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

NILESH PADHARIYA[1], ANIRBAN MONDAL[2], SANJAY KUMAR MADRIA[3]
AND MASARU KITSUREGAWA[4]
[1]Atmiya Institute of Technology and Science, Gujarat, INDIA, [2]Shiv Nadar University, Delhi, INDIA, [3]Missouri University of Science and Technology, Rolla, USA, [4]University of Tokyo, JAPAN

In mobile ad-hoc peer-to-peer (M-P2P) networks, data availability is typically low due to rampant free-riding, frequent network partitioning and mobile resource constraints. *Rare* data items are those, which get sudden bursts in accesses based on *events* as they are only hosted by a few peers in comparison to the network size. Thus, they may not be available within few hops of query-issuing peers. 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. 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 this paper are three-fold. First, it provides incentives for replication of rare data items by means of a novel licensing mechanism, thereby improving rare data availability. Second, it provides additional incentives for MPs to collaborate in groups, thereby further improving rare data availability. Third, 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.

Keywords: Mobile Peer-to-Peer, economic incentive scheme, rare data availability

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. Proliferation of mobile devices (e.g., laptops, PDAs, mobile phones) coupled with the ever-increasing popularity of the P2P paradigm (e.g., BitTorrent) strongly motivate M-P2P applications. Mobile devices wirelessly communicating in a P2P fashion 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. Notably, improving the availability of rare data items is of paramount importance because although queries for such items do not arrive on a regular basis, they become very important when a given rare item becomes very popular (i.e., when a sudden burst of queries occur for the rare data item). Since relatively very few mobile peers would typically carry information about such rare items, improving the availability of rare data items by means of replication becomes critical.

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)

Authors' addresses: Nilesh Padhariya; Atmiya Institute of Technology and Science, Gujarat, INDIA, nilesh@aits.edu.in, Anirban Mondal; Shiv Nadar University, Delhi, INDIA, anirban.mondal@snu.edu.in, Sanjay Kumar Madria; Missouri University of Science and Technology, Rolla, USA, madrias@mst.edu, Masaru Kitsuregawa; University of Tokyo, JAPAN, kitsure@tkl.iis.u-tokyo.ac.jp

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. In our application scenarios, we assume that no cellular communication is possible (valid in many situations) and they do not support any such applications as well; so we have mobile ad hoc communication to share the information across the mobile peers. Given the absence of cellular infrastructure, we need to adapt mobile ad hoc networks as infrastructure to share the rare data items and their related information. Even if we assumed cellular infrastructure at such remote areas, it would not be practically feasible to provide mobile users with an app to provide such rare data item information. This is because the queries in mobile ad hoc networks are typically unbounded since we can never know in advance what kind of rare item information would become important to the users. Moreover, any user can issue a query for information about any rare data item; hence it is not practically feasible to store such information about all possible types of rare item information in a centralized system. 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.

Information about rare data items is generated by mobile users. For example, upon seeing gas-masks being sold at a shop, a mobile user may decide to capture some information about the shop and store such information in his/her mobile device. The mobile user (mobile peer) has an incentive to carry this information because when there arrives a sudden burst of queries for information about the rare data items, the user can earn high amounts of revenue because the prices of the rare items are significantly higher than that of non-rare items. In essence, the user is holding on to information about the rare item in the hope that when the rare item becomes hot, he/she will earn high amount of revenues. As we shall see shortly, our approach does not broadcast information about rare data items across the peers. Hence, a mobile peer knows about the existence of such data items in his/her vicinity by issuing a query (search) for information about the rare data item.

In this work, we assume an environment, where all the MPs collaborate on information sharing and are trusted. Notably, any distributed trust management schemes [Qureshi et al. 2010; Rathnayake et al. 2011; Singh and Liu 2003] 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.

Our application scenarios could be collaborative or non-collaborative applications. In case of collaborative applications, incentive mechanism still remains useful for resource allocation purposes e.g., even in collaborative environments, we need incentives in order to control the flooding by messages so that limited resources such as bandwidth, mobile peers' energy can be preserved. For example, in military application in remote areas, the communication bandwidth is limited and it generates many messages and that clogs the network. Thus, each peer is given some quota of tokens and based on self-evaluation about the importance of their respective messages, peers decide to communicate.

Similar to the works in [Hara and Madria 2006; Xia and Prabhakar 2003], 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 [Saroiu et al. 2001].) Rampant free-riding further reduces data availability since a large percentage of MPs are typically free-riders [Ham and Agha 2005; Kamvar et al. 2003] 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) [Hara and Madria 2006; Khan et al. 2008] do not consider economic incentives for data hosting, licensing mechanisms, M-P2P architecture and data item rarity issues. Incentive schemes for MANETs [Buttyán and Hubaux 2003; Chen and Nahrstedt 2004; Crowcroft et al. 2003; Srinivasan et al. 2003] primarily focus on encouraging message forwarding, but they do not address replication. M-P2P incentive schemes [Wolfson et al. 2004; Xu et al. 2006] 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*. 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 [Garyfalos and Almeroth 2004; Ratsimor et al. 2003; Straub and Heinemann 2004] 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 this paper are three-fold:

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

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

—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 [Turner and Ross 2004]. The works in [Daras et al. 2003; Elrufaie and Turner 2004; Zhong et al. 2003] 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 [Hara and Madria 2006], which is the closest 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 the paper is organized as follows. Section 2 reviews related works, while Section 3 details the economic model of E-Rare. Section 4 presents the ECR and ECR+ replication schemes. Section 5 reports our performance study. Finally, we conclude in Section 6 with directions for future work.

2. RELATED WORK

This section provides an overview of existing works. 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 mobile environments such as mobile ad hoc networks (MANETs) and M-P2P networks, 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.

Economic schemes for resource allocation: Economic schemes have been discussed for resource allocation in distributed systems [Ferguson et al. 1993; Ferguson et al. 1988; Kurose and Simha 1989]. However, they do not address M-P2P issues such as node mobility, free-riding, frequent network partitioning and mobile resource constraints. The proposals in [Liu and Issarny 2004; Xue et al. 2006a; 2006b] discuss economic schemes for resource allocation in wireless ad hoc networks. However, they do not consider replication and data rarity issues. Moreover, their focus is network-centric, while our focus is data-centric.

Schemes for static P2P networks: Replication schemes have been proposed for static P2P networks [Bhagwan et al. 2003; Datta et al. 2003b] and also for traditional distributed systems [Kemme and Alonso 2000]. Incentive-based schemes for encouraging peer participation in static P2P networks involve formal game-theoretic model for incentive-based P2P file-sharing systems [Golle et al. 2001], utility functions to capture peer contributions [Ham and Agha 2005; Ramaswamy and Liu 2003], EigenTrust scores to capture participation criteria [Kamvar et al. 2003] and asymmetric incentives based on disparities between upload and download bandwidths [Liebau et al. 2005]. However, these approaches are too static to be deployed in M-P2P networks because they assume peers' availability and fixed topology. Furthermore, they do not address mobile resource constraints (e.g., energy) and data rarity issues.

Non-incentive-based replication in MANETs: The proposals in [Hara and Madria 2006; 2005] discuss replication in MANETs. **E-DCG+** [Hara and Madria 2006] creates groups of 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 [Hara and Madria 2005] aims at classifying different replica consistency levels in a MANET based on application requirements, and proposes protocols to realize them. Consistency maintenance is performed via quorums and it is based on local conditions such as location and time. Incidentally, P2P replication suitable for mobile environments has been incorporated in systems such as ROAM [Ratner et al. 2001], Clique [Richard et al. 2003] and Rumor [Guy et al. 1998]. The work in [Sakashita et al. 2010] considers the reduction of delays in P2P streaming environments, but it is focused on lower level networking layer. Notably, these proposals do not consider any economic model, M-P2P architecture, data licensing and data rarity issues.

Incentive schemes for combating free-riding in MANETs: The proposals in [Buttyán and Hubaux 2001; 2003; Chen and Nahrstedt 2004; Crowcroft et al. 2003; Srinivasan et al. 2003] address free-riding in MANETs. The work in [Buttyán and Hubaux 2001] introduces a virtual currency to stimulate node cooperation. The works in [Buttyán and Hubaux 2003; Zhong et al. 2003] use virtual currency to stimulate the cooperation of mobile nodes in forwarding messages. The auction-based iPass [Chen and Nahrstedt 2004] incentive scheme and the works in [Crowcroft et al. 2003; Srinivasan et al. 2003] also provide incentives for relaying messages. In particular, the proposals in [Crowcroft et al. 2003; Srinivasan et al. 2003] concentrate on compensating forwarding cost in terms of battery power, memory and CPU cycles. However, these works do not consider data rarity issues, data item prices and incentives for data replication i.e., they do not entice peers to host data.

Non-incentive-based replication in M-P2P networks: The work in [Mondal et al. 2006b] has proposed 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 proposal in [Mondal et al. 2006a] has discussed both collaborative replica allocation and deallocation in tandem to facilitate optimal replication and to avoid 'thrashing' conditions. However, these proposals do not consider economic schemes and data rarity issues.

Incentive schemes for combating free-riding in M-P2P networks: The proposals in [Xu et al. 2006; Wolfson et al. 2004] discuss incentive schemes for combating free-riding in M-P2P networks. The work in [Xu et al. 2006] provides incentives to MPs 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 [Wolfson et al. 2004] considers opportunistic resource information dissemination in transportation application scenarios. An MP transmits its resources to the MPs that it encounters, and obtains resources from them in exchange. The works in [Wolfson et al. 2004; Xu et al. 2006] 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, licensing-based data replication and data rarity issues are not considered in [Wolfson et al. 2004; Xu et al. 2006].

Payment schemes: A small study [Mannak et al. 2004], 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 [Straub and Heinemann 2004]. The works in [Daras et al. 2003; Elrufaie and Turner 2004; Zhong et al. 2003] discuss how to ensure secure payments using a virtual currency. Another way proposed in [Garyfalos and Almeroth 2004] describes Coupons, an incentive scheme that is inspired by the eNcentive framework [Ratsimor et al. 2003], which allows mobile agents to spread digital advertisements with embedded coupons among mobile users in a P2P manner.

Several non-repudiation [Kremer et al. 2002; Sabater and Sierra 2005] 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 [Figueiredo et al. 2004]. MoB [Chakravorty et al. 2005] 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 [Datta et al. 2003a].

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 [Chen and Nahrstedt 2004] 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. Notably, these secure payment schemes are complementary to our proposal, but they can be used in conjunction with our proposal.

3. E-RARE: AN ECONOMIC INCENTIVE MODEL FOR RARE DATA ITEMS

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 $(d,\ L,\ \tau_Q)$, 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 [Doriguzzi Corin et al. 2014]. Several existing content authoring techniques can be used for license protection and restricted distribution [Jokela 2003].

3.1 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:

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

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. In this regard, we performed many experiments for several scenarios with different values for N. Based on the results of those experiments, we found that for values of N that are higher than 5, the value of the rarity score does not change significantly. In fact, in most of our experiments, we found that the value of the rarity score does not change beyond N=3. Hence, we used N=5 in this work to capture the outlier cases.

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 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.

Incidentally, the computation of rarity scores requires a distributed collection protocol among all the peers in the network. The protocol here would be for each peer to maintain in its log a counter for the download of items at itself. Using this counter, the information about the access frequency of a given data item is sent to its owner. (Notably, the owner of a given data item is the MP who provided the first copy of the item. In other words, the owner is the MP that generated and responded to the first query on the given data item.) Observe that this protocol incurs communication overhead. We can reduce the communication overhead by piggybacking onto other essential system status messages.

3.2 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_lice-nse_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. Based on our understanding of the problem and general application scenarios, we envisaged these possible classes of service, but there could be more classes of service possible based on the application. We have tried to justify the different levels of services through price distribution related to rare data in M-P2P networks. We assume that users are versatile in terms of rare data usage, thus each price is in increasing order of its related data item usage. Furthermore, the licensing mechanism also attracts those mobile peers, who have little or no interest in rare data item usage, but they may involve in its distribution process, thereby increasing rare data availability in M-2P networks.

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. This assumption is indeed valid in practice and can be realized by using Digital Rights Management (DRM) software at the mobile devices of the peers. Notably, there are several examples of existing systems, which allow only a pre-specified number of downloads. We assume that Software such as DRM can be used to control this. 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 [Doriguzzi Corin et al. 2014]. 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.

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 I: Summary of Notations

For the sake of convenience, Table I summarizes the notations used in this paper. 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.

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:

$$P_{d,Q} = \begin{cases} (\ \lambda_d \times e^{\tau_Q / R_Q}\ ) & if \ \ R_Q \leq \tau_Q \\ 0 & otherwise \end{cases} \tag{2}$$

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:

$$F_{d,Q} \ = \ P_{d,Q} \times \Upsilon \tag{3}$$

In Equation 3, the value of $P_{d,Q}$ is computed from Equation 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:

$$PUL_{d,Q} \ = \ \mu_{P_{d,Q}} \ + \ (\ \mu_{P_{d,Q}} \ + \ \mu_{F_{d,Q}}\ ) \ / \ \mu_{P_{d,Q}} \tag{4}$$

$$FUL_{d,Q} \ = \ \mu_{F_{d,Q}} \ + \ (\ \mu_{P_{d,Q}} \ + \ \mu_{F_{d,Q}}\ ) \ / \ \mu_{P_{d,Q}} \tag{5}$$

In Equations 4 and 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. In case of the owner encountering a failure or getting disconnected from the network (e.g., due to network partitioning), we assume that there is a reasonable range of prices for data items, and these are well-known among all the peers in the network. Furthermore, the owner can connect back to the network eventually and once the owner connects, it can collect information about the price history, thereby facilitating in the computation of the prices.

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 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 5, $FUL_{d,Q}$ always exceeds $\mu_{F_{d,Q}}$.

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

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

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

3.3 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* $\mu_{P_{d,Q}}$ at the owner MP of the given data item. Observe that we intentionally set the price for relaying messages to be significantly lower than the price of a given data item for better incentivizing replica hosting in comparison to message relaying. In essence, our intent here is that replica hosting should be more incentivized than message relaying. Hence, we keep the relay price K to a low fixed value (fixed at 5% of the average price) so that MPs have some incentive to relay messages. However, since our key objective here is to incentivize replica hosting, hence the message relay incentives are secondary to our goals. Hence, we did not perform any experiments to observe the impact of varying the value of K. 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:

$$E_M = \sum_{i=1}^{p} [(ns_{P_i} \times P_i) + (ns_{F_i} \times F_i) + (ns_{PUL_i} \times PUL_i) + (ns_{FUL_i} \times FUL_i)] + (m \times K) \tag{6}$$

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_price* (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:

$$S_M = [\sum_{i=1}^{nq_P} P_i] + [\sum_{i=1}^{nq_F} F_i] + [\sum_{i=1}^{nq_{PUL}} PUL_i] + [\sum_{i=1}^{nq_{FUL}} FUL_i] + (n \times K) \tag{7}$$

In Equation 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 and 7, the revenue ω of M is computed below:

$$\omega = E_M - S_M \tag{8}$$

4. 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.

4.1 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

[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.

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.

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. This is because over the course of time, each MP would like to host data items with higher rarity score values in order to improve its chances of earning more revenues when such rare items become popular. Alternatively, if an MP decided to set the value of its TH_R threshold to the maximum rarity score value (among all the items that it hosts), it would likely decrease its future opportunities for earning revenues because it would prevent it from serving the requests of many peers. On the other hand, if an MP decided to set the value of its TH_R threshold to the minimum rarity score value (among all the items that it hosts), it would also likely decrease its future opportunities for earning revenues since it would imply that over the course of time, the MP would eventually be hosting less rare data items. Hence, we believe that setting the TH_R threshold to the average of the rarity scores of the items that it currently hosts would be reasonable for a given MP for real-world application scenarios.

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. Figure 1 depicts the algorithm used by a given MP for determining which data items to host at itself.

Algorithm *ECR_an_MP M*

/* 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 */
Receive broadcasted list B_R of the data items from M's neighbours
Merge B_R with M's own list of available data items in A_R
Sort A_R in descending order of data items' rarity scores

for each item i in A_R
    if MEM ¿ 0 and λ_i ¿ TH and i is not already resident
        Store i in MEM
        MEM = MEM - sizeof[i]
        Add i to purchased list P_R
    else
        break

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

end

Algorithm for an MP M

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, which is 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.

4.2 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 [Lubke et al. 2011] and Team-Formation [Ambroszkiewicz et al. 1998; Hsu and Liu 2005] schemes can be used in conjunction with our proposal.

Group members periodically broadcast their list of items to members within their group. In this work, we use flooding to perform the broadcasting of the information within the group. We assume that these broadcast messages are received by all the MPs within the group. In particular, we do not handle message loss explicitly, assuming that the underlying broadcast/multicast message protocol will take care of this issue. In other words, our focus is on the application side assuming that networking protocol will take care of lower level issues. 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 1 by additionally considering the number ξ of MPs (group members) that host d. The idea of using the parameter ξ is to compute an estimate of the rarity score of a given data item based on information available within the group and then trying to extrapolate this score as an average estimate of the item's rarity score across the network. 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. Note that the parameter ξ does not necessarily imply better approximation of the average number of queries. Thus, our computation of ξ represents an inevitable compromise for defining rarity in the absence of complete information in decentralized settings.

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, ECR+ computes λ_d based on more information than ECR. In ECR+, each MP computes the rarity score λ_d for each data item d (that it hosts) as follows:

$$\lambda_d \ = \ [\ \{\ (\ \eta_c - (\ \tfrac{1}{N}\sum_{i=1}^{N} \eta_i\ )\ )\ /\ (\ max\ (\ \eta_c\ ,\ \tfrac{1}{N}\sum_{i=1}^{N} \eta_i\ )\ \times\ \xi\ )\ \}\ +\ 1\ ]\ /\ 2 \tag{9}$$

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 1, observe that the range of λ_d in Equation 9 is between 0 and 1.

4.3 Illustrative example of peer groups in E-Rare

Figure 2 depicts an illustrative example of an instance of network topology in ECR+. Now we shall use Figure 2 to illustrate how groups facilitate the improvement of rare data availability. In Figure 2, 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.

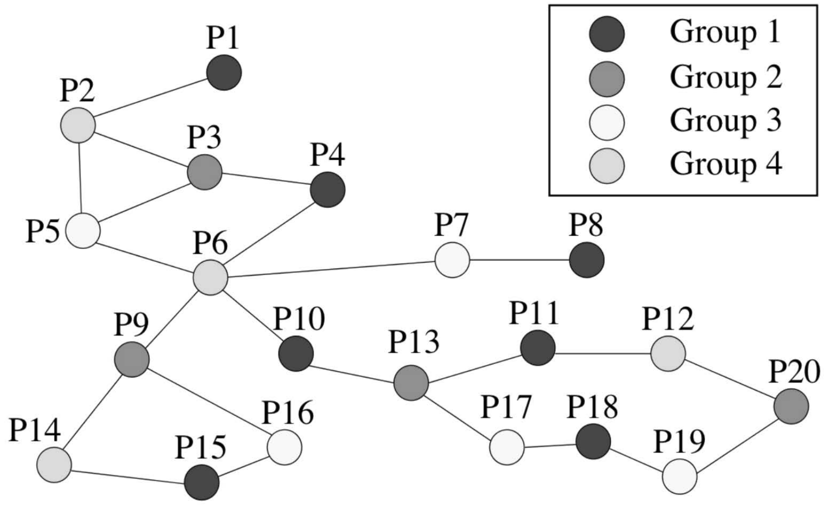
Suppose peer P18 sees high access count for an item d, which it does not own or host. Additionally, suppose d is owned and hosted by one of its group members, say P1. For simplicity, assume that no replica of d exists at any other peer in the M-P2P network[3]. In this scenario, queries on d initiated 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

---

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

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



Peer groups in E-Rare

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.

4.4 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 3. 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 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_lice-nse_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 2, 3, 4 and 5 (see Section 3) as follows:

$$DP_{d,Q} = P_{d,Q} \times (1 - \delta) \tag{10}$$

$$DF_{d,Q} = F_{d,Q} \times (1 - \delta) \tag{11}$$

$$DPUL_{d,Q} = PUL_{d,Q} \times (1 - \delta) \tag{12}$$

$$DFUL_{d,Q} = FUL_{d,Q} \times (1 - \delta) \tag{13}$$

where $0 \leqslant \delta < 1$.

4.5 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 *ECR+_Licensor_MP*

```
/* φ is an item's revenue-earning potential */
Sort all its items in descending order of φ
Compute the average value φ_avg of all its items
Select items for which φ exceeds φ_avg into a list Lic
/* Lic is the set of items for licensing */
for each item i in Lic
    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 */
Broadcast the list Lic upto its n-hop neighbours
for each item i in Lic
    Wait for replies from potential licensees
    Receive replies from potential licensees
    for each potential licensee j
        Calculate the value of Ω for j
    Sort the licensees in descending order of Ω into a list P_L
    for each licensee j in P_L
        if ( A[i]_PUL = 0 ) break
        if ( A[i]_PUL − N[j]_PUL ⩾ 0 )
            Send N[j]_PUL licenses of i to j
            A[i]_PUL = A[i]_PUL − N[j]_PUL
        else
            Send A[i]_PUL licenses of i to j
            A[i]_PUL = 0
    for each licensee j in P_L
        if ( A[i]_FUL = 0 ) break
        if ( A[i]_FUL − N[j]_FUL ⩾ 0 )
            Send N[j]_FUL licenses of i to j
            A[i]_FUL = A[i]_FUL − N[j]_FUL
        else
            Send A[i]_FUL licenses of i to j
            A[i]_FUL = 0
end
```

Algorithm for *Licensor_MP* M in ECR+

Algorithm *ECR+_Licensee_MP*

```
(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 ≤ 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
```

Algorithm for *Licensee_MP* M_E in ECR+

Figure 3 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 potential 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.

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:

$$\Omega \;=\; [\, w_1 \times energy \,] \;+\; [\, w_2 \times n_{hop} \,] \tag{14}$$

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 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

[4]Since E-Rare considers four types of item prices, the respective products of access counts and each price type are summed up to obtain the value of ϕ.

$w_1 = w_2 = 0.5$.

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 4 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.

4.6 Illustrative example of licensing in ECR+

Figure 5 depicts an illustrative example of licensing in ECR+. From Figure 5a, 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 5b depicts the license set *Lic*

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

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

$φ_{avg}$ = 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

Illustrative example of licensing in E-Rare comprising items {36, 92, 53} along with the number of available PUL and FUL licenses for each item in *Lic*. Figure 5c 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: {P1, P2, P3}.

Figure 5d 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.

5. PERFORMANCE EVALUATION

This section reports the performance of our incentive-based replication schemes by means of simulation using OMNET++ [Pongor 1993]. We assume that MPs move according to the *Random Waypoint Model* [Broch et al. 1998] 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 III summarizes the parameters used in our performance evaluation. Notably, we have looked into the literature [Hara and Madria 2005; 2006] to understand the different parameters. Based on our understanding of our application environment, we have selected these parameters.

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 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 3 or 4, 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

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 III: Performance Study Parameters

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)** in M-P2P network. 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 ($(N_S/N_Q) \times 100$), where N_S represents the number of successful queries. We define the query hop-count HC as the average hop-count incurred by the query in the successful query path. Thus, HC equals ($(1/N_Q) \sum_{i=1}^{N_Q} HC_i$), where HC_i represents the hop-count incurred by the i^{th} query. HC 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 [Hara and Madria 2006], 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.

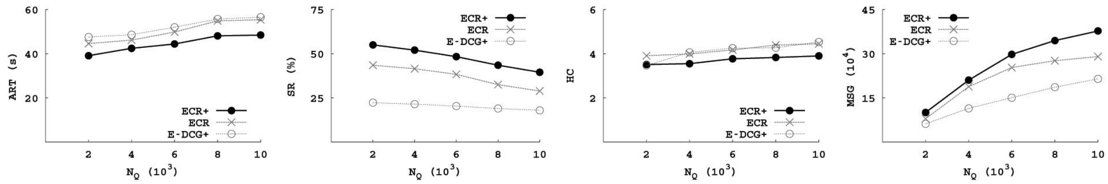
We have implemented E-DCG+ in E-Rare as follows. E-DCG+ performs the periodic broadcast to perform replication. MP obtains the data items list with their respective rarity scores. Based on rarity scores of data items and MP's available memory space, each MP hosts data items in their decending order of 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.

5.1 Performance of E-Rare

Figure 6 depicts the results of our experiments using default values of the parameters in Table III. 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 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.



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

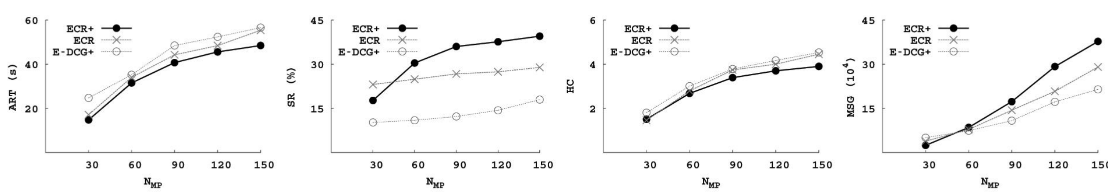
Performance of E-Rare: Data availability and communication overhead

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 6c 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 6d 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 6b), which result in decreased amount of data transfer, albeit at the cost of reduced SR.

5.2 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 7 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 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.



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

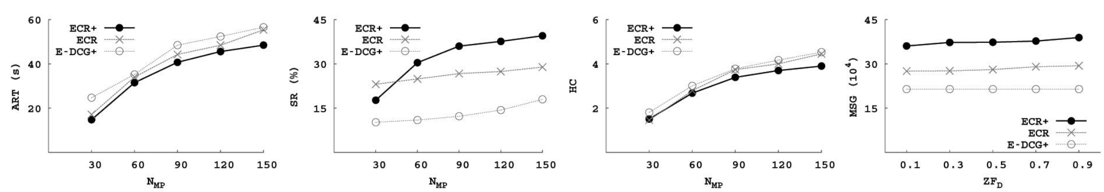
Effect of variations in the number of MPs

ECR+ performs better than ECR due to the reasons explained for Figure 6. 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. Observe that if there was no memory space constraint and no competition among copies of rare data items for memory space in the mobile peers, the value of SR for ECR+ would eventually have reached 100% as we keep increasing the number of mobile peers assuming that most of the peers are not disconnected. However, each of the mobile peers in the groups formed under the ECR+ approach have limited memory space and there is also competition among the copies of rare data items for memory space within the mobile peers in a given group. Hence, after increasing the number of mobile peers beyond a certain point, the success rate SR exhibits a plateau because of the implicit upper limit imposed upon SR since an adequate number of copies of all rare data items cannot be stored in the mobile peers of the group due to the memory space constraints of these peers and also due to competition among the rare items for the limited memory space across all mobile peers in the group.

Observe that for small values of N_{MP}, ECR performs better than ECR+ because the lower number of total MPs implies that the number of MPs in each of the groups in ECR+ is also low; hence few rare data items can be hosted by any of the groups in ECR+. However, for values of N_{MP} beyond 50, ECR+ outperforms ECR because now the groups in ECR+ have enough members to host most of the rare data items in the M-P2P network. Hence, rare data availability is also increased in this case, thereby improving SR for ECR+ as compared to that of ECR.

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

Figure 8 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 8. ART follows a pattern similar to HC for each of the three approaches.



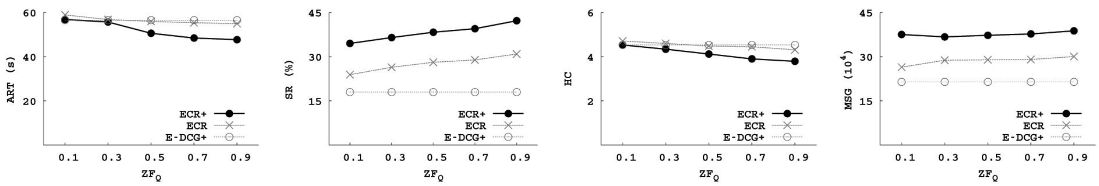
(a) ART (b) SR (c) HC (d) MSG

Effect of variations in the data distribution across rarity classes

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 8a and 8c 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.

5.4 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 9 depict the effect of variations in ZF_Q.



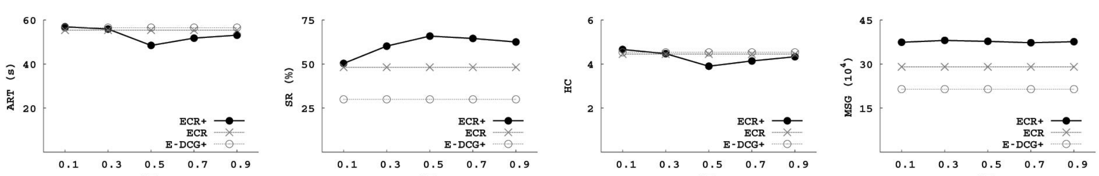
(a) ART (b) SR (c) HC (d) MSG

Effect of variations in the query distribution for rare items

The results in Figure 9 indicate that as ZF_Q increases, ART and HC both 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 6d.

5.5 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 10 depicts the results of variations in ZF_G.



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

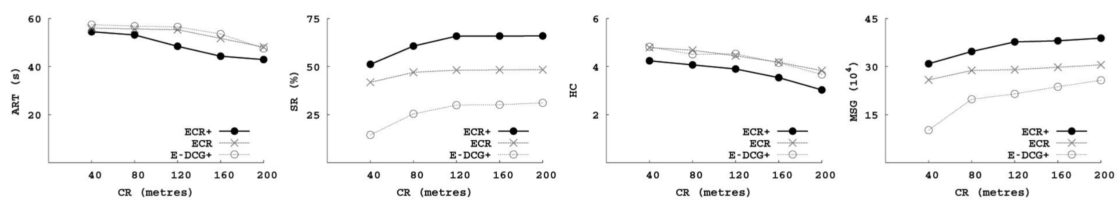
Effect of variations in the interest group sizes

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.

5.6 Effect of variations in the communication range

The results in Figure 11 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 11c 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 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.



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

Effect of variations in the communication range

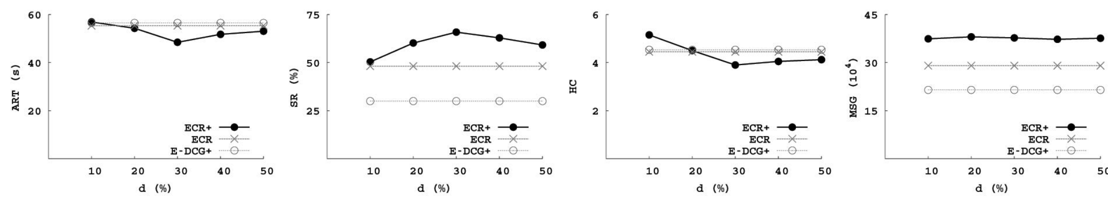
As CR increases, SR increases for all the approaches upto a certain point and then saturates. The increase in SR occurs essentially due to MPs being ‘nearer’ in effect with increase in CR, thereby making data items more accessible to query-issuing MPs. A relatively lower number of queries fail due to 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.

5.7 Effect of variations in the discount δ

The results in Figure 12 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 the value of δ is close to 30%.



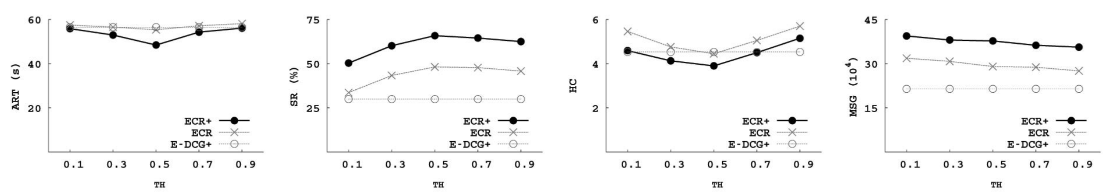
(a) ART (b) SR (c) HC (d) MSG

Effect of variations in the discount δ

5.8 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. Here, we are not trying to estimate an optimal value for f_{TH}. Instead, we are examining the impact on the performance after a certain number of queries have arrived. Incidentally, replication should be initiated when the average query response time becomes high or the average rare data availability becomes low w.r.t. the requirements of the application scenarios. Observe that the optimal time for the initiation of replication is likely to be difficult to estimate in advance in highly variable scenarios.

The total number of issued queries in our experiments 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.



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

Effect of variations in the access count threshold f_{TH}

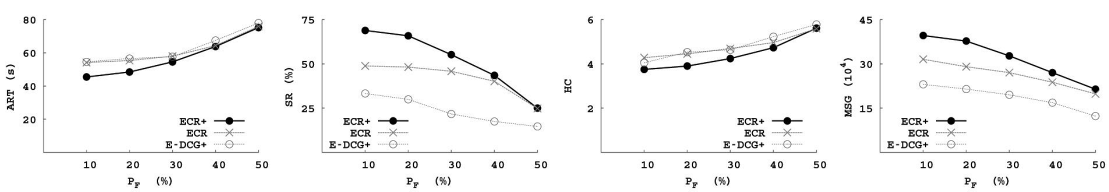
The results in Figure 13 indicate that ECR+ and ECR both perform best in terms of ART, SR and HC at $f_{TH} = 0.5$. However, as the value of f_{TH} keeps deviating away from 0.5 the performance of ECR and ECR+ both degrade. This is because at low values of f_{TH} (e.g., $f_{TH} =$ 0.1) relatively non-rare 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 13d indicate, MSG decreases slightly for both ECR and ECR+ with increase in f_{TH}.

The key insight that we obtain from the results of this experiment is that a “good” time to initiate replication would be somewhere in the middle of the query burst, and this need not necessarily be exact as in f_{TH}=0.5 (i.e., when half the expected number of queries have already come in) across all kinds of highly variable scenarios. We can obtain a reasonably good approximation about the expected number of queries based on statistical historical information in our application scenarios. Note that here, we do not make any claims on optimality of the initiation time for replication because in highly variable scenarios, such optimality is extremely challenging to derive. Our goal here is to determine a reasonably “good” time to initiate replication such that the impact of replication on rare data availability is significantly pronounced.

5.9 Effect of MP failures

We conducted an experiment to investigate the effect of MP failures[5] on the performance of E-Rare. Figure 14 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.



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

Effect of MP failures

From Figures 14a, 14b and 14c, 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 6d.

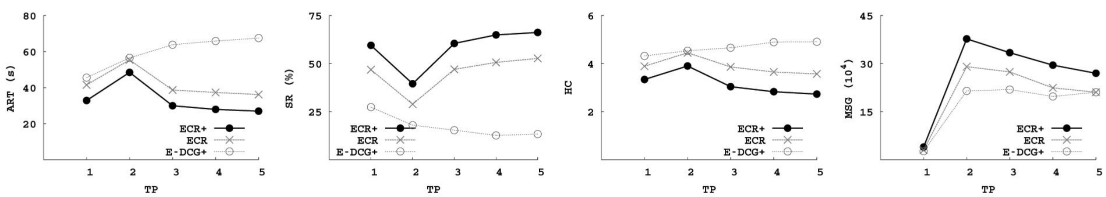
5.10 Effect of sudden bursts on a single data item

We conducted an experiment to demonstrate the effect of sudden bursts for a *single* data item. Figure 15 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

[5]MPs can fail due to reasons such as depletion of their limited energy resources.

time-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}.



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

Effect of sudden bursts on a single data item

In Figure 15, 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 15 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 6d. 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. CONCLUSION

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. Observe that in this work, we have used a greedy strategy for licensing based on heuristics. In the near future, we plan to refine this greedy strategy by using classical optimization and game-theoretic techniques, which are aimed at maximizing the global revenues, for rare data item pricing. We also plan to compare the performance of E-Rare for different economic models.

REFERENCES

Ambroszkiewicz, S., Matyja, O., and Penczek, W. 1998. Team Formation by Self-Interested Mobile Agents. *Multi-Agent Systems: Theories, Languages and Applications 1544*, 1–15.

Bhagwan, R., Moore, D., Savage, S., and Voelker, G. M. 2003. Future directions in distributed computing. A. Schiper, A. A. Shvartsman, H. Weatherspoon, and B. Y. Zhao, Eds. 153–158.

Broch, J., Maltz, D. A., Johnson, D. B., Hu, Y.-C., and Jetcheva, J. 1998. A performance comparison of multi-hop wireless ad hoc network routing protocols. In *Mobile Computing and Networking (MobiCom)*. 85–97.

Buttyán, L. and Hubaux, J.-P. 2001. Nuglets: a Virtual Currency to Stimulate Cooperation in Self-Organized Mobile Ad Hoc Networks. *Technical Report No. DSC/2001/001, Swiss Federal Institute of Technology (EPFL)*.

Buttyán, L. and Hubaux, J.-P. 2003. Stimulating cooperation in self-organizing mobile ad hoc networks. *Journal of Mobile Networks and Applications 8,* 5, 579–592.

Chakravorty, R., Agarwal, S., Banerjee, S., and Pratt, I. 2005. MoB: a mobile bazaar for wide-area wireless services. In *Mobile Computing and Networking (MobiCom)*. 228–242.

Chen, K. and Nahrstedt, K. 2004. iPass: An Incentive Compatible Auction Scheme to Enable Packet Forwarding Service in MANET. In *Distributed Computing Systems (ICDCS)*. 534–542.

Crowcroft, J., Gibbens, R., Kelly, F., and Ostring, S. 2003. Modelling Incentives for Collaboration in Mobile Ad hoc Networks. In *WiOpt*. 427–439.

Daras, P., Palaka, D., Giagourta, V., and Bechtsis, D. 2003. A Novel Peer-to-Peer Payment Protocol. In *EUROCON*. Vol. 1. 2–6.

Datta, A., Hauswirth, M., and Aberer, K. 2003a. Beyond "Web of trust": enabling P2P e-commerce. In *IEEE E-Commerce (ICEC)*. 303–312.

Datta, A., Hauswirth, M., and Aberer, K. 2003b. Updates in Highly Unreliable, Replicated Peer-to-Peer Systems. In *Distributed Computing Systems (ICDCS)*. 76–85.

Doriguzzi Corin, R., Russello, G., and Salvadori, E. 2014. *Security for Multihop Wireless Networks*.

Elrufaie, E. and Turner, D. A. 2004. Bidding in P2P Content Distribution Networks using the Lightweight Currency Paradigm. In *Information Technology: Coding and Computing (ITCC)*. 129–130.

Ferguson, D., Nikolaou, C., and Yemini, Y. 1993. An economy for managing replicated data in autonomous decentralized systems. In *Autonomous Decentralized Systems*. 367–375.

Ferguson, D., Yemini, Y., and Nikolaou, C. 1988. Microeconomic algorithms for load balancing in distributed computer systems. In *Distributed Computing Systems*. 491–499.

Figueiredo, D. R., Shapiro, J. K., and Towsley, D. 2004. Payment-based incentives for anonymous peer-to-peer systems. In *UMass CMPSCI Technical, Report 04:62*.

Garyfalos, A. and Almeroth, K. C. 2004. Coupon Based Incentive Systems and the Implications of Equilibrium Theory. In *E-Commerce Technology (CEC)*. 213–220.

Golle, P., Brown, K., and Mironov, I. 2001. Incentives for Sharing in Peer-to-Peer Networks. In *Electronic Commerce*. 264–267.

Guy, R., Reiher, P., Ratner, D., Gunter, M., Ma, W., and Popek, G. 1998. Rumor: Mobile Data Access Through Optimistic Peer-to-Peer Replication. In *ER Workshops*. 254–265.

Ham, M. and Agha, G. 2005. ARA: A Robust Audit to Prevent Free-Riding in P2P Networks. In *Peer-to-Peer Computing (P2P)*. 125–132.

Hara, T. and Madria, S. 2005. Consistency management among replicas in peer-to-peer mobile ad hoc networks. In *Reliable Distributed Systems (SRDS)*. 3–12.

Hara, T. and Madria, S. 2006. Data Replication for Improving Data Accessibility in Ad Hoc Networks. *IEEE Transactions on Mobile Computing (TMC) 5,* 11, 1515–1532.

Hsu, H.-H. and Liu, A. 2005. Multiagent-Based Multi-team Formation Control for Mobile Robots. *Journal of Intelligent and Robotic Systems 42,* 4, 337–360.

Jokela, T. 2003. Authoring tools for mobile multimedia content. In *Multimedia and Expo*. 637–640.

Kamvar, S., Schlosser, M., and Garcia-Molina, H. 2003. Incentives for Combatting Free-riding on P2P Networks. In *Technical report, Stanford University*.

Kemme, B. and Alonso, G. 2000. A New Approach to Developing and Implementing Eager Database Replication Protocols. *ACM TODS 25,* 3, 333–379.

Khan, S., Maciejewski, A., Siegel, H., and Ahmad, I. 2008. A game theoretical data replication technique for mobile ad hoc networks. In *Parallel and Distributed Processing*. 1–12.

Kremer, S., Markowitch, O., and Zhou, J. 2002. An intensive survey of fair non-repudiation protocols. *Journal of Computer Communications 25,* 17, 1606–1621.

Kurose, J. F. and Simha, R. 1989. A Microeconomic Approach to Optimal Resource Allocation in Distributed Computer Systems. *IEEE Transactions on Computers (TOC) 38,* 5, 705–717.

Liebau, N., Darlagiannis, V., Heckmann, O., and Steinmetz, R. 2005. Asymmetric Incentives in Peer-to-Peer Systems. In *AMCIS*. 260–268.

Liu, J. and Issarny, V. 2004. Service allocation in selfish mobile ad hoc networks using vickrey auction. In *Current Trends in Database Technology (EDBT)*. 385–394.

Lubke, R., Schuster, D., and Schill, A. 2011. MobilisGroups: Location-based group formation in Mobile Social Networks. In *Pervasive Computing and Communications Workshops (PERCOM Workshops)*. 502–507.

Mannak, R., de Ridder, H., and Keyson, D. V. 2004. The human side of sharing in peer-to-peer networks. In *European Union Symposium on Ambient Intelligence (EUSAI)*. 59–64.

Mondal, A., Madria, S., and Kitsuregawa, M. 2006a. CADRE: A Collaborative replica allocation and deallocation approach for Mobile-P2P networks. In *International Database Engineering and Applications Symposium (IDEAS)*. 21–28.

Mondal, A., Madria, S., and Kitsuregawa, M. 2006b. CLEAR: An Efficient Context and Location-Based Dynamic Replication Scheme for Mobile-P2P Networks. In *Database and Expert Systems Applications (DEXA)*. 399–408.

Pongor, G. 1993. OMNeT: Objective Modular Network Testbed. In *Modeling, Analysis, and Simulation On Computer and Telecommunication Systems (MASCOTS)*. 323–326.

Qureshi, B., Min, G., and Kouvatsos, D. 2010. M-Trust: A Trust Management Scheme for Mobile P2P Networks. In *Embedded and Ubiquitous Computing (EUC)*. 476–483.

Ramaswamy, L. and Liu, L. 2003. Free riding: a new challenge to peer-to-peer file sharing systems. In *System Sciences*. 10.

Rathnayake, U., Sivaraman, V., and Boreli, R. 2011. Environmental context aware trust in mobile P2P networks. In *Local Computer Networks (LCN)*. 324–332.

Ratner, D., Reiher, P., Popek, G. J., and Kuenning, G. H. 2001. Replication requirements in mobile environments. *Journal of Mobile Networks and Applications 6,* 6, 525–533.

Ratsimor, O., Finin, T., Joshi, A., and Yesha, Y. 2003. eNcentive: a framework for intelligent marketing in mobile peer-to-peer environments. In *Electronic commerce (ICEC)*. 87–94.

Richard, B., Nioclais, D. M., and Chalon, D. 2003. Clique: A Transparent, Peer-to-Peer Replicated File System. In *Mobile Data Management (MDM)*. 351–355.

Sabater, J. and Sierra, C. 2005. Review on Computational Trust and Reputation Models. *Artificial Intelligence Review 24,* 1, 33–60.

Sakashita, S., Yoshihisa, T., Hara, T., and Nishio, S. 2010. A Data Reception Method to Reduce Interruption Time in P2P Streaming Environments. In *Network-Based Information Systems (NBiS)*. 166–172.

Saroiu, S., Gummadi, P. K., and Gribble, S. D. 2001. Measurement study of peer-to-peer file sharing systems. In *Electronic Imaging*. 156–170.

Singh, A. and Liu, L. 2003. TrustMe: Anonymous Management of Trust Relationships in Decentralized P2P Systems. In *Peer-to-Peer Computing (P2P)*. 142–149.

Srinivasan, V., Nuggehalli, P., Chiasserini, C. F., and Rao, R. 2003. Cooperation in wireless ad hoc networks. In *INFOCOM*. Vol. 2. 808–817.

Straub, T. and Heinemann, A. 2004. An anonymous bonus point system for mobile commerce based on word-of-mouth recommendation. In *ACM Symposium on Applied Computing (SAC)*. 766–773.

Turner, D. A. and Ross, K. W. 2004. A Lightweight Currency Paradigm for the P2P Resource Market. In *Electronic Commerce Research (ECR)*.

Wolfson, O., Xu, B., and Sistla, A. P. 2004. An Economic Model for Resource Exchange in Mobile Peer to Peer Networks. In *Scientific and Statistical Database Management (SSDBM)*. 235–244.

Xia, Y. and Prabhakar, S. 2003. Q+Rtree: Efficient Indexing for Moving Object Databases. In *Database Systems for Advanced Applications (DASFAA)*. 175–182.

Xu, B., Wolfson, O., and Rishe, N. 2006. Benefit and Pricing of Spatio-Temporal Information in Mobile Peer-to-Peer Networks. In *System Sciences (HICSS)*. Vol. 9. 223b–223b.

Xue, Y., Li, B., and Nahrstedt, K. 2006a. 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.

Xue, Y., Li, B., and Nahrstedt, K. 2006b. Optimal resource allocation in wireless ad hoc networks: A price-based approach. *IEEE Transactions on Mobile Computing 5,* 4, 347–364.

Zhong, S., Chen, J., and Yang, Y. R. 2003. Sprite: A simple, cheat-proof, credit-based system for mobile ad-hoc networks. In *INFOCOM*. Vol. 3. 1987–1997.

Nilesh Padhariya is an associate professor in computer engineering at Atmiya Institute of Technology and Science for Diploma Studies, Gujarat, INDIA. He has completed his Ph.D. in Computer Science at Indraprstha Institute of Information Technology, Delhi (IIIT-D), INDIA. He earned his MTech in Computer Applications in 2006 from Indian Institute of Technology, Delhi (IIT-D), INDIA. His work addresses the efficient data management using effective economic incentive-based schemes in mobile peer-to-peer (M-P2P) networks. It includes the dynamic query processing and the data replication in M-P2P networks using economic schemes. His research interests include mobile computing, crowdsourcing, cloud computing, BigData, systems of engagements etc. His works have been published at prestigious conferences and peer-reviewed journals from IEEE, ACM, ScienceDirect etc. He has also received several grants for research and publications.

Anirban Mondal is an associate professor at Shiv Nadar University, India. Prior to this, he was a Senior Research Scientist at Xerox Research Centre India. His expertise is in the area of distributed systems with focus on large-scale data management and in domains such as smart cities and financial services. Prior to joining Xerox, he had been an associate professor at IIIT Delhi forthree years. He also had a long tenure of seven years at the University of Tokyo, Japan, where he worked on mobile-P2P incentive models for crowdsourcing applications, indexing of large-scale spatial data and load balancing in large-scale distributed systems. Anirban has numerous publications in key conferences/journals, where he also maintains an active level of involvement as PC Chair/Co-chair, PC member, journal reviewer and keynote/tutorial speaker. His awards include the prestigious JSPS (Japanese Society for Promotion of Science) Fellowship as well as a DST Fast Track project for Young Scientists of India. He is an ACM India Eminent Speaker. Anirban completed his Bachelor's degree in Computer Science and Engineering from IIT Kharagpur (India), and his PhD degree in Computer Science from the National University of Singapore (NUS). He also has an MBA degree from the University of Massachusetts Amherst (UMass), USA.

Sanjay Kumar Madria received his Ph.D. in Computer Science from Indian Institute of Technology, Delhi, India in 1995. He is a full professor in the Department of Computer Science at the Missouri University of Science and Technology (formerly, University of Missouri-Rolla, USA). He has published over 225+ Journal and conference papers in the areas of mobile data management, cloud and sensor computing, and cyber security and trust management. He won five best paper awards including in IEEE SRDS, and IEEE MDM conferences. He is the co-author of a book published by Springer in Nov 2003. His research is supported by several grants from federal sources such as NSF, DOE, AFRL, ARL, ARO, NIST and industries like Boeing, Unique*Soft, etc. He has been awarded JSPS (Japanese Society for Promotion of Science) visiting scientist fellowship in 2006 and ASEE (American Society of Engineering Education) fellowship at AFRL from 2008 to 2015. In 2012-13, he was awarded NRC Fellowship by National Academies. He has received faculty excellence research awards in 2007, 2009, 2011, 2013 and 2015 from his university for excellence in research. He served as an IEEE Distinguished Speaker, and currently, he is an ACM Distinguished Scientist, ACM Speaker, and IEEE Senior Member and Golden Core awardee.

Masaru Kitsuregawa received his Information Engineering Ph.D. degree from the University of Tokyo in 1983. Since then he joined the Institute of Industrial Science, the University of Tokyo, and is currently a professor. He is also a professor at Earth Observation Data Integration & Fusion Research Initiative of the University of Tokyo since 2010. He also serves Director General of National Institute of Informatics since 2013. Dr. Kitsuregawa's research interests include Database Engineering, and he had been principal researcher of Funding Program for World-Leading Innovative R&D on Science and Technology, MEXT Grant-in-Aids Program for "Info-Plosion", and METI's Information Grand Voyage Project. He had served President of Information Processing Society of Japan from 2013 to 2015. He served as a committee member for a number of international conferences, including ICDE Steering Committee Chair. He is an IEEE Fellow, ACM Fellow, IEICE Fellow and IPSJ Fellow, and he won ACM SIGMOD E.F.Codd Contriutions Award, Medal with Purple Ribbon, 21st Century Invention Award, and C&C Prize.