N time periods. Observe that this is a moving average. (We divide time into equal intervals called periods, the size of a period being application-dependent.) The results of our preliminary experiments showed that $N = 5$ is a reasonable value for our application scenarios. We have defined d as follows:

$$d \;=\; \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \left(np_i \,/\, \sum_{j=1}^{R} tp_{ij} \right) \tag{3.1}$$

where np_i is the number of unique MPs, which had connected to the broker during the i^{th} time period, while tp_{ij} is the total number of MPs in the j^{th} region of interest and R is the number of regions that broker passed through during the i^{th} time period. Whenever the brokers come within communication range of each other, they exchange information about tp_{ij}. Brokers periodically broadcast the value of tp_{ij} in their respective region so that all the MPs are aware of the value of tp_{ij}. Since $np_i < \sum_{j=1}^{R} tp_{ij}$, therefore $0 \leqslant d \leqslant 1$.

3.2.1 Query processing in E-Top

Figure 3.1 illustrates query processing in E-Top. Query-issuer QI broadcasts a top-k query Q, and waits for W time units to get replies from the potential brokers. W is computed as below:

$$W \;=\; (\,1 \,-\, d\,) \,\times\, \tau_Q \tag{3.2}$$

where d is the density of the query issuer's region (i.e., square cell), and it is computed using Equation 3.1. τ_Q is the query deadline time of Q. Notably, QI estimates the value of np_i in Equation 3.1 as the average number of unique MPs, which connected to it during recent time periods. As Equation 3.2 indicates, QI is willing to wait longer for replies from potential brokers if the density of its region is low.

Each *broker* replies to QI with information about its remaining energy En, bid price ρ_{bid}, current currency $Curr$, distance $Dist$ from QI and density d of its current location. QI computes the average location density d_{avg} as $\frac{1}{n}\sum_{i=1}^{n} d_i$, where d_i is the density for the i^{th} broker, and n is the total number of brokers that replied to QI. Now, as candidates, QI will only consider brokers, whose value of d exceeds d_{avg} because brokers in higher-density locations are likely to provide better service due to their proximity to an increased number of potential rankers. Thus, for each broker, whose density exceeds d_{avg}, QI computes a score η and selects the broker with the highest value of η for processing Q. η is computed below:

$$\eta \;=\; (\, w_1 \times En \,) \;+\; (\, w_2 \,/\, \rho_{bid} \,) \;+ (\, w_3 \,/\, Curr \,) \;+\; (\, w_4 \,/\, Dist \,) \tag{3.3}$$

where w_1 to w_4 are weight coefficients such that $0 < w_1, w_2, w_3, w_4 \leqslant 1$ and $\sum_{i=1}^{4} w_i = 1$. Thus, E-Top prefers relatively high-energy brokers because they are less likely to run out of energy, while processing the query. Lower values of bid prices are preferred by QI since it wants to obtain the query result at lower cost. Brokers with less currency are given higher preference to facilitate revenue-balancing across brokers. This prevents low-currency brokers from starvation, which may result in decreased number of brokers in the network. QI prefers relatively nearby brokers to obtain the query result in a timely manner.

Now the broker broadcasts Q with time-to-live (TTL) of n hops. (Results of our preliminary experiments showed that $n = 6$ is a reasonable value for our application scenarios.) The high value of TTL leads to the longer query path, hence it increases both the query latency and the communication overhead. But very low value of TTL also has negative impacts such as decreasing in peer participation, thereby reducing the data accuracy and the success rate. Hence, considering the impacts of very high or very low values of TTL, we considered to keep the value of TTL reasonable, which is dependent on the application scenario and the density of the region. Here, low-density region may need high TTL and vice versa.

Each ranker R has an *individual* item ranking list T_{fR}, each data item of which is associated with an item rank r and a selection probability μ. Notably, the value of r is subjective because it is autonomously assigned to an item by a given ranker. The implication is that the same item may be ranked differently at different rankers. As we shall see in Section 3.3, μ facilitates the adjustment of item selection probability based on recent payoffs assigned to a given item. Using the values of μ and r, each ranker R computes a score γ and selects items with relatively higher values of γ to send to the broker. γ is computed below:

$$\forall i \in T_{fR}: \quad \gamma_i \quad = \quad (\ w_1 \times (N_{T_{fR}} - r_i)/N_{T_{fR}}\ ) \ + (\ w_2 \times \mu_i\ ) \qquad (3.4)$$

where r_i and μ_i are the rank and the selection probability of item i respectively. $N_{T_{fR}}$ is the total number of items in T_{fR}. Here, w_1 and w_2 are weight coefficients such that $0 < w_1, w_2 \leqslant 1$ and $w_1 + w_2 = 1$. E-Top stipulates that $w_2 > w_1$ to give higher weightage to the item selection probability than to the rank of the item. As we shall see in Section 3.3, this is consistent with the overall objective of E-Top i.e., linking item re-ranking with payoffs. Moreover, these weight coefficients are application-dependant i.e., according to application's requirement, weight coefficients are set to any values in-between 0 and 1. There is no restriction on whether to choose $w_1 > w_2$ or $w_2 > w_1$, but to prioritize the item's selection probability, we have chosen $w_2 > w_1$ for our proposed application scenarios. In this work, based on our experimental results, we set $w_1 = 0.2$ and $w_2 = 0.8$ for all the MPs. Furthermore, each ranker is associated with a risk profile δ, where $0 < \delta \leqslant 1$. Only items, whose respective values of γ exceed δ, are consolidated by the ranker in a list T_R and sent to the broker. Thus, T_R is a sorted item ranking list, which is sent by an individual MP in response to a query. Hence, $T_R \subseteq T_{fR}$. Observe that as the value of δ increases, the risk of the ranker in incurring a penalty decreases.

E-Top considers that each broker has a global ranking list, which we shall henceforth designate as T_G. Here, T_G is a global standard (e.g., michelin guide) across the system for considering guideline for the items' ranks. This approach is adopted to incorporate the global rank views (such as Internet or feedback-based) about the items along with the local rankings. T_G is periodically exchanged among nearby brokers. Upon receiving the individual T_R lists from possibly multiple rankers, the broker B collates and compares them with T_G. B parses T_G in a top-down fashion as follows. If an item i in T_G occurs in at least one of the individual T_R lists, it is added to the top-k result set T_A along with the unique identifiers of the rankers that sent i. (In case i does not occur in any of the individual T_R lists, B simply traverses downwards to the next item in T_G.) B continues parsing T_G in the above manner until the result set T_A contains k items. Then B sends T_A to QI. Notably, if T_A contains less than k items, the result set is deemed to be *incomplete*, and it is not sent to QI.

Upon receiving T_A, QI pays B, which deducts its own commission before distributing the payoffs to rankers and commissions to relay MPs. (We shall discuss ranker payoffs, and broker and relay commissions in Section 3.3.) Then each ranker R re-evaluates the selection probability μ of each item in its own T_R based on received payoffs, and then re-computes the values of γ for these items.

In this work, we do not address the formation of the global list T_G because this is application-dependent. Moreover, we do not consider updates to T_G because it may exist for a long time. Furthermore, any update to T_G must be propagated to all the relevant brokers, which also increases the communication overhead.

3.3 Economic incentive schemes in E-Top: ETK and ETK+

This section discusses the ETK and ETK+ economic incentive schemes used by E-Top. We define an item i to be **relevant** to a top-k query Q if it occurs in the top-k query result set T_A. We define a **successful** ranker w.r.t. its (sent) data item i if i is relevant to Q, otherwise the ranker is deemed to be *unsuccessful*. Thus, a ranker may be successful w.r.t. item i, but unsuccessful w.r.t. another item j. Notably, in ETK and ETK+, a ranker can only participate for query Q if it hosts at least k relevant items to Q.

Incidentally, a given ranker R has no way of knowing if its sent-result would finally occur in T_A. R may maintain historical data concerning items that have occurred previously in T_A in response to similar queries. However, if a new query comes to R, no such historical data would be available at R. In such cases, R would send its individual top-k ranking list without considering the historical data.

In both ETK and ETK+, the total payment ρ_R to be distributed to the successful rankers is computed as follows:

$$\rho_R = \rho - \rho_B - \rho_{RL} \tag{3.5}$$

where ρ is the query price paid by QI to the broker, ρ_B is the broker commission and ρ_{RL} is the total amount of relay commission that the broker will pay to the relay MPs in the respective successful query paths. Notably, the value of ρ_B is application-dependent. For both ETK and ETK+, we defined ρ_B as 10% of the query price ρ. Although our schemes can be intuitively generalized to work with other values of ρ_B, results of our preliminary experiments showed that our schemes perform best when ρ_B is in the range of 5% to 15% of ρ. This is also consistent with our overall objective of providing better incentives to rankers than to brokers. For both ETK and ETK+, we define the relay commission ρ_{RL} as 1% of the query price ρ, thereby incentivizing brokers more than relay peers.

As we shall see shortly, the rewards to be assigned to the successful rankers are computed based on the value of ρ_R. Similarly, the penalties to be assigned to the unsuccessful rankers are also computed based on the value of ρ_R. The broker receives the penalty payments from the unsuccessful rankers, and sends the total amount of penalty payments back to QI. Thus, it is possible for the effective payment made by QI to the broker to be less than ρ.

Notably, as is common with currency-based approaches, there is a bootstrapping problem. That is, an MP must first earn currency by providing services, but at the beginning, no MP can request for those services because no MP has any currency yet. To address the bootstrapping problem, the system will provide some initial currency to every MP at the beginning.

3.3.1 ETK

In ETK, ρ_R is equally divided among all the relevant items. Then each ranker, which successfully sent item i, receives a reward P_i that is equal to the total reward for item i divided by the total number f_i of successful rankers w.r.t. item i. Given that the top-k result set is T_A, P_i is computed as follows:

$$\forall i \in T_A : \quad P_i \quad = \quad \tfrac{1}{f_i} \left(\tfrac{\rho_R}{k} \right) \tag{3.6}$$

The reward REW_{Rj} assigned to a given ranker Rj is the total amount that it obtains for each of its relevant items i.e., those that occur in the $T_A \cap T_{Rj}$, where T_{Rj} is the individual rank list of Rj. Given the set S_{Ranker} of rankers, the computation of REW_{Rj} follows:

$$\forall j \in S_{Ranker} : \quad REW_{Rj} \quad = \sum_{i \in (T_A \cap T_{Rj})} P_i \tag{3.7}$$

ETK defines penalties based on the notion of *opportunity cost.* This is because for all successful items, which were not sent by ranker Rj, Rj would have earned currency if it had sent those items. Hence, the penalty PEN_{Rj} assigned to Rj equals $\sum P_i$, where i represents items that occur in $T_A - T_{Rj}$. The computation of PEN_{Rj} follows:

$$\forall j \in S_{Ranker}: \ PEN_{Rj} = \psi \times \left[\sum_{i \in (T_A - T_{Rj})} P_i \right] \tag{3.8}$$

where ψ is the factor that represents the trade-off between communication overhead and peer participation. If the value of ψ is high, communication overhead would reduce because peers would be wary of sending data to the broker due to the higher penalties assigned to unsuccessful rankers. However, this would also reduce peer participation. On the other hand, if the value of ψ is low, peer participation would increase albeit at the cost of increased communication overhead due to lower disincentives for sending items that do not contribute to the top-k result. In this work, we set the value of ψ to 1.3, which implies that the penalties for sending unsuccessful items is 30% more than the reward for sending successful items. This creates disincentives for sending out unsuccessful items, while keeping the peer participation at a reasonable level. We leave the determination of an optimal value for ψ to future work.

The net payment NET_{Rj} received by Rj is the difference between its total reward and its total penalty. NET_{Rj} is computed as follows:

$$\forall j \in S_{Ranker}: \ \ NET_{Rj} = REW_{Rj} - PEN_{Rj} \tag{3.9}$$

Now, based on the payoffs received, Rj will re-evaluate the selection probability of all the items in its individual T_{Rj}. ETK performs rank-weighted increase/decrease in μ for each item, depending on whether the item is rewarded or penalized. For each item i in T_{Rj}, the value of μ_{ij} is computed as follows:

$\forall j \in S_{Ranker}, \forall i \in T_{Rj}$:

$$\mu_{ij} = \begin{cases} min(\ \mu_{ij} + \alpha_{up} \left(\frac{|T_{Rj}| - r_{ij}}{|T_{Rj}|} \right),\ 1\ ), & \text{if } i \text{ is rewarded} \\ max(\ \mu_{ij} - \alpha_{down} \left(\frac{|T_{Rj}| - r_{ij}}{|T_{Rj}|} \right),\ 0\ ), & \text{if } i \text{ is penalized} \end{cases} \tag{3.10}$$

where r_{ij} is the rank of item i in T_{Rj}. Observe that, μ_{ij} increases slightly for higher-rank items that received rewards but decreases significantly in case of a penalty. Similarly, μ_{ij} increases significantly for lower-rank items that received rewards but decreases relatively slightly in case of a penalty. Here, α_{up} and α_{down} represent the weight coefficients for assigning rewards and penalties respectively. ETK stipulates that $0 < \alpha_{up}, \alpha_{down} \leqslant 1$ and $\alpha_{up} < \alpha_{down}$ to ensure that penalties exceed rewards, thereby creating disincentives for rankers in terms of sending out items that are not relevant. In this work, we set the values of α_{up} and α_{down} to 0.1 and 0.3 respectively. We leave the determination of optimal values of α_{up} and α_{down} to future work.

3.3.2 ETK+

In ETK+, ρ_R is divided among all the items in the top-k result T_A based on their respective rank-weights i.e., each item i with its associated rank r_i has weight $w_i = (k - r_i)$, where highest to lowest rank counts are from 0 to $(k-1)$. Furthermore, total number W of weights of all items in T_A is computed as $W = \sum_{i=1}^{k} w_i = k\ (k+1)/2$. Similar to ETK, each ranker, which successfully sent item i, receives a reward P_i that is equal to the total reward for item i divided by the total number f_i of successful rankers w.r.t. item i. Thus, in ETK+, P_i is computed as follows:

$$\forall i \in T_A : \quad P_i = \frac{1}{f_i} \left(\frac{w_i}{W} \times \rho_R \right) \tag{3.11}$$

Consequently, rewards and penalties assigned to each ranker Rj are computed as in Equations 3.7 and 3.8 respectively, using the value of P_i from Equation 3.11. Hence, the net payment received by Rj is computed by Equation 3.9.

Now, each ranker Rj will re-evaluate the score (effectively the selection probability μ) of each item i in its top-k rank list T_{Rj} on the basis of its received payoffs. The effective change in the selection probability of an item depends upon two factors: (a) the notion of item selection potential w.r.t. the risk profile (δ) (b) earning potential of the ranker Rj. Item selection potential increases as the *difference* between μ and δ increases. Average selection potential for rewarded and penalized items for each ranker Rj are computed as s_j and s'_j respectively. The computations of s_j and s'_j are shown below:
$\forall j \in S_{Ranker}$:

$$s_j = \frac{1}{|T_{Rj} \cap T_A|} \left[\sum_{i \in (T_{Rj} \cap T_A)} (\mu_{ij} - \delta_j) \right] \tag{3.12}$$

$$s'_j = \frac{1}{|T_{Rj} - T_A|} \left[\sum_{i \in (T_{Rj} - T_A)} (\mu_{ij} - \delta_j) \right] \tag{3.13}$$

where T_{Rj} is the top-k rank list of Rj, T_A is the top-k result of a query Q, μ_{ij} is the selection probability of item i in T_{Rj} and δ_j is the risk profile of Rj.

Earning potential e_j of each ranker Rj is a measure of its item selection efficiency. $e_j = |\ (REW_{Rj} - PEN_{Rj})/(REW_{Rj} + PEN_{Rj})\ |$. Based on the payoff of each item i in T_{Rj}, the new (re-evaluated) value of μ_{ij} is computed as follows:
$\forall j \in S_{Ranker}, \forall i \in T_{Rj}$:

$$\mu_{ij} = \begin{cases} min(\ \mu_{ij} + \alpha_{up} \left(\frac{s_j + e_j}{2} \right)\ ,\ 1\ ), & \text{if } i \text{ is rewarded} \\ max(\ \mu_{ij} - \alpha_{down} \left(\frac{s'_j + e_j}{2} \right)\ ,\ 0\ ), & \text{if } i \text{ is penalized} \end{cases} \tag{3.14}$$

where α_{up} and α_{down} are the weight coefficients discussed in Equation 3.10.

3.3.3 Illustrative example for ETK and ETK+

Figure 3.2 illustrates the computations in ETK and ETK+. In Figure 3.2a, observe how each ranker R computes the value of γ using Equation 3.4 with $w_1 = 0.2$ and $w_2 = 0.8$. For ranker $R1$, the elements of T_{R1} are shaded in grey i.e., $T_{R1} = \{60, 51, 77\}$ because their respective values of γ exceed 0.8 ($\delta_1 = 0.8$). Figure 3.2b depicts the payoff computations with $\psi = 1.3$. Observe that ETK+ assigns higher penalties (than ETK) to rankers for sending irrelevant items e.g., ETK+ assigned 97.50 to $R3$ as compared to 78.00 in ETK. Figure 3.2c depicts the re-evaluation of the selection probability μ using Equation 3.14 with $\alpha_{up} = 0.1$ and $\alpha_{down} = 0.3$.

T_{fR1} ($\delta_1 = 0.8$)			
r	id	μ	γ
0	77	0.87	0.90
1	44	0.69	0.72
2	51	0.95	0.90
3	47	0.47	0.49
4	26	0.37	0.38
5	60	0.98	0.84
6	73	0.88	0.73

T_{fR2} ($\delta_2 = 0.5$)			
r	id	μ	γ
0	28	0.47	0.58
1	53	0.23	0.36
2	84	0.67	0.70
3	86	0.97	0.92
4	14	0.26	0.33
5	83	0.22	0.28
6	36	0.36	0.37
7	60	0.60	0.54
8	45	0.60	0.52
9	97	0.95	0.78

T_{fR3} ($\delta_3 = 0.3$)			
r	id	μ	γ
0	16	0.10	0.28
1	6	0.28	0.40
2	42	0.25	0.36
3	95	0.45	0.50
4	51	0.63	0.62
5	60	0.37	0.40
6	29	0.20	0.24
7	84	0.10	0.14
8	87	0.71	0.61
9	72	0.75	0.62

T_{R1}	T_{R2}	T_{R3}
60	86	72
51	97	87
77	84	51

T_G
id
95
52
60
13
22
84
51
23
97
87
63

T_A
60
84
51

(a) Compilation of top-k result T_A ($k = 3$)

	P_j (ETK)				P_j (ETK+)			
T_A	ρ_R / k	R1	R2	R3	ρ_R / k	R1	R2	R3
60	30	30	-	-	45	45	-	-
84	30	-	30	-	30	-	30	-
51	30	15	-	15	15	7.5	-	7.5
	REW_{Ri}	45	30	15	REW_{Ri}	52.5	30	7.5

Ranker	T_A-T_{Ri}	PEN_{Ri} (ETK)	PEN_{Ri} (ETK+)
R1	{84}	1.3 x 30 = 39.00	1.3 x 30 = 39.00
R2	{60, 51}	1.3 x (30+15) = 58.50	1.3 x (45+7.5) = 68.25
R3	{60, 84}	1.3 x (30+30) = 78.00	1.3 x (45+30) = 97.50

(b) Computation of rewards and penalties

R1 ($\delta_1 = 0.8$)			
id	μ	μ_{ETK}	μ_{ETK+}
77	0.87	0.57	0.82
44	0.69	0.69	0.69
51	0.95	1.00	0.96
47	0.47	0.47	0.47
26	0.37	0.37	0.37
60	0.98	1.00	0.99
73	0.88	0.88	0.88

R2 ($\delta_2 = 0.5$)			
id	μ	μ_{ETK}	μ_{ETK+}
28	0.47	0.47	0.47
53	0.23	0.23	0.23
84	0.67	0.75	0.71
86	0.97	0.76	0.89
14	0.26	0.26	0.26
83	0.22	0.22	0.22
36	0.36	0.36	0.36
60	0.60	0.60	0.60
45	0.60	0.60	0.60
97	0.95	0.92	0.87

R3 ($\delta_3 = 0.3$)			
id	μ	μ_{ETK}	μ_{ETK+}
16	0.10	0.10	0.10
6	0.28	0.28	0.28
42	0.25	0.25	0.25
95	0.45	0.45	0.45
51	0.63	0.69	0.69
60	0.37	0.37	0.37
29	0.20	0.20	0.20
84	0.10	0.10	0.10
87	0.71	0.65	0.53
72	0.75	0.72	0.57

(c) Updates in the selection probabilities

Figure 3.2: Illustrative example for ETK and ETK+

3.4 ETG: A peer group-based economic incentive scheme in E-Top

This section discusses the group-based ETG scheme.

3.4.1 Peer groups in ETG

We define a **peer group** as a set of MPs, which collaborate in answering a given top-k query. Recall that in our application scenarios, a query-issuing MP QI may try to find top-k restaurants with "happy hours" nearby itself. MPs that are moving nearby QI form *ad hoc groups* for answering this query. Thus, groups are formed based on region. The universe is initially divided into rectangular cells of equal area, and all the MPs moving within a particular cell constitute a group. In case there are not sufficient members in a region at a given point of time, the region can be enlarged based on some minimum spatial density threshold. Conversely, group region can be shrunk based on a maximum density threshold. This work does not specifically focus on how groups are formed, but existing works [GNVTS11] can be used in conjunction with our work for group formation purposes. Notably, ETG stipulates that each MP can belong to any one group at a given point of time, thereby ensuring that any MP obtains its payoff from not more than one group leader for a given top-k query.

In ETG, each group has a group leader, which facilitates top-k query processing within the group. A group leader should be an MP with relatively high energy, bandwidth and processing capacity. Its mobility is typically limited and it stays within the region. In our application scenarios, some of the restaurant managers in the vicinity of the query location can be the group leaders. The group leader receives a percentage of the group rewards as a commission, thereby incentivizing it to participate. In this work, we set the group leader's commission to 5% of the group reward.

Query processing in ETG proceeds via group members sending their individual list of top-k items to the group leader. The group leader selects the top-k items to be sent to the broker based on relative frequencies of items in these individual top-k lists by sorting the items in descending order of frequency. Then the group leader sends the k items with the highest frequencies to the broker. Ties in item frequencies are resolved arbitrarily by the group leader.

ETG uses either ETK or ETK+ for performing the following two economic functions in the top-k query processing. First, brokers assign payoffs to the groups based on either ETK or ETK+. (These payoffs are allocated by the group leader among the group members, as we shall describe shortly.) Second, upon receiving the payoffs, group members modify their item selection probabilities as in either ETK or ETK+. Thus, ETG works in conjunction with either of these schemes. In our performance study, we have first shown the performance of ETG in conjunction with both ETK and ETK+, and then presented the remaining results corresponding to ETG in conjunction with ETK+.

Recall that in ETK and ETK+, any given ranker can only participate for a top-k query Q if it hosts at least k items related to Q. In case of ETG, this criterion is relaxed because even if a ranker does not host k items related to Q, it can still participate in the top-k query processing as long as the group hosts at least k items related to Q. Thus, ETG increases the opportunities for rankers to contribute to the top-k query processing, thereby providing increased opportunities for rankers to earn currency and also providing additional incentives towards ranker collaboration.

Group-based collaboration provides better incentives for MPs to answer top-k queries. When an MP M acts individually in answering top-k queries, it can incur significant penalties due to sending irrelevant items to the broker. This may discourage M from answering queries. As we shall see shortly, when an MP participates in a group, both rewards as well as penalties are distributed among the group members. In effect, this encourages MPs to provide answers to top-k queries because in case its answer turns out to be irrelevant, it risks a lower amount of individual penalties due to sharing of penalties among group members. Group-based collaboration also increases the probability of obtaining rewards because collective top-k answers from the members of the group are likely to be of higher quality (i.e., more relevant and accurate) than individual answers. As we shall see shortly, this is made possible by the 'filtering' performed by the group leader on the individual top-k lists sent by the group members. In essence, group-based collaboration leads to better economy of scale and better results than MPs acting individually.

3.4.2 Illustrative example of peer groups in ETG

Figure 3.3 depicts an illustrative example of an instance of network topology in ETG. Now we shall use Figure 3.3 to illustrate the concept of groups as well as the steps involved in top-k query processing under the ETG scheme. In Figure 3.3, $P12$ and $P15$ are the query-issuers, $P1$ to $P23$ (except $P12$ and $P15$) represent the rankers, and $B1$ to $B3$ indicate the brokers. The groups corresponding to the queries of $P12$ and $P15$ are $\{G1, G2, G3\}$ and $\{G4, G5\}$ respectively. The group leaders of $G1$ to $G5$ are $P3$, $P5$, $P16$, $P21$ and $P22$ respectively.

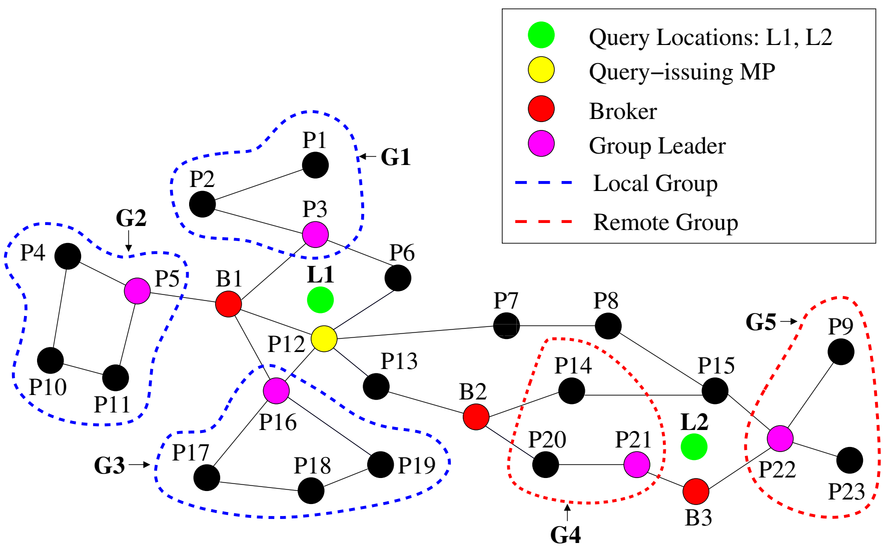


Figure 3.3: Illustrative example of peer groups in ETG

In Figure 3.3, observe that multiple brokers exist. However, given a query

Q, only one of them act as the broker for Q. Broker selection for Q is based on the value of η that is computed from Equation 3.3, as discussed earlier in Section 3.2. For example, in case of $P12$'s query, the candidate brokers are $B1$ and $B2$. For simplicity, suppose the energy, bid price and currency of $B1$ and $B2$ are equal. In this case, $B1$ will be selected as the broker because it is nearer to the query-issuer $P12$. Similarly, in case of $P15$'s query, holding all other factors constant, $B3$ will be selected as the broker because it is nearer (than $B2$) to $P15$.

Now let us examine the processing of $P12$'s query. $P12$ first sends out a broadcast query to list candidate brokers in its vicinity. Based on the respective values of the broker scores η, suppose $P12$ selects broker $B1$ as the broker for processing its query. Then $B1$ sends out the query to groups $G1$ to $G3$, which are nearby the query location. The group leaders (i.e., $P3$, $P5$, $P16$) in these groups consolidate the top-k results from their respective groups and send the results back to the broker. Upon receiving the results from the group leaders, $B1$ compares with its global top-k list to generate the final top-k list, which it sends to the query-issuer $P12$. At this stage, $B1$ also assigns payoffs to the groups. Notably, the broker's assignment of payoffs to the groups is done based on either ETK or ETK+. Then $B1$ sends the top-k results to $P12$ and obtains payment from $P12$. Finally, $B1$ sends the respective payments to the group leaders according to its assigned payoffs.

3.4.3 Allocation of payoffs among group members in ETG

Given a top-k query Q, the resulting payoff for the group has to be allocated among the group members such that they are incentivized towards group-based collaboration.

We define a group member as a **participant** in query Q if at least one of the top-k items that it sent to the group leader is selected by the group leader in the top-k list that the group leader propagates to the broker. On the other hand, we define a group member as a **contributor** to Q if at least one of the top-k items that it sent to the group leader occurs in the final top-k list that is selected by the broker. Thus, the set of contributors is a subset of the set of participants for a given top-k query. Based on the notions of contributors and participants, we propose three schemes for payoff allocation among group members. In all these proposed schemes, penalties are divided among all the participants for a query Q, while rewards are distributed only among the contributors. Thus, the three approaches differ in the way in which group rewards are allocated.

Let n_P be the number of participants for Q. Let PEN_G represent the penalty for the group corresponding to Q. The penalty PEN_j for participant j is computed as follows:

$$PEN_j = \frac{PEN_G}{n_P} \tag{3.15}$$

Notably, all our three proposed approaches for payoff allocation compute the penalty PEN_j incurred by participant j by means of Equation 3.15 above.

Observe how ETG incentivizes MP participation in groups (as opposed to MPs acting individually) by reducing potential penalties for group members in two ways. First, the 'filtering' of top-k items performed by the group leader implies that even if a group member P had sent one or more items (to the group leader), which do not occur in the final top-k result selected by the broker, P incurs no penalty for such irrelevant items as long as the group leader does not send them to the broker. In effect, being part of a group shields the MP from incurring penalties to a certain extent. Moreover, since the group leader receives top-k items from multiple group members, it has a broader (and more collective) view of the likely top-k results than individual group members. This increases the likelihood of the group leader's 'filtering' process being more effective in predicting the final top-k results than if the top-k predictions were done by individual MPs.

Second, the sharing of penalties across participants reduces the penalties incurred by those members, which sent out irrelevant items, which were selected by the group leader and which did not occur in the final top-k results. This does not incentivize group members to frivolously send out irrelevant items to the group leader because the items should have at least some chance of occurring in the final top-k result for the group leader to have selected them.

Incidentally, the equal sharing of penalties across all participants may result in increased penalties for some of the participants, especially for the contributors. For example, even if a contributor had not sent out any irrelevant items, it still has to pay the penalty due to irrelevant items being sent by some of the other participants. However, the cost of such possible additional penalties is offset by the benefit obtained by the contributor(s) in terms of avoiding potential penalties due to the group leader's filtering process. This explains the rationale for dividing penalties equally among all the participants for a given query Q.

The rationale for distributing rewards only among the contributors is two-fold. First, it discourages free-riding within the group since a peer has to contribute to the final top-k query result in order to qualify for obtaining a share of the group reward. Observe that if the group reward were to be distributed across all the participants, it would act as a disincentive for the contributors since they would earn lower amounts of currency. Second, it recognizes the contribution of the contributors to the group revenue, thereby incentivizing peer contributions to the group.

Now we shall discuss the three approaches that ETG deploys for allocation of group rewards among contributors for a given top-k query Q.

Equal allocation of payoff (EQ)

In EQ, each contributor obtains an equal share of the group reward REW_G for Q. Given that n_C is the number of contributors for Q, the reward REW_j for contributor j is computed as follows:

$$REW_j = \frac{REW_G}{n_C} \tag{3.16}$$

Notably, a major drawback of the EQ approach is that the allocation of group reward is not based on the contribution of individual contributors since it does not consider the number of items contributed by each of them.

Item contribution-based allocation of payoff (ICON)

To address the drawback of EQ, we propose ICON. In ICON, each contributor obtains a share of the group reward REW_G based on the number of items that it contributed to the final top-k query result. ICON computes the reward REW_j for contributor j as follows:

$$REW_j = \frac{|C_j|}{\sum_{g=1}^{n_C} |C_g|} \times REW_G \tag{3.17}$$

where C_j is the set of items that MP j has contributed to the final top-k result, and n_C represents the total number of contributors corresponding to Q.

ICON suffers from the drawback that the allocation of group rewards is not based on the actual revenue earned from the item (i.e., the reward that is assigned to the item). For example, suppose contributor $P1$ has contributed three items, while contributor $P2$ has contributed only one item. However, the revenue earned by the group from the one item sent by $P2$ could be higher than that of the total revenue earned from the three items contributed by $P1$.

Revenue contribution-based allocation of payoff (RCON)

To address ICON's drawback, we propose RCON. In RCON, each contributor obtains a share of the group reward REW_G based on the revenue earned from the items that it contributed to the final top-k result. RCON computes reward REW_j for contributor j as follows:

$$REW_j = \frac{\sum_{i \in C_j} \lambda_i}{\sum_{g=1}^{n_C} \left(\sum_{i \in C_g} \lambda_i \right)} \times REW_G \tag{3.18}$$

where C_j is the set of items that MP j has contributed to the final top-k result, λ_i represents the revenue earned for a given item i, and n_C represents the total number of contributors for Q.

Illustrative example of group reward allocation among contributors

Table 3.1 depicts an illustrative example of allocation of rewards among contributors. Consider a top-3 query, for which the result set comprises items {1, 2, 3}. Suppose the relevant items sent by contributors $P1$, $P2$ and $P3$ are {1, 3}, {2, 3} and {1} respectively. As Table 3.1 indicates, the rewards for items {1, 2, 3} are {60, 20, 10} currency units respectively. Hence, the total group reward REW_G is the sum of these individual item rewards i.e., 90 currency units. (For simplicity, we ignore the group leader's commission for this example.)

Item ID (i)	**Item Reward** (λ_i)	**Contri-butor** (j)	**Relevant Item Set** (C_j)	**Contributor j's Reward** (REW_j)			
				EQ	**ICON**	**RCON**	
1	60	$P1$	{1, 3}	$90/3$ $=$ **30**	$(2/5) \times 90$ $=$ **36**	$60 + 10$ $= 70$	$(70/160) \times 90$ $=$ **39.375**
2	20	$P2$	{2, 3}	$90/3$ $=$ **30**	$(2/5) \times 90$ $=$ **36**	$20 + 10$ $= 30$	$(30/160) \times 90$ $=$ **16.875**
3	10	$P3$	{1}	$90/3$ $=$ **30**	$(1/5) \times 90$ $=$ **18**	60	$(60/160) \times 90$ $=$ **33.750**
	$\sum \lambda_i =$ $REW_j =$ **90**		$\sum \lvert C_j \rvert =$ **5**			$Nett = 160$	

Table 3.1: Illustrative example of group reward allocation among contributors in ETG

For EQ, the reward is distributed equally among the contributors, hence $P1$, $P2$ and $P3$ would each obtain a reward of 30 currency units. For ICON, the number of relevant items sent by $\{P1, P2, P3\}$ are {2, 2, 1} respectively, Hence, the reward for $P1$ is $(2/(2 + 2 + 1)) * 90$ i.e., 36 currency units. In case of RCON, the rewards for P1's relevant (sent) items {1, 3} are {60, 10} currency units respectively, thereby resulting in a total of 70. Similarly, the corresponding totals for $P2$ and $P3$ are 30 and 60 respectively. Thus, RCON computes the reward of $P1$ based on the weighted average of item revenues earned. Hence, $P1$ obtains $(70/(70+30+60))*90$ i.e., a reward of 39.375 currency units.

Observe the difference in rewards obtained by $P2$ and $P3$ under ICON and RCON. In case of ICON, $P2$ obtains double the reward of $P3$, even though $P3$ contributed more revenue to the group. This highlights the drawback of ICON. Observe how RCON alleviates this drawback by assigning $P3$ a higher amount of reward than $P2$.

3.5 Performance Evaluation of E-Top

This section reports our performance evaluation by means of simulation in OMNeT++ [Pon93]. MPs move according to the *Random Waypoint Model* [BMJ+98] within a region of area 1000 metres × 1000 metres. The *Random Waypoint Model* is appropriate for our application scenarios, which generally involve random movement of peers. For example, people looking for top-

Parameter	Default Value	Variations
k	8	4, 12, 16, 20, 24
Number of MPs (N_{MP})	100	20, 40, 60, 80
Percentage of brokers (P_B)	20%	10%, 30%, 40%, 50%
Queries/time unit	10	
Communication Range (CR)	120m	40m, 80m, 160m, 200m
Percentage of MP failures (P_F)	20%	10%, 30%, 40%, 50%
Group size (quantified by S_G)	30%	10%, 20%, 40%, 50%
Bandwidth between MPs	28 Kbps to 100 Kbps	
Initial energy of an MP	90000 to 100000 energy units	
Memory space of each MP	8 MB to 10 MB	
Speed of an MP	1 meter/s to 10 meters/s	
Size of a data item	50 Kb to 350Kb	

Table 3.2: Parameters of our performance evaluation for E-Top

k restaurants generally move randomly i.e., they do not follow any specific mobility pattern. Our experiments use a total of 100 MPs. Each MP contains

20 to 25 data items. The default communication range of all MPs is a circle of 120 metre radius. Table 3.2 summarizes the parameters used in our performance evaluation.

Query-issuers are selected randomly from among all the MPs in the network. The number of such top-k queries issued in the network per time unit is 10, the query deadline τ_Q being varied randomly between 3 to 5 time units. Query price ρ is chosen randomly in the range of 100 to 500 currency units. Broker commission ρ_B and relay commission ρ_{RL} are respectively set to 10% and 1% of ρ. For ETG, group leader's commission is set to 5% of the group reward for a given query. Initial energy of an MP is selected to be randomly in the range of 90000 to 100000 energy units. Sending and receiving a message require 1.5 and 1 energy units respectively.

Recall that each ranker is associated with a risk profile δ. The number of MPs with the values of δ as 0.3 (high-risk), 0.5 (medium-risk) and 0.8 (low-risk) are 27, 43 and 30 respectively. For all our experiments, the economic parameters are set as follows: (a) the values of weight coefficients w_1 to w_4 for computing the broker score η in Equation 3.3 are each set to 0.25 (b) the values of weight coefficients w_1 and w_2 for computing the item score γ in Equation 3.4 are set to 0.2 and 0.8 respectively (c) the penalty factor ψ (see Equation 3.8) is set to 1.3 (d) the values of α_{up} and α_{down} for item selection probability re-evaluation (see Equations 3.10 and 3.14) are set to 0.1 and 0.3 respectively.

Performance metrics are average response time (ART), precision rate (PREC), query completeness rate (QCR) and communication traffic (MSG). We define a query as **completed** if the broker receives at least k items from individual rankers (or group leaders in case of ETG) within 70% of the query deadline time τ_Q. Notably, a broker may fail to receive at least k items due to reasons such as ranker unavailability and network partitioning. (Queries that are not completed are deemed to be query failures.) We compute ART only for the *completed* queries. $ART = \frac{1}{N_C} \sum_{q=1}^{N_C}(t_f - t_0)$, where t_0 is the query-issuing time, t_f is the time of the query result reaching the query-issuer, and N_C is the total number of *completed* queries. We compute ART in simulation **time units (t.u.)**.

PREC is the average precision rate over all the queries. Suppose T_{A_q} is the top-k query result and T_{G_q} is the global top-k rank list of the respective broker for a query q. To obtain PREC for q, we measure the number of items in T_{A_q} which also occur in T_{G_q}; then we divide by the number of items in T_{G_q}. Notably, PREC is computed only for *completed* queries. Thus, $PREC = \frac{1}{N_C} \sum_{q=1}^{N_C} \left(\frac{|T_{G_q} - T_{A_q}|}{|T_{G_q}|} \right) \times 100$.

QCR is the ratio of total number N_C of completed queries to the total number N_Q of queries. $QCR = (\ N_C / N_Q\ ) \times 100$. We define MSG as the total number of messages incurred for query processing during the course of the experiment. Thus, $MSG = \sum_{q=1}^{N_Q} M_q$, where M_q is the number of messages incurred for the q^{th} query.

To the best of our knowledge, none of the existing top-k query processing schemes in M-P2P environment uses incentives. Hence, for purposes of meaningful comparison, we adapt an existing non-incentive-based top-k processing scheme for MANETs. We designate this scheme as **NETK (Non-Economic Top-K)**, proposed in [HHS+10]. Although NETK does not provide incentives to the MPs. it is closest to our top-k query processing scheme.

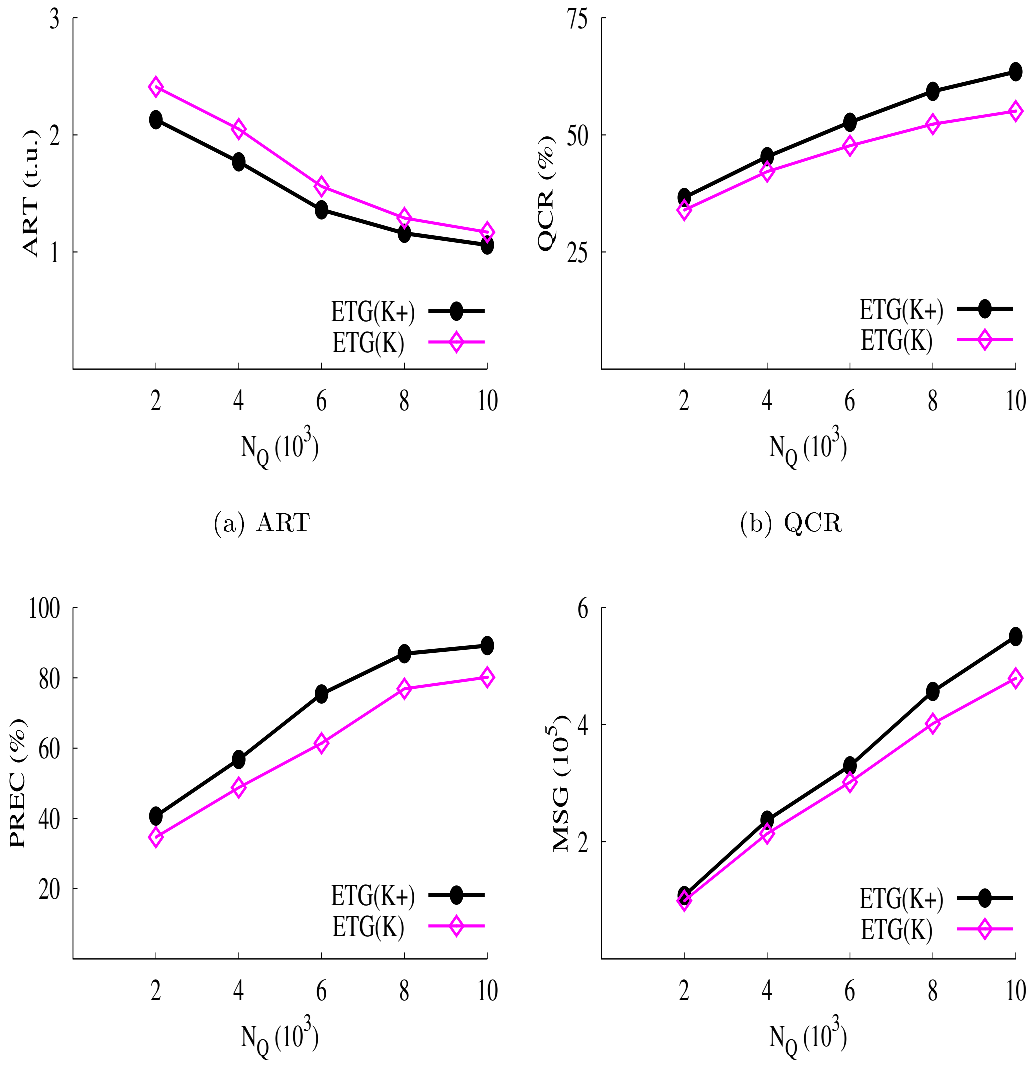
In NETK, each MP that receives a query message sends back a fixed number, R, of its holding data items with the R highest scores. If each MP finds that the total number of data items received from all its successor MPs and its own data items with the R highest scores becomes larger than k (i.e., top-k data items), it only sends k data items with the highest scores among those data items to its predecessor. Notably, NETK suffers from the serious drawback of not being able to encourage peer participation in top-k query processing since it does not provide incentives. To strengthen NETK, we adapted NETK to our scenario with $R = \lceil k/50 \rceil$ (i.e., 50% of top-k values are allowed to send towards contributing into top-k query) because at this value of R, NETK has above-average peer participation (based on the results of our preliminary experimental observations), thereby making NETK a fairly efficient approach in itself. Furthermore, NETK does not incorporate the notion of item re-ranking as no feedback has been sent back to the MPs, who participated into top-k query processing.

3.5.1 Effect of peer groups with ETK and ETK+

Recall that ETG uses either ETK or ETK+ for performing some of the economic functions (e.g., assignment of payoffs from broker to group leader) during top-k query processing. We designate these variations as ETG(K) and ETG(K+) corresponding to ETK and ETK+ respectively. Figure 3.4 depicts the results. ETG(K+) outperforms ETG(K) due to two reasons. First, ETK+'s rank-weighted payoff strategy provides better incentivization than the uniform incentivization provided by ETK. Second, ETK+ provides more effective re-evaluation of the item selection probability μ by tying μ to payoffs associated with rankers' items. In contrast, ETK does not directly link μ to payoffs. However, ETG(K+) incurs more MSG due to group communication overhead. For the remainder of this section, we show the performance of ETG in conjunction with ETK+.

3.5.2 Effect of variations in the percentage of brokers

We performed an experiment to determine the percentage P_B of brokers in the network. Figure 3.5 depicts the results. As P_B is increased from 10% to 20%, ART decreases and QCR increases for all the schemes. This is because the involvement of more brokers increases the probability that a given query is processed by at least one of the brokers. Notably, the sum total of the number of brokers and the number of rankers is fixed. Hence, when P_B is increased beyond 30%, the number of rankers reduces, thereby reducing QCR and increasing ART (due to more hop-counts required to reach the rankers). Interestingly, beyond P_B = 40%, ETG performs slightly worse QCR than



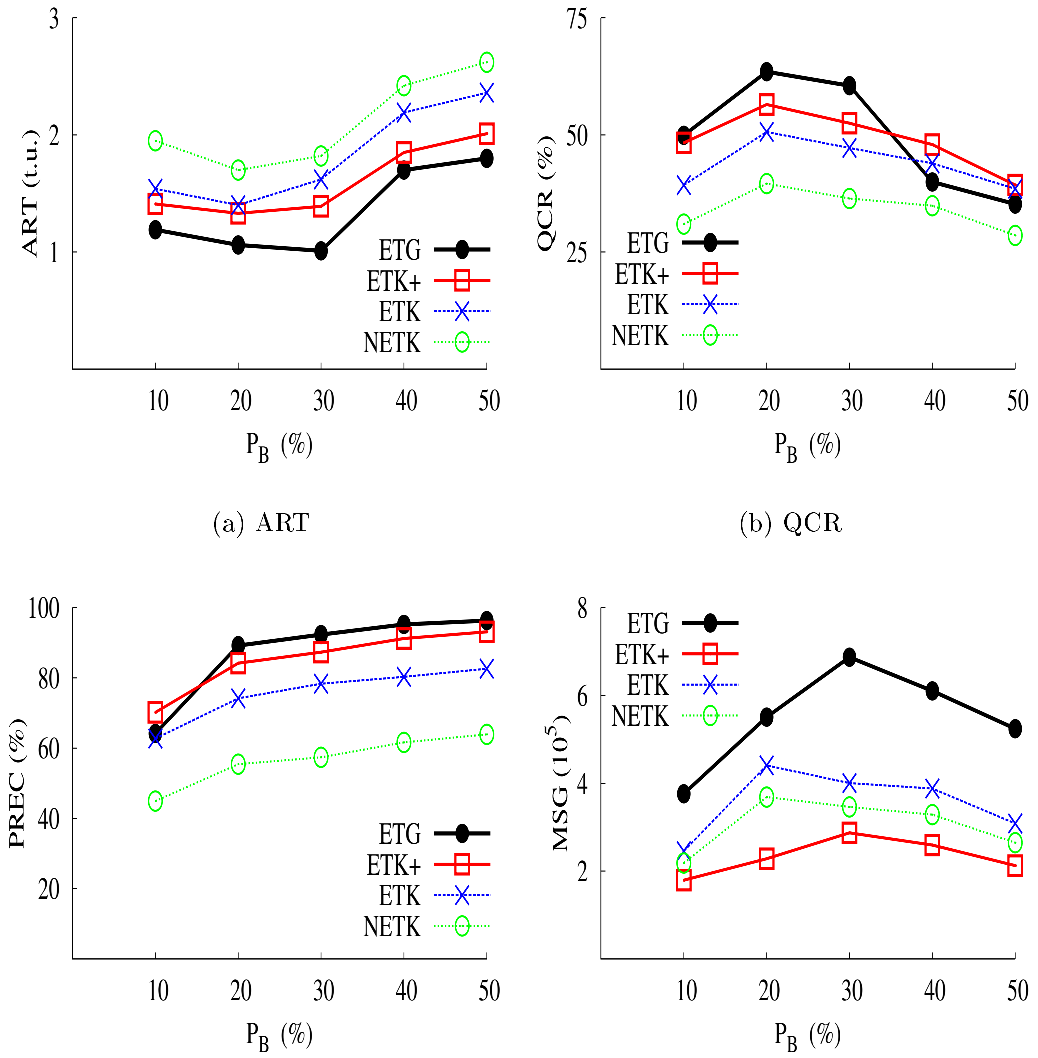
(a) ART (b) QCR

(c) PREC (d) MSG

Figure 3.4: Effect of peer groups with ETK and ETK+

both ETK and ETK+ due to a significantly decreased number of rankers, which make group formation difficult.

As P_B increases, PREC increases due to the involvement of more brokers for all the schemes. However, PREC exhibits a saturation effect beyond $P_B = 30\%$ due to reduced number of rankers. As P_B is increased till 30%, MSG increases for all the schemes due to the involvement of more brokers. However, beyond $P_B = 30\%$, MSG decreases due to reduced number of rankers. Based on the results, we set the percentage of brokers to 20% so that we can obtain good performance of E-Top with reasonable communication



(a) ART (b) QCR

(c) PREC (d) MSG

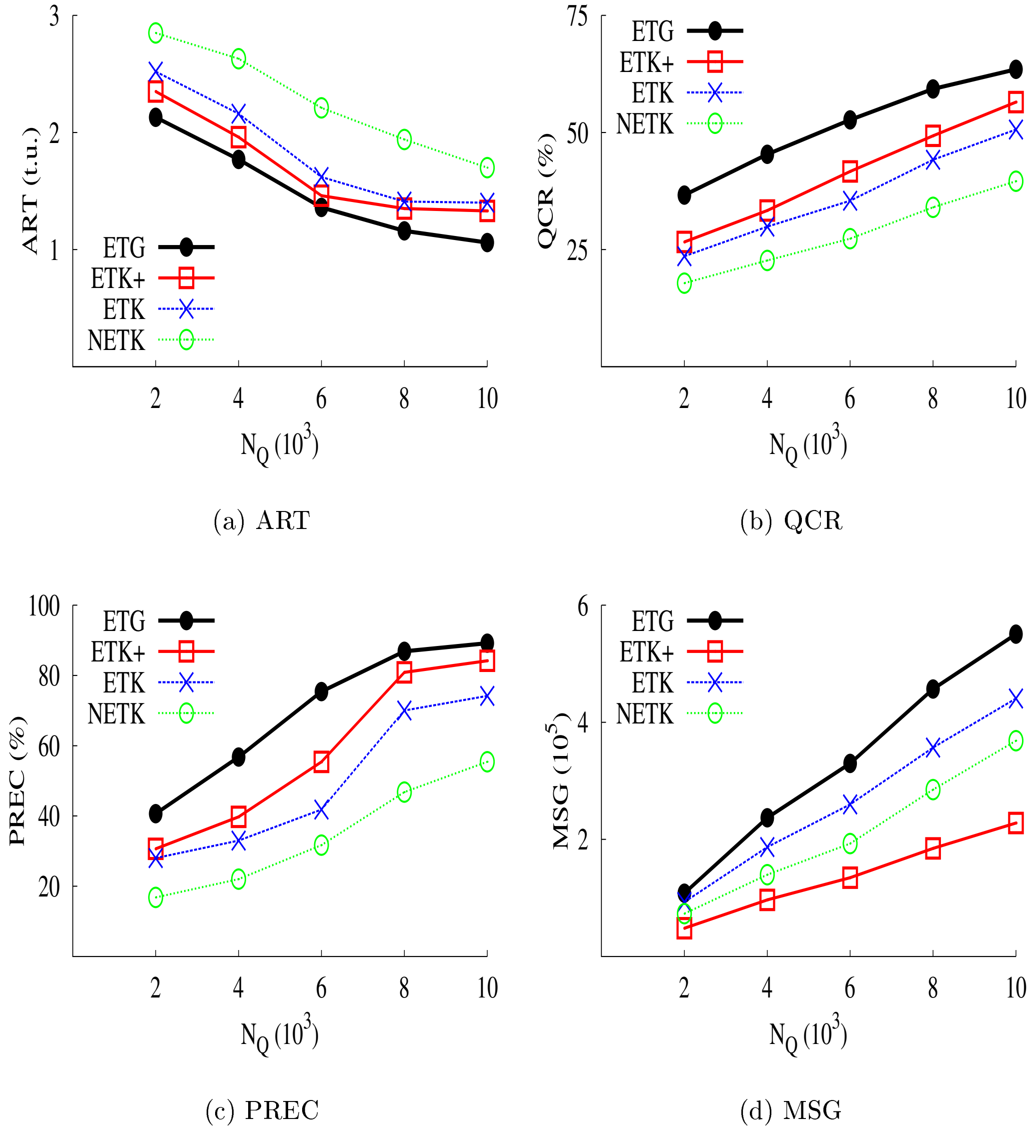
Figure 3.5: Effect of variations in the percentage of brokers

traffic.

3.5.3 Performance of ETK and ETK+

We conducted an experiment using the default values of the parameters in Table 3.2. Figure 3.6 depicts the results. As more queries are processed, performance improves for ETK, ETK+ and ETG due to incentives and effective item re-ranking. However, the performance eventually plateaus due to network partitioning and unavailability of some of the rankers. ETK+

outperforms ETK because it provides better incentivization and more effective re-evaluation of the item selection probability, as explained for the results in Figure 3.4. ETG outperforms both ETK and ETK+ due to better incentives for group-based collaboration and effective payoff sharing among group members.



(a) ART (b) QCR

(c) PREC (d) MSG

Figure 3.6: Performance of ETK & ETK+

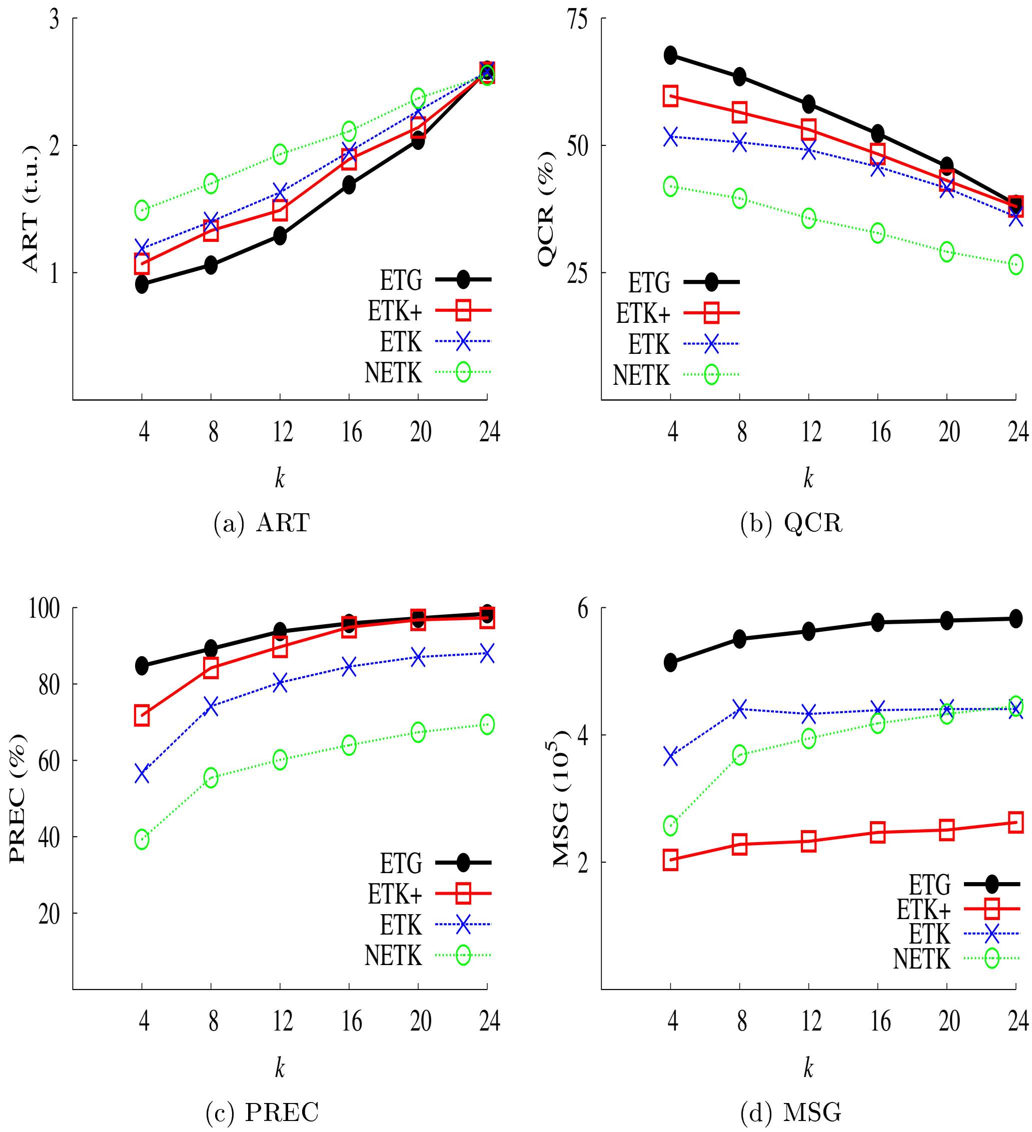
NETK performs worse than that of ETK in terms of ART, QCR and PREC due to less ranker participation (owing to the lack of incentives), which may cause inadequate items to generate the final top-k result. Since the broker does not always receive at least k items from rankers, NETK results in a significant number of incomplete query results.

MSG increases over time for all the schemes as more queries are being processed since MSG is a cumulative metric. Interestingly, NETK incurs lower MSG than the other schemes due to lower levels of ranker participation in the absence of item re-ranking. ETK+ incurs lower MSG than ETK because ETK+ assigns higher amount of penalties (as compared to ETK) to rankers that send irrelevant items, hence fewer rankers reply to the broker in case of ETK+. ETG incurs higher MSG than both ETK and ETK+ due to periodic communication between group members for exchanging their own individual top-k lists and for sharing payoffs.

3.5.4 Effect of variations in k

Figure 3.7 depicts the effect of variations in k. As k increases, QCR decreases for all the schemes because relatively fewer rankers would be capable of participating in the top-k query processing. This is because a ranker is allowed to send its top-k result only if it hosts at least k items pertaining to the query. ART increases due to longer query paths as nearby rankers are unable to provide enough relevant items. The performance gap (in terms of ART and QCR) between ETK, ETK+ and ETG keeps decreasing with increase in k due to decreased ranker participation.

As k increases, PREC increases for all the schemes due to increase in the probability of the items (sent by the rankers) being relevant to the top-k result. For example, if $k = 4$, an item will contribute to the top-k if it matches one of the four items in the broker's global top-k list T_G. However, if $k = 24$, T_G has 24 items, hence the ranker-sent item has a better chance of a 'match' with any one of the items in T_G. ETG and ETK+ exhibit comparable PREC beyond $k = 12$ because their incentives result in the same rankers sending the top-k results at these higher values of k. PREC also eventually plateaus for all the schemes after $k = 12$ due to peer mobility, frequent network partitioning and unavailability of some of the rankers.



(a) ART (b) QCR

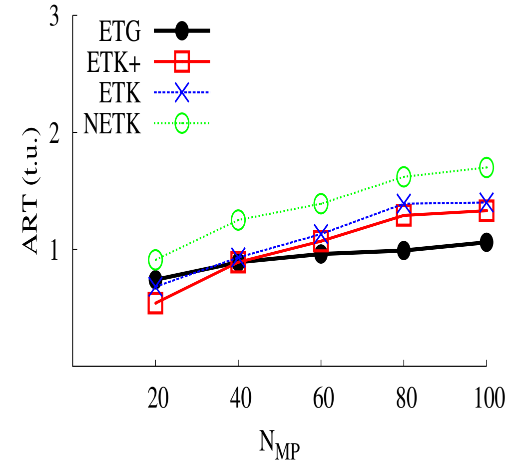
(c) PREC (d) MSG

Figure 3.7: Effect of variations in k

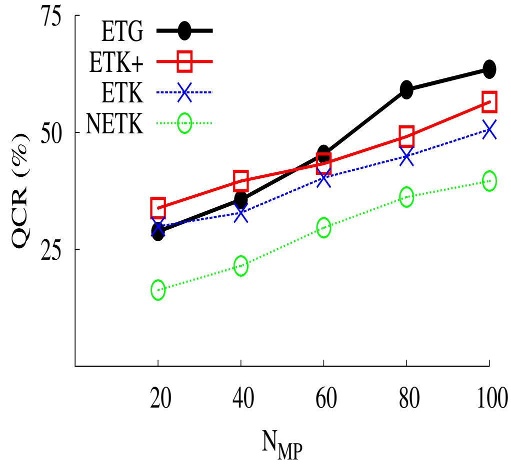
As k increases, MSG increases for all the schemes due to longer query paths arising from less ranker participation. However, MSG eventually plateaus at $k = 12$ because the increased number of hops required to reach the relevant rankers is offset by the decreased number of relevant rankers.

3.5.5 Effect of variations in the number of MPs

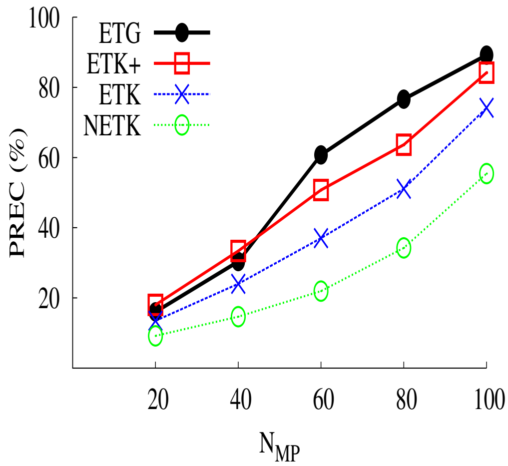
We conducted an experiment to examine the scalability of our proposed schemes. Figure 3.8 depicts the results. As the number N_{MP} of MPs increases, ART increases for all the schemes due to larger network size. As N_{MP} increases, QCR and PREC increase because larger network implies more rankers. Observe that the performance of NETK is worse than that of ETK, ETK+ and ETG due to lower levels of ranker participation in the absence of item re-ranking, as explained for the results in Figure 3.6.



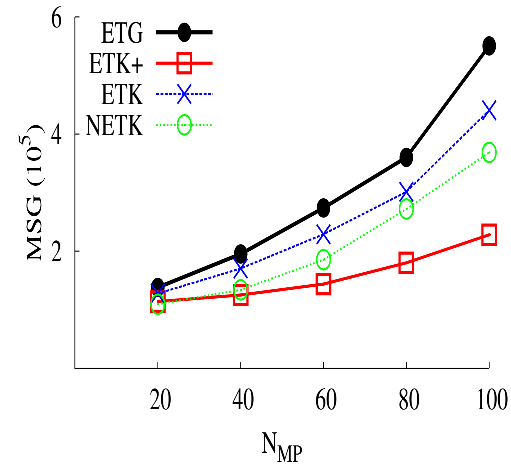
(a) ART



(b) QCR



(c) PREC



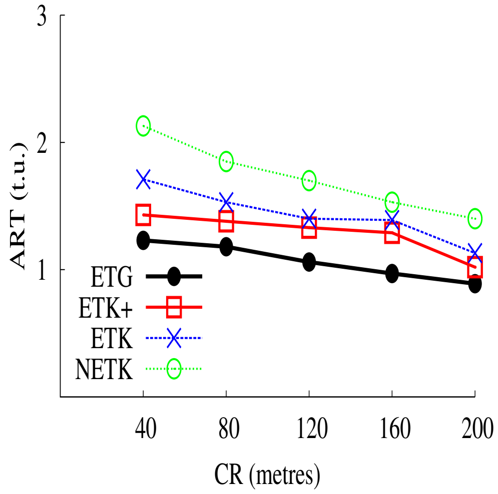
(d) MSG

Figure 3.8: Effect of variations in the number of MPs

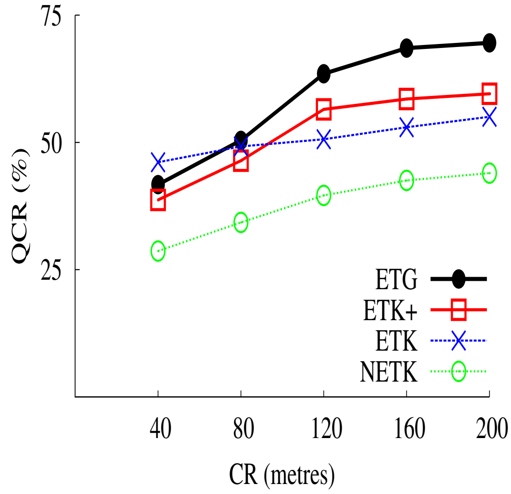
ETG outperforms ETK and ETK+ due to more pronounced effect of peer group collaboration when N_{MP} exceeds 40. However, below $N_{MP} = 40$, ETG performs slightly worse than ETK+ due to limited effect of group collaboration. MSG increases for all the schemes due to larger network size.

3.5.6 Effect of variations in the communication range

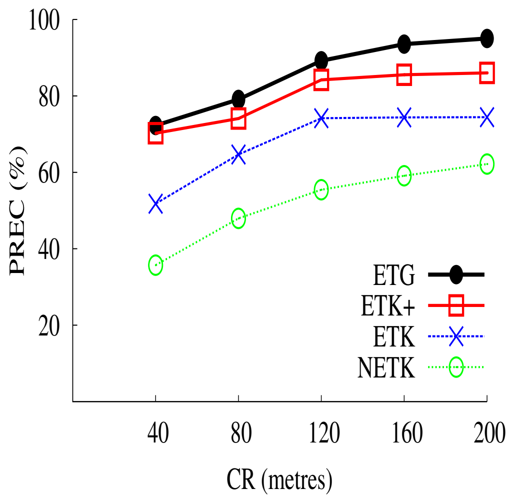
The results in Figure 3.9 depict the effect of variations in the communication range CR of the MPs. Increase in CR has the effect of bringing the MPs



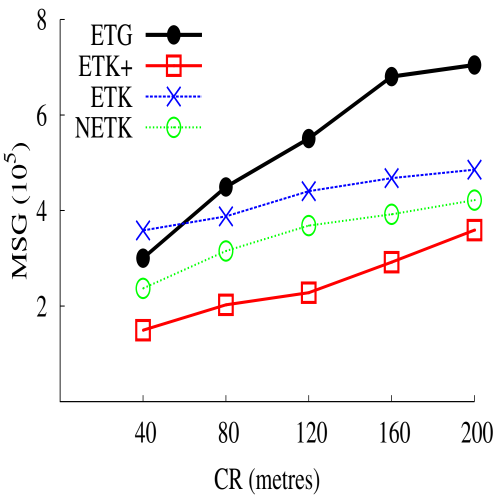
(a) ART



(b) QCR



(c) PREC



(d) MSG

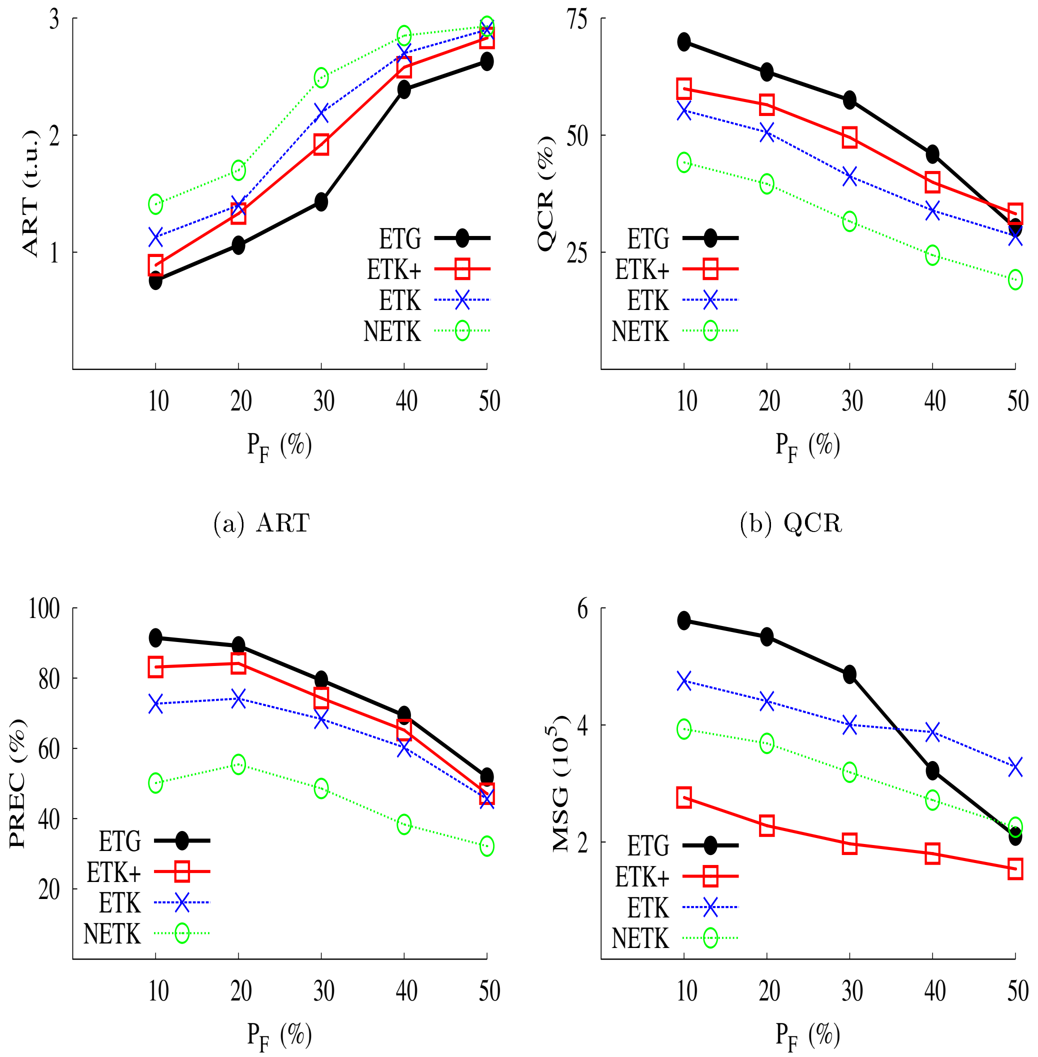
Figure 3.9: Effect of variations in the communication range

'nearer' to each other. Hence, performance improves for all the schemes due to data items becoming 'nearer' and more accessible to query-issuers. Thus, relatively fewer queries fail due to the maximum TTL criteria of 6 hops as more MPs come within range to answer queries. Observe that the rate of decrease in ART is not necessarily uniform because of deviations arising from bandwidth differences at MPs. QCR and PREC exhibit a saturation effect for all the schemes beyond CR = 160 metres due to unavailability of some of the rankers.

As CR increases, MSG increases for all the schemes because the increased reachability of the MPs increases communication among them. With increase in CR, a lower number of messages are required to reach a given MP, thereby decreasing MSG. However, more MPs become involved in the processing of a given query, thereby increasing MSG. These two opposing effects somewhat offset each other at higher values of CR, thereby explaining the reason why MSG eventually plateaus. Interestingly, ETG incurs lower MSG than ETK at CR = 40 metres. This occurs because at such low values of CR, the MPs are in effect 'far' from each other, thereby reducing the effectiveness of group collaboration. Consequently, a lower number of messages are required for group interactions.

3.5.7 Effect of MP failures

MPs can fail due to reasons such as depletion of their limited energy resources. Figure 3.10 depicts the results of the effect of MP failures. As the percentage P_F of MP failures increases, MP participation decreases, query paths become longer and fewer potential rankers remain available, thereby degrading the performance of all the schemes. Interestingly, at $P_F = 50\%$, ETK, ETK+ and ETG exhibit comparable performance due to limited MP participation making the effect of groups and item re-ranking less pronounced. As the results in Figure 3.10d indicate, MSG decreases with increase in P_F for all the schemes due to reduced communication overhead among a lower number of available MPs. Interestingly, detailed examination of the experimental logs indicated that beyond $P_F = 35\%$, ETG incurs lower MSG than ETK due to difficulties in group formation when relatively fewer MPs are available.



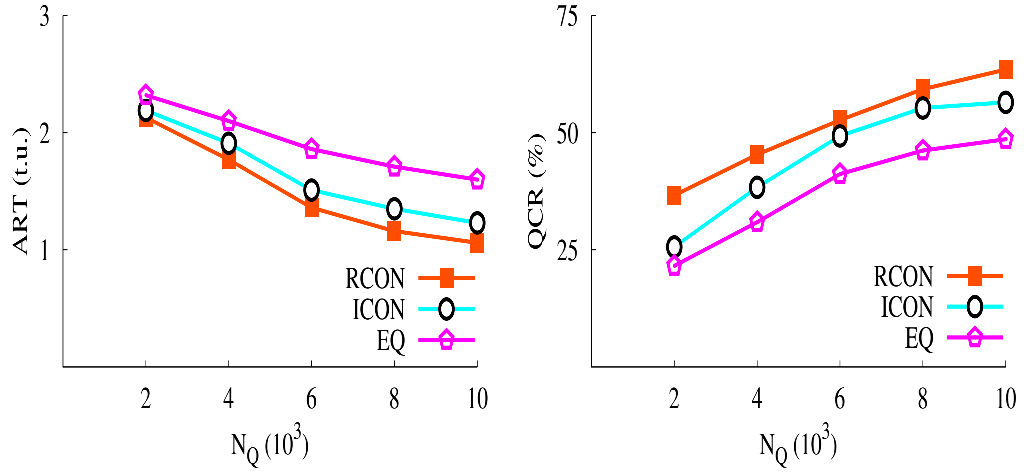
(a) ART (b) QCR

(c) PREC (d) MSG

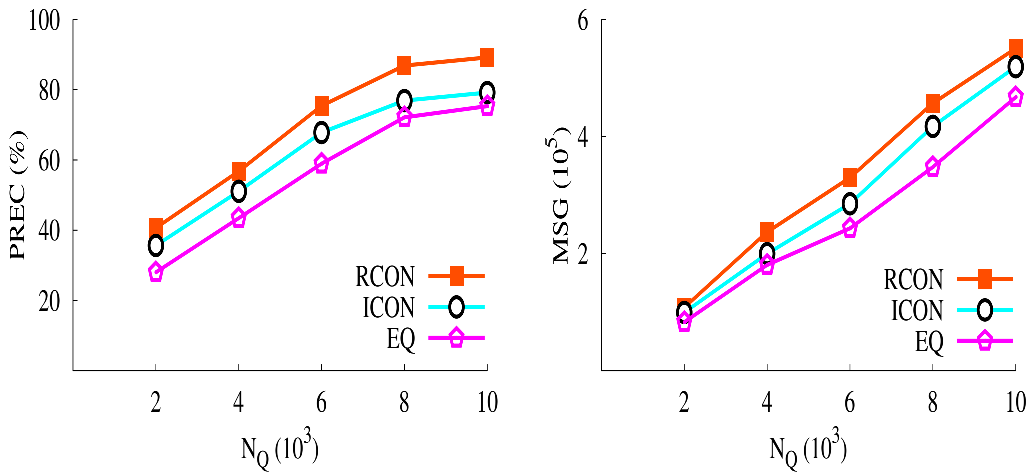
Figure 3.10: Effect of MP failures

3.5.8 Effect of different payoff allocation approaches in ETG

We conducted an experiment to investigate the relative performance of ETG with the different payoff allocation approaches, namely EQ, ICON and RCON. Figure 3.11 depicts the results. As the number of queries increases, performance improves for all the approaches partly due to better filtering by the group leader and partly because participants lower the item selection probability for penalized items. (Recall that in this work, participants in ETG decrease the item selection probability in the same manner as in the ETK+ scheme.) In effect, the filtering occurs iteratively to refine the top-k results across an increased number of queries, thereby improving both QCR and PREC.



(a) ART (b) QCR



(c) PREC (d) MSG

Figure 3.11: Effect of different payoff allocation approaches in ETG

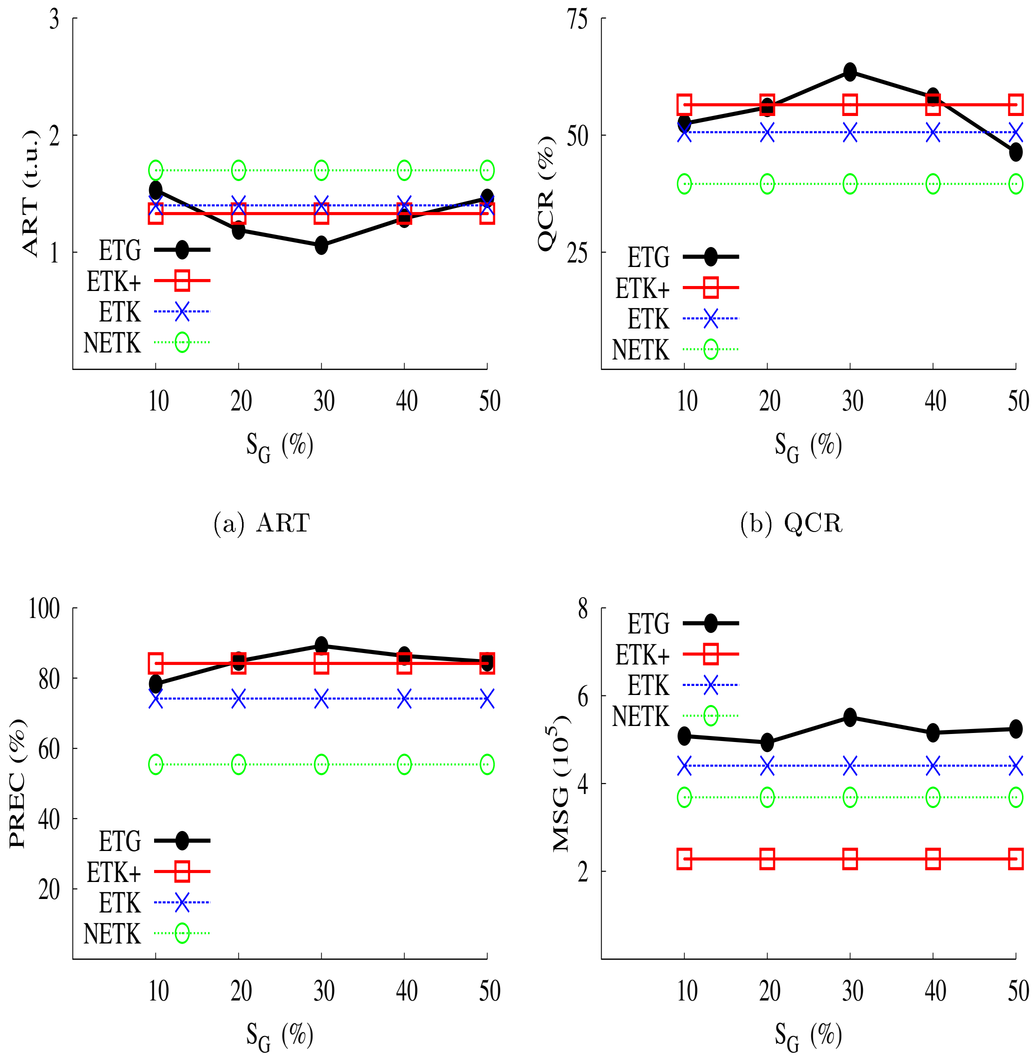
ICON outperforms EQ because it provides better incentives to contributors. Unlike EQ, it takes into account the number of relevant items sent by contributors for distributing rewards. Moreover, RCON outperforms ICON since it better incentivizes contributors by tying rewards to the revenue earned by the items. QCR and PREC saturate for all the approaches after 8000 queries have been processed due to network partitioning and unavailability of some of the rankers. RCON incurs more MSG than ICON, and ICON incurs more MSG than EQ because better incentives entail more participation towards top-k query processing.

3.5.9 Effect of variations in the group size

We define the size of a group as the number of MPs in it. We conducted an experiment to investigate the effect of variations in the group size. For simplicity, we divide the region of interest into square cells of equal area in a grid. MPs moving within any given cell constitute a group, hence each cell corresponds to a group. Thus, we vary the group size by adjusting the area of the cells. Hence, if we increase the area of each cell, the group size increases and vice versa. Notably, although the cells are of equal area, group size may vary across cells because MPs are not uniformly distributed across the region.

We define a parameter S_G to quantify the side-length of an individual cell as a percentage of the total side-length of the region. Recall that our region is 1000 metres $\times$ 1000 metres. Hence, when $S_G = 10\%$, each cell has an area of 100 metres $\times$ 100 metres, hence there will be 100 cells i.e., groups. Similarly, when $S_G = 30\%$, each cell has an area of 300 metres $\times$ 300 metres, hence there will be $11.11 \simeq 12$ cells. (Notably, the group corresponding to the last cell is likely to have fewer MPs than the first eleven cells.) Observe how the number of cells (and consequently, groups) decreases drastically with increase in S_G.

Figure 3.12 depicts the effect of variations in S_G. At low values of S_G, the number of groups is high, but each group is relatively small in size due to fragmentation. Hence, it becomes more difficult for the group leaders to obtain k relevant items for sending to the broker. Thus, group members behave more like individual rankers, thereby not fully realizing the benefits of group-based collaboration. Hence, ETG exhibits improved performance



(a) ART (b) QCR

(c) PREC (d) MSG

Figure 3.12: Effect of variations in the group size

as S_G is increased from 10% to 30% due to increase in group size.

On the other hand, at high values of S_G, group size becomes large, hence relatively fewer groups exist. However, when the group size is too large, performance degrades. In the extreme case when the group encompasses the whole region, the performance essentially reduces to that of ETK and ETK+ because all the MPs act as part of one group. This explains why the performance of ETG degrades and becomes comparable to that of ETK and ETK+, at $S_G = 40\%$ and beyond. Observe that ETG performs best at $S_G = 30\%$ when the group sizes are neither too small nor too large.

Interestingly, MSG is highest for ETG at $S_G = 30\%$ due to more group interactions in processing a larger number of successful queries.

3.6 Summary

We have proposed the E-Top system for efficiently processing top-k queries in M-P2P networks. E-Top issues economic rewards to the MPs, which send relevant data items, and penalizes peers otherwise, thereby optimizing the communication traffic. Peers use the payoffs as a means of feedback to re-evaluate the scores of their items for re-ranking purposes.

E-Top uses three economic incentive schemes, namely ETK, ETK+ and ETG. In ETK and ETK+, MPs act individually, the difference being that ETK performs equal distribution of payoffs to the rankers, while ETK+ uses a weighted distribution. ETG extends ETK and ETK+ by considering MP collaboration in groups. Our performance evaluation demonstrates that E-Top is indeed effective in improving the performance of top-k queries in terms of query response times and accuracy at reasonable communication traffic cost.

The proposed schemes in E-Top require some parameters to be tuned manually. In the near future, we plan to extend E-Top by devising an online auto-tuning method for these parameters to enhance the usability and novelty of the proposed schemes. Based on the application and the real implementation, we can fine-tune these parameters. Moreover, we plan to extend E-Top by incorporating game-theoretic techniques.

4

E-Broker: Economic Incentive-based Brokerage Schemes for Improving Data Availability in Mobile-P2P Networks

4.1 Overview

In a Mobile ad hoc Peer-to-Peer (M-P2P) network, mobile peers (MPs) interact with each other in a peer-to-peer (P2P) fashion [MMK09]. Proliferation of mobile devices (e.g., laptops, PDAs, mobile phones) coupled with the ever-increasing popularity of the P2P paradigm (e.g., Kazaa, Gnutella) strongly motivate M-P2P network applications, which facilitate MPs in sharing information *on-the-fly*. For example, an application could involve an MP looking for an available parking slot within 1 km of its current location. MPs in the vicinity can collect information about available parking slots and charges, and then they can inform the brokers. The broker can then provide the available parking slots to the query-issuing MP in terms of price or distance (from the user's current location). Note that the parking slot availability information has to be current. Incidentally, although we consider brokers, the nature of the networking environment is still *ad hoc* in the sense that the peers can move and they can change their brokers. Hence, the presence of brokers does not make our environment completely structured.

In a similar vein, a user could look for a restaurant with "happy hours" (or "manager's special hours") within 1 km of her current location. A broker can facilitate such queries by soliciting information from the peers moving in the vicinity of the query location. Similarly, an MP may want to find nearby shops selling Levis jeans in a shopping mall with criteria such as (low) price during a specific time duration. Observe that such ad hoc queries are spatio-temporal in nature (e.g., parking slot availability information), hence they cannot be answered by the broker without obtaining information from other MPs. Incidentally, such P2P interactions, which facilitate *spatio-temporal* queries among MPs, are generally not freely supported by existing wireless communication infrastructures. Notably, this research will also contribute towards CrowdDB [FKK+11], which uses human input via crowdsourcing to process queries that cannot be answered by database systems or search engines.

Our target applications mainly concern slow-moving objects e.g., mobile users in a shopping mall. Notably, our application scenarios consider tolerance to lower data quality depending upon the requirements of the peers. We measure data quality in terms of image resolution or MP3 audio quality. Moreover, observe that the inherently ephemeral nature of M-P2P environments necessitates query deadlines.

Data availability in M-P2P networks is typically lower than in fixed networks due to frequent network partitioning arising from peer movement and also due to mobile devices being autonomously switched 'off'. Rampant free-riding further reduces data availability i.e., most peers do not provide any data [HA05, KSGM03a]. (Nearly 90% of the peers in Gnutella were free-riders [AH00].) Incidentally, data availability is less than 20% even in a wired environment [SGG01]. Given the generally limited resources (e.g., bandwidth, energy, memory space) of MPs and the fact that relaying messages requires energy, the relay MPs may not always be willing to forward queries in the absence of any incentives, let alone search *pro-actively* for query results in order to ensure timeliness of data delivery. Thus, providing incentives for relay MPs to pro-actively search for query results becomes a necessity to improve data availability in M-P2P networks. Notably, many schemes such as replication-based schemes, reward-and-punish-based schemes and trust-based schemes can also be used for improving data availability, but the focus in this work is on incentive-based schemes. Observe that increased MP participation in providing service to the network would likely lead to better data availability, better data quality, higher available bandwidth and multiple paths to answer a given query.

Existing schemes for improving data availability in mobile ad-hoc networks (MANETs) [HM06] focus on replication, but they do not use economic incentives to encourage peer participation. On the other hand, incentive schemes [BH03, CN04, CGKO03, SNCR03] for MANETs primarily focus on providing incentives to relay MPs to forward messages, but they do not address the issue of creating pro-active MPs for providing value-added routing services. M-P2P incentive schemes [WXS04, XWR06] also do not incentivize relay MPs to perform value-added routing and to host data.

This work proposes the E-Broker system for improving data availability in M-P2P networks by incentivizing MPs to provide *value-added routing service*. Here, the term “value-added routing service” refers to the broker MPs enabling pro-active search for the query results by maintaining an index of the data items (and replicas) stored at other MPs (as opposed to just forwarding queries). The main contributions of E-Broker are three-fold:

1. It proposes the EIB (Economic Incentive-based Brokerage) scheme,

which incentivizes relay peers to act as information brokers for performing value-added routing and replication in M-P2P networks, thereby effectively improving data availability.

2. It proposes the EIB+ (enhanced Economic Incentive-based Brokerage) scheme, which extends the EIB scheme by incorporating three different broker scoring strategies for providing additional incentives to brokers towards providing better service. Brokers with higher scores become *preferred brokers* and they earn higher commissions than *common brokers*. EIB+ also facilitates load-sharing among the peers.

3. It experimentally determines the number of brokers, beyond which the mobile peers are better off without a broker-based architecture i.e., they can directly access data from the data-providing peers.

E-Broker also discourages free-riding in M-P2P networks. Both EIB and EIB+ use economic incentives in that every data item is associated with a *price* (in *virtual currency*). Data item price depends upon several factors such as access frequency, data quality and estimated response time of access. The query-issuer pays the price of the queried item to the data-provider, and a commission to the broker and the relay MPs in the successful query path.

Both EIB and EIB+ use a bid-based brokerage approach, in which brokers collect bids from data-providers and then create a summary of recommendations based on the query preferences specified by the query-issuer M_I. Based on the bids and the application, M_I selects a single bid depending upon the price that it wants to pay and its desired data quality. After a bid is accepted, M_I obtains the requested data item directly from the data-provider and pays the commission to the broker. Brokers also replicate frequently queried data items to earn revenues as well as to reduce the traffic.

We have evaluated the performance of EIB and EIB+ w.r.t. the non-economic **E-DCG+** replication scheme [HM06]. Notably, E-DCG+ is the closest to our schemes since it aims at improving data availability in MANETs. As a baseline, we also do performance comparison w.r.t. a non-incentive and non-broker-based **NIB** (Non-Incentive without Brokerage) scheme to show the performance gain due to brokerage. We experimentally determine that EIB and EIB+ perform best when the percentage of brokers is 20% of the total number of MPs. Moreover, EIB+ performs best when the percentage of preferred brokers is 20% of the total number of brokers. Both EIB+ and EIB outperform E-DCG+ and NIB due to economic incentives and brokerage. EIB+ performs better than EIB due to preferred brokerage and load-sharing. Furthermore, E-DCG+ outperforms NIB due to its superior replication scheme.

Both EIB and EIB+ exhibit good scalability with increasing number of MPs due to increased opportunities for replication. However, their performance degrades with increasing percentage of MP failures essentially due to reduced MP participation. With increasing workload skew, EIB+ shows better performance than the other schemes primarily due to its load-sharing mechanism. A preliminary version of this work has appeared in [MMK07c].

The remainder of this chapter is organized as follows. Section 4.2 discusses architecture of E-Broker system. Section 4.3 details the EIB brokerage scheme, while Section 4.4 discusses the enhanced brokerage scheme EIB+. Section 4.5 reports our performance study. Finally, we summarize E-Broker in Section 4.6.

4.2 Economic Incentives in E-Broker

This section discusses the economic incentives in E-Broker. These incentives are used by both the EIB and EIB+ brokerage schemes. We defer the discussion of the brokerage schemes to Sections 4.3 and 4.4. Incidentally, each MP maintains recent access statistics of data items (and replicas) hosted at itself for the purpose of computing data item prices. We assume that there could be one original version of any given data item d and multiple replicas of d hosted at different MPs. Memory space of MPs, bandwidth and data item sizes may vary.

4.2.1 Querying-related incentive issues

Each query is a request for a data item. Queries are of the form (Q_{id}, DDQ, ϵ, τ_S, τ_H, max_μ, $BType$, w_1, w_2, w_3), where Q_{id} is the unique identifier of the query, while DDQ represents the desired data quality of the query-issuer M_I. To satisfy query deadlines, M_I stops accepting bids from brokers after ϵ time units have elapsed since the time of query issue. (The significance of ϵ will become clear when we discuss our brokerage scheme in Section 4.3.) Here, τ_S and τ_H are M_I's specified soft and hard deadlines for answering the query. max_μ is the maximum price that M_I is willing to pay for the query.

$BType$ is M_I's specified broker type for the query and assumes two values i.e., 0 for a common broker and 1 for a preferred broker. As we shall see later, in case of EIB+, $BType$ can assume either value, but for EIB, $BType$ always equals 0 since EIB does not consider preferred brokers. Here, w_1, w_2 and w_3 are the query-issuer's specified weight coefficients for the query such that $0 \leq w_1, w_2, w_3 \leq 1$ and $w_1 + w_2 + w_3 = 1$. As we shall see in Section 4, these weight coefficients pertain to query response time, data quality and data item price respectively, and they are used by the broker for computing the ranking scores for the data items in the query result set (See Equation 4.4).

Given that a query Q for a data item d is issued at time t_0, if Q is answered within time $(t_0 + \tau_S)$ (i.e., within the soft deadline), M_I pays the price μ of d to the data-provider M_S. However, if Q is answered within the time interval $[t_0 + \tau_S,\ t_0 + \tau_S + \tau_H]$, M_I pays a reduced price for d to M_S, thereby penalizing M_S for delayed service. Higher delay implies more reduction in price. Finally, if Q is answered after the hard deadline τ_H, M_I does not pay any currency to M_S. This is consistent with the timeliness requirements of M-P2P environments.

Observe that there is no incentive for a data-provider to answer a query after the deadline. Hence, data-providers estimate (based on past statistics concerning network history) whether their transmitted data item will reach the query-issuer within the deadline. Based on their estimate, they decide whether to send the data. Notably, such estimation requires synchronized clocks among the MPs. For example, if an MP receives a message with a timestamp, clock synchronization among the MPs would become a necessary condition for the MP to calculate the delay. The existing clock synchronization approaches proposed in [SCHS07] can be used in conjunction with our proposed approach. However, an MP cannot absolutely know in advance whether its answer will reach the query-issuer in a timely manner because of issues such as network congestion, relay node failures and network partitioning.

Incidentally, if an MP is not able to pay the price of accessing its requested data item, its query fails and it would not be able to access its queried data item. This is in consonance with our overall objective of incentivizing free-riders to provide replica hosting, brokerage and relay services. If our scheme allowed MPs to access data items without having to pay for the access, the free-riders would have little or no incentive to provide service.

4.2.2 Price of a data item

Each data item d has a *price* μ (in *virtual currency*) that quantitatively reflects its relative importance to the M-P2P network. When an MP issues a query for a data item d, it pays the price of d to the MP serving its request. (A query request could also be satisfied by a replica.)

The price μ of d depends upon d's (recent) access frequency, average query response times (w.r.t. deadlines) for queries on d and data quality of d. An MP M_S computes the price of a data item (or replica) d stored at itself in two steps: (a) M_S first computes the price μ_{rec} of d based on accesses to d during the most recent time period. (We divide time into equal intervals called *periods*, the size of a period being application-dependent.) (b) M_S computes the moving average price μ of d based on the previous N time periods. The moving average price is necessary to take spurious spikes in accesses to d into account to ensure that d's price actually reflects its importance. M_S computes μ_{rec} of d as follows:

$$\begin{aligned}\mu_{rec} \;&=\; \textstyle\int_{t_1}^{t_2} \int_0^{\delta} (\eta\, dt \,\times\, (1/\delta^2) d\delta \,\times\, \tau \,\times\, DQ \times \\ &\qquad\qquad BA_{M_S} \,\times\, PA_{M_S}) \,/\, J_{M_S, t_j}\end{aligned} \tag{4.1}$$

where $[t_2 - t_1]$ represents a given time period, and δ is the number of hops between the query-issuer M_I and the data-provider M_S during the time of query issue. We assume that the query message maintains a counter that is incremented with each hop. Thus, M_S can know the number of hops between itself and M_I at the time of query issue by examining the query message. Furthermore, we assume that the number of hops between M_I and M_S does not change significantly between the time of query issue and the time of query retrieval. Observe how μ_{rec} decreases as δ increases due to likely increased query response times.

In Equation 4.1, η is the access frequency of the given data item d during the most recent time period. τ reflects the price reduction (i.e., penalty) due to delayed service. Given that t_0 is the time of query issue, and t_q is the time when the query results reached the query issuing MP, τ is computed as follows.

$$\tau \;=\; \begin{cases} \mu & if\, t_0 \geq t_q \geq (t_0 + \tau_S) \\ \mu \times e^{-(t_q - \tau_S)} & if\, (t_0 + \tau_S) \geq t_q \geq (t_0 + \tau_S + \tau_H) \\ 0 & otherwise \end{cases} \tag{4.2}$$

where τ_S and τ_H are the soft and hard deadlines of a given query respectively. Notably, the data-provider M_S estimates the time when the query results would reach the query-issuer M_I based on the average network conditions and historical information. Hence, in Equation 4.1, the data item price is first estimated by M_S using the value of τ based on the estimation of the time when the query results would reach M_I. Payments are done periodically, and information concerning the actual times of query results reaching the respective query-issuers is piggybacked onto the status messages that are sent periodically by the peers to the brokers. Thus, the actual price, which is paid by M_I to M_S, is based on the actual time when the query results reached a given query-issuer.

The term DQ in Equation 4.1 reflects the quality of data provided by M_S for queries on d. DQ is essentially application-dependent. For example, data quality could be determined based on MP3 audio quality or image resolution. We compute DQ as follows. Each MP maintains a copy of the table T_{DQ}, which contains the following entries: (x%, high), (y%, medium), (z%, low), where x, y, z are error-bounds, whose values are application-dependent and pre-specified by the system at design time. Essentially, we consider three discrete levels of DQ i.e., *high*, *medium* and *low*, and their values are 1, 0.5 and 0.25 respectively.

In Equation 4.1, BA_{M_S} is the bandwidth allocated by M_S for d's download. BA_{M_S} equals $(\sum B_i)/n_d$, where B_i is the bandwidth that M_S allocated for the i^{th} download of d from itself during the most recent time period, while n_d is the number of downloads of d from M_S. As BA_{M_S} increases, μ_{rec} increases because higher bandwidth implies reduced response times for queries on d. PA_{M_S} is the probability of availability of M_S. When PA_{M_S} is high, the implication is that other MPs can rely more on M_S to provide d, hence μ_{rec} increases with increase in PA_{M_S}. J_{M_S,t_j} is the job queue length at M_S during time t_j. μ_{rec} decreases with increase in the job queue of M_S because when M_S is overloaded with too many requests, M_S's response time in answering queries on d can be expected to increase due to longer waiting times of queries.

After computing μ_{rec}, M_S computes the moving average price μ of d. We use the Exponential Moving Average (EMA), which is capable of reacting quickly to changing access patterns of data items since it gives higher weights to recent access patterns relative to older access patterns. This is in consonance with the dynamically changing access patterns that are characteristic of M-P2P networks. M_S computes the price μ of d as follows:

$$\mu \ = \ (\mu_{rec} - EMA_{prev}) \ \times \ 2/(N+1) \ + \ EMA_{prev} \tag{4.3}$$

where EMA_{prev} represents the EMA that was computed for the previous time period, and N represents the number of time periods over which the moving average is computed. Our preliminary experiments suggest that $N = 5$ is a reasonably good value for our application scenarios.

4.2.3 Revenue of an MP

The revenue of an MP M is the difference between the amount of virtual currency that M earns and M spends. M earns virtual currency from accesses to its own data items and replicas that are hosted at itself, and through relay and broker commissions. Conversely, M spends currency when it queries for data items hosted at other MPs.

We incorporate commissions to incentivize relay MPs. Relay commission is a constant k. We use the price μ_{min} of the cheapest data item in the network as a guide to determining a suitable value of k. The value of k is selected to be lower than that of μ_{min} to incentivize data sharing more than relay functions. Observe that the value of μ_{min} could change over time because new items could be introduced into the network. However, based on the application, it is feasible to estimate the value of μ_{min}. Thus, the value of k is essentially application-dependent. We defer the discussion of broker commissions to Section 4.3.

Notably, every MP joining the system is provided with a small initial amount of currency for bootstrapping the system. Observe that the MPs would soon exhaust this initial amount of currency by issuing queries, and by paying the data item prices and relay commissions. Hence, after that, they would have to earn currency for issuing their own requests, and they can earn currency only by hosting items and relaying messages, thereby effectively combating free-riding. Observe how our economy-based paradigm encourages MPs to increase their revenues, thereby ensuring that they obtain better service.

4.3 EIB: An Economic Incentive-based Brokerage scheme for M-P2P networks

This section discusses our proposed EIB scheme.

4.3.1 Role of the brokers in EIB

EIB provides an incentive to the relay MPs to act as brokers by pro-actively searching for the query results as opposed to just forwarding queries. A broker obtains a commission for each query processed successfully through itself. Hence, each MP is incentivized to maintain an index of the data items (and replicas) stored at other MPs. This index is built by each MP on-the-fly in response to queries and data that it relays. Hence, indexes may differ across MPs. Brokers also provide value-added service in EIB by replicating frequently queried data items at themselves.

Notably, the mobile peers participating in the system have software installed in their mobile devices, and this software enables them to use the proposed schemes. Once they use this software, they have to follow our architecture i.e., they have to go through the brokers. Thus, when using the software, a selfish query-issuer cannot contact the data provider directly by bypassing the brokers. In this regard, the rationale behind our architecture (i.e., every query must pass through brokers) is that query-issuing peers would not want to evaluate a large number of replies coming from prospective data-providers. Moreover, such evaluation would drain their limited energy resources. Furthermore, query-issuing peers would want to have more options (e.g., price, quality) about their requested data items, and the broker is in a position to provide such options.

A data-provider may allow a broker to host a replica of some of its 'hot' data items in lieu of a royalty payment. This is possible because we use our proposed royalty-based revenue-sharing scheme [MMK09] in conjunction with EIB. Brokers have an incentive for hosting replicas of 'hot' items because they can earn revenue when those replicas are queried. Data-providers are incentivized to replicate their 'hot' items at brokers because they can earn revenue from accesses to the replicas without necessitating any expenditure of their limited energy resources for answering queries on those items. In this manner, even if a data-provider is disconnected, it can still earn revenues.

To perform replication, every data-provider periodically broadcasts a list of items that it wants to replicate. Brokers intercept this broadcast and decide whether to replicate these items based on their estimate about the future access frequencies and prices of those items. (This estimate is made based on the queries that pass through a broker.) Since brokers have limited memory space for hosting replicas, each broker tries to select only those items, which would maximize its revenue-earning potential. An item's revenue-earning potential is the product of its price and its (estimated) access frequency. Thus, EIB facilitates brokers in replicating frequently queried items, thereby reducing the querying traffic. In essence, EIB effectively converts relay MPs into brokers.

4.3.2 Illustrative example for the network topology in EIB

The architecture of EIB consists of query-issuers, relay MPs, brokers and data-providers. Figure 4.1 depicts an illustrative example of the M-P2P network topology in EIB at a certain point in time. In Figure 4.1, M_I is the query-issuer, $R1$ to $R7$ are the relay peers, $D1$ to $D4$ are the data-providers, and $B1$ to $B5$ are the brokers. Using Figure 4.1, we shall now make certain key observations. Observe that there can be multiple paths from a query-issuer to a given data-provider and these paths may pass through multiple brokers. As a single instance, a query issued by M_I for a data item hosted by $D4$ could proceed through multiple paths such as $\{M_I, B2, B3, B4, R4, D4\}$ and $\{M_I, B2, B3, B4, R5, D4\}$.

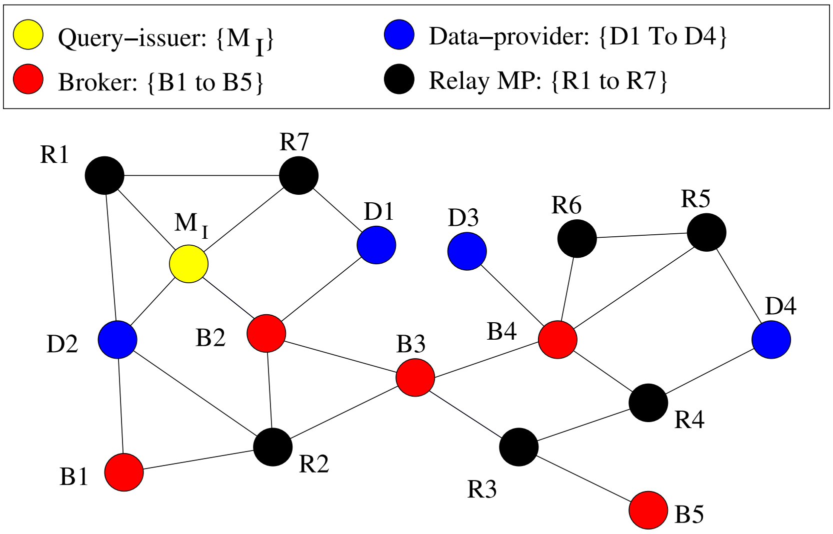


Figure 4.1: Illustrative example of an instance of network topology in EIB

Our scheme stipulates that only one MP can act as the broker in a given query path. This becomes a necessity to avoid conflicts among brokers. Hence, when multiple brokers exist in a given query path, the broker, which occurs first in the traversal starting from the query-issuer, would act as the broker for the query and make the bid to the query-issuer, while the other brokers would simply act as relay MPs. For example, in the query path $\{M_I, B2, B3, B4, R4, D4\}$, $B2$ would act as the broker since it occurs first in the traversal starting from M_I, while $B3$ and $B4$ would act as relay MPs. When an MP decides to act as the broker for a query, it appends a *broker tag* to the query message, thereby enabling other MPs in the same query path to determine that a broker has already been selected in that query path. Notably, even though EIB limits the number of brokers in a given query path to only one, the existence of multiple query paths safeguards against the unavailability of some of the brokers.

The number of relay MPs between a query-issuer and a data-provider may differ. For example, let us consider a query Q issued by M_I for a data item hosted by $D4$. In this case, the query path $\{M_I, B2, B3, B4, R4, D4\}$ has three relay MPs, namely $B3$, $B4$ and $R4$. On the other hand, the path $\{M_I, D2, B1, R2, B3, R3, R4, D4\}$ has five relay MPs, namely $D2$, $R2$, $B3$, $R3$ and $R4$. Thus, the total cost of relay commissions may vary across query paths since EIB incorporates a constant relay commission per relay MP, as discussed in Section 4.2.

It is also possible for a given data-provider to be a one-hop neighbour of a query-issuer e.g., M_I and $D2$ are one-hop neighbours. However, our architecture dictates that M_I cannot bypass the brokers for directly obtaining its queried data from $D2$. Recall that the mobile peers are able to use the proposed schemes by installing software in their mobile devices, and this software enforces that each query must follow our architecture by going through the brokers. Thus, the role of the brokers would still be relevant in such cases. For example, some other data-provider such as $D1$ may be able to provide better data quality and/or lower response time than $D2$ (e.g., due to low bandwidth between $D2$ and M_I). In essence, the brokers provide the query-issuer with different paths for accessing its requested data item d or its replica. This allows the query-issuer to choose the copy of d, which best suits its requirements in terms of response time, data quality and price. Furthermore, as discussed earlier, there may be many prospective data-providers replying to a query, and the query-issuer would not want to evaluate a large number of replies since performing such evaluation would drain its limited energy resources.

4.3.3 Value-added routing by relay MPs in EIB

An MP M_I issues a query Q using a broadcast mechanism[1] and waits until ϵ time units have elapsed (since the time of query issue) to collect the bids from all the brokers. When any given MP receives the broadcast query, it checks

[1] After a period of time, if M_I knows a broker that can serve the query, the broadcast would not be necessary.

its index. If its index does not contain the identifier of at least one MP that hosts the requested data item or if another broker (in the same query path) has already decided to act as the broker for that query[2], it simply forwards the query. Otherwise, it determines (from its index) the MPs, which can answer the query, and acts as a broker by issuing a route-finding query to locate these MPs.

Once a given broker obtains the route to one or more MPs that can serve the query, it acquires information about the price and data quality of the requested data item at each of these MPs. Thus, the broker summarizes information of the form (d, MP_{id}, μ, DQ, $Path$) in a list L_{bid}, where d is the data item being requested, MP_{id} is the unique identifier of the MP that hosts d, DQ is the data quality of d and μ is the price of d. $Path$ is a linked list data structure containing the list of MPs, which fall in the path between the broker and the data-provider. In case of multiple paths between the broker and the data-provider, $Path$ could be a pointer to a set of linked lists (or a two-dimensional array).

Observe that if the broker were to include in its bid (to the query-issuer) all the data items about which it has acquired information, communication traffic would increase and the query-issuer would have to expend its limited energy resources to evaluate all the query results. On the other hand, if the broker were to include only one data item in its bid, the query-issuer would have limited choices (in terms of query results), which could potentially not satisfy its query requirements in terms of response time, data quality and price. Hence, the broker provides a value-added service by including in its bid only *some* of the data items about which it has acquired information. The broker determines which items it will include in its bid by using the information in the list L_{bid}. For each data item in the list L_{bid}, the broker

[2]Recall that only one MP can be the broker in a given query path.

computes a score γ:

$$\gamma \ = \ (w_1/RT) \ + \ (w_2 \times DQ) \ + \ (w_3/\mu) \tag{4.4}$$

where RT represents the query response time, which is estimated by the broker based on network statistics. RT is estimated by the data item size divided by the sum of the bandwidths at the intermediate hops between the query-serving MP and the query-issuer. DQ and μ are the data quality and price of the item respectively, and they are evaluated in the same manner as discussed for Equation 4.1. In Equation 4.4, w_1, w_2 and w_3 are the query-issuer's specified weight coefficients for the query such that $0 \leq w_1, w_2, w_3 \leq 1$ and $w_1 + w_2 + w_3 = 1$. Thus, EIB takes the requirement of the query-issuer into account.

The value of γ increases with decreasing values of RT and μ because the query-issuer would want the results quickly and with lower price. The value of γ increases with increase in DQ because higher data quality commands higher bid price. The broker includes in its bid (to the query-issuer) only those items, for which the value of γ exceeds the threshold Th_γ, where Th_γ is the average value of γ for all the items in L_{bid}. Hence, Th_γ equals $(\Sigma_{i=1}^{N} \gamma_i/N)$, where γ_i is the value of γ for the i^{th} item and N is the total number of items in L_{bid}. The values of RT and DQ for each item in every bid are also provided by the broker to the query-issuer.

Corresponding to each data item included in the broker's bid, the broker also specifies the total cost of relay commissions and broker commission to inform the query-issuer about the total cost of querying. Since the broker knows the number of relay MPs in the query path, it can compute the total cost of relay commission since the amount of relay commission per MP is a constant, as discussed earlier in Section 4.2. The amount β of broker commission for a given data item d depends upon the data item price. Given a data item d of price μ, a broker computes β as $(\mu \times \alpha)$, where α is a percentage of the data item price, hence $0 \leq \alpha \leq 1$. The value of α depends upon the urgency of the query-issuer. Thus, we compute α as $e^{-\tau_S}$, where τ_S is the soft deadline of the query. Increase in τ_S implies decrease in β due to less urgency. Observe that different brokers may bid different amounts of currency for the same data item (or its replica). Incidentally, the broker's commission is significantly higher than that of the relay MPs' commissions, which incentivizes relay MPs to act as brokers by indexing more data items.

Upon receiving bids from possibly multiple brokers, the query-issuer *autonomously* evaluates each item in each of these bids. (Recall that each broker may send multiple items in its bid to provide the query-issuer with more options.) Then the query-issuer selects the item, which best suits its requirements in terms of the weight coefficients w_1, w_2 and w_3 corresponding to (estimated) response time, data quality and price respectively. In particular, EIB does not force a query-issuer to perform bid selection based on any specific algorithm. This is because we believe that query-issuers should be provided the flexibility to choose the item (in the bids) that best satisfies their requirements.

An example to illustrate a possible way in which a query-issuer could select an item from multiple bids is as follows. Suppose $w_1 > w_2 > w_3$. In this case, the query-issuer could first sort the items in all the bids in ascending order of estimated response time into a list L_{Select}. Then from L_{Select}, it could select only those items, whose estimated response time is lower than the average response time of all the items in L_{Select}. Then it could sort the remaining items in L_{Select} in descending order of data quality, and select only those items, whose data quality exceeds the average data quality of all the (remaining) items in L_{Select}. Finally, among the remaining items in L_{Select}, it could select the item with the lowest price.

Upon completion of the bid selection, the query-issuer contacts the broker corresponding to the successful bid, and requests it for the data item. The successful broker contacts the data-provider, which sends the data item to the query-issuer. Finally, upon receiving the query results, the query-issuer pays the commission to the broker and the relay commissions to the MPs in

Algorithm 4.1 EIB: Algorithm for a query-issuer

begin

Inputs: (a) Q : Query (b) d : Queried data item

(1) Broadcast its query Q for a data item d
(2) Receive all bids that arrive within ϵ time units of query issue
(3) Examine each item in every bid and autonomously select the item, which best suits query requirement
(4) Select the broker Sel corresponding to the successful bid
(5) Send message to selected broker Sel requesting selected item and provide Sel with identifier of selected data-provider M_S
(6) Obtain data item from M_S
(7) Pay the price of the item to M_S
(8) Pay the broker commission to the selected broker Sel
(9) Pay relay commissions to relay MPs in successful query path

end

Algorithm 4.2 EIB: Algorithm for a broker and relay MPs

begin

Inputs: (a) Q : Query (b) d : Queried data item

(1) Receive the broadcast query Q for data item d from query-issuer M_I
(2) if broker_tag not attached to Q
/* EIB stipulates one broker per query path */
(3) Check own index to list the identifiers of all MPs hosting d into a set Set_{M_S}
(4) if Set_{M_S} is empty
(5) Forward Q to its one-hop neighbours
(6) else
(7) for each M_S M in Set_{M_S}
(8) Issue a query to find the route(s) to M
(9) List all the routes from itself to M into a set Set_{route}
(10) if Set_{route} is empty
(11) Forward Q to its one-hop neighbours
(12) else
(13) Select the shortest route R from itself to M based on bandwidths at the intermediate hops
(14) Obtain price and data quality of d from M, and add d to a list L_{bid}
(15) Select from L_{bid} only those items, for which the value of γ exceeds Th_γ and include these items in the bid
(16) For each item included in the bid, collate all the price, M_S, response time and data quality information and the bid value β
/* The bid value β for a given data item is a percentage of the data item price. β is the broker commission for a successful bid. */
(17) Send the bid to M_I
(18) Wait for M_I's reply
(19) if M_I accepts bid
(20) Obtain identifier of selected M_S from M_I
(21) Send a message to selected M_S to send the data item to M_I
(22) Receive broker commission from M_I

end

the successful query path.

Algorithm 4.1 is executed by a query-issuer, while Algorithm 4.2 is executed by the other MPs, which can either be brokers or relay MPs.

4.4 EIB+: An Enhanced Economic Incentive-based Brokerage Scheme for M-P2P networks

This section discusses the EIB+ scheme, which extends the EIB scheme by incorporating three broker scoring strategies for further incentivizing brokers towards providing better service. EIB+ distinguishes two different types of brokers, namely *common brokers* and *preferred brokers*. Brokers with higher scores become preferred brokers and they earn higher commissions than common brokers. Furthermore, only the preferred brokers are allowed to spawn sub-brokers for load-sharing purposes, thereby further incentivizing brokers since they can earn currency from royalty-based revenue-sharing [MMK09] with the sub-brokers.

Notably, in order to become a preferred broker, a broker needs to serve a minimum threshold number of users. Thus, if a broker serves an adequate number of different users, the rating scores from different users average out, thereby implying that a broker cannot become a preferred broker by serving only one peer well because broker scores are based on averages. Even though we understand that it is difficult to synchronize the ratings for different brokers, peers can select in their region their preferred brokers. Furthermore, observe that complete synchronization of broker score ratings across different users is not practically feasible due to subjectivity in human judgment.

4.4.1 Illustrative example of network topology in EIB+

Figure 4.2 depicts an illustrative example of the M-P2P network topology in EIB+ at a certain point in time. M_I is the query-issuer, $D1$ to $D4$ are the data-providers, $R1$ to $R4$ are the relay peers, $CB1$ to $CB3$ are the common brokers, and $PB1$ to $PB2$ are the preferred brokers. $SB1$ is the sub-broker corresponding to the preferred broker $PB1$, while $SB2$ and $SB3$ are the sub-brokers corresponding to the preferred broker $PB2$. Observe that the common brokers such as $CB1$ and $CB2$ do not have any sub-brokers. Consider a query Q issued by M_I for a data item hosted by $D4$. For the query path $\{M_I, PB1, CB2, PB2, SB3, D4\}$, if Q needs to be processed by a common broker, $CB2$ would act as the broker, while $PB1$ and $PB2$ would act as relay MPs. If Q needs to be processed by a preferred broker, $PB1$ would act as the broker.

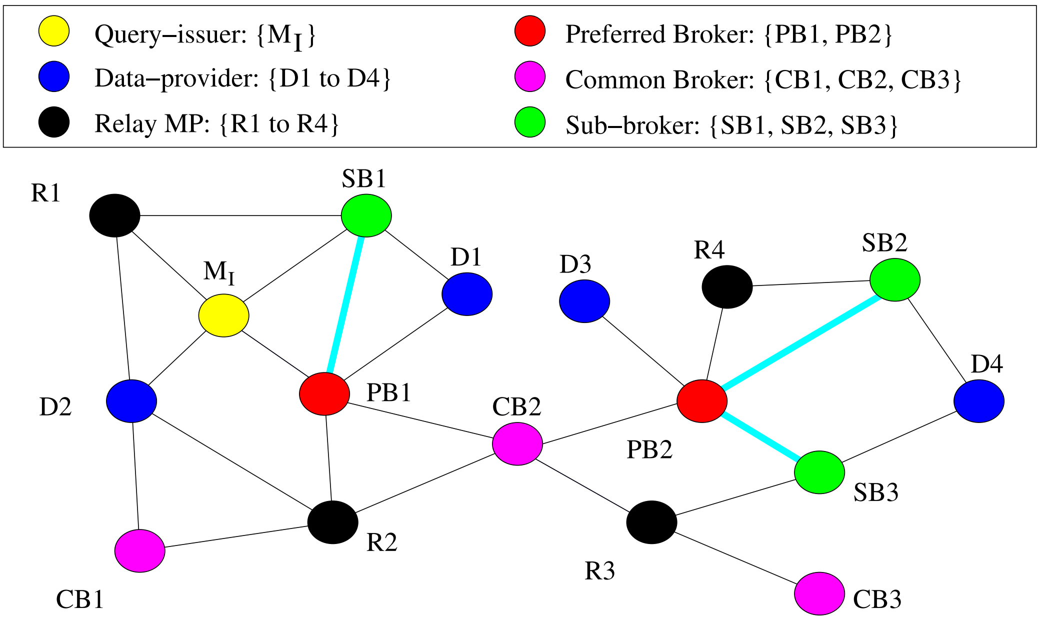


Figure 4.2: Illustrative example of an instance of network topology in EIB+

The broker type (i.e., common or preferred) specified in Q should match with at least one broker in the given query path for it to be processed in that query path. This is in consonance with adhering to the query-issuer's intentions. However, this does not necessarily result in query failures due to the possible existence of multiple brokers (which match the broker type specified in Q) in different query paths. Thus, if M_I issues a query for an item in $D4$ with the condition that it should be processed by a preferred broker, it will not be processed in the path $\{M_I, D2, CB1, R2, CB2, R3, SB3, D4\}$ since this path does not contain any preferred broker. However, it would be processed in other query paths e.g., $\{M_I, PB1, CB2, PB2, SB2, D4\}$ and $\{M_I, PB1, CB2, R3, SB3, D4\}$.

4.4.2 Strategies for assigning scores to brokers

We propose three strategies for assigning performance-based scores to brokers in EIB+.

4.4.2.1 Individual Ranking (IR) strategy

In IR, each MP assigns a score λ to each broker, with whom it has interacted within a particular time-period. Each broker returns a bid to the query-issuer M_I, and the bid contains the estimated query response time, data quality (of query result) and the total bid price for processing the query. (Total bid price refers to the sum of data item price, broker commission and relay commissions.) M_I uses this bid information to compute the value of λ for the broker that made the bid. The value of λ is computed for both successful and unsuccessful bids.

If a query is answered after the hard deadline τ_H, M_I assigns $\lambda = 0$ for that query to the corresponding broker(s) to penalize broker performance because queries answered after the deadline are not useful to M_I. Furthermore, since a broker has no incentive to bid a total price, which is higher than that of M_I's maximum specified price, the question of the total bid price exceeding the maximum specified price does not arise. λ is computed as follows:

$$\lambda \;=\; (w_1 \times \lambda_{RT}) \;+\; (w_2 \times (1 - \lambda_{DQ})) \;+\; (w_3 \times \lambda_{\mu}) \tag{4.5}$$

where λ_{RT}, λ_{DQ} and λ_{μ} quantify broker performance w.r.t. broker response time, data quality and total (bid) price respectively, and they are computed in Equations 4.6, 4.7 and 4.8 respectively. (Broker response time is the difference between the time of query issue and the time at which the broker's response arrives at M_I.) In Equation 4.5, w_1, w_2 and w_3 are weight coefficients such that $w_1+w_2+w_3 = 1$. The values of w_1, w_2 and w_3 are decided by M_I for a given query depending upon its requirements. For example, if quick response time is critical to M_I, it will assign a high value to w_1. Observe how EIB+ provides autonomy to the MPs in assigning scores to brokers based on their individual querying requirements. λ_{RT} is computed below:

$$\lambda_{RT} \ = \ (\ \tau_H \ - \ RT \ ) \ / \ \tau_H \tag{4.6}$$

where τ_H and RT are the hard deadline and the broker response time of the query respectively. Observe that the value of λ_{RT} increases as RT decreases. Thus, the objective of Equation 4.6 is to reward brokers for providing timely service. The amount of reward is based on the difference between the hard deadline of the query and the broker response time. The computation of λ_{DQ} follows:

$$\lambda_{DQ} \ = \begin{cases} (DDQ - DQ)/DDQ & if\ DQ < DDQ \\ 1 & otherwise \end{cases} \tag{4.7}$$

where DDQ and DQ are M_I's specified desired data quality and the actual data quality for the query respectively. The objective of Equation 4.7 is to penalize brokers, which provide lower quality of data than that of M_I's desired data quality. The amount of penalty is based on the difference between M_I's desired data quality and the actual data quality provided by the broker. The value of λ_{DQ} increases as queries are answered with lower data quality, hence we use the value of $(1 - \lambda_{DQ})$ in Equation 4.5 for the computation of λ. However, when $DQ \geq DDQ$, we set $\lambda_{DQ} = 1$ to reward brokers, who have performed upto (or better than) M_I's expectations of data quality.

The computation of λ_μ follows:

$$\lambda_\mu \ = \ (\ max_\mu \ - \mu \ ) \ / \ max_\mu \tag{4.8}$$

where max_μ and μ are the M_I's specified maximum price and the total price bid by the broker for the query respectively. Observe that the value of λ_μ

increases as the total bid price decreases. Thus, the objective of Equation 4.8 is to reward brokers, which can serve the queries at lower total price. The amount of such reward is based on the difference between M_I's maximum specified price and the total bid price of the given query. Thus, an MP will have an estimate about the performance of the brokers that it has interacted with. However, IR suffers from the drawback that each MP is likely to be able to interact with and assign scores to only a few brokers that are in its vicinity.

4.4.2.2 Neighbour-based Gossiping (NGS) strategy

To address the drawback of IR in terms of being able to assign scores to only a relatively few brokers, we propose the NGS strategy. In NGS, MPs gossip with their one-hop neighbours to share their respective broker scores (obtained by using IR). Thus, each MP will get to know the performance of brokers, with whom it may not have had any interaction. For example, suppose MP M_1 has interacted with only brokers $B1$, $B5$ and $B7$, while its one-hop neighbour M_2 has interacted with brokers $B1$, $B6$, $B7$ and $B8$. Thus, M_1 will obtain new information from M_2 about the performance of $B6$ and $B8$, while M_2 will obtain information from M_1 about the performance of $B5$. Gossiping facilitates neighbouring MPs to refine their information about broker scores. Since MPs are likely to obtain new information, they have an incentive to participate in gossiping.

When a given MP M obtains broker scores from its one-hop neighbours, it computes its score for each broker B_i as follows. If M has not interacted with B_i, it will simply compute its score for B_i as the average Avg of all the scores (for B_i) that it receives from its neighbours, who have interacted with B_i. On the other hand, if M has interacted with B_i, it will compute its score for B_i as the average of the score that it assigned to B_i and Avg.

4.4.2.3 K-hop neighbour-based Gossiping (K-NGS) strategy

The K-NGS strategy extends the NGS strategy by allowing gossiping among K-hop neighbours. (Recall that in NGS, gossiping is limited only to one-hop neighbours.) Thus, K-NGS facilitates MPs in assigning scores to more brokers than NGS and also uses inputs about broker scores from more MPs than NGS, thereby providing a broader and more refined picture of relative broker performance albeit at the cost of increased communication overhead. Note that under the K-NGS strategy, a given MP M computes its score for each broker B_i in the same manner as discussed for the NGS strategy.

4.4.3 Load-sharing by means of sub-brokers in EIB+

Preferred brokers in EIB+ are allowed to spawn sub-brokers for load-sharing purposes. Now let us examine the concept of sub-brokers. When a preferred broker PB becomes overloaded with too many requests, it replicates its data and index at MPs, which are willing to host its data and index. We designate such MPs as **sub-brokers**. Thus, preferred brokers dynamically create sub-brokers based on load and network performance to effectively convert relay MPs into brokers. This facilitates load-sharing among preferred brokers and sub-brokers, thereby making it likely to improve query response times due to less waiting times at the job queues of these MPs.

The preferred broker is incentivized to share its data and index with the sub-brokers because it can earn currency from such sharing. This is because we use our proposed royalty-based revenue-sharing scheme [MMK09] in conjunction with EIB+. Thus, revenues of preferred brokers are further increased due to the presence of sub-brokers. Observe how EIB+ incentivizes brokers to perform better in order to become preferred brokers.

A preferred broker PB selects its sub-brokers based on three factors, namely remaining energy, bandwidth and current value of currency. PB prefers MPs with higher remaining energy as sub-brokers because such MPs are likely to be able to serve more queries, thereby facilitating them in earning more currency and consequently, also enabling PB to earn more currency because of the royalty-based revenue-sharing scheme [MMK09]. Moreover, PB gives preference to MPs with high bandwidth because such MPs are likely to serve queries relatively quickly, thereby enabling them to earn more currency. (Recall that data item prices depend upon timeliness of query response.) Furthermore, PB prefers MPs with low current value of currency as sub-brokers because such MPs have more incentive to serve queries to earn currency than MPs, whose current values of currency are high. Notably, this also facilitates newly joined MPs (that have low currency) to seamlessly integrate themselves into the system by actively participating in the network as sub-brokers.

Notably, PB selects its sub-brokers from among its one-hop neighbours in order to minimize the communication traffic incurred for allocating replicas at sub-brokers. To select its sub-brokers, PB sends a message to its one-hop neighbour MPs requesting them to send their values of remaining energy, bandwidth and currency. Those MPs, which are interested to become sub-brokers of PB, reply to PB with the requested values. PB uses these values to compute a score S for each MP as follows.

$$S \;=\; (w_1 \times En) \;+\; (w_2 \times BA) \;+\; (w_3/Curr) \tag{4.9}$$

where En, BA and $Curr$ are the values of remaining energy, bandwidth and currency of the MP. In Equation 4.9, w_1, w_2 and w_3 are weight coefficients such that ($w_1 + w_2 + w_3 = 1$). The values of these weight coefficients are autonomously selected by a given preferred broker, hence they may differ across preferred brokers. MPs with relatively higher values of S are selected by PB as its sub-brokers. We leave the determination of the optimal number of sub-brokers per preferred broker to future work.

4.5 Performance Evaluation of E-Broker

This section reports our performance evaluation by means of simulation in OMNeT++ [Pon93]. MPs move according to the *Random Waypoint Model* [BMJ+98] within a region of area 4 km × 4 km. We believe that the Random Waypoint Model is appropriate for our application scenarios.

Parameter	Default Value	Variations
Number of MPs (N_{MP})	1000	200, 400, 600, 800
% of brokers (P_B)	20%	10%, 30%, 40%, 50%
% of preferred brokers (ψ)	20%	10%, 30%, 40%, 50%
Queries/time unit	10	
Communication Range (CR)	120m	40m, 80m, 160m, 200m
Percentage of MP failures (P_F)	20%	10%, 30%, 40%, 50%
Workload skewness (ZF_W)	0.5	0.1, 0.3, 0.7, 0.9
Bandwidth between MPs	1 Mbps to 2 Mbps (Bluetooth)	
Initial energy of an MP	90000 to 100000 energy units	
MP service capacity	1 to 5 units	
Time-to-expire of a data item	3 mins to 7 mins	
Memory space of each MP	120 MB to 150 MB	
Speed of an MP	1 meter/s to 10 meters/s	
Size of a data item	0.5 MB to 10 MB	

Table 4.1: Parameters of our performance evaluation for E-Broker

Table 4.1 summarizes our performance study parameters. A total of 8000 data items is uniformly distributed among 1000 MPs i.e., each MP owns 8 data items. For each MP, the available memory space for hosting replicas is its remaining memory space, after memory for storing its 8 data items has been allocated. Query-issuers are selected randomly from among all the MPs. Each query is a request for a single data item. The number of such queries issued in the network per time unit is 10, the query's hard deadline τ_H being varied randomly between 25 to 30 time units. The query's soft deadline τ_S is 90% of τ_H. Query price is chosen randomly in the range of 100 to 500 currency units. Broker commission and relay commission are respectively set to 10% and 1% of the query price. For query routing purposes, we use the AODV protocol until a query is intercepted by a broker. Initial energy of an MP is selected randomly between 90000 to 100000 energy units. Sending and receiving a message require 1.5 and 1 energy units respectively.

In Table 4.1, TP stands for 'replica allocation Time Period'. *Periodically*, every TP seconds, MPs broadcast a list of items that they want to replicate. Similar to existing works [HM06], we assume that network topology does *not* change significantly during replica allocation since it requires only a few seconds. The default communication range of all MPs is a circle of 120 metre radius.

For all our experiments, the weight coefficients are set as follows: (a) the values of w_1, w_2 and w_3 for computing γ in Equation 4.4 are set to 0.5, 0.25 and 0.25 respectively, (b) the values of w_1, w_2 and w_3 for computing λ in Equation 4.5 are set to 0.5, 0.25 and 0.25 respectively, (c) the values of w_1, w_2 and w_3 for computing S in Equation 4.9 are set to 0.4, 0.3 and 0.3 respectively.

Our performance metrics are **average response time** (**ART**) of queries, **data availability** (**DA**), **query hop-count** (**HC**) and **communication traffic** (**MSG**). $ART = (1/N_Q)\sum_{i=1}^{N_Q}(T_f - T_i)$, where T_i is the time of query issue, T_f is time of the query result reaching the query-issuer, and N_Q is the total number of queries. ART includes the download time, and is computed only for the successful queries. $DA = (N_S/N_Q) \times 100$, where N_S is the number of queries that were answered successfully. Thus, DA measures the percentage of successful queries. Queries may fail due to network partitioning or due to energy-depletion or unavailability of MPs that host the queried data items, or due to queries exceeding the TTL ('hops-to-live'). Preliminary experiments suggested that TTL=8 is a reasonable value for our application scenarios. Hence, we consider TTL=8 for our proposed EIB and EIB+ schemes. We define the query hop-count **HC** as the hop-count incurred by the query in the successful query path. Thus, HC is measured only for successful queries. We define MSG as the total number of messages incurred for query processing during the course of the experiment. Thus, $MSG = \sum_{q=1}^{N_Q} M_q$, where M_q is the number of messages incurred for the q^{th}

query.

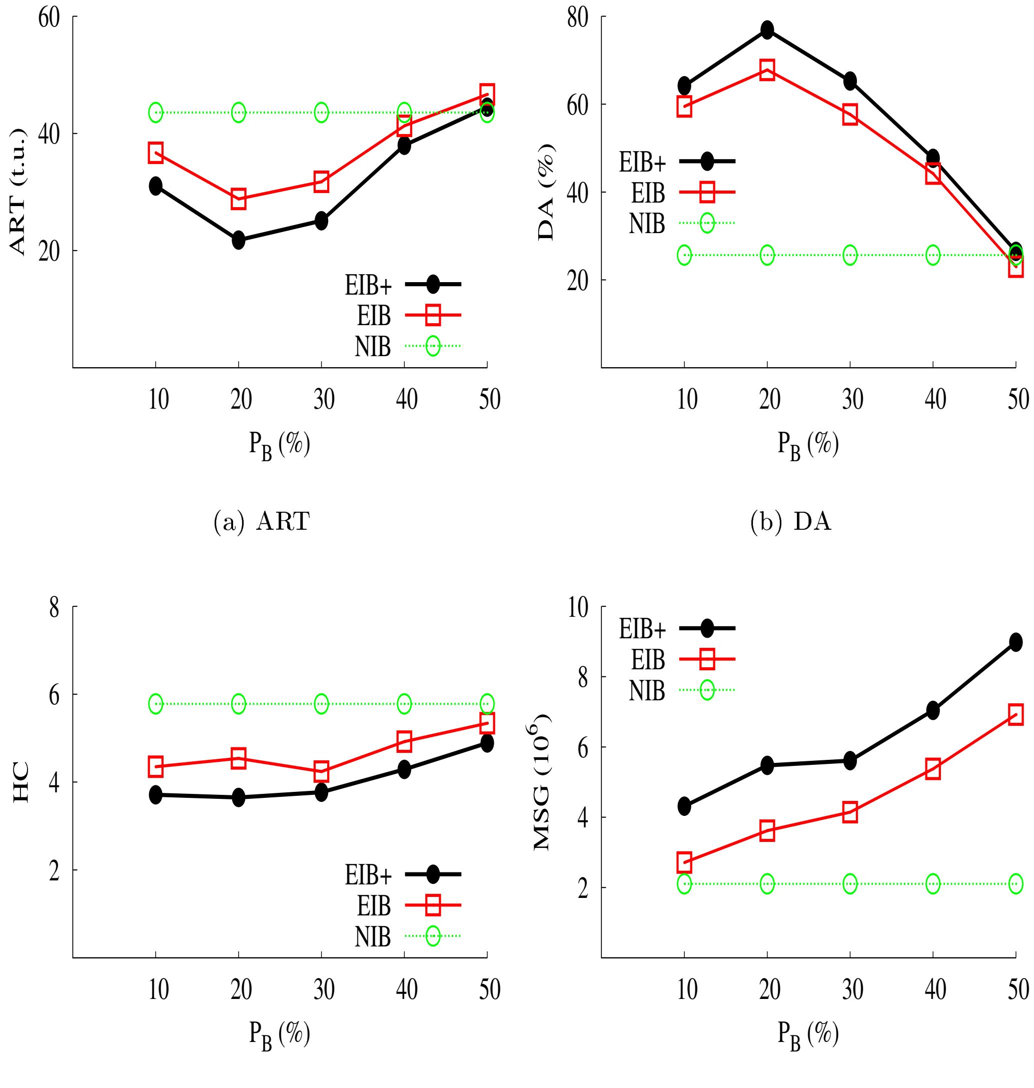
We compare the performance of our proposed broker-based **EIB** and **EIB+** incentive schemes with the non-incentive **E-DCG+** scheme [HM06]. We adapted the **E-DCG+** scheme [HM06] to our scenario. As discussed in Section 2, E-DCG+ is a non-incentive and non-economic replication scheme, and it does not provide incentives for replica hosting. E-DCG+ is executed at every replica allocation period. E-DCG+ is the closest to our scheme since it addresses replication in mobile ad-hoc networks. Furthermore, we believe that E-DCG+ is among the best approaches for meaningful performance comparison with our proposed schemes because it is the most recent approach and it has already been compared to other non-incentive schemes.

As a baseline, we also do performance comparison w.r.t. a non-incentive and non-broker-based **NIB** (Non-Incentive without Brokerage) scheme to show the performance gain due to brokerage. Notably, querying in NIB is simply AODV-based and broker commissions do not arise. Furthermore, in case of NIB, we set the TTL to be 12 i.e., 50% higher than the TTL for our proposed EIB and EIB+ schemes. NIB does not provide any incentive to a peer to forward messages. In NIB, a peer forwards a message in the multi-hop network with a probability of 0.3.

Recall that EIB+ uses three different strategies for assigning broker scores. Here, we present the performance of EIB+ in conjunction with the K-NGS strategy. We have also performed an experiment to indicate the performance of EIB+ with each of the three broker scoring strategies.

4.5.1 Determining the percentage of brokers

We performed an experiment to determine the percentage P_B of brokers in the network. Figure 4.3 depicts the results. As P_B is increased from 10% to 20%, DA improves (albeit at the cost of higher MSG) for both EIB and EIB+ because the involvement of more brokers increases the probability that a given query is processed by at least one of them. However, as P_B is increased beyond 20%, performance keeps degrading due to reduction in the number of data-providers. This is because the sum total of the number of brokers and the number of data-providers is fixed. Notably, EIB+ exhibits better performance than EIB due to the presence of preferred brokers.



(a) ART (b) DA

(c) HC (d) MSG

Figure 4.3: Determining the percentage of brokers

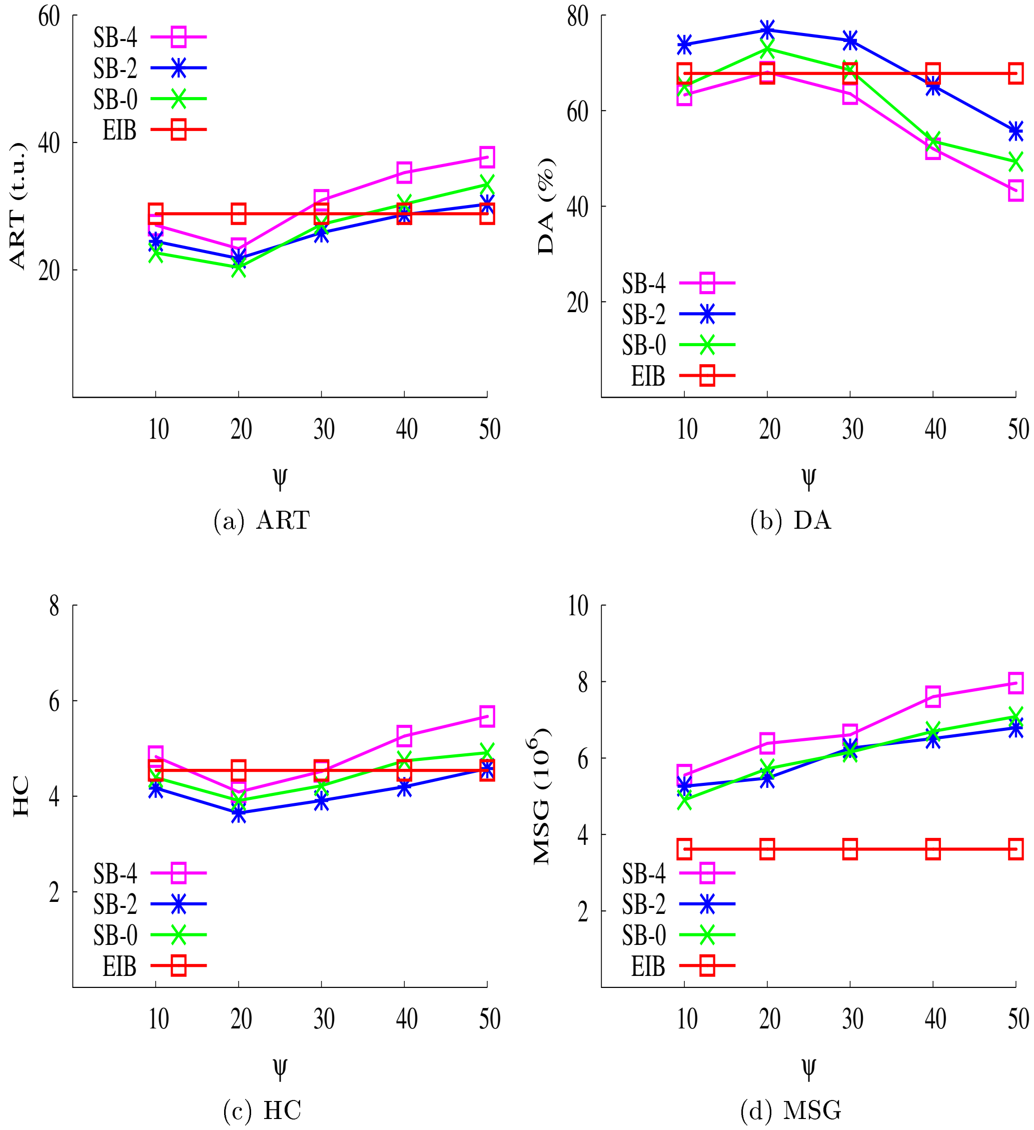
The results in Figure 4.3 suggest that there is a trade-off between the performance (in terms of ART, DA and HC) and the communication traffic. Based on our experimental results, we set the percentage of brokers to 20% so that we can obtain good performance of EIB and EIB+ with reasonable communication traffic. Observe that both EIB and EIB+ perform slightly worse than NIB when $P_B = 50\%$. A closer look at the results in Figure 4.3 suggests that performance gain of EIB over NIB occurs only when P_B is less than 48%. This is because when P_B exceeds 48%, the benefits from brokerage are offset by the additional overhead of interactions among the relatively larger number of brokers. Hence, when P_B exceeds 48%, the peers are better off without a broker-based architecture i.e., they can directly obtain the data from the data-providers.

4.5.2 Determining the percentage of preferred brokers in EIB+

We conducted an experiment to determine the percentage ψ of preferred brokers. Here, $\psi = ((N_{Pref}/N_{Total}) * 100)$, where N_{Pref} is the number of preferred brokers, while N_{Total} is the total number of brokers. For example, if $N_{Total} = 20$ and $\psi = 20\%$, the number of preferred and common brokers would be 4 and 16 respectively. For this experiment, we also varied the number SB of sub-brokers corresponding to each preferred broker. Figure 4.4 depicts the results. We use the notations SB-0, SB-2 and SB-4 to represent the scenarios for EIB+ corresponding to 0, 2 and 4 sub-brokers respectively per preferred broker.

The results in Figure 4.4 indicate that as ψ is increased from 10% to 20%, the performance of EIB+ improves slightly in the cases of SB-0, SB-2 and SB-4 due to the incentivizing effect of preferred brokerage becoming more pronounced. However, as ψ is increased to 30% and beyond, the performance of EIB+ degrades. This occurs because at higher values of ψ, more brokers are allowed to become preferred brokers, thereby implicitly reducing the level of service required to become a preferred broker. This reduces the incentive for preferred brokerage.

EIB+ performs better in the case of SB-2 (albeit at the cost of higher MSG) as compared to that of SB-0 due to load-sharing among the preferred brokers and their respective sub-brokers. However, in case of SB-4, EIB+ performs



(a) ART (b) DA

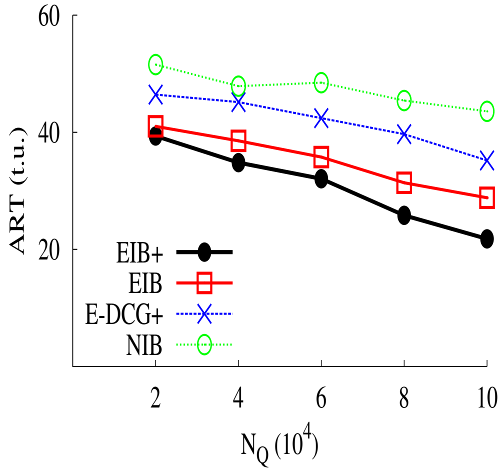
(c) HC (d) MSG

Figure 4.4: Determining the percentage of preferred brokers in EIB+

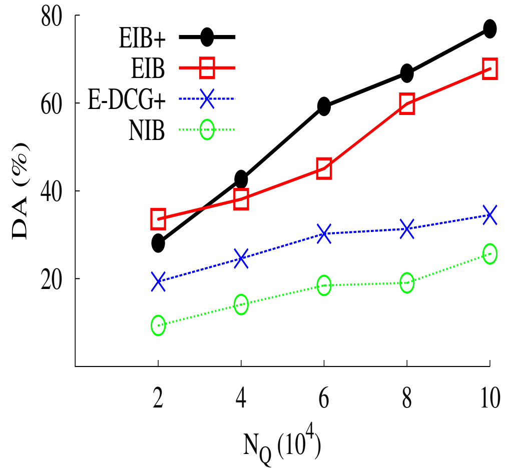
worse than for SB-2 because the relatively high overhead of data allocation among a larger number of sub-brokers reduces the performance. The results in Figure 4.4 suggest that EIB+ performs best at reasonable communication overhead when $\psi = 20\%$ (in case of SB-2). Thus, we experimentally determine ψ to be 20% and SB to be two.

4.5.3 Performance of EIB and EIB+

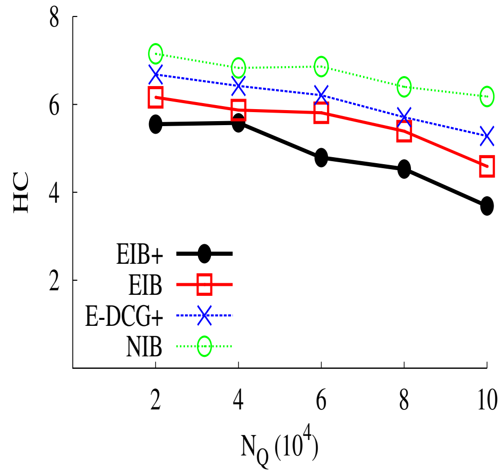
Figure 4.5 depicts the results using the default values of the parameters in Table 4.1. The results in Figure 4.5a indicate that after the first 20000 queries



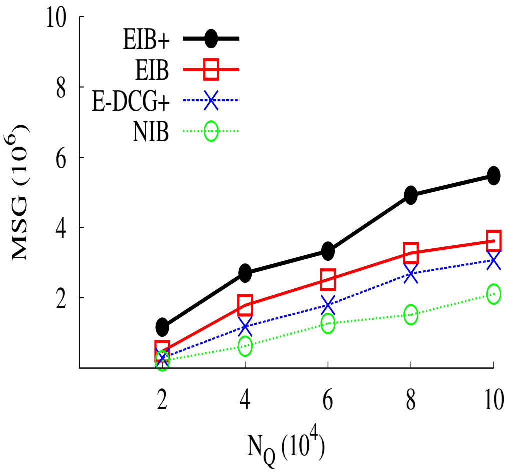
(a) ART



(b) DA



(c) HC



(d) MSG

Figure 4.5: Performance of EIB & EIB+

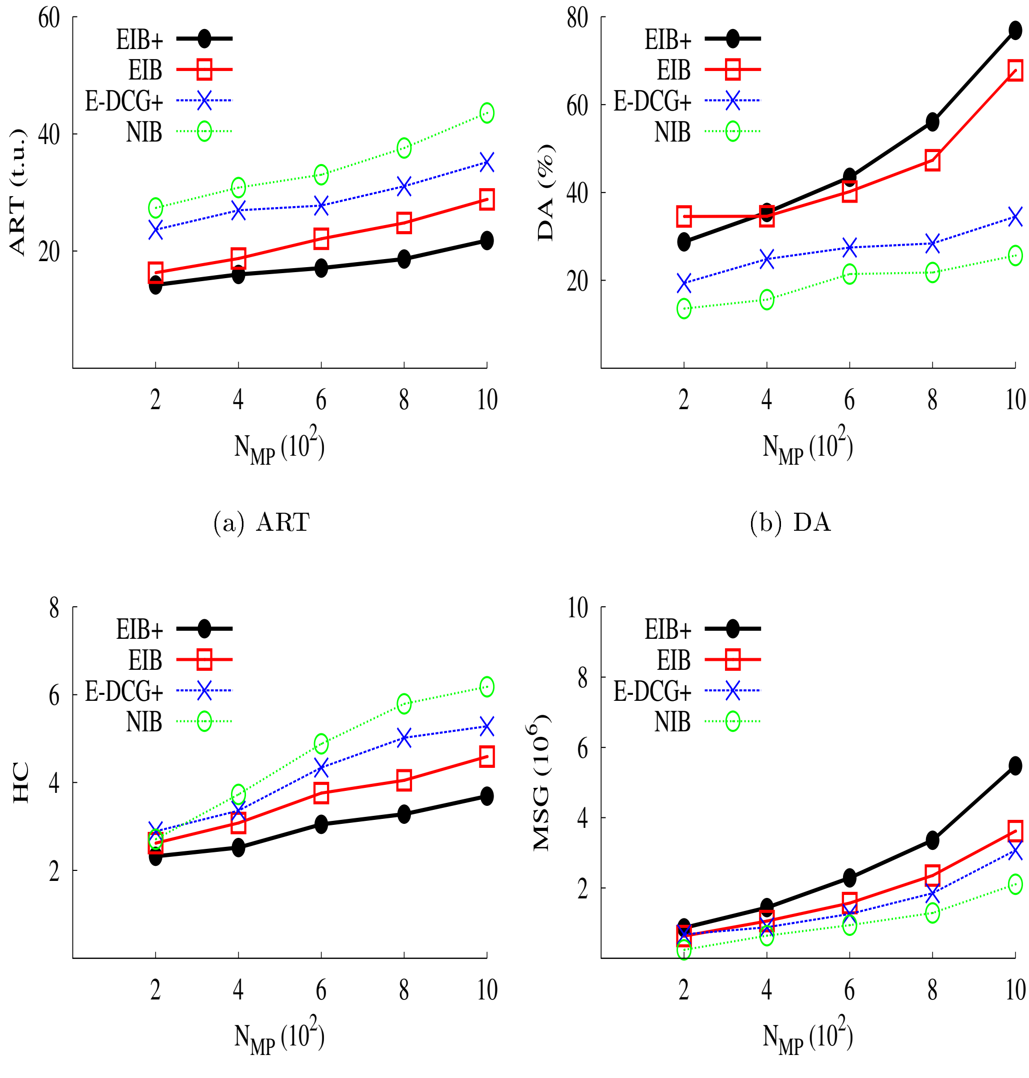
have been processed, EIB, EIB+ and E-DCG+ exhibit comparable performance because the effect of replication is not pronounced at the initial stages. However, over time as more queries are processed, performance improves in terms of ART, DA and HC for all the schemes essentially due to the effect of replication becoming more prominent. Both ART and HC eventually plateau due to reasons such as network partitioning, competition among replicas for memory space and unavailability of some of the MPs.

The results in Figure 4.5d indicate that EIB and EIB+ incur higher MSG than E-DCG+ and NIB primarily due to the additional communication overhead introduced by brokers (and sub-brokers in case of EIB+). However, we believe that the additional number of messages incurred by EIB and EIB+ is a small price to pay for the performance benefits of these schemes. EIB+ incurs higher MSG than EIB because it incorporates gossiping among neighbouring MPs for computing broker scores. E-DCG+ incurs higher MSG than NIB because in E-DCG+, every MP needs to periodically send messages to other MPs to convey replication-related information.

EIB+ outperforms EIB because it provides additional incentives to brokers for performing value-added routing by incorporating the notion of preferred brokers. Moreover, EIB+ also performs effective load-sharing between preferred brokers and sub-brokers, thereby reducing query waiting times in the job queues of the brokers. EIB performs better than E-DCG+ due to its economic incentives, which encourage MP participation. Increased MP participation implies more opportunities for replication, more memory space for hosting replicas and multiple paths for locating a data item/replica. Furthermore, unlike E-DCG+, EIB maintains indexes at the brokers (which facilitate value-added routing) and it replicates 'hot' data items at the brokers. E-DCG+ exhibits better performance than NIB because of its superior replication mechanism.

4.5.4 Effect of variations in the number of MPs

We varied the total number N_{MP} of MPs, keeping the number of queries proportional to N_{MP}. Figure 4.6 depicts the results. As N_{MP} increases, ART and MSG increase for all the schemes due to increase in network size. However, the rate of increase in ART is lower for EIB and EIB+ than for E-DCG+ and NIB due to their better incentivization of replication by means of economic incentives and brokerage. As N_{MP} increases, DA increases for all the schemes due to increased opportunities for replication. HC follows a pattern similar to that of ART, the slight deviations occurring due to bandwidth differences. Observe that when $N_{MP} = 20$, EIB+ exhibits slightly worse DA than that of EIB because the benefits provided by preferred brokers are not realized due to the existence of fewer preferred brokers.



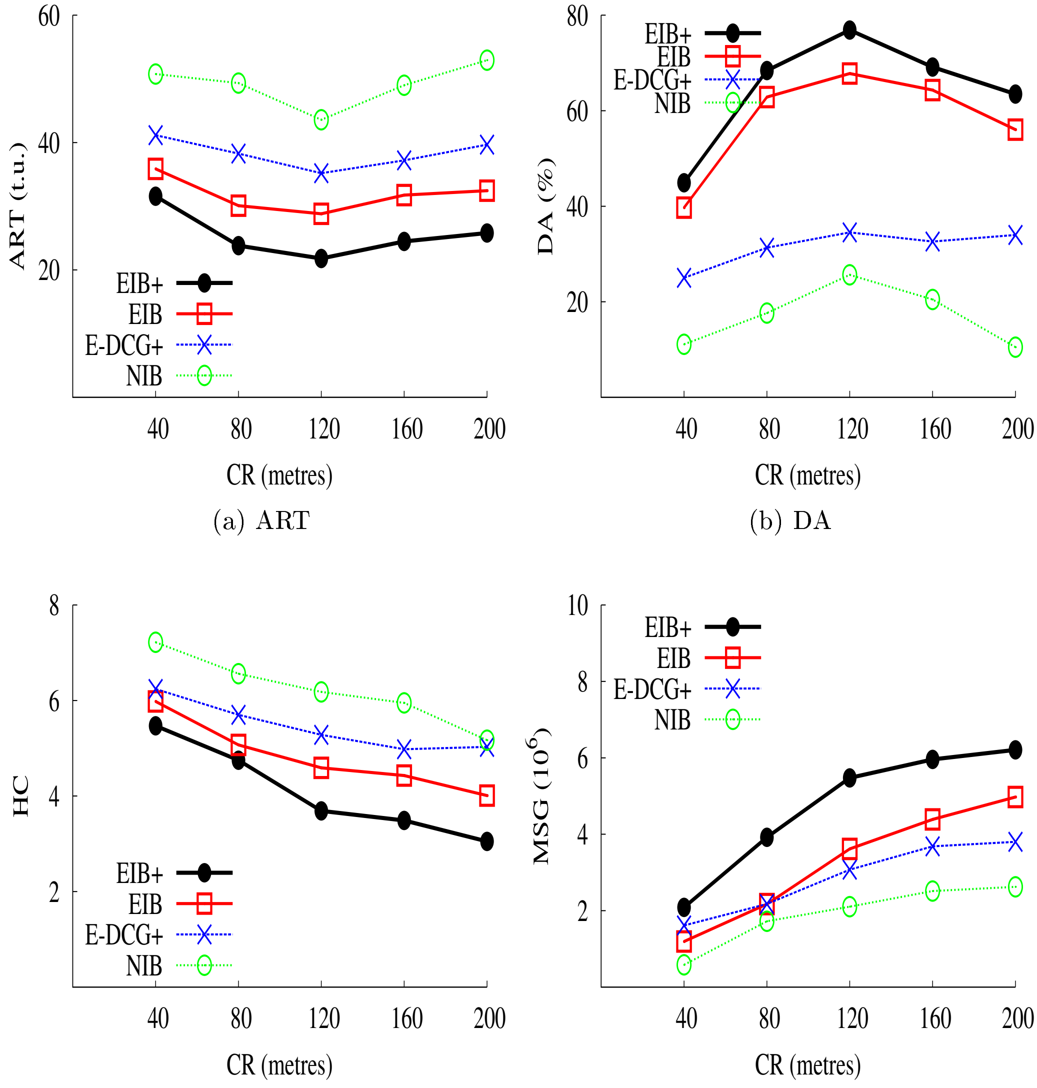
(a) ART (b) DA

(c) HC (d) MSG

Figure 4.6: Effect of variations in the number of MPs

4.5.5 Effect of variations in the communication range

The results in Figure 4.7 depict the effect of variations in the communication range CR of the MPs. Increase in CR has the effect of bringing the MPs



(a) ART (b) DA

(c) HC (d) MSG

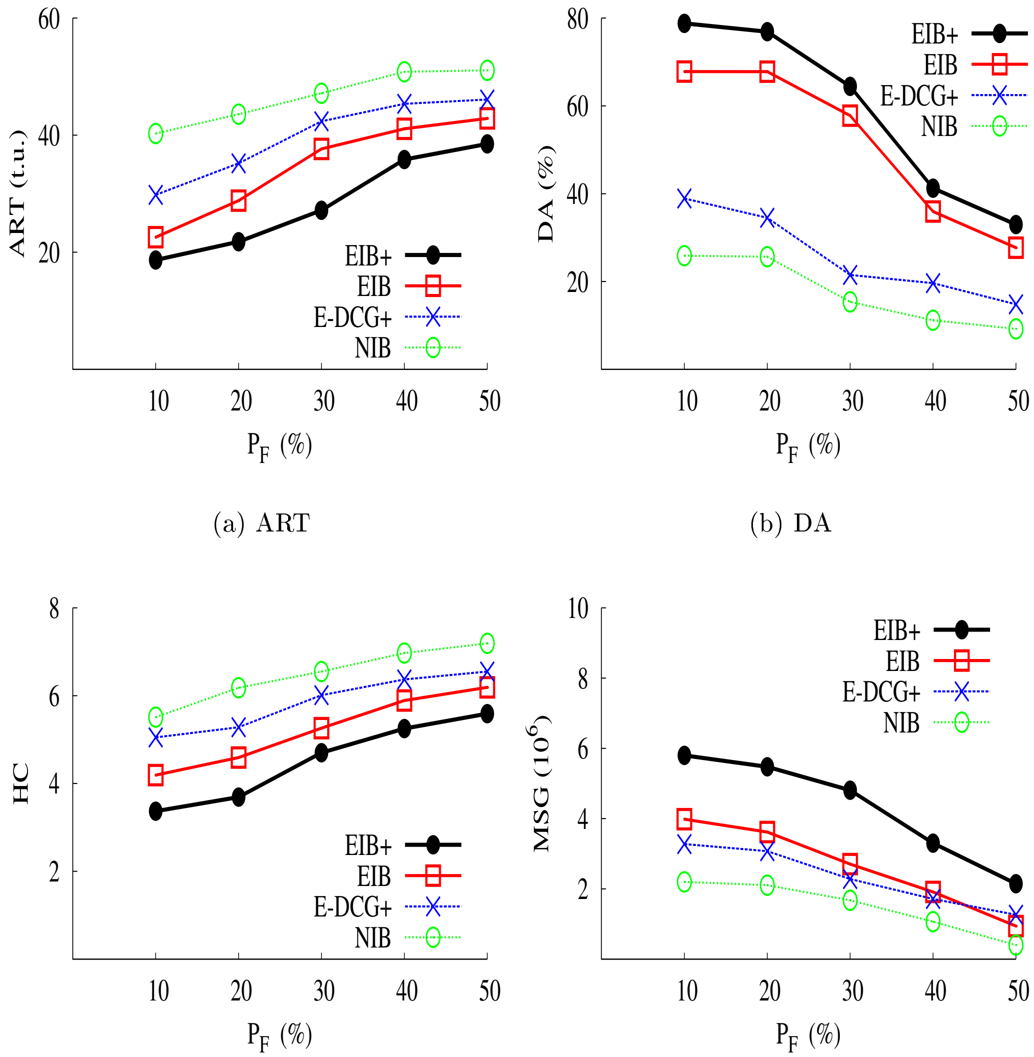
Figure 4.7: Effect of variations in the communication range

'nearer' to each other. Hence, performance improves with increase in CR for all the schemes due to data items becoming 'nearer' and more accessible to query-issuers. However, performance gains occur only until CR=120 metres. Beyond CR=120 metres, ART and DA degrade for all the schemes because the MPs become too 'close' to each other, hence a relatively larger number of MPs and brokers become involved in the processing of any given query. This results in a relatively larger number of queries waiting in the job queues of the data-providers, hence some of the query deadlines are missed. Beyond CR=120 metres, the performance gap between EIB and EIB+ keeps decreasing because the benefits of preferred brokerage become less pronounced when the MPs are already too 'near' to each other. In essence, all the schemes perform best when CR=120 metres.

As CR increases, MSG increases for all the schemes because the increased reachability causes more MPs to become involved in the processing of a given query. On the other hand, with increase in CR, a lower number of messages are required to reach a given MP. These two opposing effects somewhat offset each other at higher values of CR, thereby explaining the reason why MSG eventually plateaus for all the schemes.

4.5.6 Effect of MP failures

MPs can fail due to reasons such as depletion of their limited energy resources. Figure 4.8 depicts the results of the effect of MP failures. As the percentage P_F of MP failures increases, MP participation decreases, query paths become longer and fewer data-hosting MPs remain available, thereby degrading the performance of all the schemes. Interestingly, at $P_F = 50\%$, all the schemes exhibit comparable ART due to limited MP participation making the effect of economic incentives and brokerage less pronounced. As the results in Figure 4.8d indicate, MSG decreases with increase in P_F for all the schemes due to reduced communication overhead among a lower number of available MPs. Interestingly, at $P_F = 50\%$, EIB incurs lower MSG than E-DCG+ due to scarcity of brokers when the total number of available MPs become relatively low. However, EIB+ still incurs higher MSG than E-DCG+ due to gossiping-related communication overheads.



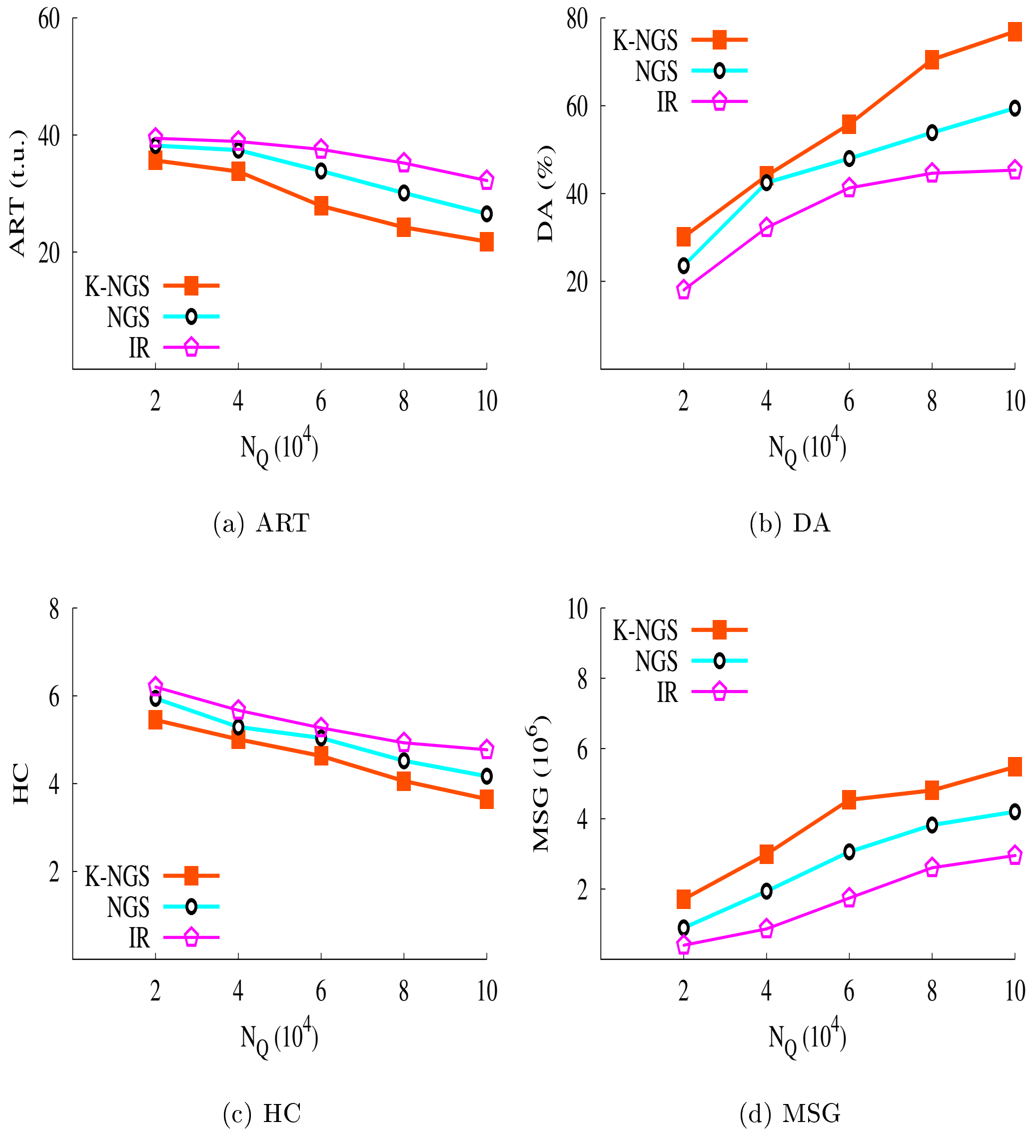
(a) ART (b) DA

(c) HC (d) MSG

Figure 4.8: Effect of MP failures

4.5.7 Effect of different strategies for assigning performance-based scores to brokers in EIB+

We conducted an experiment to investigate the relative performance of EIB+ with the different strategies, namely IR, NGS and K-NGS, for assigning performance-based scores to brokers. Figure 4.9 depicts the results. K-NGS outperforms NGS because its gossiping among k-hop neighbours better incentivizes preferred brokerage by incorporating broker scores from a larger number of MPs albeit at the cost of higher MSG. Similarly, NGS performs better than IR since its gossiping among one-hop neighbours provides better incentives for preferred brokerage than IR. The performance of all the three strategies improve over time as more queries are processed due to the reasons discussed for the results in Figure 4.5.

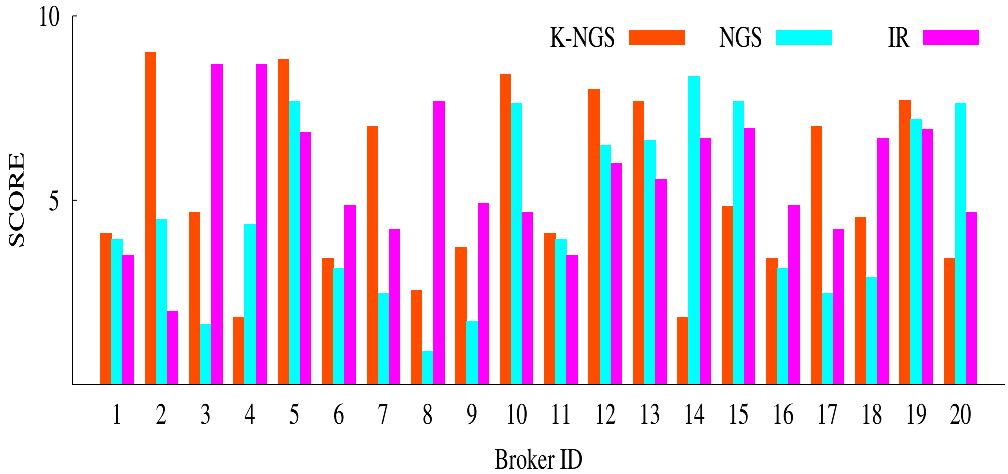


(a) ART (b) DA

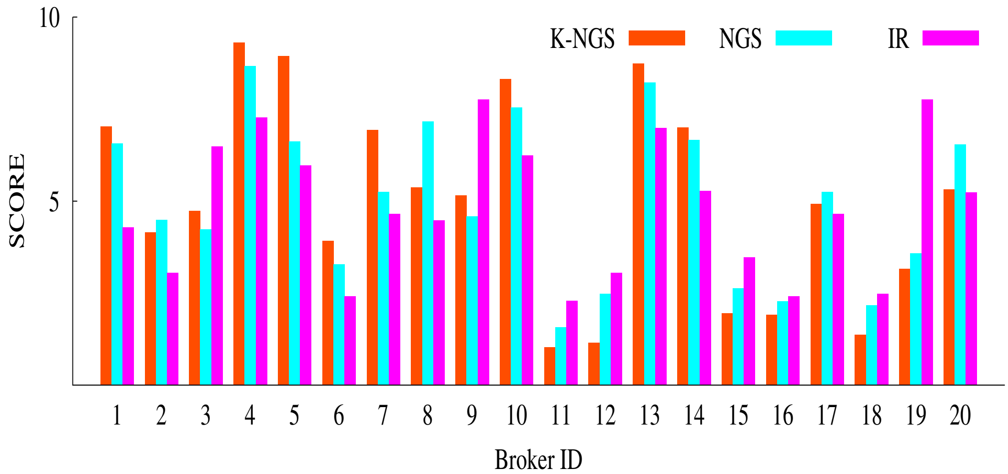
(c) HC (d) MSG

Figure 4.9: Effect of different strategies for assigning performance-based scores to brokers in EIB+

Figure 4.10 depicts the snapshots of broker scores at the time-points of 40000 and 100000 queries respectively under IR, NGS and K-NGS. The X-axis represents the unique identifiers of the brokers, while the Y-axis depicts the score of each broker. Periodically, after every 20000 queries, the scores of brokers are recorded. The scores are on a scale of 1 to 10, where a higher score indicates better performance. Each MP assigns an initial score of 5 to all the brokers at the start of every 20000-query time-period. (This periodic resetting of scores is necessary to reflect current performance of brokers.) Factors such as a broker's location, mobility pattern and current network conditions result in variation of scores across brokers.



(a) Snapshot of the broker scores at the 40000-query time-period



(b) Snapshot of the broker scores at the 100000-query time-period

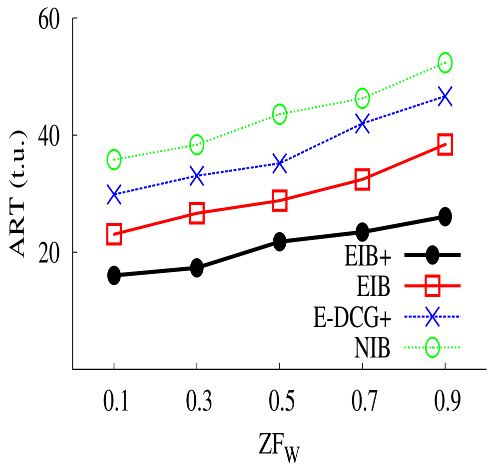
Figure 4.10: Snapshots of broker scores at the time-points of 40000 and 100000

Now let us examine the results in Figure 4.10a. We will denote the broker with ID of i as Bi. Observe that there is no clear pattern regarding any specific scoring strategy assigning higher or lower scores than the others. For example, K-NGS assigned the lowest score to $B4$, but it assigned the highest score to $B2$. Observe that $B8$ is assigned a much higher score by IR than by NGS and K-NGS. Broker scores vary across scoring strategies because they consider varying amounts of interaction with other MPs. These strategies may also assign comparable scores to any given broker e.g., $B1$ and $B11$ in the results in Figure 4.10a. This occurs when a broker's performance remains comparable in providing services to MPs at different locations. A broker's score may fall below 5 (e.g., $B1$ in Figure 4.10a) due to reasons such as connectivity to limited resources in its mobility path and limited energy.

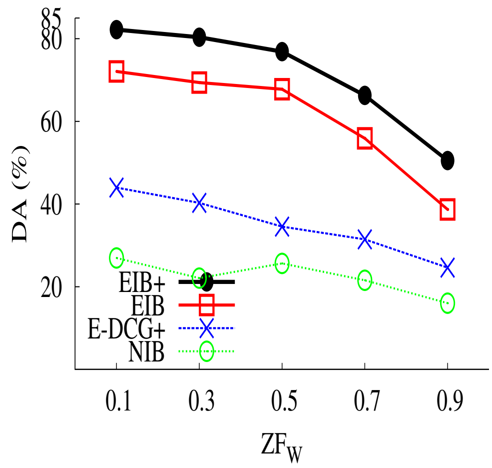
Even though broker scores may vary across scoring strategies, the results in Figures 4.10a serve as a guide for evaluating broker performance, thereby facilitating in distinguishing between common and preferred brokers. For example, in Figure 4.10a, $B5$, $B12$, $B13$ and $B19$ and in Figure 4.10b, $B4$, $B5$, $B10$ and $B13$ would be the preferred brokers, while the other brokers would be common brokers.

4.5.8 Effect of variations in the workload skew

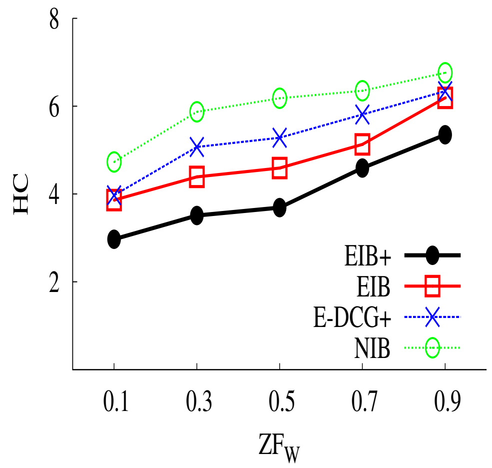
Figure 4.11 depicts the results when the zipf factor ZF_W is varied. Notably, among all the schemes, only EIB+ supports load-sharing, which occurs between preferred brokers and sub-brokers. As ZF_W increases (i.e., increasing skew in the workload), performance degrades for all the schemes. This occurs due to increased waiting times at the job queues of overloaded data-providers, thereby causing some of the queries to miss the deadlines. Observe how EIB+'s load-sharing mechanism facilitates it in outperforming the other schemes. However, the performance gap between EIB+ and the reference schemes decreases with decreasing skew due to the effect of load-sharing becoming less pronounced. As ZF_W increases, the number of query failures increase (due to queries missing their deadlines), thereby reducing MSG. However, for EIB+, MSG increases beyond $ZF_W = 0.5$ due to the interactions between the preferred brokers and their sub-brokers.



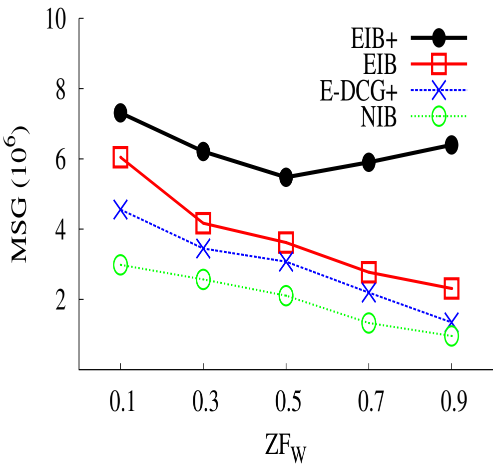
(a) ART



(b) DA



(c) HC



(d) MSG

Figure 4.11: Effect of variations in the workload skew

4.6 Summary

In M-P2P networks, data availability is typically low due to rampant free-riding, frequent network partitioning and mobile resource constraints. We have proposed the E-Broker system for improving data availability in M-P2P networks.

E-Broker incorporates two economic incentive-based brokerage schemes, namely EIB and EIB+. EIB incentivizes relay peers to act as information brokers for performing value-added routing and replication in M-P2P networks, thereby effectively improving data availability. The EIB+ scheme extends the EIB scheme by incorporating three different broker scoring strategies for providing additional incentives to brokers. EIB+ also facilitates load-sharing among the peers.

We have also evaluated the number of brokers, beyond which the peers are better off without a broker-based architecture. Our performance study indicates that the proposed schemes are indeed effective in improving query response times, data availability and query hop-counts at reasonable communication traffic cost in M-P2P networks. In the future, we plan to extend this work by using game-theoretic approaches for data item pricing.

5

E-VeT: Economic Reward/Penalty-based System for Vehicular Traffic Management

5.1 Overview

The proliferation of mobile devices with embedded GPS sensors coupled with the growth in the popularity of infotainment services for vehicles have created new avenues for improving vehicular traffic management in road networks. Thus, schemes for improving transportation system efficiency are becoming increasingly popular [AWX+12, SWYX11].

Given that vehicles generally tend to autonomously select shorter routes with lower traffic congestion, a relatively large number of vehicles often choose the same 'popular' (i.e., relatively shorter and congestion-free) routes, thereby causing congestion. Such traffic congestion typically results in increased vehicular fuel consumption and delayed arrival at destinations. Thus, coordination among the routes allocated to different vehicles becomes a necessity to reduce traffic congestion. However, vehicles trying to coordinate their routes among themselves in a vehicle-to-vehicle (V2V) manner in vehicular ad hoc networks (VANETs) would cause privacy concerns, communication traffic congestion and selfish behaviors.

Incentives have been proposed for stimulating content sharing in mobile-P2P networks [PMG$^+$11, WXS04, XWR06]. However, these works do not address traffic management and vehicular routing issues such as congestion. Moreover, a P2P traffic information system for dynamic route guidance has been discussed in [RSKM09]. However, these works do not incentivize vehicles in following system-assigned traffic routes.

This work proposes the E-VeT system for efficiently managing the vehicular traffic in road networks using economy-based reward/penalty schemes. In this work, the cost of traversing a path in the road network corresponds to the time required for the path traversal, unless otherwise specified. Hence, we shall use the terms "**path cost**" and "**path time-cost**" interchangeably. Observe that defining path cost in terms of time encompasses factors such as path distance, the speed limit relevant to the path and the path's traffic congestion.

In E-VeT, base stations collaboratively facilitate dynamic vehicular route assignments for mitigating the traffic congestion, thus reducing the average time of arrivals and fuel consumption. However, vehicles may not follow the paths assigned by the base stations e.g., when they can find lesser-cost paths. To incentivize vehicles towards following the system-assigned paths, E-VeT uses **rewards/penalties (payoffs)**, which are in terms of *real currency*. Hence, these payoffs can be used towards paying road taxes, car registration, and license/toll fees.

This work assumes that all the vehicles fall under the purview of the E-VeT reward/penalty framework, which could be implemented as part of a government-mandated program for facilitating traffic management. Note that the proposed scheme is a government-mandated system, it is always operational, but only the price changes dynamically based on the congestion. Thus users, will not know the pricing scheme and congestion information well ahead of time. Though the system suggests and offers options to users, they still have a choice of paying the penalty and taking the higher-priced paths; the objective is not to force users for explicit load-balancing. Since we have a reward/penalty system, it is using incentives for load-balancing. It is also different and better than randomization where users can get one of the options, which they have to follow, and they have no choice to alter the option they received. Thus, in our scheme, we preserve the notion that a user is the final entity to decide the path taken.

This work can be seen as a further extension to the initial proposal for routing of the VS-scheme for parking introduced in [AWX+12]. In the VS-scheme, a central authority (CA) makes an optimal assignment, and penalizes vehicles severely for deviating from it. Furthermore, in the VS-scheme, the CA guarantees that each vehicle v will pay a travel-cost to slots that is not higher than v's cost in equilibrium. Since in an optimal assignment some vehicles may travel longer than in equilibrium, the CA compensates them in dollars so that the total cost that v pays is not higher than v's travel-time in equilibrium. The CA also charges vehicles that travel less in the optimum assignment than in the equilibrium assignment. This dollar-charge is equivalent to the saving in travel-time.

Our work here is mainly focused on routing in V2V different from parking of vehicles in [AWX+12] in terms of policies for route allocation of vehicles based on revenues, modeling the pricing problem for revenue generation and finding a suitable reward/penalty scheme that adapts to changing behavior of drivers over period of time. In addition, the performance metrics directly focus on the impact of different revenue allocation schemes on the average fuel saving, average time of arrival and the number of messages exchanged among others.

In summary, our proposed schemes differ from existing proposals [Bra96, Mor10, Xu06, Yan12, Iss11] in mainly two ways. First, we introduce a reward/penalty framework for controlling the traffic congestion. Second, users' good behavior (i.e., following the system advice) is considered in the congestion control decision-making in the sense that the system remembers past behavior and rewards/penalty earned in the past. Thus, our scheme is user-centric and it inspires users to earn rewards so that they can get preferred assignment of paths when needed by redeeming rewards.

The contributions of E-VeT are three-fold:

1. It proposes an $\mathrm{R}^2\mathrm{A}$ (Revenue-based Route Allocation) scheme, which rewards vehicles for following system-assigned longer-time paths, and charges a fee for following system-assigned shorter-time paths. Furthermore, it penalizes (charges much higher fee) vehicles for any deviations from the system-assigned paths.

2. It presents the $\mathrm{R}^2\mathrm{A}^+$ (extended $\mathrm{R}^2\mathrm{A}$) scheme by incorporating the notion of *revenue-scales* for further incentivizing vehicles based on their past system usage.

3. It discusses a route allocation algorithm, which gives lesser-time paths as a preference to vehicles that have earned higher revenue based on the scheme used i.e., either $\mathrm{R}^2\mathrm{A}$ or $\mathrm{R}^2\mathrm{A}^+$.

Note that both $\mathrm{R}^2\mathrm{A}$ and $\mathrm{R}^2\mathrm{A}^+$ schemes are designed to ensure fairness in the sense that vehicles pay when they travel faster, and they earn currency when they travel slower. Both schemes penalize vehicles, which deviate from system-assigned paths, thereby incentivizing them to adhere to the system-assigned paths. Furthermore, when vehicles follow the system-assigned paths, they are rewarded either in terms of time-savings (i.e., lower time-cost routes being allocated) or in terms of real currency (i.e., payments for following longer time-cost routes).

$\mathrm{R}^2\mathrm{A}$ and $\mathrm{R}^2\mathrm{A}^+$ differ in that while $\mathrm{R}^2\mathrm{A}$ assigns payoffs to vehicles based on every individual journey, $\mathrm{R}^2\mathrm{A}^+$ performs the payoff assignment based on the *consistency* of a given vehicle in following the system-assigned paths across *multiple* journeys. To achieve this, R^2A^+ uses a set of pre-defined *revenue-scales* and provides better payoffs to the vehicles that are associated with higher revenue-scales. This entices vehicles to consistently follow the system-assigned routes. Our performance study shows that the proposed schemes are indeed effective in managing vehicular traffic in road networks by reducing the average time of arrival and fuel consumption.

The remainder of this chapter is organized as follows. Section 5.2 presents the architecture of E-VeT, while Section 5.3 discusses the proposed R^2A and R^2A^+ economic reward/penalty-based schemes and the route allocation algorithm. Section 5.4 provides the proof of correctness. Section 5.5 reports the performance study. Finally, we conclude in Section 5.6.

5.2 Architecture of E-VeT

This section discusses the architecture of E-VeT. The architecture of E-VeT consists of the road network, checkpoints, base stations and vehicles. E-VeT envisages the road network as an overlay graph, where each vertex represents a checkpoint, and each edge represents a route connecting these checkpoints. Here, a checkpoint is a landmark such as a major road intersection, a hospital or a well-known tourist spot. Thus, the journey of each vehicle comprises a traversal of a set of such checkpoints, and we designate the route between two checkpoints as a **path**. We define the **source** and **destination** of a given vehicle's journey as the checkpoints that are nearest to the starting point and the desired end-point of the journey respectively. A **base station** is a powerful, reliable and static node. For simplicity, we assume that each checkpoint is associated with a single base station and vice versa.

When a given vehicle V approaches a checkpoint C, it sends information about its destination to the base station B corresponding to C. Upon receiving this information from multiple vehicles in its vicinity, B executes a route allocation algorithm (discussed later in Section 5.3) and assigns a path to each vehicle for travelling to the next checkpoint. We shall henceforth refer to the path assigned to a given vehicle by a base station as the **system-assigned path**. Notably, the route allocation algorithm is executed by the corresponding base station at every checkpoint (that falls along a given vehicle's route) until it is routed to its destination checkpoint.

In E-Vet, base stations assign rewards/penalties (i.e., payoffs) to the vehicles. As we shall see in Section 5.3, E-VeT performs route allocation by providing preference to vehicles, which have earned more payoffs, thereby incentivizing vehicles to follow system-assigned paths. Payoff allocation to the vehicles is performed on a checkpoint-to-checkpoint basis. Suppose vehicle V traverses the path from checkpoint C1 to checkpoint C2. Let us refer to the base stations corresponding to C1 and C2 as $B1$ and $B2$ respectively. Here, $B2$ performs the payoff allocation to V, after communicating with the base station B at the checkpoint that was previously traversed by V. If V had followed the system-assigned path, B=$B1$, otherwise B could be any of the neighboring base stations of $B2$.

Observe how base stations collaborate with each other to facilitate revenue-based dynamic vehicular routing for reducing traffic congestion. Such collaboration becomes a necessity for coordinating smooth traffic flow among vehicles, which do not directly interact with each other to preserve their privacy. Incidentally, traversal of a path between two given checkpoints is associated with a cost, which we shall discuss now.

5.2.1 Computation of path cost

Recall that in E-VeT, the path cost corresponds to the time required for traversing the path. The cost t_j of traversing path j is computed by a given base station in two steps (a) Compute the path cost t_{rec_j} for the current time-period (b) Compute t_j as the exponential moving average of the path costs over the most recent time-periods to account for fluctuations in path usage. t_{rec_j} depends upon factors such as path distance, speed limit of the path and path congestion. Thus, path cost can change temporally depending upon path congestion. Since path congestion is related to path flow, let us first compute the path flow $F_{j,t}$ for path j as follows:

$$F_{j,t} = NL_{j,t}/NE_{j,t} \tag{5.1}$$

where $NE_{j,t}$ and $NL_{j,t}$ are respectively the number of vehicles that entered or left path j during time-period t. We assume that $NE_{j,t}$ and $NL_{j,t}$ are both non-zero i.e., there are always vehicles on the road. Let us henceforth refer to the path flow as *flow*. Consistent with real-world scenarios, we consider that bi-directional flow values may differ e.g., the flow value from a given checkpoint X to a checkpoint Y may differ from that of the flow value from Y to X. However, such differences in flow values do not impact our proposed schemes.

t_{rec_j} is computed as follows:

$$t_{rec_j} = \begin{cases} (\, D_j/S_{max} \,) & \text{if } F_{j,t} = 1 \\ (\, D_j/S_{max} \,) \,/\, F_{j,t} & \text{otherwise} \end{cases} \tag{5.2}$$

where D_j and S_{max} are the distance and speed limit of path j respectively. Observe that the term (D_j/S_{max}) in Equation 5.2 concerns the congestion-free path cost (i.e., $F_{j.t}$=1). Moreover, t_{rec_j} increases with decrease in $F_{j,t}$ because more congested paths typically entail higher path costs.

Using the value of t_{rec_j}, the computation of t_j according to the exponential moving average (EMA) formula follows:

$$t_j = (\,(t_{rec_j} - EMA_{prev}) \times 2/(T+1)\,) + EMA_{prev} \tag{5.3}$$

where EMA_{prev} is the EMA that was computed for the previous time-period and T is the number of time-periods considered in the moving average computation. Results of our preliminary experiments suggest that T=5 is suitable for traffic management application scenarios. Notably, EMA gives higher weights to recent time-periods, hence it is appropriate for dynamically changing path costs that may occur in traffic management application scenarios.

5.2.2 Illustrative example for road network topology in E-VeT

Figure 5.1 depicts an illustrative example of the road network topology in E-VeT at a certain point in time. In Figure 5.1, the checkpoints C1 to C6 are connected by weighted paths P1 to P8, whose respective path costs are shown in parentheses. Here, path costs are indicated in case of congestion-free

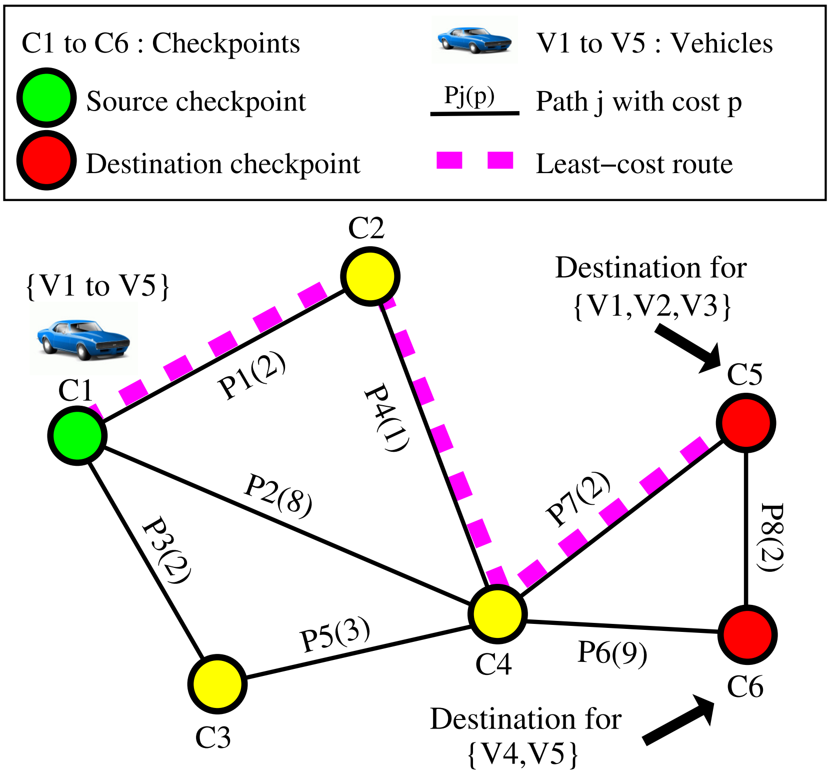


Figure 5.1: Example for E-VeT road network topology

paths. Observe that path costs are essentially *dynamic* in that they increase as path congestion increases. For simplicity, all the vehicles V1 to V5 are associated with the same source checkpoint C1. The destination checkpoints for {V1, V2, V3} and {V4, V5} are C5 and C6 respectively. For simplicity, this example assumes that bi-directional path costs are equal.

Assume that each vehicle prefers to take the least-cost path to its respective destination. Observe that all the vehicles V1 to V5 would have to traverse C4 on their way to their respective destinations. From C1 to C4, three paths are possible, namely {P1, P4}, {P2} and {P3, P5} with path costs of 3 (i.e., 2+1), 8 and 5 (i.e., 2+3) respectively. Thus, all the vehicles would want to take the least-cost path {P1, P4} to C4. However, all five vehicles taking path P1 would increase traffic congestion there, thus effectively increasing the path cost.

For reducing path congestion, coordination among vehicular routes becomes a necessity for ensuring smooth traffic flow. As we shall see in the next section, such coordination in E-VeT is performed by the base stations, which incentivize vehicles towards following system-assigned paths.

5.3 Revenue-based route allocation in E-VeT

This section discusses our proposed R^2A and R^2A^+ economic reward/penalty-based schemes. Based on the scheme used (i.e., either R^2A or R^2A^+), E-VeT assigns payoffs to the vehicles. These payoffs are used as inputs for E-VeT's incentive-based route allocation algorithm, which is also presented in this section.

5.3.1 The R^2A scheme

Recall that we define path costs in terms of time. For computing rewards and penalties in R^2A, we define the notions of *lower-cost paths* and *higher-cost paths* as follows. Consider the existence of multiple possible paths between two given checkpoints. Each path, whose cost is below the median path-cost of all these paths, is defined to be in the set of *lower-cost paths*. Conversely, each path, whose cost equals or exceeds the median path-cost of all the paths between the two given checkpoints, is defined to be in the set of *higher-cost paths*. R^2A rewards vehicles, which follow system-assigned higher-cost paths, and charges a fee for following system-assigned lower-cost paths. Furthermore, it penalizes vehicles for any deviations from the system-assigned paths, thereby incentivizing them to adhere to the system-assigned paths. Thus, vehicles pay (in the form of fees) when they travel faster, while they get paid (in terms of rewards) when they travel slower, thereby achieving fairness.

Given that a path is the route between two given checkpoints, a given vehicle V has to traverse *multiple* paths during its journey from its source to its destination. Thus, its *revenue* from a given journey equals the difference between the rewards and the fees/penalties over all these paths. During a given journey, suppose V follows (a) r_1 system-assigned higher-cost paths (b) r_2 system-assigned lower-cost paths and (c) r_3 paths that are *not* assigned by the system. Then V's revenue REV from the given journey is computed as follows:

$$REV = \sum_{l=1}^{r_1} REW_l - \left(\sum_{m=1}^{r_2} Fee_m + \sum_{n=1}^{r_3} LY_n \right) \tag{5.4}$$

where REW_l is the reward for the l^{th} system-assigned higher-cost path followed by V. Fee_m is the fee that is charged for the m^{th} system-assigned lower-cost path followed by V, while LY_n is the penalty for the n^{th} non-system-assigned path followed by V. Thus, given that the revenue of V from its p^{th} journey is REV_p, its total revenue equals $\sum_{p=1}^{P} REV_p$, where P is the total number of journeys performed by V.

Now let us see how the rewards and penalties are computed for V for a given path. The reward depends upon the cost difference between the system-assigned path and the corresponding median-cost path. Thus, the base station computes the reward REW_{R^2A} earned by V for a given path as follows:

$$REW_{R^2A} \;\; = \;\; (\, t_j - t_{median} \,) \times \lambda \tag{5.5}$$

where t_j is the path cost of the path j that the system assigned to V, and t_{median} is the median-cost path between the two checkpoints associated with path j. λ is a parameter for converting time-cost to dollar-cost. Notably, the value of λ is a constant, which is fixed by the system. Observe that REW_{R^2A} increases with increase in the difference between t_j and t_{median} because higher rewards should be provided to vehicles for incentivizing them to follow relatively higher-cost system-assigned paths.

In Equation 5.5, both t_j and t_{median} are computed using Equations 5.2 and 5.3. Thus, both t_j and t_{median} are system-estimated time-costs based on current conditions of congestion as opposed to being actual times. Observe that if actual times had been used, vehicle users would have an incentive to spend significantly large amounts of time on the system-assigned path for obtaining increased amount of rewards. Furthermore, observe that if there is only one path between two checkpoints, the reward would be zero because both the median-cost path and the system-assigned path would be the same in this case.

To better understand the computations of t_j and t_{median}, let us refer to Figure 5.1. For simplicity, assume that in Figure 5.1, all paths are congestion-free. Suppose a vehicle needs to find a path j between the two checkpoints C1 and C4. Here, the median-cost path is {P3, P5} with path cost of 5, hence t_{median}=5. If the system assigned the path P2 to V, t_j=8 because the path cost of P2 equals 8. On the other hand, if the system had assigned the path {P1, P4} to V, t_j would have been 3.

The fee Fee_{R^2A} is charged to V for following a given system-assigned lower-cost path j. It depends upon the difference between the cost of taking the median path and the cost of taking path j. Notably, we consider the cost of the median path to effectively handle scenarios involving outliers. Thus, given that t_j is the cost of the system-assigned lower-cost path j and t_{median} is the cost of taking the median path between the two checkpoints corresponding to j, Fee_{R^2A} is computed as follows:

$$Fee_{R^2A} = (t_{median} - t_j) \times \lambda \tag{5.6}$$

Both t_j and t_{median} are computed using Equations 5.2 and 5.3. Referring to Figure 5.1, consider the paths between C1 and C4. Suppose the system has assigned the path {P1, P4} to V, hence t_j=3. Here, t_{median}=5 because it corresponds to the cost of taking the path with the median cost {P3, P5} between C1 and C4. Notably, the significance of λ in Equation 5.6 is essentially the same as in Equation 5.5.

A penalty is incurred by V when it deviates from the system-assigned path. E-Vet assigns different values of LY depending upon whether the opted path is in the set of lower-cost paths or in the set of higher-cost paths. When the opted path is in the set of higher-cost paths, the system shall levy no penalty, hence LY=0. Conversely, when the opted path is in the set of lower-cost paths, the system shall levy a penalty proportional to the difference between t_{median} and $t_{j'}$, where t_{median} is the median-cost path and $t_{j'}$ is the cost of the opted path, as shown in Equation 5.4.

$$LY_{R^2A} = \begin{cases} (t_{median} - t_{j'}) \times \lambda & \text{if } t_{j'} < t_{median} \\ 0 & \text{otherwise} \end{cases} \tag{5.7}$$

where t_{median} is the median-cost path, and $t_{j'}$ is the path cost of the path j' taken by the user. Here, λ is the dollar-cost for penalty, and it is system-defined.

Observe that R^{2}A assigns payoffs without taking into account the past system usage of a given vehicle in following the system-assigned paths across its multiple journeys. Thus, it suffers from the drawback of not being capable of incentivizing *consistent* behavior by the vehicles in adhering to the system-assigned paths.

5.3.2 Illustrative example for R^{2}A

Figure 5.2 depicts an illustrative example for the computation of rewards and penalties in R^{2}A for a given vehicle V. The values of the reward REW_{R^2A}

C_{next}	t_j
C1	1.15
C2	1.07
C3	1.03
C4	0.90
C5	0.34
C6	0.17

λ	10.00
t_{median}	0.97

(a) Path costs to next checkpoints

REW_{R^2A}	Fee_{R^2A}	LY_{R^2A}	REV_{R^2A}
0	0	0.00	0.00
1.00	0	0.00	1.00
0	0	0.00	0.00
0	0	0.70	-0.70
0	0	6.30	-6.30
0	0	8.00	-8.00

(b) System-assigned path C2

REW_{R^2A}	Fee_{R^2A}	LY_{R^2A}	REV_{R^2A}
0	0	0.00	0.00
0	0	0.00	0.00
0	0	0.00	0.00
0	0.70	0.00	-0.70
0	0	6.30	-6.30
0	0	8.00	-8.00

(c) System-assigned path C4

Figure 5.2: Illustrative example for the computation of rewards/penalties in R^{2}A

and the penalties Fee_{R^2A} and LY_{R^2A} are computed using Equations 5.5, 5.6 and 5.7. Suppose C_{curr} is the checkpoint at which V is currently located, while C1 to C6 are the possible next-checkpoints for its journey towards its destination, as indicated in the column C_{next} of Figure 5.2a. The second column t_j in the same figure refers to the cost of traversing the path from C_{curr} to C_{next}.

Observe that C6 is the checkpoint that is associated with the minimum-cost path from C_{curr} to any of the next-check-points. Figure 5.2a also indicates the values of λ and t_{median} that will be used for computing the rewards and penalties. For simplicity, in this example, we show the computation of LY_{R^2A} (see Equation 5.7) using only the current time-period instead of averaging the values over the past T time-periods.

Figures 5.2b and 5.2c depict the rewards and penalties when the system-assigned next-checkpoints for V are C2 and C4 respectively. In Figure 5.2, observe that V earns rewards only when it follows the system-assigned path to C2, which is in the set of higher-cost paths, hence its revenue REV_{R^2A} is positive only in this case. The reward for V in this case is computed using Equation 5.5. Thus, REW_{R^2A}=((1.07 − 0.97) × 10) i.e., 1.00. Furthermore, observe that in Figure 5.2, the value of Fee_{R^2A} is 0 for all the cases because the system-assigned path to C2 is in the set of higher-cost paths, thereby making the penalty Fee_{R^2A} inapplicable.

Observe that here the applicable penalty for deviating from the system-assigned path and opting for a lower-cost path are computed using Equation 5.7. For example, the value of LY_{R^2A} when the next-checkpoint is C1 is zero because C1 is associated with a higher-cost path, whereas the value of LY_{R^2A} is -6.30 in the case of C5, where C5 is associated with the lower-cost path. The values of REV_{R^2A} are computed using Equation 5.4, which are also indicated in Figure 5.2.

5.3.3 The R^{2}A$^+$ scheme

The R^{2}A$^+$ scheme extends the R^{2}A scheme by incorporating the notion of **revenue-scales** for taking into account a given vehicle's consistency in adhering to the system-assigned paths across multiple journeys. R^{2}A$^+$ defines M revenue-scales, each of which is associated with a range of revenues. Then it associates a given vehicle with a revenue-scale based on the vehicle's revenue. Suppose M=4, where revenue-scales {1, 2, 3, 4} correspond to revenue ranges {0-1000, 1001-2000, 2001-3000, 3001-4000} respectively. Here, the vehicle with revenue of 2500 units is associated with revenue-scale 3.

R^{2}A$^+$ uses these revenue-scales for distributing the payoffs. Vehicles, which are associated with higher revenue-scales, earn better payoffs. This provides an additional incentive to the vehicles to consistently follow the system-assigned paths so that they can earn adequate currency to qualify for higher revenue-scales, at which their payoffs would improve. Given M revenue-scales with a given vehicle being associated with revenue-scale m, R^{2}A$^+$ computes the payoffs as follows:

$$\begin{aligned} REW_{R^2A^+} &= (m/M) \times REW_{R^2A} \\ Fee_{R^2A^+} &= (m/M) \times REW_{R^2A} \\ LY_{R^2A^+} &= (m/M) \times LY_{R^2A} \end{aligned} \tag{5.8}$$

where REW_{R^2A}, Fee_{R^2A} and LY_{R^2A} are computed using Equations 5.5, 5.6 and 5.7 respectively. Similar to R^{2}A, R^{2}A$^+$ computes a given vehicle's revenue using Equation 5.4.

5.3.4 Route allocation algorithm

A given base station performs the route allocation to all the vehicles that are moving towards its corresponding checkpoint. When a vehicle approaches a checkpoint, it communicates to the corresponding base station the following information: its destination, its previous checkpoint, the checkpoint assigned to it at the previous checkpoint and its revenue. Notably, this communication is done by the tamper-resistant software module in the vehicle to the base station, thereby ensuring that a vehicle cannot provide false information to the base station concerning its system-assigned checkpoint.

Upon receiving the information from all the vehicles during a system-defined time-period, the base station computes the payoffs of the vehicles based on either R^{2}A or R^{2}A$^+$. Moreover, it uses the route allocation algorithm to assign paths to the vehicles. Then it performs the following actions for each vehicle: (a) updates their revenues based on whether they followed the system-assigned paths, (b) sets its value as the last visited base station, and (c) sets the cost of the assigned path in the vehicle's system.

Notably, the route allocation algorithm uses a *path flow threshold*, which we designate as PF_{th}. The implication of PF_{th} is that the route allocation algorithm would assign vehicles to a given path only upto the value of PF_{th}, thereby not allowing a path to go beyond a given level of traffic congestion. Once the threshold PF_{th} is reached for a given path (when path flow keeps decreasing due to traffic congestion), the path is considered to be full, hence the algorithm would not assign any more vehicles to that path for that time interval.

The value of the PF_{th} threshold can be decided based on the acceptable time of travel between two checkpoints C1 and C2 during different time intervals. Recall that flow in a given path P is the ratio of the number of vehicles exiting P to the number of vehicles entering P during a given time interval. Thus, for example, the flow threshold can be set to 0.6 and 0.8 during peak hours and non-peak hours respectively for maintaining smooth traffic flow. The acceptable threshold can be determined based on the distance between the two given checkpoints and the average possible speed (i.e., (maximum speed limit + minimum speed limit) / 2), which can provide the ideal time of travel between the two checkpoints. This can then be calculated for peak hours and off-peak hours based on the threshold selected and the time determined should be within the acceptable limit set by the travel authority for smooth flow of traffic at different times.

Algorithm 5.1 discusses how a base station identifies the top-k preferred checkpoints for a given vehicle, given the overlay graph G(V, E) of the road network, and the respective destination of each vehicle as input. First, it determines the least-cost k paths to the destination of the vehicle based on the path cost using the approach in [Epp98]. Then, for each of these least-cost k paths, it identifies the next checkpoint that the vehicle needs to traverse for following that path to its destination, and stores these checkpoints in a list PL. Now, for each checkpoint in PL, it computes the path cost. Finally, it sorts the checkpoints in PL in ascending order of the path costs. Thus, the sorted PL list essentially reflects the preference of the vehicle towards its route assignment in terms of path cost minimization.

Algorithm 5.2 presents the economy-based route allocation algorithm in E-VeT using the preference list for each vehicle, as generated by the Algorithm 5.1. Incidentally, vehicles with relatively higher revenues are likely to be those that have either adhered more frequently to their system-assigned

Algorithm 5.1 Greedy algorithm for identifying the preference list of next-checkpoints for a given vehicle

begin

Input: (a) G(V,E): Overlay graph of the road network
(b) *dest*: Destination of the vehicle
Output: PL: Sorted list of top-k preferred next-checkpoints for the vehicle

(1) Determine least-cost k paths to *dest* based on path cost
(2) for each least-cost k path
(3) Identify corresponding next-checkpoint and add it to list PL

(4) for each checkpoint in PL
(5) Compute the path cost

(6) Sort the checkpoints in PL in ascending order of the path costs

end

Algorithm 5.2 Algorithm for route allocation in E-VeT

begin

Input: (a) Destination, previous checkpoint and the checkpoint assigned by the previous checkpoint's base station for each vehicle i
(b) Preference list PL_i for each vehicle i
(c) Revenue of each vehicle
Output: Assignment of next-checkpoint to each vehicle

(1) Sort the vehicles in descending order of revenue into a list L_V
(2) for each vehicle i in L_V
(3) for each checkpoint C_j in PL_i
/* The path from the base station to checkpoint j is designated as path j
(4) if (path flow$>PF_{th}$)
(5) Assign vehicle i to path j
(6) Recompute the path flow
(7) **break**

end

paths or taken longer-time paths more often. Thus, observe how the route allocation Algorithm 5.2 incentivizes vehicles with higher revenues by providing them with preference in route allocation. Furthermore, from Line 4 of Algorithm 5.2, notice how the algorithm assigns vehicles to a given path depending upon the PF_{th} threshold criterion.

5.4 Proof of correctness of E-VeT

The proof aims to show that the mechanism is designed in a way that the dominant strategy for all the vehicles is to follow the system assigned path.

Definition. For a given vehicle V, we define two possible decisions, namely (a) follow the system-assigned path and (b) not follow the system-assigned path. We denote the former and latter as f and $\bar{f}$ respectively. Let us denote the corresponding payoffs for the former and the latter cases as P_f and $P_{\bar{f}}$ respectively.

To prove the correctness of the mechanism, we shall prove that $P_f > P_{\bar{f}}$ i.e., $(P_f - P_{\bar{f}} > 0)$ for all possible cases. (We shall also derive the minimum value of penalty that should be levied to hold the above condition.)

Proof: The following two cases arise:

Case 1: *The system-assigned path j is among the higher-cost paths.* $(t_j > t_{median})$

In this case, observe that the fee Fee_{R^2A} is not applicable. Hence, $P_f = REW_{R^2A} - \lambda t_j$, where REW_{R^2A} is computed using Equation 5.5 and t_j represents the path cost as computed using Equations 5.2 and 5.3. Moreover, when V follows a non-system-assigned path j', $P_{\bar{f}} = -LY_{R^2A} - \lambda t_{j'}$, where LY_{R^2A} is computed using Equation 5.7, and $t_{j'}$ is the cost of path j', which is computed using Equations 5.2 and 5.3.

Now, let us find the minimal value LY_{R^2A} of penalty for which $P_f \geqslant P_{\bar{f}}$: (Equating it to get minimal value)

$$\begin{aligned} P_f \geqslant P_{\bar{f}} &\implies (t_j - t_{median})\lambda - \lambda t_j \geqslant -LY_{R^2A} - \lambda t_{j'} \\ &\implies \lambda(t_{j'} - t_{median}) \geqslant -LY_{R^2A} \\ &\implies LY_{R^2A} \geqslant \lambda(t_{median} - t_{j'}) \end{aligned}$$

The penalty algebraically becomes negative when the vehicle deviates from the system by opting for a path higher than the median cost path. In such a scenario, the system levies no penalty on the vehicle keeping LY to be zero.

Case 2: *The system-assigned path j is among the lower-cost paths.* $(t_j < t_{median})$

In this case, observe that the fee Fee_{R^2A} becomes applicable and REW_{R^2A} is not applicable. Hence, $P_f = -Fee_{R^2A} - \lambda t_j$, where Fee_{R^2A} is computed using Equation 5.6 and t_j represents the path cost as computed using Equations 5.2

and 5.3. Moreover, when V follows a non-system-assigned path j', $P_{\bar{f}} = -LY_{R^2A} - \lambda t_{j'}$, where LY_{R^2A} is computed using Equation 5.7, and $t_{j'}$ is the cost of path j', which is computed using Equations 5.2 and 5.3.

Now, let us find the minimal value LY_{R^2A} of penalty for which $P_f \geqslant P_{\bar{f}}$: (Equating it to get minimal value)

$$\begin{aligned} P_f \geqslant P_{\bar{f}} &\implies -(t_{median} - t_j)\lambda - \lambda t_j \geqslant -LY_{R^2A} - \lambda t_{j'} \\ &\implies \lambda(t_{j'} - t_{median}) \geqslant -LY_{R^2A} \\ &\implies LY_{R^2A} \geqslant \lambda(t_{median} - t_{j'}) \end{aligned}$$

The penalty algebraically becomes negative when the vehicle deviates from the system by opting for a path higher than the median cost path. In such a scenario, the system levies no penalty on the vehicle keeping LY to be zero.■

5.5 Performance Study

This section reports the performance evaluation of our proposed R^{2}A and R^{2}A$^+$ schemes by means of simulation. We consider a universe of 30 km by 30 km, which is divided into 10 regions of equal area. Table 5.1 summarizes the parameters used in the performance study.

We consider a total of 200 checkpoints. The number of checkpoints in each region is determined using a Zipf distribution with a zipf factor ZF_C of 0.5 (i.e., high skew) over 10 buckets, where each bucket corresponds to one of the 10 regions. Then for each region, the required number of checkpoints is randomly selected from the points within that region. Moreover, we consider a total of 25000 vehicles, which are homogeneous in terms of gas mileage and speed. The number of journeys for each vehicle during the course of our experiment is randomly chosen to be between 2 and 6. Furthermore, the source checkpoint for a given vehicle for each journey is chosen randomly from the 200 checkpoints.

For selecting the destination checkpoint for a given journey for a vehicle, we make the observation that in real-world scenarios, destinations for journeys

Parameter	Default	Variations
Number of journeys (N_J) (10^4)	10	2, 4, 6, 8
Number of vehicles (N_V) (10^3)	25	5, 10, 15, 20
Number of checkpoints (N_C)	200	40, 80, 120, 160
Skew in checkpoint distribution (ZF_C)	0.5	0.1, 0.3, 0.7, 0.9
Skew in destination (ZF_D)	0.5	0.1, 0.3, 0.7, 0.9
Percentage of users who are not revenue-conscious (P_U)	40	20, 60, 80, 100

Table 5.1: Parameters of the performance study

are typically skewed across regions. In other words, some of the 'popular' regions would contain destinations for a disproportionately large number of journeys, while other regions would contain destinations for only a relatively smaller number of journeys. Thus, given the total of 100000 journeys (performed by different vehicles) in our experiments, we first select the destination region for each journey using a Zipf distribution with zipf factor ZF_D=0.5 (i.e., high skew) over 10 buckets corresponding to the 10 regions. Then given a destination region, we randomly select any *one* of the checkpoints contained in that region as the destination for a given journey.

Performance metrics are **average fuel savings (AFS)**, **average time savings (ATS)**, **success rate (SR)** and communication cost in terms of the **total number of messages (MSG)**. AFS and ATS are both computed based on the differences in fuel consumption and journey time respectively between the minimum-cost route and the system-assigned route. AFS is computed as follows:

$$AFS = \frac{1}{N_J}\sum_{i=1}^{N_J}(FC_{minP_i} - FC_{SA_i}) \tag{5.9}$$

where FC_{minP_i} and FC_{SA_i} are the fuel consumption corresponding to the minimum-cost route and the system-assigned route respectively for the i^{th} journey. Observe that AFS is computed as the average value of fuel savings across the total number N_J of journeys.

Similarly, ATS is computed as follows:

$$ATS = \frac{1}{N_J} \sum_{i=1}^{N_J} (TC_{minP_i} - TC_{SA_i}) \tag{5.10}$$

where TC_{minP_i} and TC_{SA_i} are the time consumption for the minimum-cost route and the system-assigned route respectively for the i^{th} journey. ATS is computed as the average value of time savings across N_J journeys.

The success rate SR depends upon the number of journeys that completed within $x\%$ of the time required when following the minimum-cost route. Here, required time when using the minimum-cost route is estimated by the route's cost divided by vehicle speed. Our experiments use x=20%. For example, suppose the time required when using the minimum-cost route for a given journey is 20 minutes. Then, only the journeys, which were completed within 24 (i.e., 20×1.2) minutes are deemed to be **successful**. Thus, SR is computed as the ratio between the number of successful journeys and the total number of journeys.

Finally, MSG=$\sum_{i=1}^{N_J} MSG_i$, where MSG_i is the number of messages for the i^{th} journey. Thus, MSG is a cumulative metric over the total of N_J journeys. Notably, the interaction between a vehicle and the base station at any given checkpoint incurs two messages. The first message is from the approaching vehicle to the base station, while the second message is sent by the base station to the vehicle for informing it about the system-assigned path.

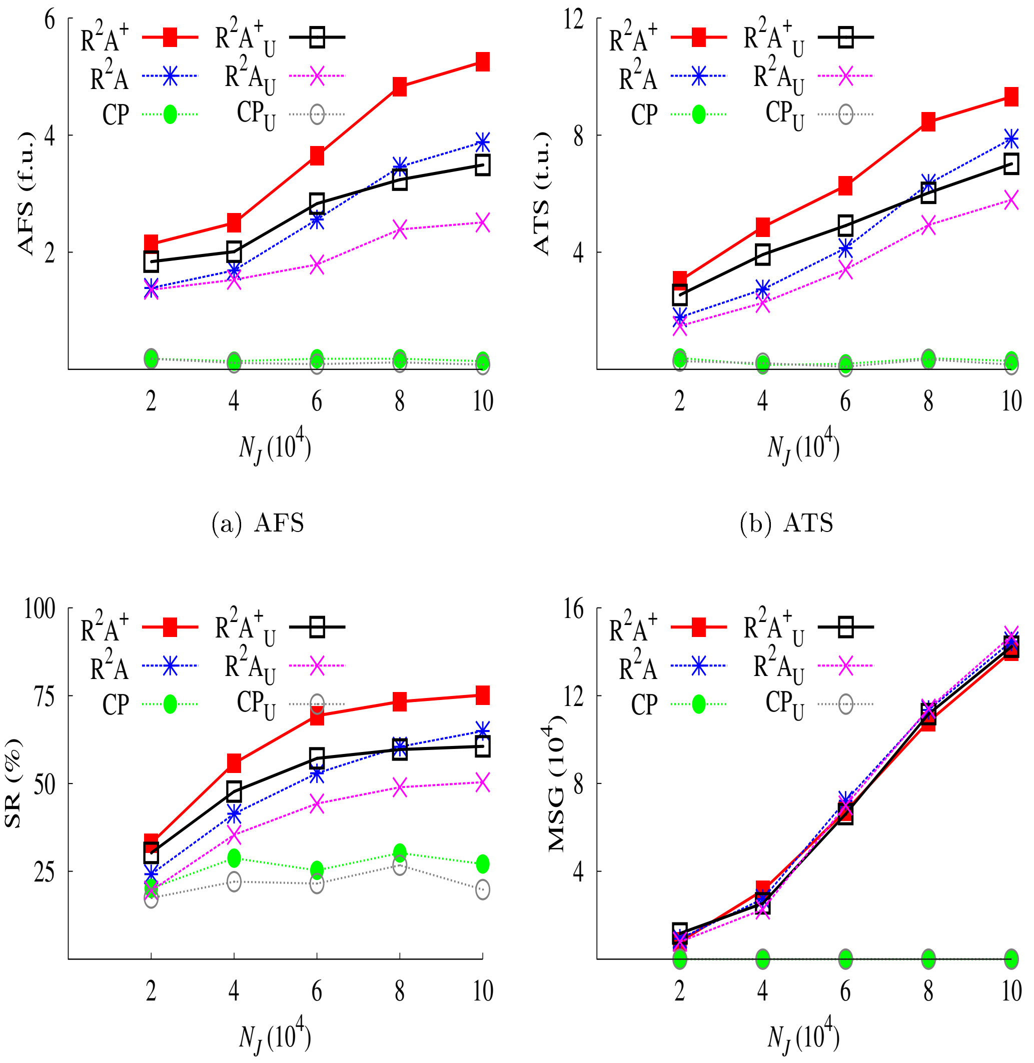
Recall that in Section 5.3.1, we defined the notion of a lower-cost path between two given checkpoints as a path, whose cost is below the median path-cost of all the paths connecting the two checkpoints. As reference, we use a scheme in which only the lower-cost routes carry a fee (akin to toll-road fees), while other paths do not entail any fees. The fee for the lower-cost paths is computed using Equation 5.6. We shall henceforth designate this scheme as the **Congestion-Pricing (CP)** scheme. CP does not provide any rewards to vehicles for taking longer time-cost routes. Moreover, in CP, the base stations do not coordinate vehicular traffic routing, hence they do not provide any economic incentives to the vehicles towards following the system. Furthermore, CP does not necessitate any interactions between base stations and vehicles. In essence, CP charges fees to vehicles when they travel faster, but it does not reward vehicles when they travel slower, and CP does not incorporate the reward/penalty mechanism of E-Vet.

Notably, in case of R^2A, R^2A^+ and CP, the implicit assumption is that every vehicle is trying to maximize its revenue. However, in practice, there could be a percentage of vehicle users, who do not care about maximizing their revenue i.e., these users are not revenue-conscious. Hence, we also examine the performance when 40% of the users are not revenue-conscious, while the other 60% are trying to maximize their revenue. In the experimental results, we designate the performance of R^2A, R^2A^+ and CP under the above condition as $\mathbf{R^2A_U^+}$, $\mathbf{R^2A}_U$ and $\mathbf{CP}_U$ respectively.

5.5.1 Performance of E-VeT

We conducted an experiment using the default values of the parameters in Table 5.1. Figure 5.3 depicts the results. As the number N_J of journeys increases, performance in terms of AFS (measured in **fuel units (f.u.)**), ATS (measured in **time units (t.u.)**) and SR improves for both R^2A and R^2A^+ due to their effective economic reward/penalty-based route allocation, which reduces traffic congestion, thereby resulting in both fuel and time savings as well as higher success rates. Both R^2A and R^2A^+ outperform CP essentially due to their economic reward/penalty-based route allocation approach. In contrast, in case of CP, the vehicles acting selfishly in trying to follow the least-cost routes to their respective destinations cause traffic congestion. This further increases both fuel consumption and average time of journey.

R^2A^+ performs better than R^2A because it provides additional incentives for the vehicles to *consistently* follow the system-assigned paths across multiple journeys. Moreover, R^2A outperforms R^2A_U, and R^2A^+ outperforms $R^2A_U^+$.



(a) AFS (b) ATS

(c) SR (d) MSG

Figure 5.3: Performance of E-VeT

This is because in case of $\mathrm{R^2A}_U$ and $\mathrm{R^2A}_U^+$, 40% of users are not revenue-conscious, thereby reducing the effectiveness of these schemes. Furthermore, $\mathrm{R^2A}_U^+$ outperforms $\mathrm{R^2A}$ (in terms of AFS, ATS, SR) upto a certain number of journeys primarily because $\mathrm{R^2A}_U^+$ better incentivizes vehicles in consistently following the system-assigned paths. However, as the number of journeys exceeds 80,000, $\mathrm{R^2A}$ performs better than $\mathrm{R^2A}_U^+$. This occurs because as the number of journeys increases beyond a certain point, the implication is that a larger absolute number of users are not revenue-conscious, thereby reducing the effectiveness of $\mathrm{R^2A}_U^+$.

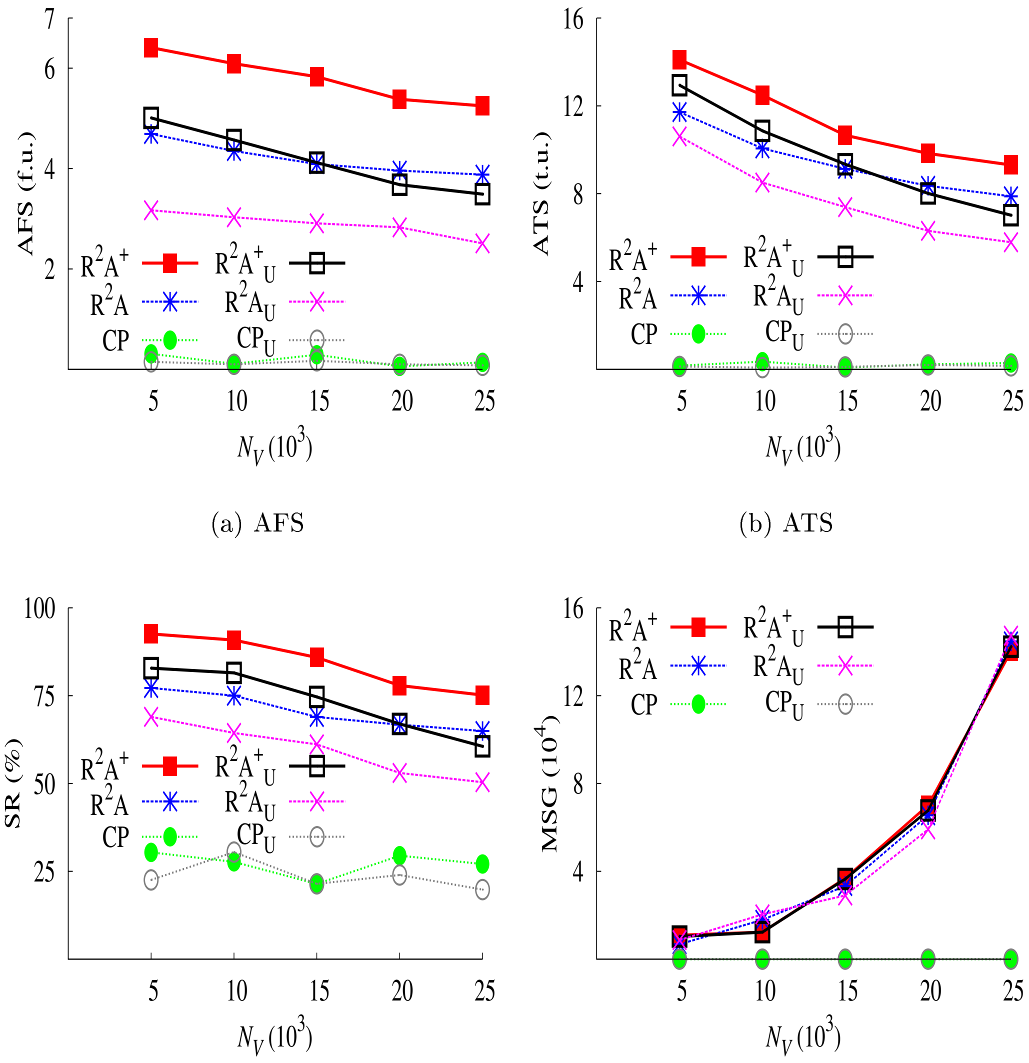
Observe that MSG equals zero for CP in all cases because in CP, the vehicles do not need to interact with the base stations. On the other hand, MSG increases over the number of journeys in our proposed schemes since it is a cumulative metric. MSG is comparable in case of R^{2}A, R^{2}A$^+$, R^{2}A$_U$ and R^{2}A$_U^+$ as these schemes involve similar interactions between the vehicles and the base stations at the corresponding checkpoints. These interactions occur through the software in each vehicle, regardless of whether the users are revenue-conscious.

5.5.2 Effect of varying the number of vehicles

We conducted an experiment to examine the scalability of E-VeT w.r.t. the number of vehicles. Figure 5.4 depicts the results of varying the number N_V of vehicles. As N_V increases, performance in terms of AFS, ATS and SR degrades for both R^{2}A and R^{2}A$^+$ due to increased traffic congestion arising from a larger number of vehicles. Observe that the performance degradation is only slight essentially due to the effective incentive-based route allocation performed by both R^{2}A and R^{2}A$^+$. As N_V increases, MSG increases for both R^{2}A and R^{2}A$^+$ because the number of interactions between the base stations and the vehicles increases, given the increase in the number of vehicles. Furthermore, the explanation for the relative performance of R^{2}A$_U^+$ and R^{2}A is essentially the same as in case of Figure 5.3.

5.5.3 Effect of varying the number of checkpoints

We conducted an experiment to investigate the effect of varying the number N_C of checkpoints. Figure 5.5 depicts the results. As N_C increases, the route assignments are performed at a relatively larger number of checkpoints, thereby implying more opportunities for fine-tuning the route assignments. Hence, performance (in terms of AFS, ATS and SR) improves for all our proposed schemes. This performance gain comes at the cost of higher MSG because vehicles and base stations exchange messages at a larger number of

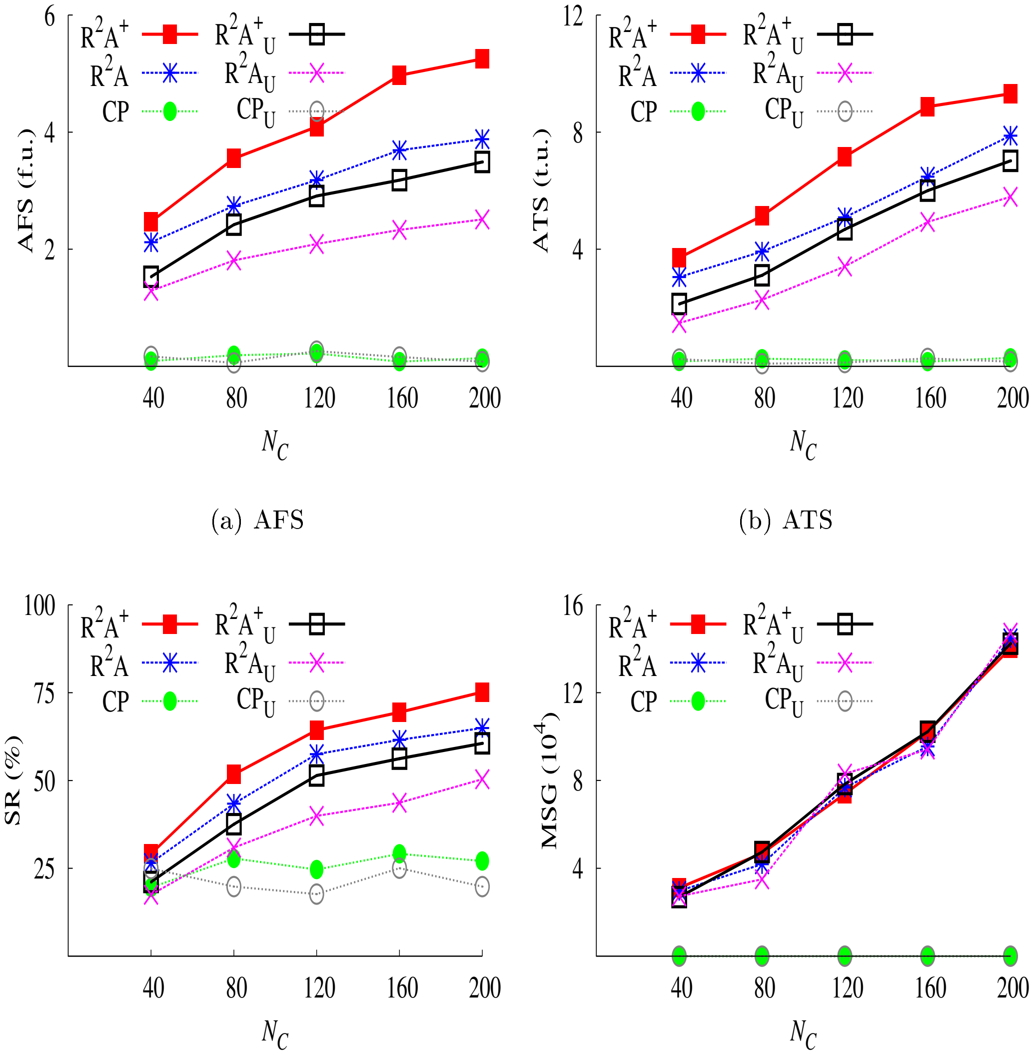


(a) AFS (b) ATS

(c) SR (d) MSG

Figure 5.4: Effect of variations in the number of vehicles

checkpoints. However, we believe that the increase in MSG is a small price to pay for the performance gain in reducing the overall traffic congestion. Furthermore, the results indicate that R^{2}A performs better than R^{2}A^+_U. This occurs because 40% of the users are not revenue-conscious in case of R^{2}A^+_U, which reduces its effectiveness in reducing traffic congestion.



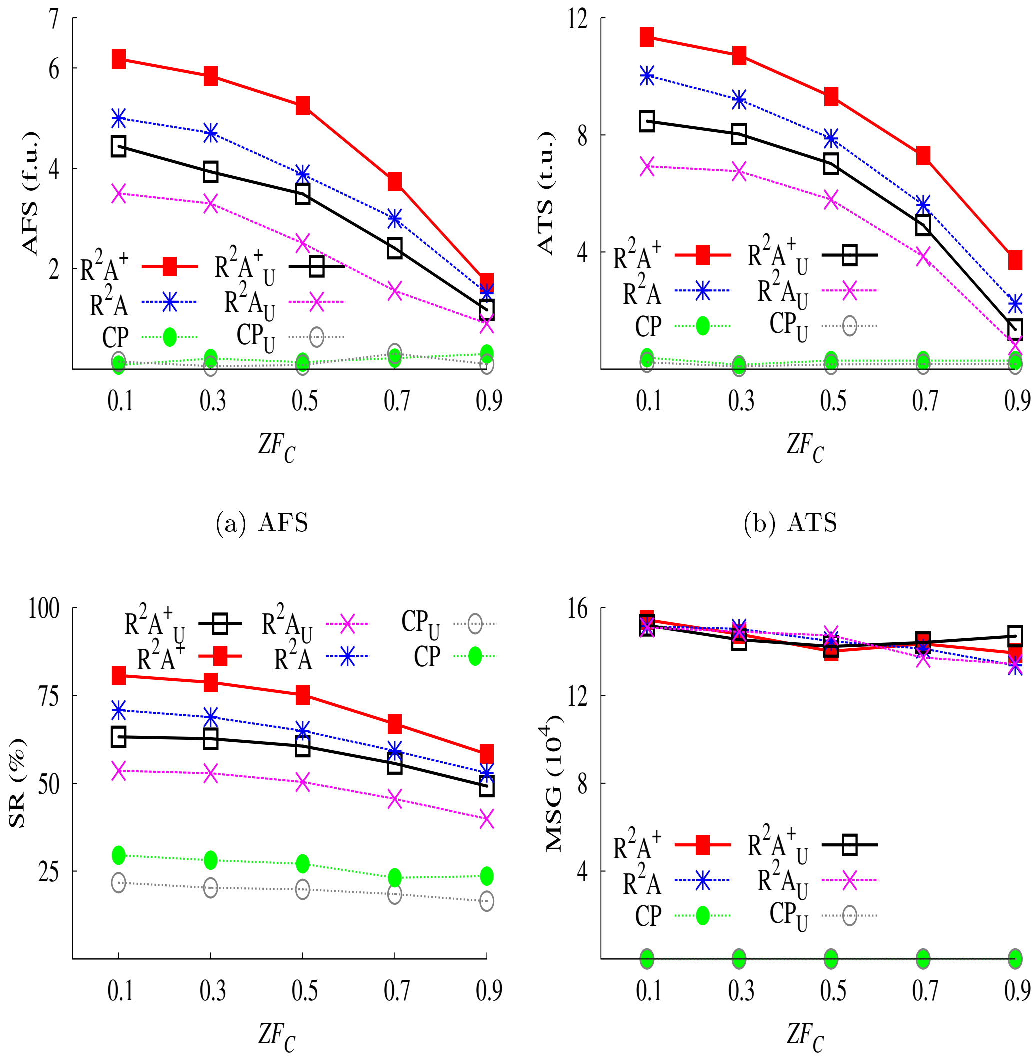
(a) AFS (b) ATS

(c) SR (d) MSG

Figure 5.5: Effect of variations in the number of checkpoints

5.5.4 Effect of variations in skew in checkpoint distribution

Recall that ZF_C is the zipf factor, which quantifies the skew in the distribution of checkpoints across the 10 regions considered in our experiments. Figure 5.6 depicts the results of varying ZF_C. As ZF_C increases, the implication is that a disproportionately large number of checkpoints occur in only a few of the regions, while other regions contain only a relatively small numbers. Hence, in most of the regions, there are fewer opportunities for



(a) AFS (b) ATS

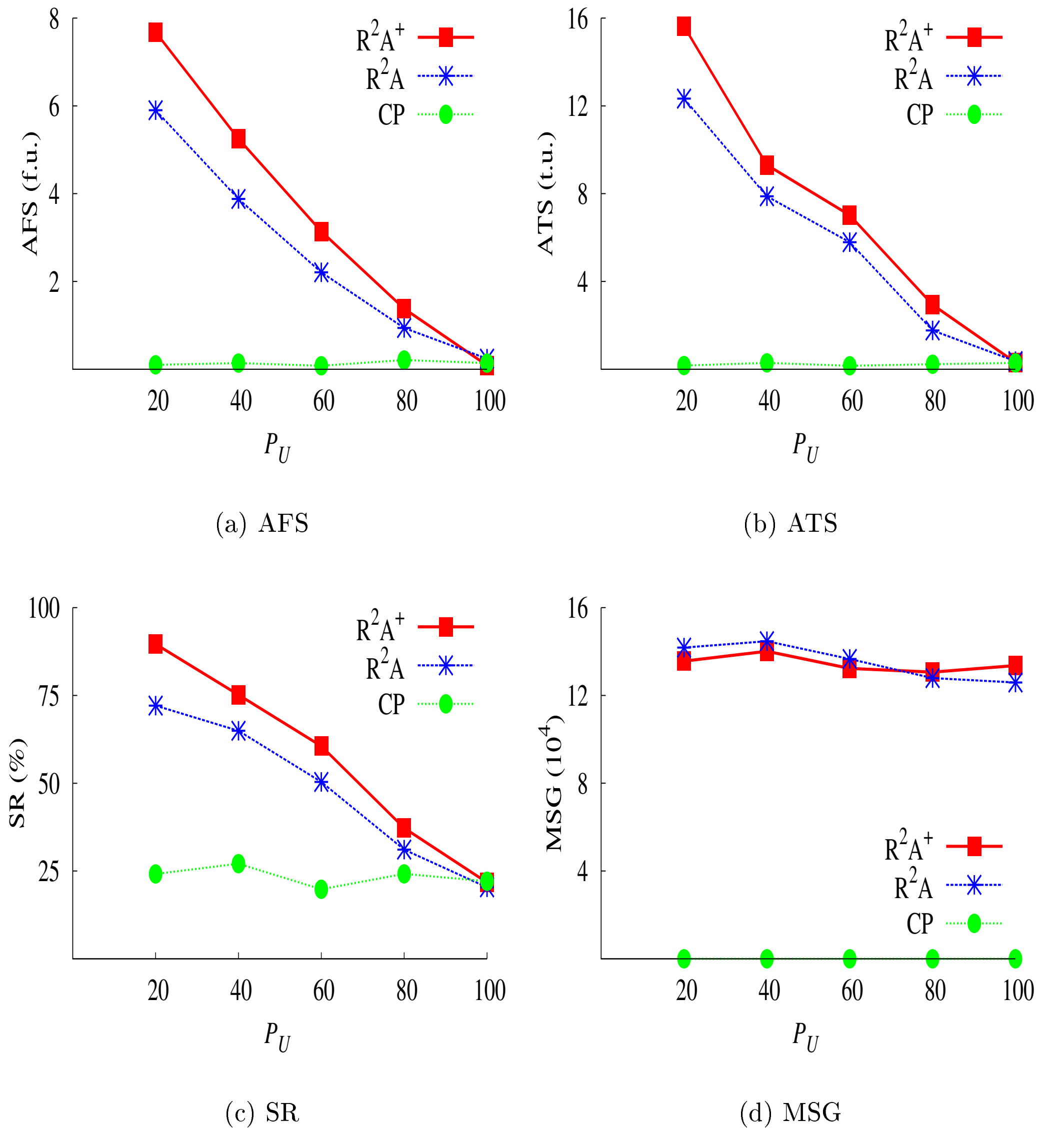
(c) SR (d) MSG

Figure 5.6: Effect of variations in skew in the distribution of checkpoints

fine-tuning the route assignments. Thus, given that vehicular journeys cut across different regions, the performance (in terms of AFS, ATS and SR) of our proposed schemes degrades as the checkpoint distribution becomes more skewed. However, MSG remains comparable for our proposed schemes because it depends upon the number of checkpoints, regardless of the skew in the distribution of checkpoints.

5.5.5 Effect of variations in the percentage of users who are not revenue-conscious

Recall that the R^2A_U, $R^2A_U^+$ and CP_U schemes were defined to take into account that a certain percentage of the users are not revenue-conscious i.e., they do not care about maximizing their revenue. Let us designate this per-



(a) AFS (b) ATS

(c) SR (d) MSG

Figure 5.7: Effect of variations in the percentage of users who are not revenue-conscious

centage as P_U. In Figure 5.7 as P_U increases, the performance of our proposed schemes degrade because an increasing number of users become indifferent to the revenue-maximization objective, which is used by our schemes to incentivize vehicles to follow the system-assigned paths, thereby reducing the overall effectiveness of our schemes. Interestingly, when P_U=100% the performance of our proposed schemes becomes comparable to that of CP because at that point, 100% of the users are not revenue-conscious, thereby implying that the users are not incentivized by our proposed schemes. However, MSG remains comparable across variations in P_U because the interactions between the vehicles and the base stations occur through the software in each vehicle, regardless of whether the users are revenue-conscious.

5.6 Summary

We have proposed the E-VeT system for efficiently managing vehicular traffic in road networks using economy-based reward/penalty schemes. E-VeT aims at reducing traffic congestion by enabling base stations to collaboratively facilitate dynamic vehicular route assignment. Our proposed R^2A scheme rewards vehicles for following system-assigned longer-time paths, and charges a fee for following system-assigned shorter-time paths. R^2A^+ extends R^2A by incorporating the notion of *revenue-scales* for additionally considering a given vehicle's past history in following system-assigned paths across multiple journeys. Thus, in E-VeT, vehicles earn revenues based on either the R^2A scheme or the R^2A^+ scheme. The route allocation algorithm used by E-VeT provides preference to vehicles earning higher revenues by assigning them to shorter-time paths, thereby incentivizing them to follow the system-assigned paths. Our performance study shows that the proposed schemes are indeed capable of effective traffic management in road networks by reducing the average time of arrival and fuel consumption.

In the near future, we will validate the results of spatio-temporal variations of vehicular traffic in VANETs [BK09] within our framework of economic reward/penalty schemes.

6

E-Rare: Economic Incentive Schemes for Improving Availability of Rare Data in Mobile-P2P Networks

6.1 Overview

In a Mobile ad hoc Peer-to-Peer (M-P2P) network, mobile peers (MPs) interact with each other in a peer-to-peer (P2P) fashion. Proliferation of mobile devices (e.g., laptops, PDAs, mobile phones) coupled with the ever-increasing popularity of the P2P paradigm (e.g., KaZaa) strongly motivate M-P2P applications. Mobile devices wirelessly communicating in a P2P fashion (as Microsoft's Zune [Zun06]) facilitate M-P2P applications by sharing information on-the-fly.

This work focusses on handling *rare* data items in an M-P2P environment. *Rare* data items are those, which get sudden bursts in accesses based on *events* as they are only hosted by only a few peers in comparison to the network size. Thus, they may not be available within few hops of query-issuing peers. The sudden burst in accesses to rare items generally occurs within a given time-frame (associated with the event), before and after which such items are rarely accessed.

Some application scenarios are as follows. Suppose a group of college students in the course of an expedition in a remote forest, where communication infrastructures (e.g., base stations) do not exist. When there is a sudden decrease in temperature and gusty winds, they need to look for information about shops selling sweaters and wind-cheaters in a nearby town, photos of such clothing and so on. In a similar vein, suppose a group of adventure tourists *unexpectedly* encounters a cave during their journey. They would like to find information about where to buy gas-masks and associated safety equipment along with instructional tutorials on how to use this equipment and so on. Similarly, when a motorist driving in a mountainous region, sees a rare animal, she may wish to find additional information about living habits. Additionally, due to the sudden onset of a heat wave, a group of botanists on an expedition in a forest may want to find information such as non-drinking water sources and pictures of the locations of such water sources. In these application scenarios, M-P2P interactions can facilitate the MPs in finding the required information.

Such M-P2P interactions for effective sharing of rare data are currently not freely supported by existing wireless communication infrastructures. Observe how the sudden urgent demand of several MPs for information concerning rare items (e.g., protective clothing or gas-masks) arises due to the occurrence of *events* such as the sudden onset of harsh weather conditions or the users unexpectedly encountering a cave.

In this work, we assume an environment, where all the MPs collaborate on information sharing and are trusted. Notably, any distributed trust management schemes [QMK10, RSB11, SL03] can be used in conjunction with our proposed work for managing trust. Furthermore, we assume that there is no connection between the seller of the rare items and the MPs who own/host information about them. Thus, these MPs are not agents of any sellers of the rare items. They provide the information that they have collected from their own use or based on their general interest in some types of rare items. Thus, the scope of our proposed model is restricted to information exchange about rare items among the MPs within the M-P2P network (such as in crowdsourcing) as opposed to the buying/selling of the *actual* rare items.

Similar to the works in [HM06, XP03], our target applications mainly concern slow-moving objects e.g., adventure tourists in a forest. Our application scenarios assume data accesses to occur within soft real-time deadlines and as such, we do not address scenarios where real-time access is required. Additionally, given our assumption concerning slow-moving objects, a query-issuing MP may still be wandering in the region for say, the next 2-3 minutes. In our work, a user specifies a TTL (hence a soft real-time) for her query, and if the answer is not found within the TTL, the query fails.

Data availability in M-P2P networks is typically lower than in fixed networks due to frequent network partitioning arising from peer movement, mobile resource constraints (e.g., bandwidth, energy, memory space) and mobile devices being autonomously switched 'off'. (Incidentally, data availability is less than 20% even in a wired environment [SGG01].) Rampant free-riding further reduces data availability since a large percentage of MPs are typically free-riders [HA05, KSGM03a] i.e., they do not provide any data. Availability of rare data is further exacerbated since they are generally stored at relatively few MPs, which may be several hops away from query-issuers. Thus, economic models become a necessity to combat free-riding and to incentivize MPs to host replicas for improving rare data availability in M-P2P networks.

Existing replication schemes for improving data availability in mobile ad hoc networks (MANETs) [HM06, KMSA08] do not consider economic incentives for data hosting, licensing mechanisms, M-P2P architecture and data item rarity issues. Incentive schemes for MANETs [BH03, CN04, CGKO03, SNCR03] primarily focus on encouraging message forwarding, but they do not address replication. M-P2P incentive schemes [WXS04, XWR06] do not address the replication of rare data items.

This work proposes **E-Rare**, a novel *economic incentive model* for improving rare data availability by means of licensing-based replication in M-P2P networks. E-Rare comprises two replication schemes, namely ECR and ECR+, both of which use its incentive model for improving rare data availability. The key difference between these schemes is that in ECR, the MPs act individually towards replication, while for ECR+, the MPs perform replication in groups. In both these schemes, a given MP issues queries specifying its desired data item, its location and query deadline.

In E-Rare, each data item d is associated with four types of prices, which provide different rights to the query-issuing MP M_I concerning the usage of d. The first two price types entitle M_I to obtain information about d at different levels of detail (e.g., information about a few shops selling gas-masks versus complete catalogues of more shops selling gas-masks), but they do not provide M_I the right (or license) to enable downloads of d from itself. In contrast, the third and fourth price types concern licensing for partial and full use downloads, and are aimed towards enabling and incentivizing replication by means of data licensing. Notably, all four price types depend upon factors such as item rarity score and timeliness of query response. In ECR, the item rarity score depends upon the variability in the access counts of d during recent periods of time. Here, we assume that time is divided into equal intervals, each of which is designated as a *time-period*. Notably, our proposed approach requires synchronized clocks among the MPs, and the existing clock synchronization approaches proposed in [CW04, SCHS07] can be used. In ECR+, the item rarity score additionally depends upon the number of MPs which host d.

E-Rare requires a query-issuing MP M_I to pay any *one* of these four prices for its requested data item to the MP M_S serving its request, depending upon the price type associated with its query. Furthermore, it requires M_I to pay a constant *commission* to each relay MP in the successful query path from which it eventually downloads the data, thereby enticing them to forward queries quickly. Note that even though the M_I has to pay a constant relay commission to each relay MP, it does not necessarily imply that a shortest path should be applied because the total payment made by the M_I includes both the item price and the relay commissions. For example, using the shortest path would result in the M_I paying a lower amount for relay commissions, but it could end up paying a higher total cost because the item price may be higher at the mobile peer (in that path) from which M_I would need to eventually download the item.

Observe how E-Rare effectively combats free-riding because free-riders would have to earn currency for issuing their own requests, and they can earn currency only by means of hosting items and relaying messages. Notably, given that E-Rare associates rare data items with prices, it is possible for an M_I to avoid accessing the items because of their prices or if the M_I has not earned adequate revenue by hosting items or by relaying messages. Furthermore, in E-Rare, item prices increase with rarity, thereby providing free-riders with higher incentive [GA04, RFJY03, SH04] to host rare items for maximizing their revenues. By enticing free-riders to pool in their energy and bandwidth resources to host rare items, E-Rare improves rare data availability due to replication.

In ECR+, a **peer group** is defined as a set of MPs working together such as an adventure tour expedition group. MPs provide *discounts* only to the MPs within their group, thereby incentivizing MP participation in the group. These discounts are applicable to all the four price types discussed earlier. Notably, group members need not necessarily be one-hop neighbors i.e., they may be scattered across the network due to peer movements.

The main contributions of E-Rare are three-fold:

1. It provides incentives for replication of rare data items by means of a novel licensing mechanism, thereby improving rare data availability.

2. It provides additional incentives for MPs to collaborate in groups,

thereby further improving rare data availability.

3. A detailed performance evaluation has been done to show the improvement in query response times and availability of rare data items in M-P2P networks.

Incidentally, virtual currency incentives are suitable for P2P environments due to the high transaction costs of real-currency micro-payments [TR04]. The works in [DPGB03,ET04,ZCY03] discuss how to ensure secure payments using a virtual currency. Notably, these secure payment schemes are complementary to our proposal, but they can be used in conjunction with our proposal.

We have performed a detailed performance evaluation of both ECR and ECR+. As a baseline reference, we have also compared against an existing non-incentive and non-economic replication E-DCG+ scheme for MANETs, proposed in [HM06], which is closure to our scenario. We have used average response times of queries, query success rates, query hop-counts and the number of messages as performance metrics. ECR+ outperforms ECR due to its group-based incentives (such as discounts), which facilitate collaborative replication among MPs. ECR outperforms E-DCG+ essentially due to its economic licensing scheme, which incentivizes MP participation in the creation of multiple copies of rare items. Both ECR and ECR+ incur more messages than E-DCG+ because in case of E-DCG+, a large percentage of unsuccessful queries result in decreased amount of data transfer, albeit at the cost of reduced query success rates.

The results also indicate that both ECR and ECR+ exhibit good scalability with increasing number of MPs due to increased opportunities for replication. Moreover, ECR+'s performance improves with increasing group size due to increased replication opportunities. However, beyond a certain point, further increase in group size does not significantly improve performance due to saturation. Both ECR and ECR+ perform best when the communication range is neither too high nor too low. This is because when the communication range is large (i.e., in effect, the MPs are 'nearer' to each other), the effect of gains in query response times is offset by the overheads of higher number of incoming queries at MPs that host data items and increased relay propagation latencies. Conversely, when the communication range is too small, query response times increase because more hops are required for answering queries.

ECR+ performs best when the discount is neither too high nor too low. This is because when the discount is too high, MPs hosting rare items have reduced incentives to join the group due to reduction in their earnings from license prices. On the other hand, when the discount is too low, MPs trying to obtain licenses for replicating rare items have reduced incentive to participate in the group. The results also demonstrate that both ECR and ECR+ perform best when replication is performed neither too early nor too late. Finally, the performance of both ECR and ECR+ degrades with increasing percentage of MP failures due to reduced opportunities for replication.

The remainder of this chapter is organized as follows. Section 6.2 details the economic incentive model of E-Rare for rare data items. Section 6.3 discusses the ECR and ECR+ replication schemes. Section 3.5 reports our performance evaluation. Finally, we summarize E-Rare in Section 6.5 with directions for future work.

6.2 Architecture of E-Rare

This section discusses our proposed economic incentive model E-Rare for improving the availability of rare data items in M-P2P networks.

In E-Rare, a given query-issuing MP M_I issues a query Q of the form (d, L, τ_Q), where d is the queried data item. Data item d is described as a combination of keywords. We assume that each device in M-P2P network has the capability to match keywords to data items stored in their devices L represents the query location, and is of the form $\{(x,\ y),\ rad\}$. Here, $(x,\ y)$

represents the spatial coordinates associated with a given query Q, while rad represents the radius. For example, M_I may query for an item d within 1 km of its current location L. τ_Q is the deadline time of Q. The ephemerality of M-P2P environments necessitates timely responses, and consequently query deadlines. Notably, the query-issuer does not specify an explicit rarity score for its queried item because rarity scores of any given item can change dynamically depending upon accesses, and these scores vary across the MPs. Hence, the query-issuer does not necessarily know the rarity scores of data items that are hosted at other MPs. In essence, we want the rarity scores to be kept transparent from the query-issuers.

This work assumes that the only way that a MP can obtain a data item is by purchasing it. Thus, a given MP cannot obtain a data item while relaying it for other MPs. This assumption is justifiable in practice because each data item is protected through copyright protection, encryption and security mechanisms [DCRS14]. Several existing content authoring techniques can be used for license protection and restricted distribution [Jok03].

6.2.1 Computation of the rarity score λ_d

Now let us discuss how the rarity score λ_d of a data item d is computed in E-Rare. λ_d depends upon the variability in the access count of d during the past N periods of time. Observe that the value of λ_d should increase with the variability in access count of d over the last N periods in consonance with our definition of rarity, which incorporates sudden bursts in accesses for rare items. For example, information about gas-masks and associated safety equipment is heavily accessed only during a specific time-frame associated with a rare event, while at other times, such information may not be accessed at all. The computation of λ_d follows:

$$\lambda_d \;=\; [\;\{\;(\,\eta_c - (\,\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\eta_i\,)\,)\;/\;max\,(\,\eta_c\,,\;\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\eta_i\,)\;\}\;+\;1\;]\;/\;2 \qquad (6.1)$$

where N is the number of time-periods over which λ_d is computed. Here, η_c refers to the access count of data item d for the current period, while η_i represents the access count of d for the i^{th} time-period. Our preliminary experiments revealed that $N = 5$ is suitable for our application scenarios.

Notably, the term $(\frac{1}{N} \sum_{i=1}^{N} \eta_i)$ represents the average access count of d during the last N time-periods. Thus, when the current period's access count exceeds the past average access count, the term $\{ \ (\eta_c - (\frac{1}{N} \sum_{i=1}^{N} \eta_i)) \ / \ max \, (\eta_c \, , \frac{1}{N} \sum_{i=1}^{N} \eta_i) \ \}$ lies between 0 and 1. On the other hand, when the current access count falls below the past average access count, this term lies between -1 and 0. Hence, in Equation 6.1, we add 1 to this term and divide by 2, thereby making the value of λ_d between 0 and 1. Observe that the value of λ_d may differ among the MPs for the same data item since it is associated with sudden bursts at each MP.

Based on the value of λ_d, a given data item is classified into one of the following three classes: *rare*, *medium-rare* and *non-rare*. Each class is associated with a range of λ_d. For *rare* items, $0.7 \leq \lambda_d \leq 1$; for *medium-rare* items, $0.5 \leq \lambda_d < 0.7$; and for *non-rare* items, $0 < \lambda_d < 0.5$. These rare data classes are determined based on our experimental results. The ranges for each class are pre-specified system constants that are known to all the MPs.

6.2.2 Types of item prices in E-Rare

Each query Q for any given item d is associated with *any one* of four types of prices, which provide different rights to the query-issuing MP M_I concerning the usage of d. We designate these prices as *partial_use_price* $P_{d,Q}$, *full_use_price* $F_{d,Q}$, *partial_use_license_price* $PUL_{d,Q}$ and *full_use_license_price* $FUL_{d,Q}$. M_I pays one of these four prices to the query-serving MP M_S, depending upon the type of price associated with its query.

Paying the *partial_use_price* $P_{d,Q}$ entitles M_I to obtain some basic or partial information about its queried data item d, while paying *full_use_price* $F_{d,Q}$ entitles M_I to obtain more detailed information about d. For example, in case of our application scenario concerning a sudden spike in the demand for gas-masks and associated safety equipment, paying $P_{d,Q}$ would entitle M_I only to information about a few shops selling such equipment and their respective prices at these shops. However, paying $F_{d,Q}$ provides M_I with more detailed information such as complete catalogues of more shops selling these items, contact addresses and telephone numbers of these shops, how to order these items (e.g., by phone) and instructional materials demonstrating how to use these items. Notably, the payments of $P_{d,Q}$ or $F_{d,Q}$ pertain to M_I's *sole use* of d i.e., M_I does not obtain the right to host d at itself for downloads by other MPs. Thus, M_I cannot earn currency by hosting d.

For obtaining the right to earn currency by hosting d and allowing downloads of d at itself, M_I needs to pay either of the two license prices for d. In E-Rare, an MP may purchase two types of licenses, which we designate as partial use license (PUL) and full use license (FUL) respectively. When an MP M purchases a single PUL for a data item d from d's original owner, it obtains the right to provide d to any *one* query-issuing MP, which issues a *partial_use_price* query for d. Thus, purchasing n_P PULs for d enables M_I to earn currency from n_P downloads of d pertaining to *partial_use_price* queries. Similarly, purchasing n_F FULs for d enables M_I to earn currency from n_F downloads of d pertaining to *full_use_price* queries. Observe that being collaborative and trusted, MPs in possession of an item would not exceed the number of pre-specified downloads. Furthermore, we assume that each data item is protected through copyright protection and license security mechanism. Several existing content authoring techniques can be used for license protection and restricted distribution. They can be encrypted using traditional public-private key encryption technqiues [DCRS14]. Notably, observe that in E-Rare, when an MP pays a one-time license price to get a given item, it is not allowed to fulfill as many queries as it wants, although allowing an MP to do so would improve data availability. The rationale behind this is to protect the original owner's benefit i.e., to incentivize the original owner to create and maintain information about rare items.

Observe how the data licensing mechanism of E-Rare provides an economic means of incentivizing data replication because the data owners can earn currency from the license prices. We assume that the initial number of licenses is fixed by the owner of data items, and that decides the number of licensees. We also assume that there are enough peers interested to ask for licenses of a given item from licensor. In general, the number of licensees can be updated by the owner on the regular feedback received from other peers within the group (in case of ECR+) based on the query response time, and their availability. Furthermore, if the owner of an item d replicates d without charging a license price, competition with the MPs hosting replicas of d would be likely to reduce its earnings from hosting d. The licensing mechanism also improves rare data availability by guarding against the possibility of unavailability of the rare item owner.

For the sake of convenience, Table 6.1 summarizes the notations used in E-Rare. Notably, each data item d is associated with a score λ_d, which quantifies its rarity, and therefore influences item prices. The remainder of this section discusses the computation of the four price types and the computation of MP revenues in E-Rare.

Symbol	Significance
d	A data item
λ_d	Rarity score of d
M_I	Query-issuing MP
M_S	Query-serving MP
$P_{d,Q}$	The *partial_use_price* of d
$F_{d,Q}$	The *full_use_price* of d
$PUL_{d,Q}$	The *partial_use_license_price* of d
$FUL_{d,Q}$	The *full_use_license_price* of d

Table 6.1: Summary of notations in E-Rare

Computation of the ***partial_use_price*** $P_{d,Q}$

The *partial_use_price* $P_{d,Q}$ of a data item d for a given query Q depends on the rarity score λ_d of d and the response time of the query Q w.r.t. the query deadline. Notably, $P_{d,Q}$ should increase with increase in λ_d because rare items should command higher prices. Furthermore, for rewarding faster service, $P_{d,Q}$ should be higher for queries answered considerably earlier than the query deadline than for queries answered very close to the deadline. Thus, given that τ_Q and R_Q represent the query deadline and the query response time respectively, $P_{d,Q}$ should increase with increase in the ratio (τ_Q / R_Q). $P_{d,Q}$ is computed as follows:

$$P_{d,Q} = \begin{cases} (\ \lambda_d \times e^{\tau_Q/R_Q}\ ) & if \;\; R_Q \leq \tau_Q \\ 0 & otherwise \end{cases} \tag{6.2}$$

When making the purchase, the buyer is provided with a list of pricing options e.g., if 3 minutes delay, \$10; if 7 minutes delay, \$2 etc. Thus, the buyer has some expectation about the total price which he will be paying for the purchase. Observe that for queries answered after the deadline, $P_{d,Q}$ is set to zero because the query results may no longer be useful to the query-issuer. Observe how $P_{d,Q}$ decreases with decreasing the rarity score λ_d. Furthermore, for the very first query on d, we assume R_Q=τ_Q for bootstrapping purposes. Hence, in this special case, $P_{d,Q}$=$\lambda_d \times e$.

Computation of the *full_use_price* $F_{d,Q}$

Intuitively, we can understand that the *full_use_price* $F_{d,Q}$ of a data item d for a given query Q should always exceed its *partial_use_price* $P_{d,Q}$ because it provides more information to the query-issuer. We compute $F_{d,Q}$ as follows:

$$F_{d,Q} = P_{d,Q} \times \Upsilon \tag{6.3}$$

In Equation 6.3, the value of $P_{d,Q}$ is computed from Equation 6.2. Here, Υ always exceeds 1 to ensure that $F_{d,Q}$ always exceeds $P_{d,Q}$. The value of Υ depends on the difference between the value proposition to the user provided by partial and full access to the information, and hence, it is application-dependent. In this work, based on the results of our preliminary experiments, we set $\Upsilon = 1.3$.

Computation of *license_prices* $PUL_{d,Q}$ and $FUL_{d,Q}$

The license prices, $PUL_{d,Q}$ and $FUL_{d,Q}$, for a single PUL and FUL respectively are computed as follows:

$$PUL_{d,Q} = \mu_{P_{d,Q}} + (\mu_{P_{d,Q}} + \mu_{F_{d,Q}}) / \mu_{P_{d,Q}} \tag{6.4}$$

$$FUL_{d,Q} = \mu_{F_{d,Q}} + (\mu_{P_{d,Q}} + \mu_{F_{d,Q}}) / \mu_{P_{d,Q}} \tag{6.5}$$

In Equations 6.4 and 6.5, $\mu_{P_{d,Q}}$ and $\mu_{F_{d,Q}}$ are the *average* values of $P_{d,Q}$ and $F_{d,Q}$ respectively at the original owner of d since both $P_{d,Q}$ and $F_{d,Q}$ vary across queries. Thus, the owner of d computes $\mu_{P_{d,Q}}$ and $\mu_{F_{d,Q}}$ by averaging the individual values of $P_{d,Q}$ across all the queries (for d) that it answered corresponding to the *partial_use_price* and the *full_use_price* respectively during recent time-periods.

Intuitively, $PUL_{d,Q}$ should exceed $\mu_{P_{d,Q}}$ because it enables the query-issuer to earn currency from hosting item d. In Equation 6.4, observe that $PUL_{d,Q}$ always exceeds $\mu_{P_{d,Q}}$ because the second term is always a positive number that is greater than 1. This is because $\mu_{F_{d,Q}} > \mu_{P_{d,Q}}$, as discussed earlier. Similarly, in Equation 6.5, $FUL_{d,Q}$ always exceeds $\mu_{F_{d,Q}}$.

For the sake of convenience, we have summarized the four price types in E-Rare in Table 6.2.

	Partial (Information)	**Full (Information)**
Use	$P_{d,Q} = \begin{cases} (\lambda_d \times e^{\tau_Q/R_Q}) & if \;\; R_Q \leq \tau_Q \\ 0 & otherwise \end{cases}$	$F_{d,Q} = P_{d,Q} \times \Upsilon$
License	$PUL_{d,Q} = \mu_{P_{d,Q}} + (\mu_{P_{d,Q}} + \mu_{F_{d,Q}})/\mu_{P_{d,Q}}$	$FUL_{d,Q} = \mu_{F_{d,Q}} + (\mu_{P_{d,Q}} + \mu_{F_{d,Q}})/\mu_{P_{d,Q}}$

Table 6.2: Summary of item price types in E-Rare

6.2.3 Revenue of an MP

Revenue of an MP M is the difference between the amount of virtual currency that it earns and the amount that it spends. M earns currency from accesses to data items that it hosts and by relaying messages. M spends currency by accessing items hosted at other MPs, and by paying commissions to the relay MPs corresponding to its queries. Given that E-Rare has four types of item prices, the revenue of M is the sum of the *net earnings* of an MP corresponding to each of these four price types and the *net earnings* due to message relay commissions.

In E-Rare, message relay commission is a constant K, which is a small percentage of the average *partial_use_price* $\mu_{P_{d,Q}}$. This is in consonance with E-Rare's objective of providing greater incentives to MPs for hosting items than for relaying messages. In this work, we set K to be 5% of the average value of $P_{d,Q}$. Note that the value of $P_{d,Q}$ varies across queries, however the prices of items in the application can be used as a guideline to estimate an approximate average value of $P_{d,Q}$. In this work, relay MPs have to relay as this is a part of the protocol i.e., they do not decide whether they want to relay the data.

Suppose M hosts p data items. For queries served by M, let the access counts of the i^{th} item corresponding to the *partial_use_price*, *full_use_price*, *partial_use_license_price* and *full_use_license_price* be ns_{P_i}, ns_{F_i}, ns_{PUL_i} and ns_{FUL_i} respectively. Moreover, let the corresponding prices for these accesses be P_i, F_i, PUL_i and FUL_i respectively. Furthermore, suppose M relays m messages. Thus, the total earnings E_M of M is computed as follows:

$$\begin{aligned} E_M \;=\; & \textstyle\sum_{i=1}^{p} [\,(\, ns_{P_i} \times P_i \,) \;+\; (\, ns_{F_i} \times F_i \,) \;+ \\ & \qquad\qquad (\, ns_{PUL_i} \times PUL_i \,) \;+\; (\, ns_{FUL_i} \times FUL_i \,)\,] \\ & + (\, m \times K \,) \end{aligned} \tag{6.6}$$

In the above equation, the first and second terms represent M's earnings corresponding to *partial_use_price* and *full_use_price* respectively, while the third and fourth terms relate to M's earnings from licensing. Note that M can earn license prices (corresponding to PUL and FUL) only for the items that it owns. The fifth term represents M's earnings from relay commissions.

Let the number of queries issued *successfully*[1] by M corresponding to the *partial_use_price*, *full_use_price*, *partial_use_license_price* (PUL) and *full_use_license_pri* (FUL) be nq_P, nq_F, nq_{PUL} and nq_{FUL} respectively. Moreover, let the i^{th} item's price paid by M to obtain the query result corresponding to its desired price type be P_i, F_i, PUL_i and FUL_i respectively. Furthermore, suppose M paid relay commissions for n messages in the course of issuing different queries. Thus, the total spending S_M of M is computed as follows:

$$\begin{aligned} S_M \;=\; & [\,\textstyle\sum_{i=1}^{nq_P} P_i\,] + [\,\sum_{i=1}^{nq_F} F_i\,] + \\ & [\,\textstyle\sum_{i=1}^{nq_{PUL}} PUL_i\,] + [\,\sum_{i=1}^{nq_{FUL}} FUL_i\,] + (\,n \times K\,) \end{aligned} \tag{6.7}$$

In Equation 6.7, the first and second terms represent M's spending on the queries that it issued corresponding to *partial_use_price* and *full_use_price* respectively. The third and fourth terms relate to M's spending due to purchases of licenses (i.e., PUL and FUL). The fifth term represents M's spending due to relay commissions.

Hence, using Equations 6.6 and 6.7, the revenue ω of M is computed below:

$$\omega \;=\; E_M - S_M \tag{6.8}$$

6.3 Economy-based Replication schemes for E-Rare

This section discusses two economy-based replication schemes, namely **ECR** and **ECR+**, for improving rare data availability. They are based on the incentive-model discussed in the previous section.

[1] A successful query is one for which M receives the query results before the deadline. For unsuccessful queries, M does not spend any currency.

6.3.1 ECR: Individual-based replication scheme

In ECR, each MP M autonomously decides the items to host at itself on a periodic basis. These items could be either the items that M owns or the items for which it sees high demand (as in case of rare items) based on the messages that it relays. M tries to obtain such high-demand items from its neighbors. Thus, this method helps in combating free-riders by attracting them to host items and earn incentives. Initially, when the system starts, for uniformity, reduced replication-related overhead and later for performance comparison purposes, we initialize the replication period, which is the same for all the MPs. Thus, the replication period is independent of data items.

Algorithm 6.1 ECR: Algorithm for an MP M

```
begin
    /* MEM is an M's memory space */
    /* TH is a rarity score threshold for M */
    /* λ_i is a rarity score for data item i */
(1) Receive broadcasted list B_R of the data items from M's neighbours
(2) Merge B_R with M's own list of available data items in A_R
(3) Sort A_R in descending order of data items' rarity scores

(4) for each item i in A_R
(5)     if MEM > 0 and λ_i > TH
(6)        Store i in MEM
(7)        MEM = MEM - sizeof[i]
(8)        Add i to purchased list P_R
(9)     else
(10)       break

(11) for each item i in P_R
(12)     Pay partial/full use price of i to its sender-MP

end
```

For determining which data items to host at itself, M autonomously sets a rarity threshold score TH_R. (Thus, TH_R can vary across MPs.) M computes TH_R as an average rarity score of the rare items that it currently hosts. M proceeds to fill up its available memory space by first sorting its own items in descending order of their rarity scores and hosting only those items, whose rarity scores exceed TH_R. Then, if M has available memory space, M creates a list of items, for which it has seen high demand (based on its intercepted relay messages). M sends a message to its neighbors to enquire whether they have some of these items and the associated item rarity scores. Upon receiving replies from its neighbors, M tries to replicate at itself only those items, whose rarity scores exceed TH_R, by paying either of the license price types to the corresponding neighbor(s). M's remaining memory space (if any) is then progressively filled up one-by-one with its own items based on descending order of their rarity scores.

Discussion on ECR

Note that resource constraints include memory space and energy of the mobile devices, and ECR uses an incentive-based replication mechanism, where peers earn currency from items that are downloaded from them. This facilitates efficient allocation of limited available memory space for replicas among the MPs because the peers are incentivized to host items (or replicas) that are more likely to maximize their revenues.

Furthermore, our proposed model requires a query-issuing peer to pay a constant commission to each relay MP in the successful query path from which it eventually downloads the data, thereby enticing them to forward queries quickly. Since sending and receiving messages tax the limited energy resources of the mobile peers, this addresses the energy constraint by ensuring that the peers preserve their energy by forwarding the important messages that are associated with a higher possibility of downloads.

As such, we do not handle node mobility explicitly. However, in our simulations, we model node-mobility in terms of the Random Waypoint (RWP) Model appropriate for our application scenarios such as adventure tourists (or archaeologists) moving randomly.

Note that the deletion of items (or replicas) at a peer is autonomous. A peer does not necessarily have to delete items, whose access count falls below a certain threshold. For example, if the item is rare and thus higher-priced, a peer may still decide to continue hosting it in the expectation of earning high amount of revenues when the rare item gets accessed due to the occurrence of some rare event. Observe that the hosting of rare items is important to the network as a whole for maintaining the data availability when a sudden burst of queries comes in for the rare items.

6.3.2 ECR+: Group-based replication scheme

Now we shall discuss ECR+, which extends the ECR scheme by incorporating the notion of peer groups for improving the availability of rare data items in E-Rare.

We define a **peer group** as a set of MPs working together such as an adventure tour expedition group. Notably, group members need not necessarily be one-hop neighbors i.e., they may be scattered across the network due to peer mobility. For the sake of convenience, we shall henceforth refer to a peer group as a **group**. As we shall see shortly, MPs provide *discounts* only to other MPs within their group to incentivize MP participation in the group. As such, group formation schemes are outside of the scope of this work. Notably, existing group formation schemes such as MobilisGroups [LSS11] and Team-Formation [AMP98, HL05] schemes can be used in conjunction with our proposal.

Group members periodically broadcast their list of items to members within their group. We assume that these broadcast messages are received by all the MPs within the group. Each MP M's broadcast message constitutes a list, which contains entries of the form $\{MP\_id, data\_id, \lambda_d, price, acc\_count\}$, where MP_id is the unique identifier of M, $data\_id$ is the identifier of the data item d that it hosts, λ_d is the rarity score of d, $price$ is the price of d, and acc_count is the average access count of d at M over the last N time-periods. Notably, as we shall see shortly, MP_id and $data\_id$ facilitates MPs in determining the number of group members that host a given item d. Furthermore, the rarity score guides the MPs in replicating rare items. Additionally, the price and access count information for each item facilitates replication by guiding the MPs in evaluating the revenue-earning potential of each item.

Given that nodes in a group are scattered across the network, messages between group members will often pass through non-group members too. Thus, when the message packets hop through the network, the intermediate non-group nodes can also see the data items and the associated hosts from the packets. This facilitates the discovery of members of the groups (and the rare items that they host) by peers, which are outside of these groups. However, we do not assume that each peer has, at any point of time, complete information about all the data items when they send out queries.

Incidentally, the periodic message exchanges among the MPs to share information about the items that they host do not matter in the calculation of revenue. These messages are sent periodically as status messages, as required by our proposed ECR+ scheme for keeping the peers informed about the information hosted at other peers. Since every peer incurs this cost of sending these messages and every peer sends a comparable number of such messages, it basically neutralizes (i.e., cancels out) and it does not have any relative effect in the calculation of peer revenues.

Computation of the rarity score λ_d in ECR+

For computing the rarity score λ_d of a data item d in ECR+, we extend Equation 6.1 by additionally considering the number ξ of MPs (group members) that host d. A given MP is able to compute the value of ξ because it knows how many other MPs host d within its group due to the periodic broadcast messages, in which each MP includes the list of data items that it hosts[2]. This is in contrast with the case of ECR, where a given MP cannot compute the value of ξ due to the lack of such broadcast messages. Thus,

[2] Recall our assumption that the broadcast messages are received by all the MPs within the group.

ECR+ computes λ_d based on more information than ECR. However, since it is also possible for members outside the group to host d, ξ is essentially an *approximate* quantification of how many MPs host d. Thus, our computation of ξ represents an inevitable compromise for defining rarity in the absence of complete information in decentralized settings. In ECR+, each MP computes the rarity score λ_d for each data item d (that it hosts) as follows:

$$\lambda_d \;=\; [\,\{\,(\,\eta_c - (\,\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\eta_i\,)\,)\;/\;(\,max\,(\,\eta_c\,,\,\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\eta_i\,)\,\times\,\xi\,)\,\}\,+\,1\,]\,/\,2 \tag{6.9}$$

where N is the number of time-periods over which λ_d is computed. Here, η_c refers to the access count of data item d for the current period, while η_i represents the access count of d for the i^{th} time-period. Similar to the case of ECR, the value of λ_d may differ among the MPs for the same data item since it is associated with sudden bursts at each MP. Thus, the value of λ_d for a given item may differ across group members in ECR+. Furthermore, as in Equation 6.1, observe that the range of λ_d in Equation 6.9 is between 0 and 1.

6.3.3 Illustrative example of peer groups in E-Rare

Figure 6.1 depicts an illustrative example of an instance of network topology in ECR+. Now we shall use Figure 6.1 to illustrate how groups facilitate the improvement of rare data availability. In Figure 6.1, the groups {P1, P4, P8, P10, P11, P15, P18}, {P2, P6, P12, P14}, {P3, P9, P13, P20} and {P5, P7, P16, P17, P19} are shown in different colors. Observe that group members need not be one-hop neighbours e.g., P1 and P10 are not one-hop neighbours.

Suppose peer P18 sees high access count for an item d, which it does not own or host. Additionally, suppose d is owned and hosted by one of its group members, say P1. For simplicity, assume that no replica of d exists at any

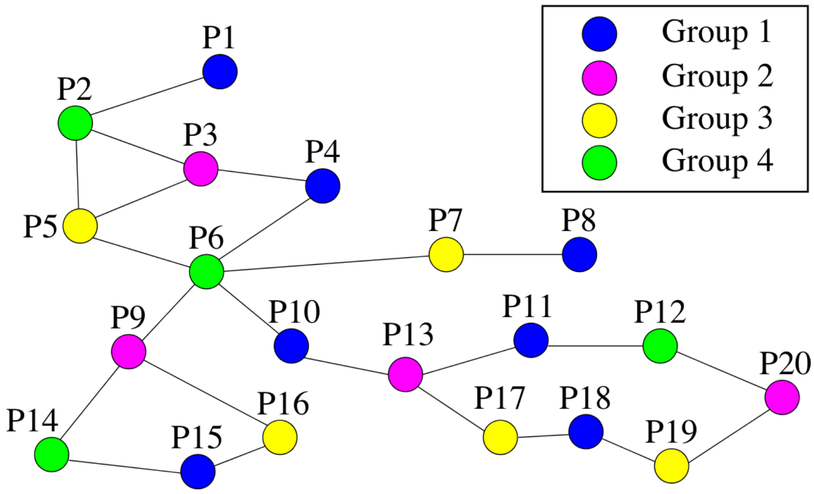


Figure 6.1: Peer groups in E-Rare

other peer in the M-P2P network[3]. In this scenario, queries on d initiated nearby P18 may fail due to exceeding the TTL (in terms of the maximum number of hops allowed for a query) because of the distance from P1, which hosts the queried item d. Furthermore, queries may also exceed the query deadline time due to incurring high query response times. Observe that this is likely to decrease M-P2P data availability.

Now suppose P18 licenses d from P1 and hosts d at itself. (Notably, P18 knows that d is owned by P1 because group members periodically exchange messages to share information about the items that they own and/or host.) Thus, subsequent queries on d, which are initiated nearby P18, can either be *locally* served by P18 if response time is an issue or served by P1 if price is an issue. Notably, P1 has an incentive to license d to P18 because it can earn currency from the license price. Furthermore, P18 has an incentive to license d from P1 because it can earn currency by serving queries on d, which it obtains at a discounted price. Herein lies the motivation for licensing among group members.

[3]For rare items, relatively few replicas exist in the network.

6.3.4 Discounts for group members in ECR+

For effective incentivization within a group, ECR+ incorporates the notion of *discounts*, which pertains to all the four price types that were previously discussed in Section 6.2. MPs provide discounts only to other MPs within their group, hence the notion of discounts acts as an incentive towards MP participation in a group. A group member that sees relatively high access count for a data item d, which is not hosted at itself, can obtain licenses for d at a *discounted price* from any of its group members owning d. Given that group members may be scattered across the network, such licensing among group members brings the data closer to the source of the queries, thereby resulting in faster query response times, improved rare data availability and reduced query-related communication overhead.

In ECR+, the incentive for MPs to join a group is quantified by the discount δ. If the value of δ is too high, MPs hosting rare items would be reluctant to join the group. This is because their revenue-earning potential would decrease due to reduced earnings because of relatively high discounts. However, MPs querying for the rare items would be incentivized to join the group because they can obtain their desired items at lower prices due to discounts. On the other hand, if the value of δ is too low, MPs hosting rare items would have better incentive to join the group because of increased revenue-earning potential from license prices. However, MPs querying for the rare items would have lower incentive to join the group due to lower discounts. Observe that when $\delta = 0$, the effect of discounts is nullified.

In effect, when the value of δ is too high or too low, rare data availability is not maximized due to reduction in the incentivization effect of groups. Hence, we shall experimentally determine suitable values of δ for maximizing rare data availability in Section 3.5.

Recall that δ applies to all the four item price types. We designate the **discounted** *partial_use_price* $P_{d,Q}$, *full_use_price* $F_{d,Q}$, *partial_use_license_price* $PUL_{d,Q}$ and *full_use_license_price* $FUL_{d,Q}$ as $DP_{d,Q}$, $DF_{d,Q}$, $DPUL_{d,Q}$

and $DFUL_{d,Q}$ respectively. Hence, to incorporate the effect of discounts, we extend Equations 6.2, 6.3, 6.4 and 6.5 (see Section 6.2) as follows:

$$DP_{d,Q} = P_{d,Q} \times (1-\delta) \tag{6.10}$$

$$DF_{d,Q} = F_{d,Q} \times (1-\delta) \tag{6.11}$$

$$DPUL_{d,Q} = PUL_{d,Q} \times (1-\delta) \tag{6.12}$$

$$DFUL_{d,Q} = FUL_{d,Q} \times (1-\delta) \tag{6.13}$$

where $0 \leqslant \delta < 1$.

6.3.5 Group-based data licensing in ECR+

In ECR+, group-based data licensing can be facilitated in two ways. MPs with adequate resources (e.g., energy, bandwidth, memory space) can request for rare items from group members so that they can earn currency by hosting and serving queries on those items. This type of licensing provides incentives to free-riders towards hosting replicas of rare items. This is because free-riders need to earn currency, without which they would not be able to issue any requests of their own.

In contrast, MPs owning rare items can also off-load their items to group members in the network for licensing purposes. An MP may use this mechanism for licensing when its resources, such as energy or bandwidth, are not adequate to serve queries on its owned items. Moreover, an MP may use this when it is about to leave the network. In this manner, an MP can earn currency from its items by means of licensing even if it becomes offline. This type of licensing also provides incentives to MPs towards replicating their items.

Interestingly, both these mechanism of licensing facilitate replication of rare data from owners to free-riders, thereby improving rare data availability. In the absence of a licensing mechanism, rare items would become inaccessible once their owners run out of energy (or leave the network), thereby reducing rare data availability.

Algorithm 6.2 ECR+: Algorithm for *Licensor_ MP M*

begin

/* ϕ is an item's revenue-earning potential */
(1) Sort all its items in descending order of ϕ
(2) Compute the average value ϕ_{avg} of all its items
(3) Select items for which ϕ exceeds ϕ_{avg} into a list Lic
/* Lic is the set of items for licensing */

(4) for each item i in Lic
(5) Decide $A[i]_{PUL}$ and $A[i]_{FUL}$ for i
/* $A[i]_{PUL}$ and $A[i]_{FUL}$ are number of available PUL and FUL licenses of i*/

(6) Broadcast the list Lic upto its n-hop neighbours

(7) for each item i in Lic
(8) Wait for replies from potential licensees
(9) Receive replies from potential licensees

(10) for each potential licensee j
(11) Calculate the value of Ω for j

(12) Sort the licensees in descending order of Ω into a list P_L

(13) for each licensee j in P_L
(14) if ($A[i]_{PUL} = 0$) **break**
(15) if ($A[i]_{PUL} - N[j]_{PUL} \geqslant 0$)
(16) Send $N[j]_{PUL}$ licenses of i to j
(17) $A[i]_{PUL} = A[i]_{PUL} - N[j]_{PUL}$
(18) else
(19) Send $A[i]_{PUL}$ licenses of i to j
(20) $A[i]_{PUL} = 0$

(21) for each licensee j in P_L
(22) if ($A[i]_{FUL} = 0$) **break**
(23) if ($A[i]_{FUL} - N[j]_{FUL} \geqslant 0$)
(24) Send $N[j]_{FUL}$ licenses of i to j
(25) $A[i]_{FUL} = A[i]_{FUL} - N[j]_{FUL}$
(26) else
(27) Send $A[i]_{FUL}$ licenses of i to j
(28) $A[i]_{FUL} = 0$

end

Figure 6.2 depicts the algorithm for a licensor MP M. In Lines 1-3, observe how M selects the items with higher revenue-earning potential ϕ for licensing. This is because such items better incentivize potential licensees towards item hosting because they can earn higher amount of revenue by hosting these items. Here, ϕ is computed as the product of item access count and item price[4]. Note that rare items will have higher revenue-earning poten-

[4]Since E-Rare considers four types of item prices, the respective products of access

Algorithm 6.3 ECR+: Algorithm for *Licensee_MP* M_E

begin

(1) Receive broadcast message from potential licensor M
/* Broadcast message contains the item set *Lic* for licensing */

(2) Sort all items in *Lic* in descending order of ϕ
/* ϕ is an item's revenue-earning potential */

(3) for each item i in *Lic*
/* *Spc* is the peer's available memory space */
(4) while $Spc > 0$
/* $size_i$ is the size of i */
(5) if ($size_i \leq Spc$)
(6) Add i to a set *Acq*
(7) $Spc = Spc - size_i$

(8) for each item i in *Acq*
(9) Decide N_{PUL} and N_{FUL} for i
/* N_{PUL} and N_{FUL} are required number of PUL and FUL licenses of i */

(10) for each item i in *Acq*
(11) Send bid to M with details of energy, hop-distance, N_{PUL} and N_{FUL} to M
(12) Wait for reply from M
(13) if (bid is successful)
(14) Obtain item i from M (with corresponding licensing rights)
(15) Send payment to M

end

tial because their prices are higher than that of non-rare items. Moreover, rare items have high access counts during periods of sudden burst. Recall that we consider a cooperative environment where all the mobile peers are trusted entities. In such cooperative and trusted environments, peers would be truthful about revealing their access counts on every data item.

As indicated in Lines 4-5, M autonomously decides the number of PUL and FUL licenses that are to be made available for each item. This work considers peer autonomy in determining the values of A_{PUL} and A_{FUL}, hence MPs are allowed to autonomously decide the number of licenses that they want to make available. In Line 6, the broadcast message also contains the values of A_{PUL}, A_{FUL} and the (discounted) prices for each item in *Lic*. This information facilitates potential licensees in determining whether to obtain license(s) for a given item.

---

counts and each price type are summed up to obtain the value of ϕ.

As Lines 10-12 indicate, ECR+ prefers potential licensees with higher value of Ω. Here, Ω quantifies the quality-of-service potential of licensees (that bid for hosting the items). Thus, MPs with higher values of Ω would be likely to provide better service in terms of improving rare data availability. Ω is computed as below:

$$\Omega \;=\; [\,w_1 \times energy\,] \;+\; [\,w_2 \times n_{hop}\,] \tag{6.14}$$

where $energy$ and n_{hop} are the potential licensee's energy level and its distance from M (in terms of hop-counts). As $energy$ increases, Ω increases because higher-energy MPs are more likely to provide better data availability. Ω also increases with increase in n_{hop} because licensing a given item to an MP, which is located at a farther distance from M, is likely to better spread the item across the region, thereby improving data availability. Furthermore, M prefers to license its items to MPs that are farther way to reduce competition. In other words, if M licenses its items to nearby MPs, the accesses for those items would get divided between M and those MPs, thereby resulting in reduced revenues for M due to competition. In Equation 6.14, w_1 and w_2 are weight coefficients such that $w_1, w_2 \geqslant 0$ and $w_1 + w_2 = 1$. In this work, for simplicity, we set $w_1 = w_2 = 0.5$.

As Lines 13-28 indicate, M distributes PUL and FUL licenses for each item to potential licensees, starting from those with higher values of Ω until its number of available PUL and FUL licenses becomes zero.

Figure 6.3 depicts the algorithm for a licensee MP M_E. In Lines 1-2, upon receiving the broadcast message from licensor M, M_E sorts the items in the broadcast message in descending order of their revenue-earning potential ϕ. As Lines 2-7 suggest, M_E prefers items with higher revenue-earning potential ϕ because it can earn more revenue by hosting such items per unit of its memory space since its memory space is limited. Thus as Lines 3-7 indicate, M_E greedily *simulates* the filling up of its memory space by items with higher value of ϕ.

As Lines 8-9 suggest, M_E autonomously decides the number of PUL and FUL licenses that it wants to acquire for each item. This work considers peer autonomy in determining the values of N_{PUL} and N_{FUL}, hence MPs are allowed to autonomously decide the number of licenses that they want to acquire. Furthermore, in case M_E does not have adequate currency to make the payment, it is allowed to make the payment after it has earned currency by hosting these items. Observe that allowing deferred payments can be justified by the fact that potential licensors and licensees are members of the same group. Hence, if a licensee fails to make the payment within a reasonable time-frame, it would risk getting removed from the group. This policy of allowing deferred payments allows free-riders, which may initially not have enough currency to acquire licenses for items, to seamlessly integrate into participating in the network.

In Lines 10-15, M_E sends its bid to the corresponding licensor M for each of its desired items along with details of its energy, distance (hop-counts) from M, N_{PUL} and N_{FUL} to M. For those items, concerning which M_E's bid is successful, M_E obtains the items with corresponding licensing rights from M and pays the (discounted) license prices of these items to M. In case M_E does not have adequate currency to pay M, it informs M about a deadline time by which it would make its payment.

6.3.6 Illustrative example of licensing in ECR+

Figure 6.2 depicts an illustrative example of licensing in ECR+. From Figure 6.2a, observe how the items are sorted in order of revenue-earning potential ϕ and only the items above the average value of ϕ are selected to be licensed by licensor MP M. Figure 6.2b depicts the license set *Lic* comprising items $\{36, 92, 53\}$ along with the number of available PUL and FUL licenses for each item in *Lic*. Figure 6.2c indicates the number of PUL and FUL licenses demanded by each of the MPs corresponding to each item. For simplicity, suppose the list of potential licensees in descending order of Ω is as follows:

ID – Unique identifier of data item	**P1, P2, P3** – Mobile Peers	$\mathbf{N_{PUL}}$ – Demanded PUL licenses
φ – Revenue-earning potential	**Ω** – Potential of licensee	$\mathbf{N_{FUL}}$ – Demanded FUL licenses
φ_{avg} – Average value of **φ** across all items of licensor	$\mathbf{A_{PUL}}$ – Available PUL licenses	$\mathbf{S_{PUL}}$ – Supplied PUL licenses
	$\mathbf{A_{FUL}}$ – Available FUL licenses	$\mathbf{S_{FUL}}$ – Supplied FUL licenses

ID	φ
36	2000
92	1600
53	1200
09	800
21	40
84	80

$\varphi_{avg} = 953.33$

(a) Licensor's item set

ID	φ	A_{PUL}	A_{FUL}
36	2000	15	5
92	1600	20	3
53	1200	25	3

(b) Item set to be licensed

		P1		P2		P3	
ID	φ	N_{PUL}	N_{FUL}	N_{PUL}	N_{FUL}	N_{PUL}	N_{FUL}
36	2000	0	0	15	5	15	5
92	1600	15	3	8	2	10	1
53	1200	6	0	11	2	0	0

(c) Required number of licenses by the licensees

List of potential licensees in descending order of Ω = { P1, P2, P3 }

		P1		P2		P3	
ID	φ	S_{PUL}	S_{FUL}	S_{PUL}	S_{FUL}	S_{PUL}	S_{FUL}
36	2000	-	-	15	5	0	0
92	1600	15	3	5	0	0	0
53	1200	11	2	6	0	-	-

(d) Supplied number of licenses to the licensees

Figure 6.2: Illustrative example of licensing in E-Rare

{P1, P2, P3}.

Figure 6.2d shows the number of supplied licenses to each MP. Observe that P1 does not demand any PUL licenses for item 36, hence M iterates to the MP with the next highest value of Ω i.e., the MP P2. Since P2 demands 15 PUL licenses for item 36 and M has 15 available PUL licenses for this item, M sends all 15 licenses to P2. Now since M has no more available PUL licenses for item 36, the MP P3 with the next highest value of Ω receives no PUL licenses, although it demanded 15 PUL licenses.

For item 92, the total number of available PUL licenses is 20, and P1 demands 15 PUL licenses. Thus, M gives 15 PUL licenses to P1. Now observe that P2's demand is for 8 PUL licenses, while the current number of available PUL licenses is now only 5 (because the other 15 licenses have already been assigned to P1). Hence, P2 acquires only 5 PUL licenses for item 92, although it originally demanded 8 PUL licenses for this item. Furthermore, since there are now no more remaining available PUL licenses for item 92, P3 is not able to acquire any licenses for this item. Notably, although we explained this illustrative example using PUL licenses, the explanation for FUL licenses is essentially similar.

Notably, our proposed algorithms in ECR+ do not have a notion of optimum selection because we are basically using heuristics. We have provided possible algorithms for achieving our purpose, but as such, we do not make any claims concerning optimality.

6.4 Performance Evaluation of E-Rare

This section reports the performance of our incentive-based replication schemes by means of simulation using OMNET++ [Pon93]. We assume that MPs move according to the *Random Waypoint Model* [BMJ$^+$98] within a region of area 1000 metres $\times$ 1000 metres. The *Random Waypoint Model* is appropriate for our application scenarios, which generally involve random movement of users such as adventure tourists looking for information about gas-masks and associated safety equipment in an unfamiliar forest. Our experiments use a total of 150 MPs. The default communication range of all MPs is a circle of 120 metre radius. Table 6.3 summarizes the parameters used in our performance evaluation.

Recall that E-Rare considers three classes of items (i.e., *rare*, *medium-rare* and *non-rare*) based on item rarity score λ_d, and each item class is associated with a range of rarity scores. For *rare* items, $0.7 \leq \lambda_d \leq 1$; for *medium-rare* items, $0.5 \leq \lambda_d < 0.7$; and for *non-rare* items, $0 < \lambda_d < 0.5$. The number of items in each of these classes is determined using a Zipf distribution with zipf factor ZF_D over three buckets, each bucket corresponding to one of the rarity classes. Notably, we set the default value of ZF_D to 0.7 (i.e., high skew) to ensure that the majority of the items in the network are rare in that they will be assigned relatively high rarity scores. Thus, for each item d, we randomly assign its value of λ_d based on the lower and upper bounds

Parameter	Default value	Variations
No. of MPs (N_{MP})	150	30, 60, 90, 120
Zipf factor for distribution of rare data (ZF_D)	0.7	0.1, 0.3, 0.5, 0.9
Zipf factor for distribution of queries across rarity classes (ZF_Q)	0.7	0.1, 0.3, 0.5, 0.9
Zipf factor for distribution of MPs across interest groups (ZF_G)	0.5	0.1, 0.3, 0.7, 0.9
Communication Range (CR)	120 m	40 m, 80 m, 160 m, 200 m
Discount (D)	30%	10%, 20%, 40%, 50%
Access count threshold for determining timing of initiating replication (f_{TH})	0.5	0.1, 0.3, 0.7, 0.9
Percentage of MP failures (P_F)	20%	10%, 30%, 40%, 50%
Queries/second	10	
Bandwidth between MPs	28 Kbps to 100 Kbps	
Size of a data item	250 Kb to 1.75 Mb	
Memory space of each MP	5 MB to 25 MB	
Speed of an MP	1 metre/s to 10 metres/s	
Size of message headers	220 bytes	

Table 6.3: Parameters of our performance evaluation for E-Rare

of its item class.

Rare items are assigned to 1-2 MPs, *medium-rare* items are assigned to 3-4 MPs, and *non-rare* items are assigned to 5-6 MPs. Thus, given a data item d, we first examine its class to determine the number of MPs to which d should be assigned. For example, if an item is *medium-rare*, it will get assigned to N MPs (Here, N is either 5 or 6, as determined by a random number generator.) Now a set of N MPs will be randomly selected from among those MPs that have adequate memory space for replication, and d will be assigned to these MPs. Observe that since item sizes vary, the available memory space for replication will vary across MPs over time.

Each query is a request for a single data item. 10 queries per second are issued in the network. Items to be queried are randomly selected from all the items in the entire network. The query-issuing MP is selected randomly from among all the MPs in the network, the constraint being that an MP cannot issue a query for an item already hosted at itself. The number of queries directed to each class of items (i.e., *rare*, *medium-rare* and *non-rare*) is determined by a Zipf distribution with a zipf factor ZF_Q. We set the default value of ZF_Q to 0.7 to ensure that a relatively high percentage of queries are directed towards rare items. This is consistent with our application scenarios, which involve sudden bursts in accesses to rare items. Furthermore, recall that queries in E-Rare are associated with one of the following prices, namely *partial_use_price* $P_{d,Q}$, *full_use_price* $F_{d,Q}$, *partial_use_license_price* $PUL_{d,Q}$ and *full_use_license_price* $FUL_{d,Q}$. The percentage of queries corresponding to $P_{d,Q}$, $F_{d,Q}$, $PUL_{d,Q}$ and $FUL_{d,Q}$ are 30%, 30%, 20% and 20% respectively. Thus, each query is randomly associated with one of the price types.

For our proposed peer group-based economic scheme ECR+, we use 10 groups for our experiments. We determine the number of MPs in each group by using a Zipf distribution with a zipf factor ZF_G over 10 buckets. Thus, the number of MPs vary across groups. Hence, in our experiments, we have varied the value of ZF_G to study the impact of variations in group sizes on the performance of ECR+. The MPs are randomly assigned to the groups. Furthermore, an MP is assigned to only one peer group to ensure that all groups are mutually disjoint.

The timing of initiation of replication can have significant impact on the performance of our proposed approaches. If replication is initiated early based on a relatively small number of queries for an item, it may result in relatively non-rare items getting replicated. Consequently, data availability would degrade because the rare items would not have a chance to get replicated due to memory space constraints at the MPs. On the other hand, if replication is initiated late after looking at a relatively large number of queries for an item, data availability may suffer because the delay in initiating replication could make the overall impact of replication much less pronounced. This is because a significant number of query failures could already have occurred before replication had been initiated. Hence, we introduce the access count threshold f_{TH}, which quantifies the time when replication is initiated.

We define f_{TH} as follows: $f_{TH} = R_q/T_q$, where R_q is the number of issued queries after which replication had been initiated and T_q is the total number of queries. The total number of issued queries in our experiments is 10,000. If replication had been initiated after the first 1000 queries had been issued in the system, the value of f_{TH} would be $(1000/10000) = 0.1$.

Our performance metrics are **average response time (ART)** of a query, query **success rate (SR)**, **hop-count (HC)** of a query and communication cost in terms of total number of **messages (MSG)**. ART equals ($(1/N_Q)\ \sum_{i=1}^{N_Q}(T_f\ -\ T_i)$), where T_i is the query issuing time, T_f is the time of the query result reaching the query issuing MP, and N_Q is the total number of queries. ART includes data download time, and is computed only for successful queries. Notably, unsuccessful queries die after TTL ('hops-to-live') of 6 hops. (Preliminary experiments suggested that TTL $=$ 6 is a reasonable value for our application scenarios.) Since a relatively high percentage of queries are directed towards rare items (which are hosted at relatively few MPs), queries can fail due to the TTL criterion. Queries can also fail due to MPs running out of energy or due to network partitioning.

The query success rate SR equals ($(N_S/N_Q)\ \times\ 100$). We define the query hop-count HC as the average hop-count incurred by the query in the successful query path. Thus, HC equals ($(1/N_Q)\ \sum_{i=1}^{N_Q} HC_i$) and is measured only for successful queries. MSG equals ($\sum_{i=1}^{N_Q} MSG_i$), where MSG_i is the total number of messages during the course of the experiment.

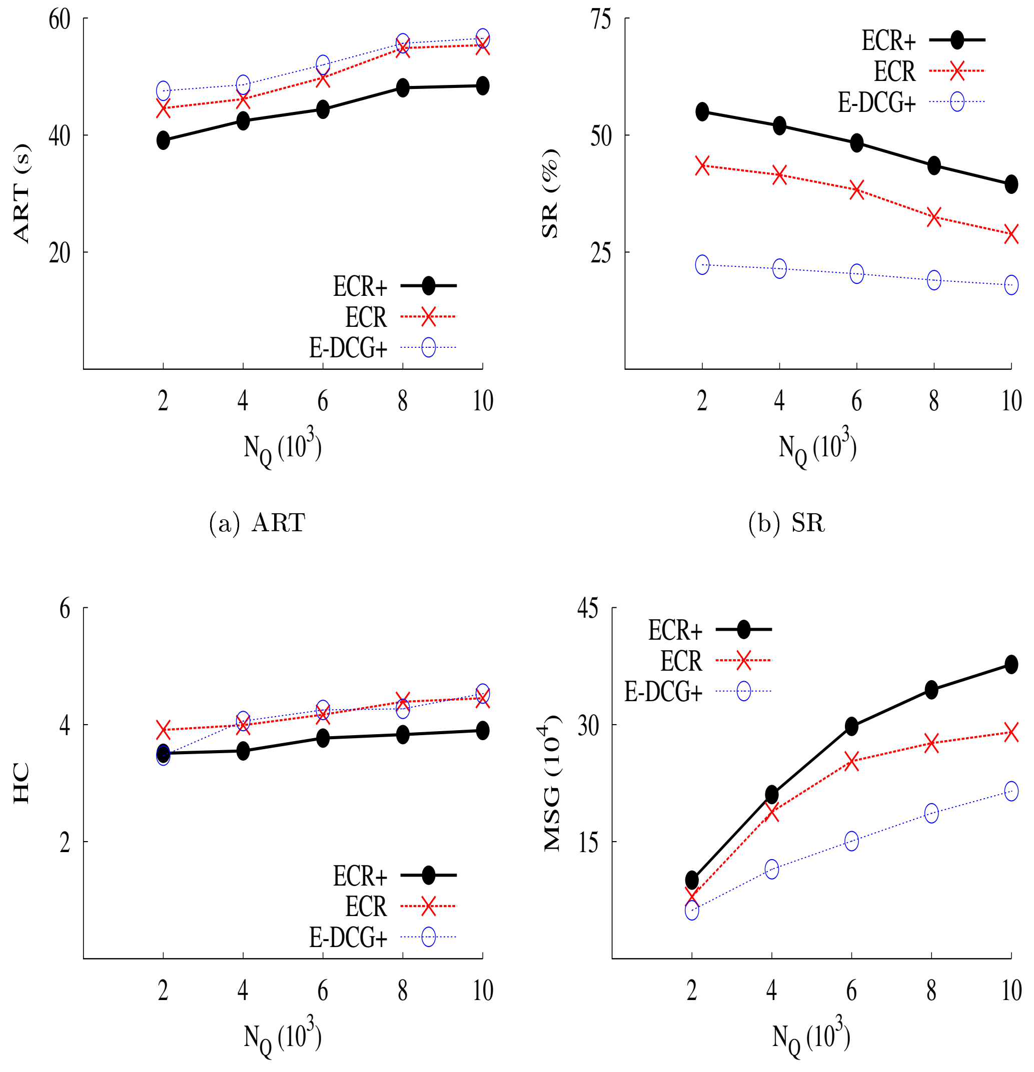
Incidentally, none of the existing proposals for M-P2P networks address economic incentives towards replication of rare data items. We compared our proposed incentive-based E-Rare schemes with an existing non-incentive E-DCG+ scheme for MANETs, proposed in [HM06], to our scenario. E-DCG+ is a non-incentive and non-economic replication scheme, and it does not provide incentives for replica hosting. E-DCG+ is executed at every replica allocation period. E-DCG+ is the closest to our scheme since it addresses replication in mobile ad-hoc networks. Furthermore, we believe that E-DCG+ is among the best approaches for meaningful performance comparison with our proposed schemes because it is the most recent approach and it has already been compared to other non-incentive schemes. Moreover, E-DCG+ does not incorporate the notion of licensing mechanism to distribute rare data items in mobile environment.

We have implemented E-DCG+ in E-Rare as follows. E-DCG+ performs the periodic broadcast to perform replication. MP obtains the data items list with their respective rarity scores. Based on rarity scores of data items and MP's available memory space, each MP hosts data items in their decending order of rarity scores till memory space becomes full. Here, MP does not obtain any incentives to host replicas, hence E-DCG+ provides freedom to MPs, whether they want to host new data items or to revise hosted data items. For the sake of experiments, we have set the MP's decision probability to host the data items to 0.7 with relocation period of 200 seconds.

Notably, in case of ECR+, group members exchange messages *periodically* every 200 seconds to inform each other concerning the items that they host. For all the approaches, querying proceeds by means of broadcast using AODV protocol.

6.4.1 Performance of E-Rare

Figure 6.3 depicts the results of our experiments using default values of the parameters in Table 6.3. For all the approaches, ART and HC increase over time, while SR decreases over time. This occurs because as more queries are answered, the energy of MPs keeps on decreasing, thereby resulting in an increasing number of MPs running out of energy. This results in longer query paths to data items, or data items becoming inaccessible. Furthermore, a relatively high percentage of queries are directed towards rare items (due to the zipf factor ZF_Q being set to 0.7), and each of these rare items are initially hosted only by 1-2 MPs. This causes the MPs hosting the rare items



(a) ART (b) SR

(c) HC (d) MSG

Figure 6.3: Performance of E-Rare: Data availability and communication overhead

to become overloaded, thereby resulting in increased query waiting times in their job queues, and this further contributes to increase in ART.

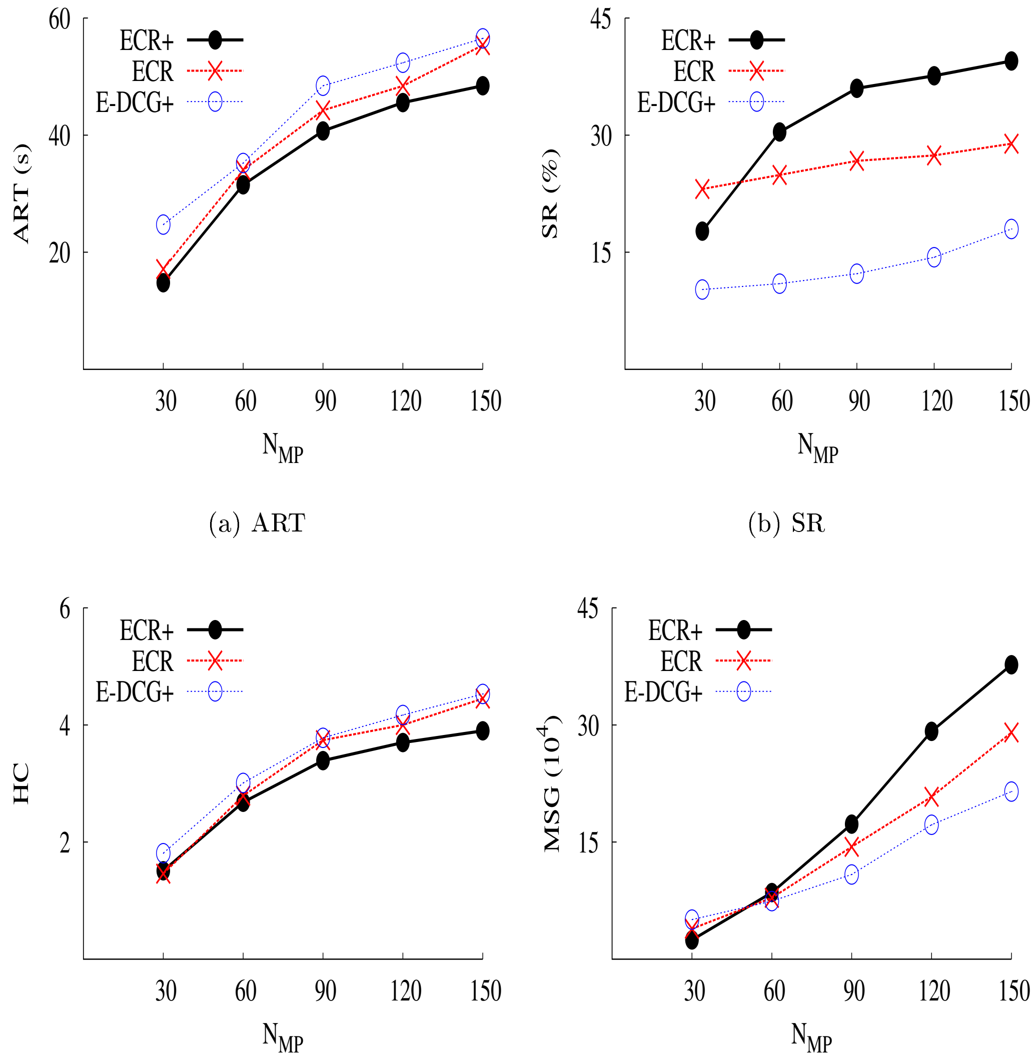
ECR outperforms E-DCG+ in terms of ART, SR and HC essentially due to its economic licensing scheme, which incentivizes MP participation in the creation of multiple copies of rare items. Increased MP participation also implies more opportunities for replication, more memory space for hosting replicas and multiple paths for locating a data item/replica. In contrast, since E-DCG+ considers neither any economic scheme nor any licensing mechanism, it does not facilitate replication. Thus, rare items become inaccessible when their host MPs run out of energy, thereby explaining the reason for SR being significantly lower for E-DCG+ as compared to that of ECR.

ECR+ outperforms ECR due to its group-based incentives (such as *discounts*), which facilitate collaborative replication among MPs. Such collaborative replication enables better spreading of the copies of frequently requested rare items throughout the network, thereby improving the probability of obtaining queried rare items within relatively fewer hops. Interestingly, the results in Figure 6.3c suggest that although HC follows a pattern similar to ART, some deviations occur. These deviations occur essentially due to bandwidth differences at MPs.

As the results in Figure 6.3d indicate, MSG increases over time for all the approaches due to more queries being answered. (Recall that MSG is the total number of messages during the course of the experiment.) Observe that after the first 6000 queries have been processed, MSG does not keep increasing linearly for ECR and ECR+. This is because depletion of the energy of some of the MPs implies that in effect, queries get forwarded to a reduced number of MPs. Moreover, ECR+ exhibits higher MSG than ECR due to additional messages for group interactions.E-DCG+ incurs least MSG due to a large percentage of unsuccessful queries (as suggested by the results in Figure 6.3b), which result in decreased amount of data transfer, albeit at the cost of reduced SR.

6.4.2 Effect of variations in the number of MPs

To test E-Rare's scalability, we varied the total number N_{MP} of MPs, keeping the number of queries proportional to N_{MP}. Figure 6.4 depicts the results. As N_{MP} increases, ART increases for all three approaches due to increase in network size. As N_{MP} increases, SR increases for both ECR and ECR+ due to increased number of copies of rare data items because of more replication opportunities provided by a larger number of MPs. With increasing value of



(a) ART (b) SR

(c) HC (d) MSG

Figure 6.4: Effect of variations in the number of MPs

N_{MP}, HC follows similar pattern as that of ART for all the three approaches essentially due to increase in network size. The pattern of HC deviates slightly from that of ART due to bandwidth differences at MPs. For all the three approaches, MSG increases with increasing value of N_{MP} because larger network sizes incur higher number of messages.

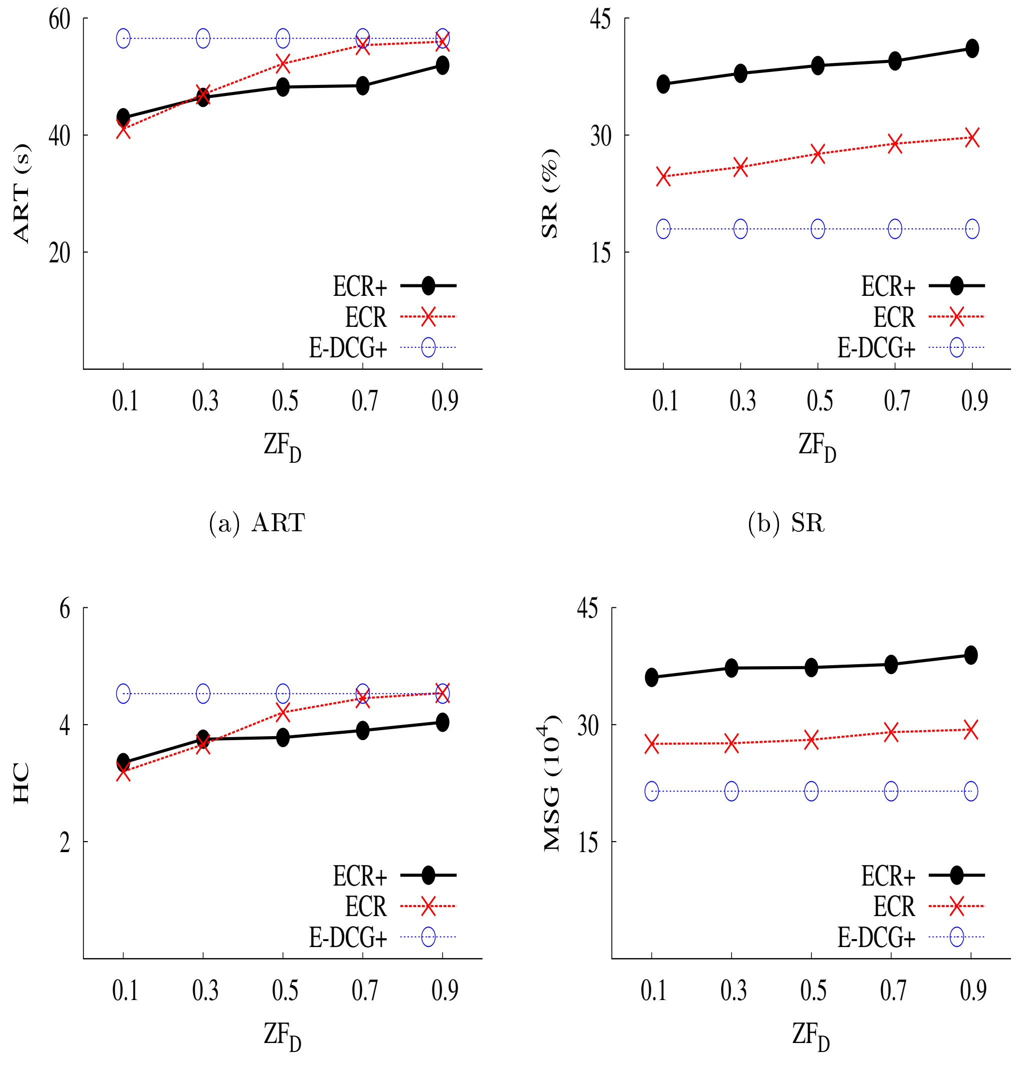
ECR+ performs better than ECR due to the reasons explained for Figure 6.3. Furthermore, observe that as N_{MP} increases, the performance gap between ECR+ and ECR in terms of SR also increases. This occurs because as N_{MP} increases, group sizes also increase, thereby making the effect of group-based collaborative replication performed by ECR+ more pronounced. The eventual plateau in SR for ECR+ occurs because SR is upperlimited by the number of copies of rare data items in a group due to memory space constraints of group members.

6.4.3 Effect of variations in the data distribution across rare item classes

Figure 6.5 shows the results of the effect of variations in the data distribution across rare item classes. Recall that ZF_D is a zipf factor for the Zipf distribution of data items in three different data classes i.e., *non-rare*, *medium-rare* and *rare*. Higher values of ZF_D imply that there are more rare items in the data distribution. ZF_D does not affect E-DCG+ since it does not consider rarity issues. Hence, E-DCG+ exhibits comparable performance across all the results in Figure 6.5. ART follows a pattern similar to HC for each of the three approaches.

Observe that as ZF_D increases, ART increases for both ECR and ECR+ because of the increase in the number of rare items. Since rare items are available at relatively lower number of MPs, queries incur more hops, thereby resulting in increased ART.

As the results in Figures 6.5a and 6.5c indicate, ECR+ performs slightly worse than ECR in terms of ART and HC for values of ZF_D that are lower than 0.3. This occurs because lowly skewed data distributions do not necessitate replication. However, as the value of ZF_D increases beyond 0.3, ECR+ exhibits improved ART and HC as compared to ECR because the effect of ECR+'s group-based replication becomes more pronounced. Thus, in case of ECR+, more MPs are able to replicate more number of rare data items. This also results in better SR for ECR+ because it can satisfy more queries. ECR+ exhibits higher MSG than ECR due to the additional messages arising for group interaction as well as from increased traffic owing to more successful queries.



(a) ART (b) SR

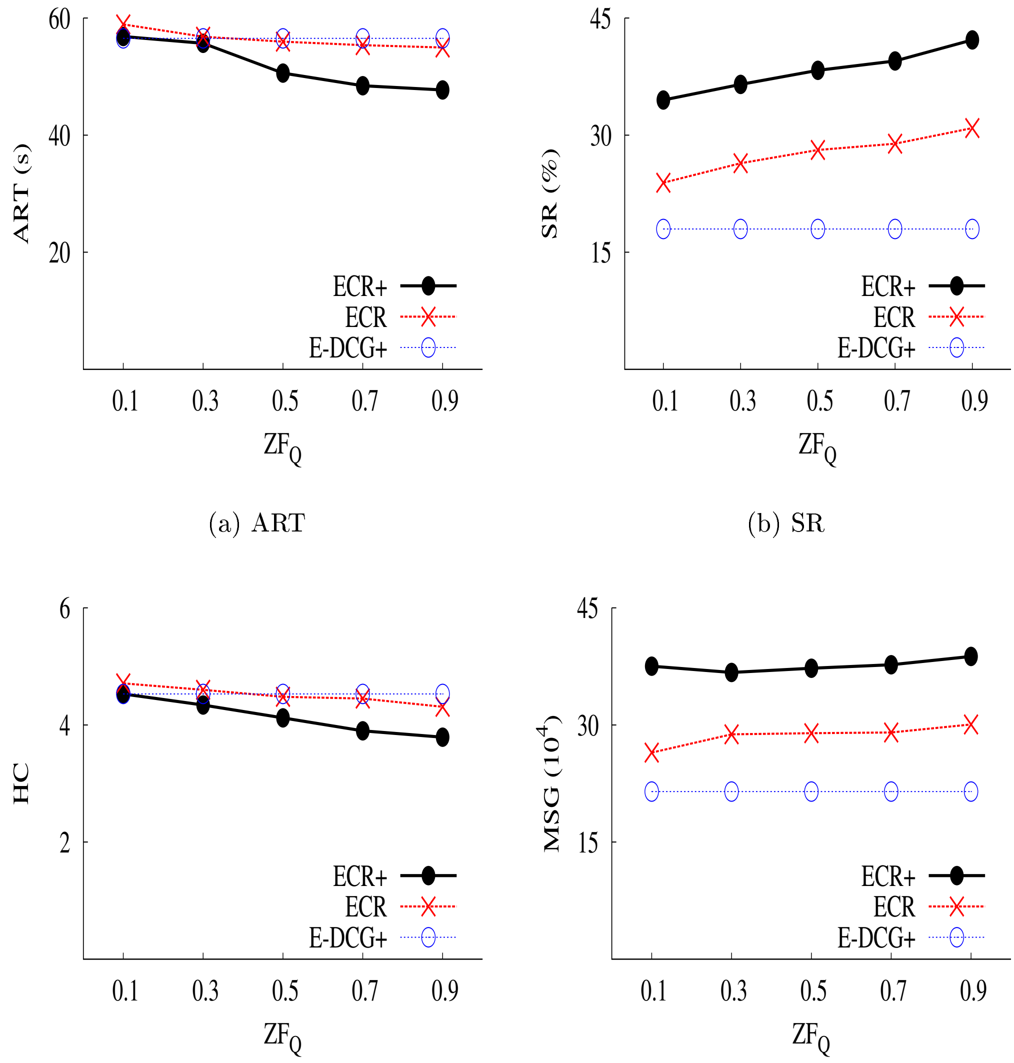
(c) HC (d) MSG

Figure 6.5: Effect of variations in the data distribution across rarity classes

6.4.4 Effect of variations in the query distribution across rare item classes

Recall that ZF_Q is used to determine how the queries are distributed over the different classes of data items (in terms of rarity) i.e., *rare*, *medium-rare* and *non-rare*. Higher values of ZF_Q imply that more number of queries are directed to *rare* data items. The results in Figure 6.6 depict the effect of variations in ZF_Q.

The results in Figure 6.6 indicate that as ZF_Q increases, ART and HC both



(a) ART (b) SR

(c) HC (d) MSG

Figure 6.6: Effect of variations in the query distribution for rare items

decrease for ECR and ECR+ because of the more pronounced effect of rare data replication in response to query workloads with higher skew. However, beyond $ZF_Q = 0.7$, a saturation effect occurs because of the number of rare item replicas becoming stable beyond this value. This occurs primarily due to competition among the MPs for limited available memory space for storing replicas. Moreover, ECR+ exhibits lower ART and HC than that of ECR because of group-based incentives and discounts. E-DCG+ exhibits comparable performance for different values of ZF_Q since it does not consider rarity issues. As ZF_Q increases, SR increases for both ECR and ECR+

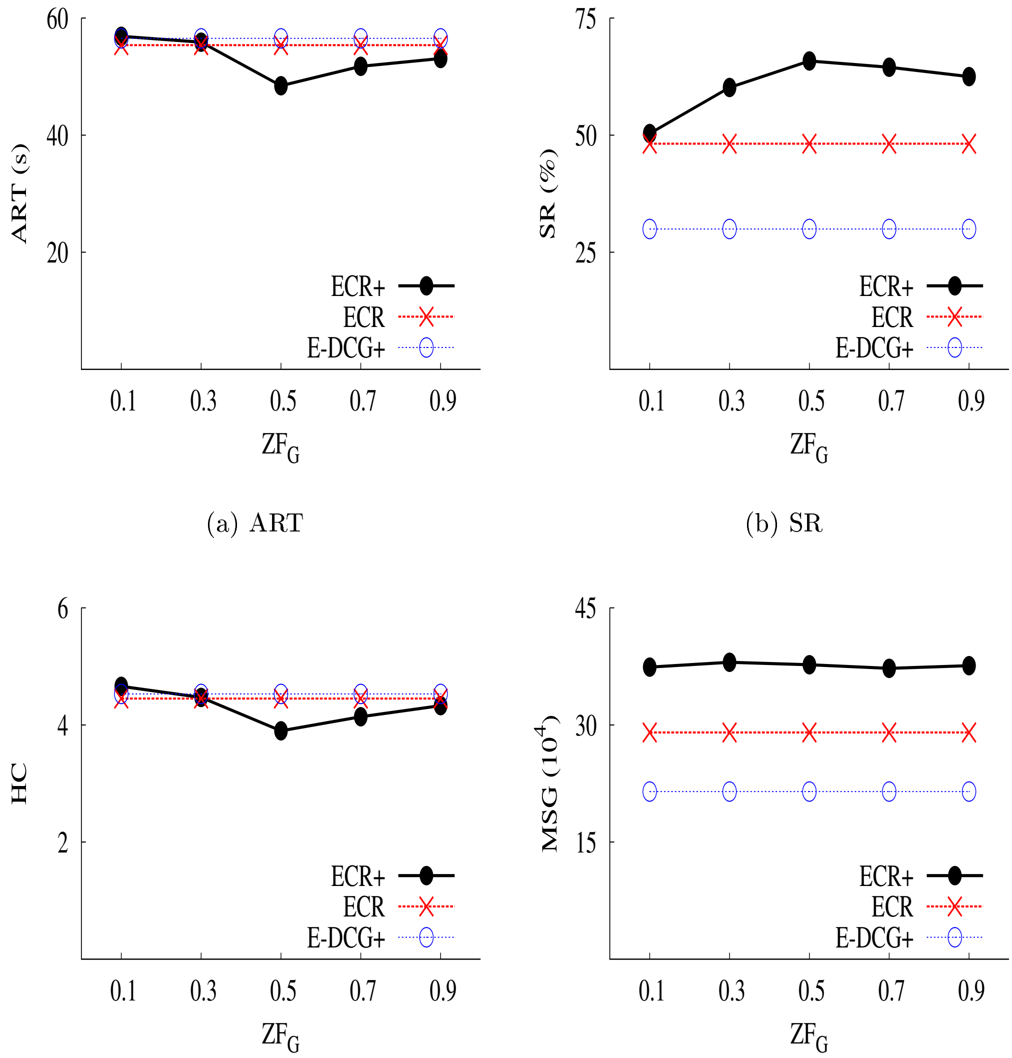
due to more replication in response to more highly skewed workloads. This results in more rare data item requests being satisfied. Furthermore, ECR+ outperforms ECR in terms of SR due to group-based incentives. ECR+ incurs more messages than ECR due to the reasons explained for Figure 6.3d.

6.4.5 Effect of variations in group sizes

We consider 10 different groups. The number of MPs may vary across groups. We conducted an experiment to examine variations in the group sizes (in terms of the number of MPs in different groups). Recall that ZF_G is the zipf factor, which determines the number of MPs assigned to each of the 10 groups. When $ZF_G = 0.1$, each group has a comparable number of MPs. At higher values of ZF_G, some groups contain a disproportionately large number of MPs, while other groups contain relatively few MPs. Figure 6.7 depicts the results of variations in ZF_G.

As ZF_G increases, ART, HC and SR improve for ECR+ due to some of the groups becoming larger, thereby creating more opportunities for replication within the group. However, this performance improvement occurs only upto $ZF_G = 0.5$. At values of ZF_G beyond 0.5, the increase in group size does not create any additional opportunities for replication. Moreover, at these higher values of ZF_G, some of the groups become too small in size, thereby hindering replication. This explains why the performance of ECR+ degrades beyond $ZF_G = 0.5$. Overall, the results indicate that ECR+ performs best (in terms of ART, SR and HC) when $ZF_G = 0.5$.

MSG is comparable across different of values of ZF_G because the increase in the sizes of some of the groups is offset by the decreased sizes of other groups, thereby implying comparable overall communication cost. Observe that ZF_G has no effect on the performance of ECR and E-DCG+ since these approaches do not consider groups.



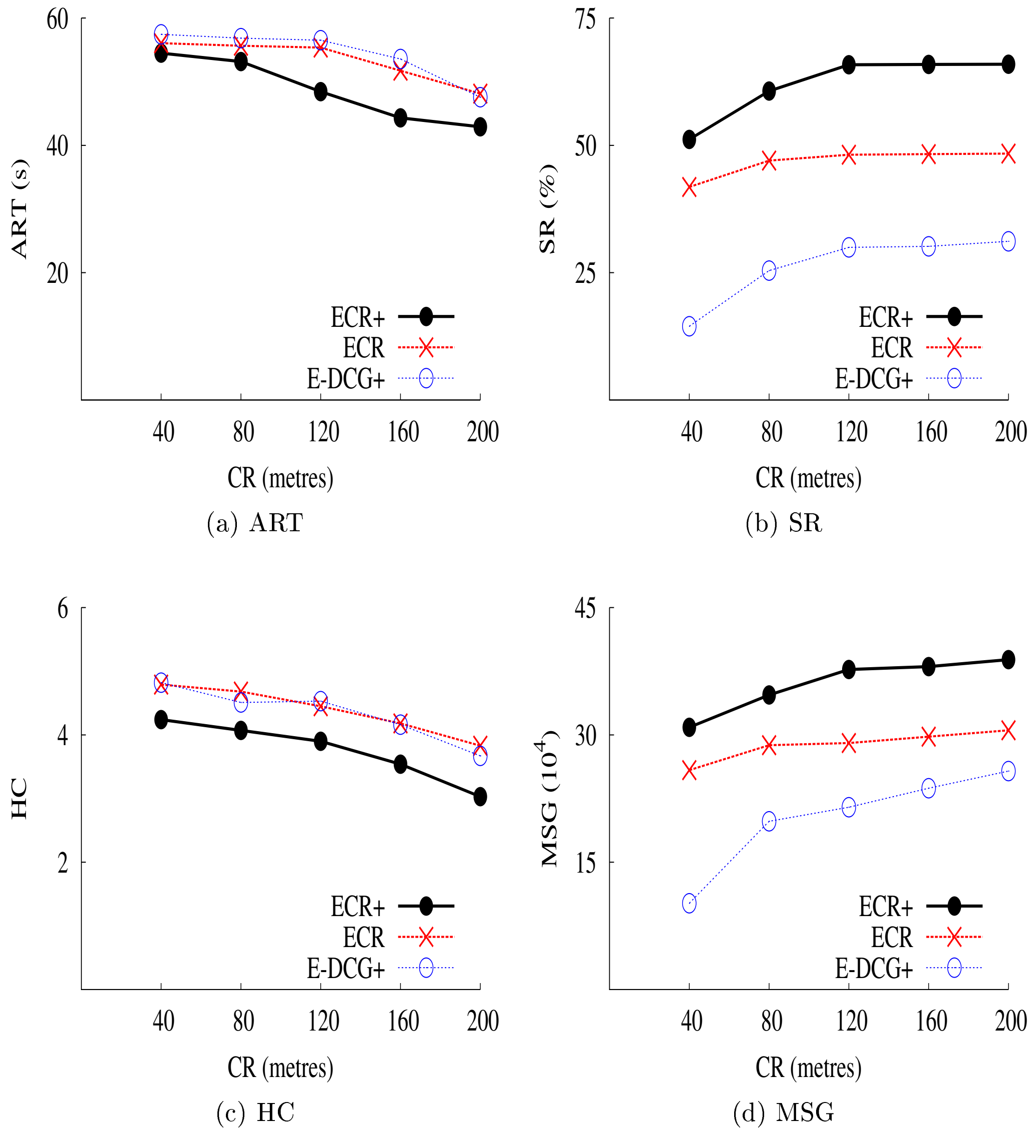
(a) ART (b) SR

(c) HC (d) MSG

Figure 6.7: Effect of variations in the interest group sizes

6.4.6 Effect of variations in the communication range

The results in Figure 6.8 depict the effect of variations in the communication range CR of MPs. Overall, increase in CR has the effect of bringing the MPs ‘nearer’ to each other. As CR increases, both ART and HC decrease for all the approaches due to the reduction in the number of hops between MPs. Interestingly, the results in Figure 6.8c suggest that although ART roughly follows a pattern similar to HC, some deviations occur. These deviations occur because at higher values of CR, an MP needs to process



(a) ART (b) SR (c) HC (d) MSG

Figure 6.8: Effect of variations in the communication range

more incoming queries, thereby resulting in higher waiting times for queries at the job queues of MPs. Consequently, the relay propagation latency also increases slightly with an increase in CR. Furthermore, deviations occur due to bandwidth differences at MPs. Beyond CR = 160 metres, ART plateaus for ECR+ because the gains in ART are offset by the overheads of higher number of incoming queries at MPs that host data items and increased relay propagation latencies.

As CR increases, SR increases for all the approaches upto a certain point and then saturates. The increase in SR occurs essentially due to MPs being 'nearer' in effect with increase in CR, thereby making data items more accessible to query-issuing MPs. A relatively lower number of queries fail due to the maximum TTL criteria of 6 hops because more MPs come within the range to answer a given query. However, beyond a certain point (e.g., CR = 120 for ECR+), any additional increase in CR does not contribute to significant improvement in SR because there is an upperlimit on the replication of rare items due to memory space constraints of the MPs.

As CR increases, MSG increases for all the approaches because the increased reachability of the MPs increases communication among them. With increasing value of CR, there are two opposing effects for MSG. First, increase in CR implies a lower number of messages to reach a given MP. Second, increase in CR also implies that more MPs become involved in the processing of a given query, thereby increasing the communication overhead. These two opposing effects somewhat offset each other at higher values of CR, thereby explaining the reason why MSG eventually plateaus.

6.4.7 Effect of variations in the discount δ

The results in Figure 6.9 depict the effect of variations in the discount δ in case of ECR+. Observe that δ has no effect on the performance of ECR and E-DCG+ since these approaches do not consider discounts. As δ increases, the performance of ECR+ also improves in terms of ART, SR and HC. This is because the effect of group-based incentives becomes more pronounced with increase in discounts. Higher discounts better incentivize MPs querying for the rare items as they can obtain their desired items at lower prices due to discounts, thereby increasing the level of participation and collaboration in the group. However, at values of δ beyond 30%, ECR+'s performance starts degrading slightly. This is because MPs hosting rare items become reluctant to join the group when the value of δ is high. Their revenue-earning potential would decrease due to reduced earnings because of relatively high discounts. In essence, our experimental results show that ECR+ performs best when

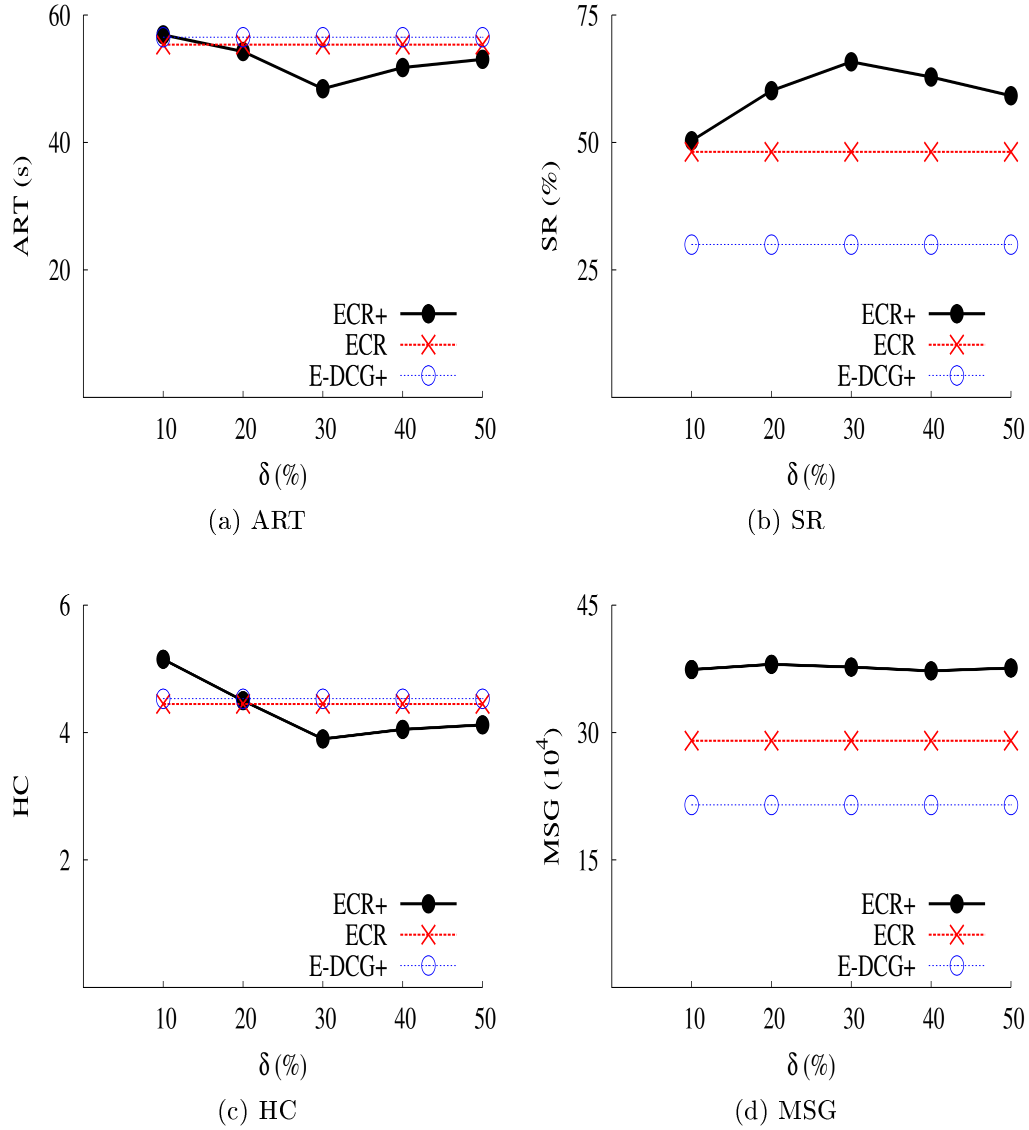
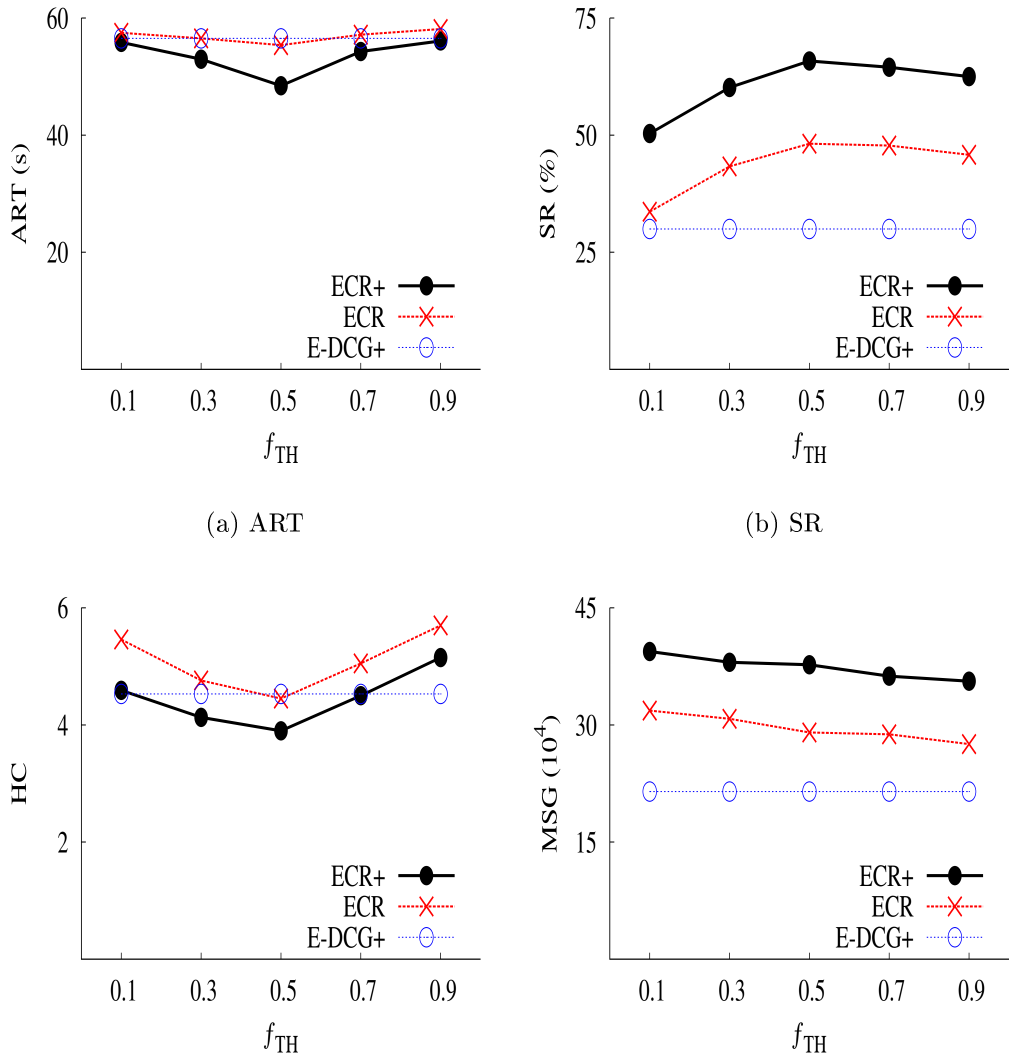


Figure 6.9: Effect of variations in the discount δ

the value of δ is close to 30%.

6.4.8 Effect of variations in the access count threshold f_{TH}

The aim of this experiment is to examine the effect of varying the initiation time of replication on the performance of ECR and ECR+. We quantify the time when replication is initiated by a parameter f_{TH}, which reflects the access count threshold. The total number of issued queries in our experiments



(a) ART (b) SR

(c) HC (d) MSG

Figure 6.10: Effect of variations in the access count threshold f_{TH}

is 10,000. When f_{TH} equals 0.1, it means that replication was initiated after the first 1000 issued queries. Similarly, when f_{TH} equals 0.7, it implies that replication was initiated after the first 7000 issued queries.

The results in Figure 6.10 indicate that ECR+ and ECR both perform best in terms of ART, SR and HC at $f_{TH} = 0.5$. However, as the value of f_{TH} keeps deviating away from 0.5 the performance of ECR and ECR+ both degrade. This is because at low values of f_{TH} (e.g., $f_{TH} = 0.1$) relatively non-rare items get replicated early on, thereby not providing the opportunity for the replication of rare items due to memory space constraints at the

MPs. Moreover, at high values of f_{TH} (e.g., $f_{TH} = 0.7$), the impact of replication on rare data availability becomes much less pronounced because a significant number of query failures already occurred before replication had been initiated.

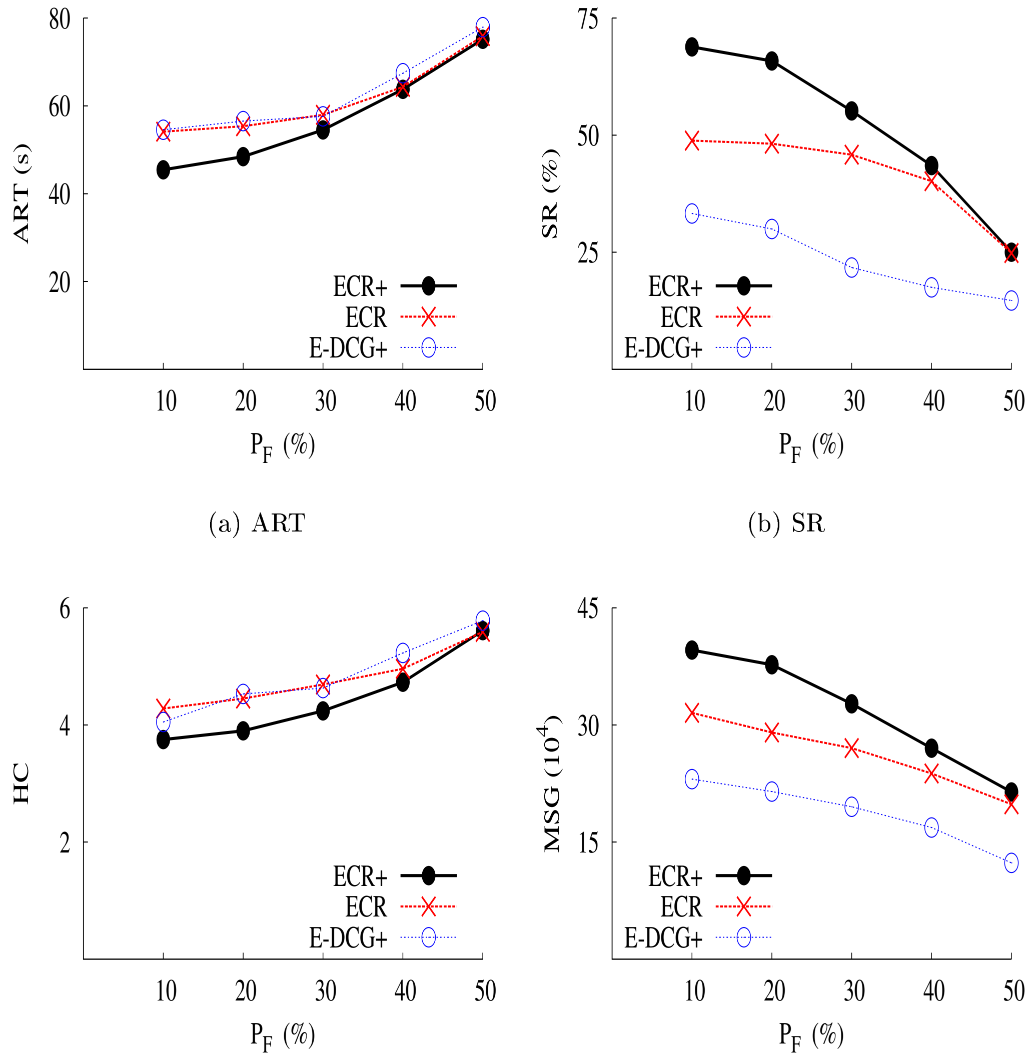
As f_{TH} increases, it implies that replication is initiated at a later point of time, thereby resulting in a lower number of replication-related messages. Hence, as the results in Figure 6.10d indicate, MSG decreases slightly for both ECR and ECR+ with increase in f_{TH}.

6.4.9 Effect of MP failures

We conducted an experiment to investigate the effect of MP failures[5] on the performance of E-Rare. Figure 6.11 depicts the results. As the percentage P_F of MP failures increases, the performance of all the approaches degrade in terms of ART, SR and HC. This is because a higher percentage of MP failures implies a decrease in overall participation in the network, thereby also decreasing the opportunities for replication of rare data items. As more MPs fail, query paths become longer, thereby increasing both ART and HC. Furthermore, SR decreases due to the failure of MPs that host rare data items.

From Figures 6.11a, 6.11b and 6.11c, observe that the performance gap between ECR and ECR+ keeps decreasing with increase in P_F. Moreover, beyond $P_F = 40\%$, both ECR and ECR+ exhibit comparable performance. This occurs due to the effect of groups becoming less pronounced when there are relatively fewer available MPs in the network. For all the three approaches, MSG decreases with increase in P_F due to reduced communication overhead arising from the decrease in the number of available MPs. Moreover, ECR+ exhibits higher MSG than ECR due to the reasons explained for Figure 6.3d.

[5]MPs can fail due to reasons such as depletion of their limited energy resources.



(a) ART (b) SR

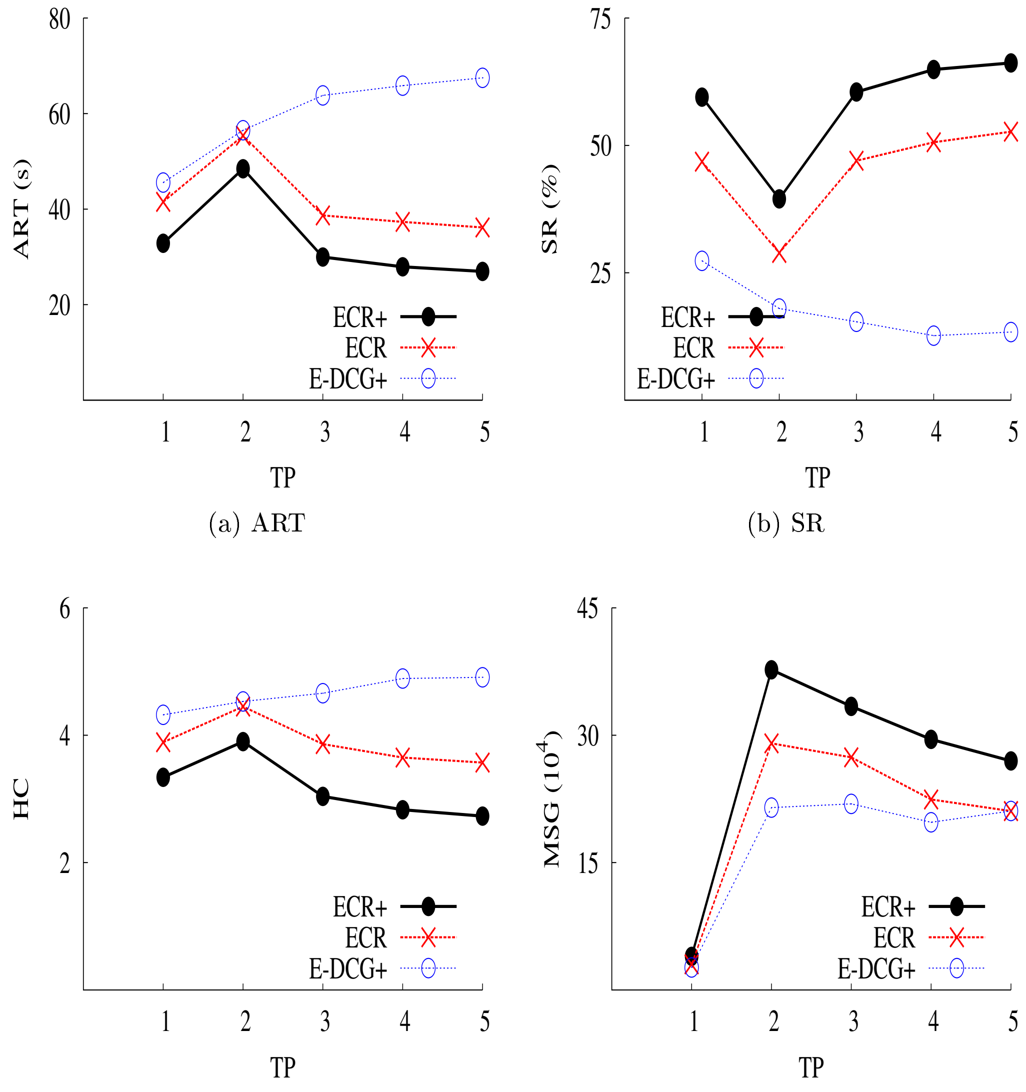
(c) HC (d) MSG

Figure 6.11: Effect of MP failures

6.4.10 Effect of sudden bursts on a single data item

We conducted an experiment to demonstrate the effect of sudden bursts for a *single* data item. Figure 6.12 depicts the results.

We quantify the sudden burst for an item d in terms of a parameter, which we designate as P_{SB}. The value of P_{SB} for an item d is defined as $((Q_d/Q_{total}) \times 100)$, where Q_d is the number of queries directed to d and Q_{total} is the total number of queries during a given time-period. Thus, when $P_{SB} = 15\%$ for item d, it means that 15% of the total number of queries during a given time-



(a) ART (b) SR

(c) HC (d) MSG

Figure 6.12: Effect of sudden bursts on a single data item

period is being directed at d. For this experiment, we consider five equal time-periods, the value of Q_{total} being 2000 for each of these time-periods. For example, when $P_{SB} = 15\%$ and $Q_{total} = 2000$, $Q_d = 300$. We set the values of P_{SB} for d these five time-periods as {15%, 45%, 45%, 45%, 45%} respectively. Thus, the number of queries for d during the five time-periods were {300, 900, 900, 900, 900}.

In Figure 6.12, TP indicates the time-points. Time-period 1 occurs between $TP = 0$ and $TP = 1$. Time-period 2 occurs between $TP = 1$ and $TP = 2$, and so on. The results in Figure 6.12 show that for both ECR and ECR+, performance degraded during the second time-period (i.e., between $TP = 1$ and $TP = 2$) in terms of ART, SR and HC. This is because at the end of the first time-period, replicas had been allocated corresponding to the 300 queries (for d), which had been issued during time-period 1. However, during time-period 2, the sudden burst of 900 queries (i.e., a threefold increase in the number of queries) overwhelmed this initial allocation of replicas. However, at the end of time-period 2, both ECR and ECR+ allocate more replicas to deal effectively with the sudden burst in accesses to d. Hence, beyond time-period 2, the effect of replication by both ECR and ECR+ becomes more prominent, due to which performance keeps gradually improving for both these schemes.

Notably, the results also indicate that the performance of both ECR and ECR+ exhibits a saturation effect during time-periods 4 and 5. This occurs primarily due to competition among replicas for the limited available memory space. For E-DCG+, the performance severely degrades during the second time-period due to the absence of replication when the sudden burst of queries come in for d. Beyond $TP = 2$, ART and HC both exhibit a saturation effect for E-DCG+ primarily because many queries get dropped, due to which SR decreases for E-DCG+.

MSG increases over time for all the approaches because it is cumulative. For ECR and ECR+, MSG increases over time also due to increased communication for licensing and replication of rare data items in response to the sudden burst. Moreover, ECR+ exhibits higher MSG than ECR due to the reasons explained for Figure 6.3d. Observe that MSG is lower for E-DCG+ than for ECR and ECR+ primarily because E-DCG+ does not perform replication and many queries get dropped (i.e., query failures occur) in case of E-DCG+.

6.5 Summary

In M-P2P networks, data availability is typically low due to rampant free-riding, frequent network partitioning and mobile resource constraints. We have proposed E-Rare, a novel economic incentive model for improving the availability of rare data by means of licensing-based replication in M-P2P networks.

E-Rare comprises two replication schemes, namely ECR and ECR+, both of which use its incentive model for improving rare data availability. In ECR, the MPs act individually towards replication, while for ECR+, the MPs perform replication in groups. Our performance evaluation demonstrates that the peer-group-based strategy of ECR+ outperforms the individual-based strategy used by ECR in terms of query response times and availability of rare data items in M-P2P networks. In the near future, we plan to use game-theoretic approaches for rare data item pricing and compare the performance of E-Rare for different economic models.

7

Summary

This dissertation examines the problem M-P2P challenges free-riding and data accessibility for realizing M-P2P applications. Furthermore, we analyse and integrate the various economic incentive-based schemes to combat free-riding and to increase peer participation, thereby leading to the increased data availability for reduction of response time in M-P2P networks.

The exponential growth of the mobile networks and the interactions of mobile peers, inspired researchers and developers to analyse and built the mobile applications based on peer-to-peer communications, which enormously reduces the Internet traffic, while providing faster and better communication platform. This strongly motivate M-P2P network applications. Mobile devices wirelessly communicating in a P2P fashion facilitate M-P2P applications by enabling information sharing *on-the-fly*. Moreover, the proliferation of mobile devices with embedded GPS sensors coupled with the growth in the popularity of infotainment services for vehicles have created new avenues for improving vehicular traffic management in road networks.

First, we have proposed E-Top for improving efficient top-k query processing in M-P2P networks, because peers in M-P2P networks are used to have top-k queries. For example, someone wants to find the top-k restaurants with “happy hours” (or “manager's special hours”) within 1 km of her current location. Here, top-k is determined based on the parameters (e.g., star rating, price and distance from the point of query reference) selected by the user. Similarly, another application could involve a parking lot, where MPs can collect information about available parking slots and charges, and then they can inform the brokers. The parking slot availability information has to be current and therefore, the broker can compare such current information with its current list of parking slots. The broker can then provide the top-k available slots to the query-issuing MP in terms of price or distance (from the MP's current location). Similarly, an MP may want to find the top-k stores selling Levis jeans in a shopping mall with criteria such as (low) price during a specific time duration.

Second, we have proposed E-Broker to incentivize brokers for providing *value-added routing service* in M-P2P networks to increase data availability, thereby improving query response time. Here, the term “value-added routing service” refers to the broker MPs enabling pro-active search for the query results by maintaining an index of the data items (and replicas) stored at other MPs (as opposed to just forwarding queries).

Third, E-VeT provides efficient vehicular traffic management in road networks using economy-based reward/penal schemes. This work aims to identify that how effectively system can assign the paths to the vehicles to manage vehicular traffic by reducing the average time of arrival and fuel consumption.

Forth, E-Rare focusses on handling *rare* data items in an M-P2P environment. Here, *Rare* data items are those, which get sudden bursts in accesses based on *events* as they are only hosted by only a few peers in comparison to the network size. Thus, they may not be available within few hops of query-issuing peers. The sudden burst in accesses to rare items generally occurs within a given time-frame (associated with the event), before and after which such items are rarely accessed. For example, during course of a remote forest expedition, sudden unexpected decrease in temperature and gusty winds rise the need to look for the information related to shops selling sweaters and wind-cheaters. Such M-P2P interactions for effective sharing of rare data are currently not freely supported by existing wireless communication infrastructures.

In this dissertation, E-Top proposed peer-based ETK and ETK+; and group-based ETG economic schemes to perform efficient top-k query processing in M-P2P networks. In addition, E-Broker has presented the basic and advanced incentivized economic schemes EIB and EIB+ respectively for brokers' value-added routing and replication services in M-P2P networks. Furthermore, we have proposed individual and peer-group incentive-based replication schemes ECR and ECR+ respectively in E-Rare system for improving rare data availability by means of licensing-based replication in M-P2P networks. Finally, E-VeT provides the proposed economic reward/penalty-based route allocation schemes R^2A and R^2A^+ to assign payoffs to the vehicles for effective traffic management in vehicular networks.

We have implemented our all above proposed schemes and conducted extensive performance studies on OMNeT++ [Pon93] based simulator using sample and real datasets. The results of our experiments demonstrate the overall effectiveness of our proposed schemes for E-Top, E-Broker, E-VeT and E-Rare systems.

To this end, we believe that our contributions have successfully addressed some of the issues concerning the efficient data management in M-P2P networks. We hope that ongoing research work in this field will not only continue to increase the availability of data, but also improve the overall performance of M-P2P applications. In near future, we plan to use game-theoretic approaches for data item pricing along with the advanced economic incentive-schemes for all our proposed models.

Bibliography

[AH00] Eytan Adar and Bernardo A Huberman. Free Riding on Gnutella. *First Monday*, 5(10-2), 2000.

[Ama05] Amazon. Mechanical Turk. `http://www.mturk.com/mturk/welcome`, November 2005.

[AMM03] Usman Arshad, Cecilia Mascolo, and Melo. Exploiting Mobile Computing in Health-care. In *Proceedings of the 23rd International Conference on Distributed Computing Systems (ICDCS'03)*, May 2003.

[AMP98] Stanisaw Ambroszkiewicz, Olaf Matyja, and Wojciech Penczek. Team Formation by Self-Interested Mobile Agents. *Multi-Agent Systems. Theories, Languages and Applications*, 1544:1–15, 1998.

[AR10] Farag Azzedin and Ahmad Ridha. Feedback behavior and its role in trust assessment forÂ peer-to-peer systems. *Journal of Telecommunication Systems*, 44(3-4):253–266, 2010.

[AWX+11] Daniel Ayala, Ouri Wolfson, Bo Xu, Bhaskar Dasgupta, and Jie Lin. Parking slot assignment games. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '11, pages 299–308, New York, NY, USA, 2011. ACM.

[AWX$^+$12] Daniel Ayala, Ouri Wolfson, Bo Xu, Bhaskar DasGupta, and Jie Lin. Pricing of Parking for Congestion Reduction. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '12, pages 43–51, New York, NY, USA, 2012. ACM.

[Azz10a] F. Azzedin. Trust-based taxonomy for free riders in distributed multimedia systems. In *High Performance Computing and Simulation (HPCS), 2010 International Conference on*, pages 362–369, 2010.

[Azz10b] Farag Ahmed Azzedin. Classifying and tracking free riders in multimedia-based systems. *Journal of Universal Computer Science*, 16(10):1368–1387, 2010.

[BBS10] Riccardo Boero, Giangiacomo Bravo, and Flaminio Squazzoni. Trust and Partner Selection in Social Networks: An Experimentally Grounded Model. *arXiv preprint arXiv:1008.4705*, 2010.

[BCFN03] B. Bakos, G. Csucs, L. Farkas, and J.K. Nurminen. Peer-to-peer protocol evaluation in topologies resembling wireless networks. An experiment with Gnutella query engine. In *Networks, 2003. ICON2003. The 11th IEEE International Conference on*, pages 673–680, 2003.

[BGJL06] J. Burgess, Brian Gallagher, D. Jensen, and B.N. Levine. MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks. In *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, pages 1–11, 2006.

[BH01] L. Buttyan and J. Hubaux. Nuglets: a Virtual Currency to Stimulate Cooperation in Self-Organized Mobile Ad Hoc Networks. Technical report, Swiss Federal Institute of Technology, Lausanne, 2001.

[BH03] Levente Buttyán and Jean-Pierre Hubaux. Stimulating cooperation in self-organizing mobile ad hoc networks. *Journal of Mobile Networks and Applications*, 8(5):579–592, October 2003.

[BK09] Fan Bai and Bhaskar Krishnamachari. Spatio-temporal variations of vehicle traffic in VANETs: facts and implications. In *Proceedings of the sixth ACM International workshop on VehiculAr InterNETworking*, VANET '09, pages 43–52, New York, NY, USA, 2009. ACM.

[BMJ+98] Josh Broch, David A. Maltz, David B. Johnson, Yih-Chun Hu, and Jorjeta Jetcheva. A performance comparison of multi-hop wireless ad hoc network routing protocols. In *Proceedings of the 4th annual ACM/IEEE International Conference on Mobile computing and networking*, MobiCom '98, pages 85–97, New York, NY, USA, 1998. ACM.

[BMSV03] Ranjita Bhagwan, David Moore, Stefan Savage, and Geoffrey M. Voelker. *Future directions in distributed computing*. Springer-Verlag, Berlin, Heidelberg, 2003.

[Bra96] Richard M. Bradford. Pricing, routing, and incentive compatibility in multiserver queues. *European Journal of Operational Research*, 89(2):226–236, 1996.

[CABP05] Rajiv Chakravorty, Sulabh Agarwal, Suman Banerjee, and Ian Pratt. MoB: a mobile bazaar for wide-area wireless services. In *Proceedings of the 11th annual International Conference on Mobile computing and networking*, MobiCom '05, pages 228–242, New York, NY, USA, 2005. ACM.

[CGKO03] J. Crowcroft, R. Gibbens, F. Kelly, and S. Ostring. Modelling incentives for collaboration in mobile ad hoc networks. In *WiOpt*, 2003.

[Chu08] Yon Dohn Chung. A cache invalidation scheme for continuous partial match queries in mobile computing environments. *Journal of Distributed and Parallel Databases*, 23(3):207–234, 2008.

[CLC04] Chi-Yin Chow, Hong Va Leong, and Alvin Chan. Peer-to-Peer Cooperative Caching in Mobile Environments. In *Proceedings of the 24th International Conference on Distributed Computing Systems Workshops - W7: EC (ICDCSW'04) - Volume 7*, ICDCSW '04, pages 528–533, Washington, DC, USA, 2004. IEEE Computer Society.

[CN04] Kai Chen and Klara Nahrstedt. iPass: An Incentive Compatible Auction Scheme to Enable Packet Forwarding Service in MANET. In *Proceedings of the 24th International Conference on Distributed Computing Systems (ICDCS'04)*, ICDCS '04, pages 534–542, Washington, DC, USA, 2004. IEEE Computer Society.

[CW04] Guangtong Cao and Jennifer L. Welch. Accurate Multihop Clock Synchronization in Mobile Ad Hoc Networks. In *Proceedings of the 2004 International Conference on Parallel Processing Workshops*, ICPPW '04, pages 13–20, Washington, DC, USA, 2004. IEEE Computer Society.

[DCRS14] R. Doriguzzi Corin, Giovanni Russello, and Elio Salvadori. *Security for Multihop Wireless Networks*. 2014.

[DGMS85] Susan B. Davidson, Hector Garcia-Molina, and Dale Skeen. Consistency in a partitioned network: a survey. *ACM Computing Survey*, 17(3):341–370, September 1985.

[DHA03a] A. Datta, M. Hauswirth, and K. Aberer. Beyond "Web of trust": enabling P2P e-commerce. In *E-Commerce, 2003. CEC 2003. IEEE International Conference on*, pages 303–312, 2003.

[DHA03b] Anwitaman Datta, Manfred Hauswirth, and Karl Aberer. Updates in Highly Unreliable, Replicated Peer-to-Peer Systems. In *Proceedings of the 23rd International Conference on Distributed Computing Systems*, ICDCS '03, pages 76–, Washington, DC, USA, 2003. IEEE Computer Society.

[DKTM11] Nikos Dimokas, Dimitrios Katsaros, Leandros Tassiulas, and Yannis Manolopoulos. High performance, low complexity cooperative caching for wireless sensor networks. *Journal of Wireless Networks*, 17(3):717–737, April 2011.

[DPGB03] P. Daras, D. Palaka, V. Giagourta, and D. Bechtsis. A novel peer-to-peer payment protocol. In *EUROCON 2003. Computer as a Tool. The IEEE Region 8*, volume 1, pages 2–6 vol.1, 2003.

[Eag09] Nathan Eagle1. txteagle: Mobile Crowdsourcing. *Lecture Notes in Computer Science*, 5623:447–456, 2009.

[EdV11] Carsten Eickhoff and Arjen P. de Vries. How Crowdsourcable is your Task? In *Workshop on Crowdsourcing for Search and Data Mining (CSDM)*, 2011.

[EHTA11] Timofey Ermilov, Norman Heino, Sebastian Tramp, and Sören Auer. OntoWiki mobile: knowledge management in your pocket. In *Proceedings of the 8th extended semantic web Conference on The semantic web: research and applications - Volume Part I*, ESWC'11, pages 185–199, Berlin, Heidelberg, 2011. Springer-Verlag.

[Epp98] David Eppstein. Finding the k shortest paths. *SIAM Journal on computing*, 28(2):652–673, 1998.

[ET04] Elharith Elrufaie and David A. Turner. Bidding in P2P Content Distribution Networks using the Lightweight Currency Paradigm. In *Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04) Volume 2 - Volume 2*, ITCC '04, pages 129–, Washington, DC, USA, 2004. IEEE Computer Society.

[FC05] Michal Feldman and John Chuang. Overcoming free-riding behavior in peer-to-peer systems. *ACM SIGecom Exchanges*, 5(4):41–50, 2005.

[FKK+11] Michael J. Franklin, Donald Kossmann, Tim Kraska, Sukriti Ramesh, and Reynold Xin. CrowdDB: answering queries with crowdsourcing. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, SIGMOD '11, pages 61–72, New York, NY, USA, 2011. ACM.

[FNY93] D. Ferguson, C. Nikolaou, and Y. Yemini. An economy for managing replicated data in autonomous decentralized systems. In *Autonomous Decentralized Systems, 1993. Proceedings. ISADS 93., International Symposium on*, pages 367–375, 1993.

[FST04] Daniel R Figueiredo, Jonathan K Shapiro, and Don Towsley. Payment-based incentives for anonymous peer-to-peer systems. *Computer Science Department, University of Massachusetts*, 2004.

[FYN88] D. Ferguson, Y. Yemini, and C. Nikolaou. Microeconomic algorithms for load balancing in distributed computer systems. In *Distributed Computing Systems, 1988., 8th International Conference on*, pages 491–499, 1988.

[GA04] Anargyros Garyfalos and Kevin C. Almeroth. Coupon Based Incentive Systems and the Implications of Equilibrium Theory.

In *Proceedings of the IEEE International Conference on E-Commerce Technology*, CEC '04, pages 213–220, Washington, DC, USA, 2004. IEEE Computer Society.

[GBM01] P. Golle, K.L. Brown, and I. Mironov. Incentives for Sharing in Peer-to-Peer Networks. In *Electronic Commerce*, 2001.

[GK02] Koichi Goto and Yahiko Kambayashi. A new passenger support system for public transport using mobile database access. In *Proceedings of the 28th International Conference on Very Large Data Bases*, VLDB '02, pages 908–919. VLDB Endowment, 2002.

[GK03] Koichi Goto and Yahiko Kambayashi. Integration of electronic tickets and personal guide system for public transport using mobile terminals. In *Proceedings of the 2003 ACM SIGMOD International Conference on Management of data*, SIGMOD '03, pages 642–646, New York, NY, USA, 2003. ACM.

[Gnu] Gnutella. `http://www.gnutella.com/`.

[GNVTS11] S. Gutierrez-Nolasco, N. Venkatasubramanian, C. Talcott, and M.-O. Stehr. Tailoring group membership consistency for mobile networks. In *Collaboration Technologies and Systems (CTS), 2011 International Conference on*, pages 548–554, 2011.

[Gro03] Christian Grothoff. Resource allocation in peer-to-peer networks. *Wirtschaftsinformatik*, 45(3):285–292, 2003.

[GRR+98] R. Guy, P. Reiher, D. Ratner, M. Gunter, W. Ma, and G. Popek. Rumor: Mobile Data Access Through Optimistic Peer-to-Peer Replication. In *ER Workshops*, 1998.

[HA05] MyungJoo Ham and Gul Agha. ARA: A Robust Audit to Prevent Free-Riding in P2P Networks. In *Proceedings of the*

Fifth IEEE International Conference on Peer-to-Peer Computing, P2P '05, pages 125–132, Washington, DC, USA, 2005. IEEE Computer Society.

[HC06] Jiun-Long Huang and Ming-Syan Chen. On the Effect of Group Mobility to Data Replication in Ad Hoc Networks. *IEEE Transactions on Mobile Computing (TMC)*, 5(5):492–507, 2006.

[HC07] Seung-won Hwang and Kevin Chen-chuan Chang. Optimizing top-k queries for middleware access: A unified cost-based approach. *ACM Transactions on Database Systems (TODS)*, 32(1):5, March 2007.

[HHS+10] Takahiro Hara, Ryo Hagihara, Yuya Sasaki, Masako Shinohara, and Shojiro Nishio. Efficient Top-k Query Processing in Mobile Ad Hoc Networks. *Journal of Next-Generation Computing*, 1(2):249–267, 2010.

[HL05] HarryChia-Hung Hsu and Alan Liu. Multiagent-Based Multiteam Formation Control for Mobile Robots. *Journal of Intelligent and Robotic Systems*, 42(4):337–360, 2005.

[HM05] T. Hara and S.K. Madria. Consistency management among replicas in peer-to-peer mobile ad hoc networks. In *Reliable Distributed Systems, 2005. SRDS 2005. 24th IEEE Symposium on*, pages 3–12, 2005.

[HM06] T. Hara and S.K. Madria. Data Replication for Improving Data Accessibility in Ad Hoc Networks. *IEEE Transactions on Mobile Computing (TMC)*, 5(11):1515–1532, 2006.

[HSHN09] Ryo Hagihara, Masako Shinohara, Takahiro Hara, and Shojiro Nishio. A Message Processing Method for Top-k Query for Traffic Reduction in Ad Hoc Networks. In *Proceedings of* *the 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware*, MDM '09, pages 11–20, Washington, DC, USA, 2009. IEEE Computer Society.

[HXW+05] Haibo Hu, Jianliang Xu, Wing Sing Wong, Baihua Zheng, Dik Lun Lee, and Wang-Chien Lee. Proactive Caching for Spatial Queries in Mobile Environments. In *Proceedings of the 21st International Conference on Data Engineering*, ICDE '05, pages 403–414, Washington, DC, USA, 2005. IEEE Computer Society.

[Iss11] A. Issariyanukula. Financial and Technical Feasibility of Dynamic Congestion Pricing as a Revenue Generation Source in Indiana- Exploiting the Availability of Real-Time Information and Dynamic Pricing Technologies. Technical report, 2011.

[JCCL10] Ha Rim Jung, Byung Ku Cho, Yon Dohn Chung, and Ling Liu. On processing location based top-k queries in the wireless broadcasting system. In *Proceedings of the 2010 ACM Symposium on Applied Computing*, SAC '10, pages 585–591, New York, NY, USA, 2010. ACM.

[Jok03] T. Jokela. Authoring tools for mobile multimedia content. In *Multimedia and Expo*, pages 637–640, 2003.

[JPNS08] Mary Magdalene Janef, R. Parameswaran, R Nadarajan, and Maytham Safar. PINE-guided cache replacement policy for location-dependent data in mobile environment. In *Proceedings of the 1st International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '08, pages 16:1–16:5, New York, NY, USA, 2008. ACM.

[JYLK02] Il-dong Jung, Young-ho You, Jong-hwan Lee, and Kyungsok Kim. Broadcasting and caching policies for location-dependent queries in urban areas. In *Proceedings of the 2nd International* *workshop on Mobile commerce*, WMC '02, pages 54–60, New York, NY, USA, 2002. ACM.

[KA00] Bettina Kemme and Gustavo Alonso. A new approach to developing and implementing eager database replication protocols. *ACM Transactions on Database Systems (TODS)*, 25(3):333–379, September 2000.

[Kaz06] KaZaa. `http://www.kazaa.com/`, 2006.

[KKMM10] Masaru Kitsuregawa, Sanjay Kumar Madria, and Anirban Mondal. A Collaborative Replication Approach for Mobile-P2P Networks. *Journal of Handheld Computing Research*, 1(2):75–97, April 2010.

[KL01] Hyunchul Kang and Sangmin Lim. Bandwidth-Conserving Cache Validation Schemes in a Moblie Database System. In *Proceedings of the Second International Conference on Mobile Data Management*, MDM '01, pages 121–132, London, UK, UK, 2001. Springer-Verlag.

[KMS07] Ajey Kumar, Manoj Misra, and A. K. Sarje. A weighted cache replacement policy for location dependent data in mobile environments. In *Proceedings of the 2007 ACM symposium on Applied computing*, SAC '07, pages 920–924, New York, NY, USA, 2007. ACM.

[KMSA08] S.U. Khan, A.A. Maciejewski, H.J. Siegel, and I. Ahmad. A game theoretical data replication technique for mobile ad hoc networks. In *Parallel and Distributed Processing, 2008. IPDPS 2008. IEEE International Symposium on*, pages 1–12, 2008.

[KMZ02] Steve Kremer, Olivier Markowitch, and Jianying Zhou. An intensive survey of fair non-repudiation protocols. *Journal of Computer Communications*, 25(17):1606–1621, 2002.

[KS89] James F. Kurose and Rahul Simha. A Microeconomic Approach to Optimal Resource Allocation in Distributed Computer Systems. *IEEE Transactions on Computers (TOC)*, 38(5):705–717, May 1989.

[KSGM03a] S. Kamvar, M. Schlosser, and H. Garcia-Molina. Incentives for Combatting Free-riding on P2P Networks. In *Euro-Par*, 2003.

[KSGM03b] Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. The Eigentrust algorithm for reputation management in P2P networks. In *Proceedings of the 12th International Conference on World Wide Web*, WWW '03, pages 640–651, New York, NY, USA, 2003. ACM.

[KSM10] Ajey Kumar, Anil K. Sarje, and Manoj Misra. Prioritised Predicted Region based Cache Replacement Policy for location dependent data in mobile environment. *Journal of Ad Hoc and Ubiquitous Computing*, 5(1):56–67, 2010.

[LC11] Yinan Li and Ing-Ray Chen. Adaptive per-user per-object cache consistency management for mobile data access in wireless mesh networks. *Journal of Parallel and Distributed Computing*, 71(7):1034–1046, 2011.

[LCLC04] Kam-Yiu Lam, Reynold Cheng, BiYu Liang, and Jo Chau. Sensor node selection for execution of continuous probabilistic queries in wireless sensor networks. In *Proceedings of the ACM 2nd International workshop on Video surveillance & sensor networks*, VSSN '04, pages 63–71, New York, NY, USA, 2004. ACM.

[LDHS05] N. Liebau, V. Darlagiannis, O. Heckmann, and R. Steinmetz. Asymmetric Incentives in Peer-to-Peer Systems. In *AMCIS*, 2005.

[LGS12] Yunzhao Li, Don Gruenbacher, and Caterina Scoglio. Reward only is not enough: Evaluating and improving the fairness policy of the P2P file sharing network eMule/eDonkey. *Journal of Peer-to-Peer Networking and Applications (PPNA)*, 5(1):40–57, 2012.

[LHC12] Dong Li, Mei Hu, and Rui Chen. Quadtree-based management on semantic cache for mobile computing. *Journal of Computer Applications in Technology*, 43(4):393–399, 2012.

[LI04] Jinshan Liu and Valérie Issarny. Service allocation in selfish mobile ad hoc networks using vickrey auction. In *Proceedings of the 2004 International Conference on Current Trends in Database Technology*, EDBT'04, pages 385–394, Berlin, Heidelberg, 2004. Springer-Verlag.

[LLKL09] Wookey Lee, James Jung-Hoon Lee, Young-Kuk Kim, and Carson Kai-Sang Leung. AnchorWoman: top-k structured mobile web search engine. In *Proceedings of the 18th ACM Conference on Information and knowledge management*, CIKM '09, pages 2089–2090, New York, NY, USA, 2009. ACM.

[LSS11] R. Lubke, D. Schuster, and A. Schill. MobilisGroups: Location-based group formation in Mobile Social Networks. In *Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pages 502–507, 2011.

[LW08] Yan Luo and Ouri Wolfson. Mobile p2p databases. *Encyclopedia of GIS*, pages 671–677, 2008.

[LXL10] Xingjie Liu, Jianliang Xu, and Wang-Chien Lee. A Cross Pruning Framework for Top-k Data Collection in Wireless Sensor Networks. In *Proceedings of the 2010 Eleventh International Conference on Mobile Data Management*, MDM '10, pages

157–166, Washington, DC, USA, 2010. IEEE Computer Society.

[MdRK04] Ronald Mannak, Huib de Ridder, and David V. Keyson. The human side of sharing in peer-to-peer networks. In *Proceedings of the 2nd European Union symposium on Ambient intelligence*, EUSAI '04, pages 59–64, New York, NY, USA, 2004. ACM.

[MK05] A. Mondal and M. Kitsuregawa. Effective Dynamic Replication in Wide-Area Network Environments: A Perspective. In *Database and Expert Systems Applications, 2005. Proceedings. Sixteenth International Workshop on*, pages 287–291, 2005.

[MK06] Anirban Mondal and Masaru Kitsuregawa. Privacy, Security and Trust in P2P environments: A Perspective. In *Proceedings of the 17th International Conference on Database and Expert Systems Applications*, DEXA '06, pages 682–686, Washington, DC, USA, 2006. IEEE Computer Society.

[MK10] Anirban Mondal and Masaru Kitsuregawa. EARL: An Economic incentive scheme for Adaptive Revenue-Load-based dynamic replication of data in Mobile-P2P networks. *Journal of Distributed and Parallel Databases*, 28:1–31, 2010.

[MLK04] Anirban Mondal, Yi Lifu, and Masaru Kitsuregawa. On Improving the Performance Dependability of Unstructured P2P Systems via Replication. In Fernando Galindo, Makoto Takizawa, and Roland TraunmÃŒller, editors, *Database and Expert Systems Applications*, volume 3180 of *Lecture Notes in Computer Science*, pages 601–610. Springer Berlin Heidelberg, 2004.

[MM08] Sanjay Kumar Madria and Anirban Mondal. Economic-based Incentive Schemes for Dynamic Data Management in Mobile

P2P Computing. In *Proceedings of the The Ninth International Conference on Mobile Data Management*, MDM '08, pages 228–, Washington, DC, USA, 2008. IEEE Computer Society.

[MMK06a] A. Mondal, S.K. Madria, and M. Kitsuregawa. CADRE: A Collaborative replica allocation and deallocation approach for Mobile-P2P networks. In *Database Engineering and Applications Symposium, 2006. IDEAS '06. 10th International*, pages 21–28, 2006.

[MMK06b] A. Mondal, S.K. Madria, and M. Kitsuregawa. EcoRep: An Economic Model for Efficient Dynamic Replication in Mobile-P2P networks. In *COMAD*, 2006.

[MMK06c] Anirban Mondal, SanjayKumar Madria, and Masaru Kitsuregawa. CLEAR: An Efficient Context and Location-Based Dynamic Replication Scheme for Mobile-P2P Networks. In StÃ©phane Bressan, Josef KÃŒng, and Roland Wagner, editors, *Database and Expert Systems Applications*, volume 4080 of *Lecture Notes in Computer Science*, pages 399–408. Springer Berlin Heidelberg, 2006.

[MMK07a] Anirban Mondal, Sanjay Kumar Madria, and Masaru Kitsuregawa. ConQuer: A Peer Group-Based Incentive Model for Constraint Querying in Mobile-P2P Networks. In *Proceedings of the 2007 International Conference on Mobile Data Management*, MDM '07, pages 134–139, Washington, DC, USA, 2007. IEEE Computer Society.

[MMK07b] Anirban Mondal, Sanjay Kumar Madria, and Masaru Kitsuregawa. LEASE: an economic approach to leasing data items in mobile-P2P networks to improve data availability. In *Proceedings of the 5th International Conference on Databases in net-*

worked information systems, DNIS'07, pages 222–231, Berlin, Heidelberg, 2007. Springer-Verlag.

[MMK07c] Anirban Mondal, SanjayKumar Madria, and Masaru Kitsuregawa. ABIDE: A Bid-Based Economic Incentive Model for Enticing Non-cooperative Peers in Mobile-P2P Networks. In Ramamohanarao Kotagiri, P.Radha Krishna, Mukesh Mohania, and Ekawit Nantajeewarawat, editors, *Advances in Databases: Concepts, Systems and Applications*, volume 4443 of *Lecture Notes in Computer Science*, pages 703–714. Springer Berlin Heidelberg, 2007.

[MMK09] Anirban Mondal, SanjayKumar Madria, and Masaru Kitsuregawa. An economic incentive model for encouraging peer collaboration in mobile-P2P networks with support for constraint queries. *Journal of Peer-to-Peer Networking and Applications (PPNA)*, 2(3):230–251, 2009.

[Mor10] E.F. Morgul. Simulation based evaluation of dynamic congestion pricing algorithms and strategies. Master's thesis, Rutgers University, USA, 2010.

[MYM10] Anirban Mondal, Kuldeep Yadav, and Sanjay Kumar Madria. EcoBroker: An Economic Incentive-based Brokerage Model for Efficiently Handling Multiple-item Queries to Improve Data Availability via Replication in Mobile-P2P Networks. In *Proceedings of the 6th International Conference on Databases in Networked Information Systems*, DNIS'10, pages 274–283, Berlin, Heidelberg, 2010. Springer-Verlag.

[NHW12] S.M. Nasution, E.M. Husni, and A.I. Wuryandari. Prototype of train ticketing application using Near Field Communication (NFC) technology on Android device. In *System Engineering* *and Technology (ICSET), 2012 International Conference on*, pages 1–6, 2012.

[NWT12] ThaoP. Nghiem, AgustinusBorgy Waluyo, and David Taniar. A pure peer-to-peer approach for kNN query processing in mobile ad hoc networks. *Journal of Personal and Ubiquitous Computing*, 4:1–13, 2012.

[PC05] Wen-Chih Peng and Ming-Syan Chen. Design and Performance Studies of an Adaptive Cache Retrieval Scheme in a Mobile Computing Environment. *IEEE Transactions on Mobile Computing (TMC)*, 4(1):29–40, 2005.

[PCR+10] Zhipeng Peng, Huajun Chen, Jinghai Rao, Ying Liu, Lei Wang, and Jian Chen. Semantic-based Mobile Mashup Platform. In *9th International Semantic Web Conference (ISWC2010)*, November 2010.

[PGVA08] Prasanna Padmanabhan, Le Gruenwald, Anita Vallur, and Mohammed Atiquzzaman. A survey of data replication techniques for mobile ad hoc network databases. *Journal on Very Large Data Bases (VLDB)*, 17(5):1143–1164, 2008.

[PMG+11] Nilesh Padhariya, Anirban Mondal, Vikram Goyal, Roshan Shankar, and SanjayKumar Madria. EcoTop: An Economic Model for Dynamic Processing of Top-k Queries in Mobile-P2P Networks. In JeffreyXu Yu, MyoungHo Kim, and Rainer Unland, editors, *Database Systems for Advanced Applications*, volume 6588 of *Lecture Notes in Computer Science*, pages 251–265. Springer Berlin Heidelberg, 2011.

[Pon93] György Pongor. OMNeT: Objective Modular Network Testbed. In *Proceedings of the International Workshop on Modeling, Analysis, and Simulation On Computer and Telecommunication Systems*, MASCOTS '93, pages 323–326, San Diego, CA, USA, 1993. Society for Computer Simulation International.

[PSS06] Bambang Parmanto, Andi Saptono, and Lijing Song. Information summarization and transcoding of biomedical information resources for mobile handheld devices. *Journal of Mobile Multimedia*, 2(1):37–51, 2006.

[QMK10] B. Qureshi, Geyong Min, and D. Kouvatsos. M-Trust: A Trust Management Scheme for Mobile P2P Networks. In *Embedded and Ubiquitous Computing (EUC), 2010 IEEE/IFIP 8th International Conference on*, pages 476–483, 2010.

[RAP12] Maneesha V. Ramesh, Sruthy Anand, and Rekha P. Mobile Software Platform for Rural Health Enhancement. In *Proceedings of the 2012 International Conference on Advances in Mobile Network, Communication and Its Applications*, MNCAPPS '12, pages 131–134, Washington, DC, USA, 2012. IEEE Computer Society.

[RAT12] M.V. Ramesh, T.A. Anu, and H. Thirugnanam. An intelligent decision support system for enhancing an m-health application. In *Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on*, pages 1–5, 2012.

[RFJY03] Olga Ratsimor, Tim Finin, Anupam Joshi, and Yelena Yesha. eNcentive: a framework for intelligent marketing in mobile peer-to-peer environments. In *Proceedings of the 5th International Conference on Electronic commerce*, ICEC '03, pages 87–94, New York, NY, USA, 2003. ACM.

[RHTA10] Christoph Rieß, Norman Heino, Sebastian Tramp, and Sören Auer. EvoPat - pattern-based evolution and refactoring of RDF knowledge bases. In *Proceedings of the 9th International semantic web Conference on The semantic web - Volume Part I*, ISWC'10, pages 647–662, Berlin, Heidelberg, 2010. Springer-Verlag.

[RL03] L. Ramaswamy and Ling Liu. Free riding: a new challenge to peer-to-peer file sharing systems. In *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on*, pages 10 pp.–, 2003.

[RNC03] Bruno Richard, Donal Mac Nioclais, and Denis Chalon. Clique: A Transparent, Peer-to-Peer Replicated File System. In *Proceedings of the 4th International Conference on Mobile Data Management*, MDM '03, pages 351–355, London, UK, UK, 2003. Springer-Verlag.

[RPR96] David Ratner, Gerald J Popek, and Peter Reiher. The ward model: A scalable replication architecture for mobility. In *In Workshop on Object Replication and Mobile Computing*. Citeseer, 1996.

[RR09] S Krishna Mohan Rao and A Venugopal Reddy. Data Dissemination in Mobile Computing Environment. *BVICAM'S Journal of Information Technology, January-July*, 1(1):57–60, 2009.

[RRPK01] David Ratner, Peter Reiher, Gerald J. Popek, and Geoffrey H. Kuenning. Replication requirements in mobile environments. *Journal of Mobile Networks and Applications*, 6(6):525–533, November 2001.

[RSB11] U. Rathnayake, V. Sivaraman, and R. Boreli. Environmental context aware trust in mobile P2P networks. In *Local Computer Networks (LCN), 2011 IEEE 36th Conference on*, pages 324–332, 2011.

[RSKM09] Jedrzej Rybicki, Björn Scheuermann, Markus Koegel, and Martin Mauve. PeerTIS: a peer-to-peer traffic information system. In *Proceedings of the sixth ACM International workshop on VehiculAr InterNETworking*, VANET '09, pages 23–32, New York, NY, USA, 2009. ACM.

[Sch09] Claude Schenk, Eric; Guittard. Crowdsourcing: What can be Outsourced to the Crowd, and Why ? Technical report, HAL, 2009.

[SCHS07] Jang-Ping Sheu, Chih-Min Chao, Wei-Kai Hu, and Ching-Wen Sun. A Clock Synchronization Algorithm for Multihop Wireless Ad Hoc Networks. *Journal of Wireless Personal Communications*, 43(2):185–200, October 2007.

[SF04] D. Schoder and Kai Fischbach. Peer-to-peer paradigm. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, pages 1 pp.–, 2004.

[SGG01] Stefan Saroiu, P Krishna Gummadi, and Steven D Gribble. Measurement study of peer-to-peer file sharing systems. In *Electronic Imaging 2002*, pages 156–170. International Society for Optics and Photonics, 2001.

[SH04] Tobias Straub and Andreas Heinemann. An anonymous bonus point system for mobile commerce based on word-of-mouth recommendation. In *Proceedings of the 2004 ACM symposium on Applied computing*, SAC '04, pages 766–773, New York, NY, USA, 2004. ACM.

[SIC08] Mohamed A. Soliman, Ihab F. Ilyas, and Kevin Chen-Chuan Chang. Probabilistic top-k and ranking-aggregate queries. *ACM Transactions on Database Systems (TODS)*, 33(3):13:1–13:54, September 2008.

[SL03] Aameek Singh and Ling Liu. TrustMe: Anonymous Management of Trust Relationships in Decentralized P2P Systems. In

Proceedings of the 3rd International Conference on Peer-to-Peer Computing, P2P '03, pages 142–, Washington, DC, USA, 2003. IEEE Computer Society.

[SNCR03] V. Srinivasan, P. Nuggehalli, C. F Chiasserini, and R.R. Rao. Cooperation in wireless ad hoc networks. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 2, pages 808–817 vol.2, 2003.

[SPFT09] Vasilis Sourlas, Georgios S. Paschos, Paris Flegkas, and Leandros Tassiulas. Caching in content-based publish/subscribe systems. In *Proceedings of the 28th IEEE Conference on Global telecommunications*, GLOBECOM'09, pages 1401–1406, Piscataway, NJ, USA, 2009. IEEE Press.

[SS05] Jordi Sabater and Carles Sierra. Review on Computational Trust and Reputation Models. *Artificial Intelligence Review*, 24(1):33–60, September 2005.

[SWYX11] Leon Stenneth, Ouri Wolfson, Philip S. Yu, and Bo Xu. Transportation mode detection using mobile phones and GIS information. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '11, pages 54–63, New York, NY, USA, 2011. ACM.

[SYHN10] S. Sakashita, T. Yoshihisa, T. Hara, and S. Nishio. A Data Reception Method to Reduce Interruption Time in P2P Streaming Environments. In *Network-Based Information Systems (NBiS), 2010 13th International Conference on*, pages 166–172, 2010.

[SZS05] Sai Sun, Xiaofang Zhou, and Heng Tao Shen. Semantic caching for multiresolution spatial query processing in mobile environments. In *Proceedings of the 9th International Conference on Advances in Spatial and Temporal Databases*, SSTD'05, pages 382–399, Berlin, Heidelberg, 2005. Springer-Verlag.

[TR04] D. A. Turner and Keith W. Ross. A Lightweight Currency Paradigm for the P2P Resource Market. In *Seventh International Conference on Electronic Commerce Research*, June 2004.

[TZF12] Narisu Tao, Konglin Zhu, and Xiaoming Fu. Thumb: A real-time resource information sharing application over mobile phones. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, pages 528–530, 2012.

[VLdOC+10] P. B. Velloso, R. P. Laufer, D. de O Cunha, O. C.M.B. Duarte, and G. Pujolle. Trust management in mobile ad hoc networks using a scalable maturity-based model. *IEEE Transactions on Network and Service Management (TNSM)*, 7(3):172–185, September 2010.

[Whi09] Paul Whitla. Crowdsourcing and Its Application in Marketing Activities. *Contemporary Management Research*, 5:15–28, 2009.

[WXS04] Ouri Wolfson, Bo Xu, and A. Prasad Sistla. An Economic Model for Resource Exchange in Mobile Peer to Peer Networks. In *Proceedings of the 16th International Conference on Scientific and Statistical Database Management*, SSDBM '04, pages 235–, Washington, DC, USA, 2004. IEEE Computer Society.

[WXTL07] Minji Wu, Jianliang Xu, Xueyan Tang, and Wang-Chien Lee. Top-k Monitoring in Wireless Sensor Networks. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 19(7):962–976, July 2007.

[XLN06a] Yuan Xue, Baochun Li, and K. Nahrstedt. Optimal resource allocation in wireless ad hoc networks: A price-based approach. *IEEE Transactions on Mobile Computing (TMC)*, 5(4):347–364, 2006.

[XLN06b] Yuan Xue, Baochun Li, and Klara Nahrstedt. Channel-relay price pair: towards arbitrating incentives in wireless ad hoc networks: Research Articles. *Journal of Wireless Communications & Mobile Computing - Special Issue on Ad Hoc Wireless Networks*, 6(2):235–245, March 2006.

[XP03] Yuni Xia and Sunil Prabhakar. Q+Rtree: Efficient Indexing for Moving Object Databases. In *Proceedings of the Eighth International Conference on Database Systems for Advanced Applications*, DASFAA '03, pages 175–, Washington, DC, USA, 2003. IEEE Computer Society.

[XTL+11] Yang Xiao, Daisuke Takahashi, Jing Liu, Hongmei Deng, and Jingyuan Zhang. Wireless telemedicine and m-health: technologies, applications and research issues. *Journal of Sensor Networks*, 10(4):202–236, 2011.

[Xu06] S. Xu. Development and Test of Dynamic Congestion Pricing Model. Master's thesis, MIT, USA, 2006.

[XWR06] Bo Xu, Ouri Wolfson, and Naphtali Rishe. Benefit and Pricing of Spatio-Temporal Information in Mobile Peer-to-Peer Networks. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences - Volume 09*, HICSS '06, pages 223.2–, Washington, DC, USA, 2006. IEEE Computer Society.

[Yan12] L. Yang. *Stochastic Traffic Flow Modeling and Optimal Congestion Pricing*. PhD thesis, University of Michigan, USA, 2012.

[YMH+09] Tingxin Yan, Matt Marzilli, Ryan Holmes, Deepak Ganesan, and Mark Corner. mCrowd: a platform for mobile crowdsourcing. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, SenSys '09, pages 347–348, New York, NY, USA, 2009. ACM.

[ZCY03] Sheng Zhong, Jiang Chen, and Yang Richard Yang. Sprite: A simple, cheat-proof, credit-based system for mobile ad-hoc networks. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 3, pages 1987–1997. IEEE, 2003.

[ZCY06] Jianpei Zhang, Yan Chu, and Jing Yang. A category on the cache invalidation for wireless mobile environments. In *Proceedings of the 2006 International Conference on Advanced Web and Network Technologies, and Applications*, APWeb'06, pages 939–942, Berlin, Heidelberg, 2006. Springer-Verlag.

[Zun06] Microsoft Zune. `http://www.microsoft.com/presspass/presskits/zune/default.mspx`, November 2006.

[ZYV08] Demetrios Zeinalipour-Yazti and Zografoula Vagena. Distributed top-k query processing in wireless sensor networks. In *Mobile Data Management, 2008. MDM'08. 9th International Conference on*, pages 227–227. IEEE, 2008.