# EcoRare: An Economic incentive scheme for Efficient Rare Data Accessibility in Mobile-P2P networks

Anirban Mondal[1] Sanjay Kumar Madria[2] Masaru Kitsuregawa[1]

[1] Institute of Industrial Science
University of Tokyo, JAPAN
{anirban, kitsure}@tkl.iis.u-tokyo.ac.jp

[2] Department of Computer Science
University of Missouri-Rolla, USA
madrias@umr.edu

**Abstract.** We propose EcoRare, a novel economic incentive scheme for improving the availability of **rare data** in Mobile-P2P networks. EcoRare combats free-riding and facilitates creation of multiple copies of rare data due to its novel selling mechanism. Experiments show that EcoRare indeed improves accessibility to rare data in Mobile-P2P networks.

## 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., Kazaa) strongly motivate M-P2P applications. M-P2P applications facilitate mobile users in sharing information *on-the-fly* in a P2P manner. Mobile devices with support for wireless device-to-device P2P communication are beginning to be deployed such as Microsoft's Zune [13].

This work focusses on improving the availability of *rare* data items. *Rare* items are those, which get sudden bursts in accesses based on *events*. In the absence of related events, rare items are generally of little interest to most M-P2P users, hence they are typically hosted by relatively few MPs, who obtain them from content distributors. For example, in case of an unexpected event such as a gas emission, several users would be interested in knowing where gas-masks are available. This would lead to sudden bursts in accesses to data concerning gas-masks, and users would *urgently* need this data in *real-time*. Peers, which host data concerning gas-masks, obtain such data as advertisements from content providers (e.g., shops selling gas-masks). Observe that the sudden bursts in accesses to rare items generally occurs with a given time-frame, before and after which rare items are accessed rarely. Similarly, in case of a sudden snowfall, many people would want to know in *real-time* where shovels are available, thereby resulting in a sudden burst of accesses to data concerning shovels. Peers hosting data about shovels obtain such data as advertisements from shops that sell shovels. Such M-P2P interactions for effective sharing of rare data are currently not freely supported by existing wireless communication infrastructures.

In the same vein, during a sudden heavy snowfall in an area usually having negligible snowfall, many cars are likely to slip and require some repairs due to the snow since most cars in areas of negligible snowfall are not typically equipped with snow-related features such as snow tires. Thus, many drivers would require *real-time* information about car repairs. Peers hosting data about car repairs obtain them as advertisements from car-repair garages. Moreover, when an art exhibition takes place in a city, many people attending the exhibition may become interested in obtaining more knowledge about paintings, which would result in sudden bursts in accesses to data concerning paintings. Peers, who host data about rare paintings, may have obtained the data from art connoiseurs and shops selling art-related items. Observe how such sudden interest of several users in paintings or in car repairing arises due to *events*. Incidentally, our target applications mainly concern slow-moving objects e.g., people in art exhibitions.

Data availability in M-P2P networks is typically lower than in fixed networks due to frequent network partitioning arising from user movement and mobile devices switching 'off' when their generally limited energy is drained. Since a large percentage of MPs are typically free-riders [14] (i.e., they do not provide any data), the limited energy of a typically small percentage of MPs, which provide data, is quickly drained, thereby further reducing M-P2P data availability. 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. Notably, existing M-P2P incentive schemes [26] do not consider rare data. Hence, we propose **EcoRare**, a novel *economic incentive scheme*, in which MPs collaborate effectively with each other to improve the availability of rare data.

In EcoRare, each data item is associated with two (*virtual-currency*) prices, namely the *use_price* $\rho_u$ and the *sell_price* $\rho_s$. Paying the *use_price* entitles the query issuing MP $M_I$ to obtain some basic information about its queried data item $d$, but $M_I$ does not obtain the ownership of $d$ i.e., it cannot sell $d$. On the other hand, by paying the *sell_price*, $M_I$ obtains more detailed information about $d$ and it obtains ownership rights of $d$, hence it can sell $d$. Thus, in EcoRare, a given data can have multiple owners. Such multiple owners hosting the same data guard the data availability against the unavailability of some of the owners. Observe that if $M_I$ pays the *use_price*, it cannot sell $d$ because it does not have complete information concerning $d$. Intuitively, given a data $d$, the *sell_price* of $d$ is always higher than its *use_price*.

To put things into perspective, let us refer to our application scenarios. When $M_I$ pays the *use_price* for obtaining information about gas-masks and shovels, it obtains information about only a few shops selling these items and the respective prices of these items at those shops. However, when $M_I$ pays the *sell_price*, it obtains additional information including complete catalogues of more shops selling these items, information about how to order these items (e.g., by phone) and how to get these items delivered to $M_I$. Observe that in the above scenarios, the peer $M_S$ that hosts the queried data essentially acts as an agent for the content-provider. Thus, the prices paid by $M_I$ to $M_S$ may be viewed as a *commission* for giving $M_I$ information about the queried data. Similarly, $M_S$ may be regarded as disseminating advertisements on behalf of the content-provider.

In case of paintings, paying the *use_price* entitles $M_I$ to obtain some basic information concerning the art shops, where replicas of the paintings can be purchased. However, when $M_I$ pays the *sell_price*, it obtains additional information such as how to order the paintings and get them delivered, digital images of collection of paintings by the same artist, educational documentary videos about the historical and cultural aspects of the paintings and so on. For the car-repairing application scenario, the *use_price* allows $M_I$ to know only a few addresses of car repair garages, while the *sell_price* enables $M_I$ to get additional information e.g., a video showing how to change a tire.

In essence, EcoRare requires a data-requestor to pay either the *use_price* or the *sell_price* to the data-provider. EcoRare also requires query issuing MPs to pay a constant relay cost to each MP in the successful query path to entice relay MPs in forwarding queries quickly. Thus, EcoRare effectively combats free-riding because free-riders would have to earn currency for issuing their own requests, and they can earn currency only by hosting items and relaying messages. As we shall see later, rare items are higher-priced than non-rare items in terms of both *use_price* and *sell_price*. This provides free-riders with higher incentive [23, 8, 20] to host rare items for maximizing their revenues. By enticing free-riders to pool in their energy and bandwidth resources to host rare items, EcoRare improves the availability and lifetimes of rare items due to multiple ownership and the creation of multiple copies. Incidentally, querying in EcoRare proceeds via broadcast which is limited to 8 hops to optimize communication overheads.

EcoRare considers both read-only and updatable data items. A rare item owner can send the subsequent updates under a contract for a period of time to the MPs, which have paid it the *sell price* for the data. For our application scenarios, in case gas-masks or shovels go out of stock at some of the shops, the previous owner of the data can send this information to the MP(s) that it had previously sold the data to. Updates can also be in the form of appending new information e.g., for the car-repair application scenario, the data owner could provide the addresses of more car garages, where a slot for car repairing is currently available. Moreover, the current buyer can also get the direct feed from the content provider (e.g., a car-repair garage).

Incidentally, virtual currency incentives are suitable for P2P environments due to the high transaction costs of real-currency micro-payments [24]. The works in [5, 6, 28] 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.

The main contributions of EcoRare are three-fold:

1. It combats free-riding and effectively involves free-riders to improve the availability (and lifetimes) of rare data.
2. It facilitates the creation of multiple copies of rare items in the network since its selling mechanism allows a given data to have multiple owners.
3. It indeed improves query response times and availability of rare data items in M-P2P networks, as shown by a detailed performance evaluation.

To our knowledge, this is the first work to propose an M-P2P economic incentive scheme for rare items.

## 2 Related Work

Economic models for resource allocation in distributed systems [7, 15] do not address unique M-P2P issues such as node mobility, free-riding and frequent network partitioning. Economic schemes for resource allocation in wireless ad hoc networks [17, 27] do not consider free-riding. Schemes for combating free-riding in static P2P networks [9, 11, 14, 16, 18] are too static to be deployed in M-P2P networks as they assume peers' availability and fixed topology.

Schemes for improving data availability in mobile ad hoc networks (MANETs) (e.g., the **E-DCG+** approach [12]) primarily focus on replica allocation, but do not consider economic incentive schemes and M-P2P architecture. Interestingly, the proposals in [25, 26] consider economic schemes in M-P2P networks. However, they focus on data dissemination with the aim of reaching as many peers as possible, while we consider on-demand services. Furthermore, the works in [25, 26] do not consider free-riders and data item rarity issues.

The works in [2, 28] use virtual currency to stimulate the cooperation of mobile nodes in forwarding messages, but they do not consider prices of data items and data item rarity issues. The proposals in [4, 22] concentrate on compensating forwarding cost in terms of battery power, memory, CPU cycles, but they do not entice free riders to host data.

P2P replication suitable for mobile environments has been incorporated in systems such as ROAM [19], Clique [21] and Rumor [10]. However, these systems do not incorporate economic schemes. MoB [3] 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.

## 3 EcoRare: An M-P2P Economic Scheme for Rare Data

This section discusses the economic scheme of EcoRare. In particular, we present the formulae for *use_price*, *sell_price* and the revenue of an MP.

### Computation of the *use_price* $\rho_u$

Table 1 summarizes the notations used in this paper. Using the notations in Table 1, the *use_price* $\rho_u$ of a data item $d$ is computed below:

$$\begin{aligned} \rho_u &= \int_{t_1}^{t_2} ( \lambda \times e^{\tau_D/\tau_R} \; dt ) \quad if \;\; \tau_R \le \tau_D \\ &= 0 \quad otherwise \end{aligned} \tag{1}$$

where $[t_2 - t_1]$ is a given time period. $\tau_R$ and $\tau_D$ are the query response times and the query deadlines respectively. The term $e^{\tau_D/\tau_R}$ implies that faster response times (w.r.t. the query deadline) for queries on $d$ command higher price, which is consistent with ephemeral M-P2P environments. For queries answered after the deadline $\tau_D$, $\rho_u = 0$ as the query results may no longer be useful to the user.

Rarity score $\lambda$ of a data item $d$ quantifies its rarity. $\lambda$ depends upon the number of MPs which host $d$, and the variability in access frequencies of $d$

| Symbol | Significance |
|---|---|
| $d$ | A given rare data item |
| $\rho_u$ | The *use_price* of $d$ |
| $\rho_s$ | The *sell_price* of $d$ |
| $\lambda$ | Rarity score of a data item |
| $\tau_D$ | Time of query deadline |
| $\tau_R$ | Time of query response |
| $n_{r_i}$ | Number of read access to $d$ during the $i^{th}$ time period |
| $N_{Copies}$ | The total number of copies of a data item in the network |

**Table 1.** Summary of Notations

during the past $N$ periods of time e.g., shovels for removing snow are heavily accessed only during a specific time-frame associated with the sudden snowfall, while at other times, they may not be accessed at all. $\lambda$ is computed below:

$$\lambda \ = \ ( \ \theta \times (max_{\eta_r} - min_{\eta_r}) \times NP \ ) \ / \ ( \sum_{i=1}^{N} \eta_i \ ) \qquad (2)$$

where $N$ is the number of time periods over which $\lambda$ is computed. $\theta = ( \ max_{\eta_w}$ - $min_{\eta_w} + 1 \ )$, where $max_{\eta_w}$ and $min_{\eta_w}$ are the maximum and minimum number of write accesses to $d$ during any of the last $N$ time periods. Thus, for read-only items, $\theta = 1$. $max_{\eta_r}$ and $min_{\eta_r}$ are the maximum and minimum number of read-accesses to $d$ during any of the past $N$ time periods. $NP$ is the ratio of the total number of MPs in the network to the number of MPs that host $d$. Observe that $\lambda$ decreases with increasing number of MPs hosting $d$. This is because a data item becomes less rare when it is hosted at more MPs, thus its rarity score $\lambda$ also decreases. Each MP knows the value of $NP$ since every MP *periodically* broadcasts its list of data items so that each MP knows the data items that are hosted at other MPs. $\eta_i$ is the total access frequency (i.e., total read accesses for read-only items and the sum of reads and writes for read-write items) during the $i^{th}$ time period. Thus, the term $(\sum_{i=1}^{N} \eta_i)$ represents the total access frequency of $d$ during the last $N$ time periods.

**Computation of the *sell_price* $\rho_s$**

Using the notations in Table 1, the *sell_price* $\rho_s$ of an item $d$ is computed below:

$$\rho_s \ = \ \rho_u \times ( \ ( \sum_{i=1}^{N} n_{r_i} \ ) \ / \ N_{Copies} \ ) \qquad (3)$$

Observe that $\rho_s$ of a given item $d$ depends upon the *use_price* $\rho_u$ of $d$. This helps in ensuring that $\rho_s$ always exceeds $\rho_u$, which is necessary in case of our application scenarios. Here, $n_{r_i}$ is the number of read-accesses to $d$ during the $i^{th}$ time-period. Thus, the term ( $\sum_{i=1}^{N} n_{r_i}$ ) represents the total number of reads to $d$ during the last $N$ time-periods. $N_{copies}$ is the total number of copies of $d$ in the network. Every MP knows the value of $N_{copies}$ since every MP *periodically* broadcasts its list of data items so that each MP knows the data items that are hosted at other MPs. Notably, the number of read-accesses to $d$ must always exceed the number of copies of $d$ hosted in the network. This is because MPs have limited memory, hence they will not host data items, whose number of read accesses is relatively low. Due to memory space constraints, each MP tries to host data items with relatively higher number of read accesses to maximize its revenue from hosting the items. Thus, the term ( ( $\sum_{i=1}^{N} n_{r_i}$ ) / $N_{copies}$ ) essentially represents a factor, which is always greater than 1. This guarantees that the *sell_price* $\rho_s$ of any given item $d$ always exceeds its *use_price* $\rho_u$.

### Revenue of an MP

**Revenue** of an MP is defined as the difference between the amount of virtual currency that it earns (by hosting data items and providing relay services) and the amount that it spends (by requesting data items and paying relay costs to the MPs in the successful query path).

Suppose MP $M$ hosts $p$ data items, and for the $i^{th}$ item, the number of queries (served by $M$) corresponding to *use_price* is $h_{u_i}$, while the $i^{th}$ item's *use_price* is $\rho_{u_i}$. On the other hand, let the price of the $i^{th}$ item and its access frequency corresponding to *sell_price* be $\rho_{s_i}$ and $h_{s_i}$ respectively. Now suppose the number of items in the network not hosted by $M$ is $q$. $M$ issues $r_{u_i}$ queries corresponding to *use_price* for the $i^{th}$ such item, while the $i^{th}$ item's *use_price* is $\rho_{u_i}$. Moreover, let the price of the $i^{th}$ item and its access frequency corresponding to *sell_price* be $\rho_{s_i}$ and $r_{s_i}$ respectively for the items requested by $M$.

Assume MP $M$ relays $m$ messages and requires to pay relay commissions for $n$ messages in the course of issuing different queries. Thus, $M$'s earnings from relaying messages equals $(m \times K)$, while $M$'s spending on relay cost is $(n \times K)$. Here, $K$ is a constant, which is a small percentage of the *use_price* $\rho_u$ of the data item corresponding to the relay message. In this work, we consider $K$ to be 5% of $\rho_u$. Observe that EcoRare provides higher incentive to MPs for hosting items than for relaying messages. This is in consonance with EcoRare's aim of improving the availability of rare items by encouraging MPs to host such items. Revenue $\omega$ of an MP is computed below:

$$\omega = \Big( \sum_{i=1}^{p} (\rho_{u_i} \times h_{u_i}) + \sum_{i=1}^{p} (\rho_{s_i} \times h_{s_i}) \Big) - \Big( \sum_{i=1}^{q} (\rho_{u_i} \times r_{u_i}) + \sum_{i=1}^{q} (\rho_{s_i} \times r_{s_i}) \Big) + \big( (m-n) \times K \big)$$

In the above equation, the first and second terms represent MP $M$'s earnings, while the third and fourth terms indicate $M$'s spending. The last term relates to the earnings and spending of $M$ due to relay commissions.

## 4 Data Selling mechanism of EcoRare

This section discusses the data selling mechanism of EcoRare such that multiple copies of the rare data are created, which improves accessibility to rare data.

In EcoRare, data selling works in the following two ways:

- *Query-based selling:* When a query-issuing MP $M_I$ pays the *sell_price* of a data item $d$ to the host MP $M_S$ of $d$, $M_I$ obtains a copy of $d$ and also becomes an owner of $d$, thus $M_I$ can now host and re-sell $d$.
- *Push/Pull-based selling:* When a data-providing MP's energy becomes low or when it is about to leave the network, it may decide to sell its rare data items to earn currency (i.e., *a push-based selling mechanism*). On the other hand, due to EcoRare's economic model, free-riders need to earn currency, without which they would not be able to issue any requests of their own. Hence, free-riders may want to buy one or more rare data items from the data-providing MPs so that they can earn currency by hosting these items (i.e., *a pull-based selling mechanism*). Periodically, data-providing MPs (including MPs that have obtained data via the selling and re-selling mechanisms) broadcast their list of rare items to facilitate both push and pull-based selling mechanisms.

Observe that both the *query-based* and the *push/pull-based* selling mechanisms are essentially equivalent to the data provider selling the data item $d$ such that both become owners of $d$. Thus, EcoRare allows multiple owners for any given item $d$, and each owner is allowed to host and sell/re-sell $d$. Moreover, both these mechanisms effectively create multiple copies of a given item $d$, thereby improving the accessibility to rare items. In the absence of a selling mechanism, rare items would become inaccessible once their original providers run out of energy. Furthermore, multiple copies of rare items also facilitate shorter query paths, thereby further improving both query response times and data availability.

Given that both these selling mechanisms have many similarities, we present their algorithms in a unified manner. In these algorithms, we shall refer to the data-providing MPs as the **sellers**, while the query-issuing MPs (in case of the query-based selling mechanism) or the free-riders (in case of the push/pull-based mechanism) shall be referred to as the **buyers**.

**Algorithm *EcoRare_Seller_MP***
(1) Sort all its data items in descending order of $\phi$ /* $\phi$ is an item's revenue-earning potential */
(2) Compute the average value $\phi_{avg}$ of all its data items
(3) Select items for which $\phi$ exceeds $\phi_{avg}$ into a list *Sell*
  /* *Sell* is the candidate set of items for selling */
(4) Broadcast the list *Sell* upto its $n$-hop neighbours
(5) for each data item $d$ in *Sell*
(6)   Wait for replies from potential buyers
(7)   Receive replies from buyers into a list *Sell_List*
(8)   if *Sell_List* is non-empty
(9)    for each MP $M$ in *Sell_List*
(10)     Sell $d$ to $M$ by sending $d$ to $M$ and granting $M$ ownership of $d$
(11)     Obtain the *sell_price* of $d$ from $M$
**end**

**Fig. 1.** EcoRare algorithm for seller-MP

**Algorithm *EcoRare_Buyer_MP***
$Sell_i$: Candidate data items for sale from seller-MP $i$
$Spc$: Its available memory space
(1) for each seller-MP $i$
(2) Receive broadcast message from $i$ containing the set $Sell_i$
(3) Append all data items in $Sell_i$ to a set $bigSell$
(4) Sort all data items in $bigSell$ in descending order of $\phi$ /* $\phi$ is an item's revenue-earning potential */
(5) for each data item $d$ in $bigSell$
(6) while $Spc > 0$
(7) /* $size_d$ is the size of $d$ */
(8) if ( $size_d \leq Spc$ )
(9) Add $d$ to a set $Buy$
(10) Spc = Spc - $size_d$
(11) for each data item $d$ in set $Buy$
(12) Send the *sell_price* of $d$ as payment to the corresponding data-provider of $d$
(13) Receive $d$ (including re-sell rights) for $d$ from the corresponding data-provider of $d$
**end**

**Fig. 2.** EcoRare algorithm for buyer-MP

Figure 1 depicts the algorithm for a seller-MP $M_{Sell}$. In Line 1 of Figure 1, $\phi$ = $(\lambda_d \times \eta_d)$, where $\lambda_d$ (computed by Equation 2) is the rarity score of item $d$ and $\eta_d$ is $d$'s access frequency. Observe that $\phi$ reflects the revenue-earning potential of $d$ since $d$'s price increases with increase in both $\lambda_d$ and $\eta_d$. Notably, items with relatively high update frequencies are not considered for sale. This is because the sale of frequently updated items would incur significant energy and bandwidth overheads for the seller MP due to continuous update transmissions. As Lines 1-3 indicate, $M_{Sell}$ sells items with relatively higher values of $\phi$ to maximize its revenue from sales, while optimizing its energy consumption for transmitting sold items to buyers. In Line 4, $M_{Sell}$'s broadcast message contains not only the items for sale, but also the price, rarity score and recent access history of each item. Moreover, this broadcast is limited only upto $n$-hops to reduce communication overheads. For our application scenarios, we found a value of $n$ = 4 to be reasonable. In Lines 5-11, $M_{Sell}$ receives replies from potential buyers, and sells items to these buyers in lieu of the respective *sell_prices* of the items.

Figure 2 depicts the algorithm for a buyer-MP $M_{Buy}$. In Lines 1-4 of Figure 2, $M_{Buy}$ receives broadcast messages from different seller-MPs, collates this broadcast information and sorts the items for sale in descending order of $\phi$, which is the revenue-earning potential of an item. In Lines 5-10, observe how $M_{Buy}$ prefers to select items with higher revenue-earning potential $\phi$ for hosting for maximizing its revenue per unit of its memory space since its memory space is limited. Thus, $M_{Buy}$ greedily *simulates* the filling up of its memory space by items with higher value of $\phi$. In Lines 11-13, having selected the items to buy, $M_{Buy}$ obtains the items from their corresponding sellers and pays the *sell_price* of these items to the sellers. In case $M_{Buy}$ 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. This policy of allowing deferred payments allows free-riders, which may initially not have enough currency to buy items, to seamlessly integrate into participating in the network.

## 5 Performance Evaluation

This section reports our performance evaluation. MPs move according to the *Random Waypoint Model* [1] within a region of area 1000 metres ×1000 metres. The *Random Waypoint Model* is appropriate for our application scenarios, which involve random movement of users. A total of 100 MPs comprise 15 data-providers and 85 free-riders (which provide no data). Communication range of all MPs is a circle of 100 metre radius. A time-period is 180 seconds, and *periodically*, every 180 seconds, MPs exchange messages to inform each other concerning the items that they host. Table 2 summarizes our performance study parameters.

| Parameter | Default value | Variations |
|---|---|---|
| No. of MPs ($N_{MP}$) | 100 | 20,40,60,80 |
| Zipf factor (ZF) for queries | 0.9 | |
| Queries/second | 20 | |
| Bandwidth between MPs | 28 Kbps to 100 Kbps | |
| Probability of MP availability | 50% to 85% | |
| Size of a data item | 50 Kb to 350 Kb | |
| Memory space of each MP | 1 MB to 1.5 MB | |
| Speed of an MP | 1 metre/s to 10 metres/s | |
| Size of message headers | 220 bytes | |

**Table 2.** Performance Study Parameters

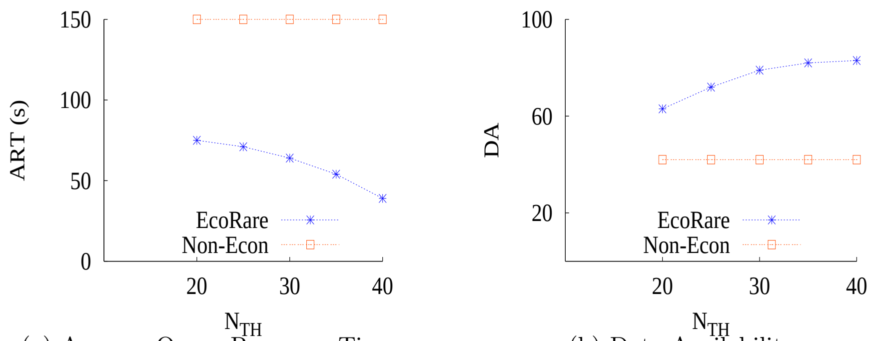
Each data-provider owns 10 data items of varying sizes with different rarity scores. We create three classes of items, namely *rare*, *medium-rare* and *non-rare*. The number of items in each of these classes is decided using a highly skewed Zipf distribution (zipf factor = 0.9) over three buckets to ensure that the majority of items in the network are *rare* in that they will be assigned high rarity scores.

We assign the rarity score $\lambda_d$ to each item $d$ as follows. For *rare* items, $0.7 \leq \lambda_d \leq 1$; for *medium-rare* items, $0.5 \leq \lambda_d < 0.7$; for *non-rare* items, $0 < \lambda_d < 0.5$. Hence, for each item $d$, we randomly assign the value of $\lambda_d$ based on the lower-bounds and upper-bounds of $d$'s class. Among the 15 data-providers, each *rare* item is assigned randomly to only one MP, each *medium-rare* item is randomly assigned to 1-2 MPs, while each non-rare item is randomly assigned to 3-4 MPs. Notably, the actual value of $\lambda_d$ used for computing item prices during the experiments is based on Equation 2, hence the above method is only used to ensure that the majority of items in the network are rare. Furthermore, 90% of the items are read-only, while the other 10% are read-write items. We decide which items are read-only based on a random number generator.

Each query is a request for a single data item. 20 queries/second are issued in the network. Items to be queried are randomly selected from all the items in the entire network. Number of queries directed to each class of items (i.e., *rare*, *medium-rare* and *non-rare*) is determined by a highly skewed Zipf distribution with Zipf factor of 0.9 to ensure that the vast majority of queries are directed towards *rare* items. This is consistent with our application scenarios, which involve sudden bursts in accesses to rare items. 90% of the queries are for read-only items. Although selling is also applicable to updatable items (albeit with low update frequencies), since most queries in our experiments are for read-only items, the selling mechanism impacts MP participation significantly. 70% of queries correspond to *sell_price*, while the rest correspond to *use_price*.

Performance metrics are **average response time** (**ART**) of a query and **data availability** (**DA**). 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. (Unsuccessful queries die after TTL of 8 hops, hence they are not considered for ART computations.) $DA$ equals $((N_S/N_Q) \times 100)$, where $N_S$ is the number of successful queries and $N_Q$ is the total number of queries. Queries can fail due to MPs being switched 'off' due to running out of energy or due to network partitioning.

Existing M-P2P economic incentive schemes do not consider data rarity issues. As reference, we adopt an economic scheme designated as **Econ**. Econ is an economic incentive-based model, which uses data item pricing formulae similar to that of EcoRare, the only difference being that Econ's item pricing formulae do not consider rarity scores $\lambda$ of data items. Thus, Econ does not give preference to rare items. Like EcoRare, Econ performs querying via broadcast.



(a) Average Query Response Time (b) Data Availability

**Fig. 3.** Effect of economic incentives

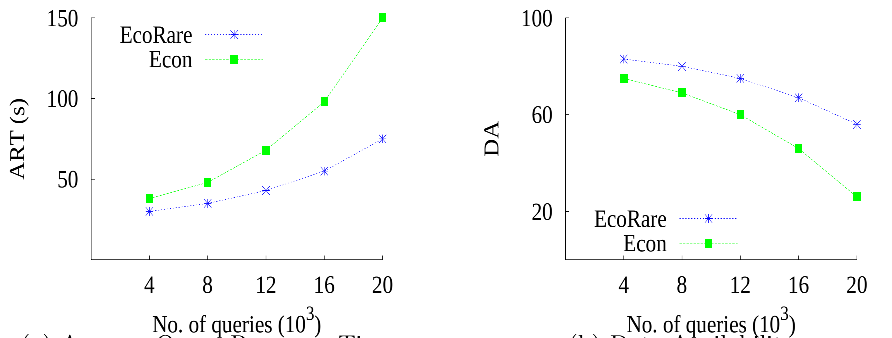
### Effect of economic incentives

To examine the effect of economic incentives, let us first compare EcoRare with a **non-economic non-incentive scheme (Non-Econ)**. Non-Econ provides no incentives for MP participation and it does not prefer rare items. We define threshold revenue $Rev_{TH}$ as 50% of the average revenue of the 15 data-providers in the network. The results in Figure 3 indicate that when the number $N_{TH}$ of MPs above $Rev_{TH}$ increases, ART and DA improve for EcoRare. This is due to more free-riders providing service as MP revenues increase. To earn revenue for issuing their own requests, free-riders host items (by paying *sell_price* to data-providers) to increase their revenues.

Higher MP participation results in multiple paths to locate queried items. Hence, queries can be answered within lower number of hops, thereby decreasing ART. The existence of multiple query paths better preserves accessibility to rare items, thereby improving DA. In contrast, Non-Econ shows constant performance since it is independent of revenue, and it is outperformed by EcoRare because it does not provide any incentives for improving MP participation.

**Performance of EcoRare**

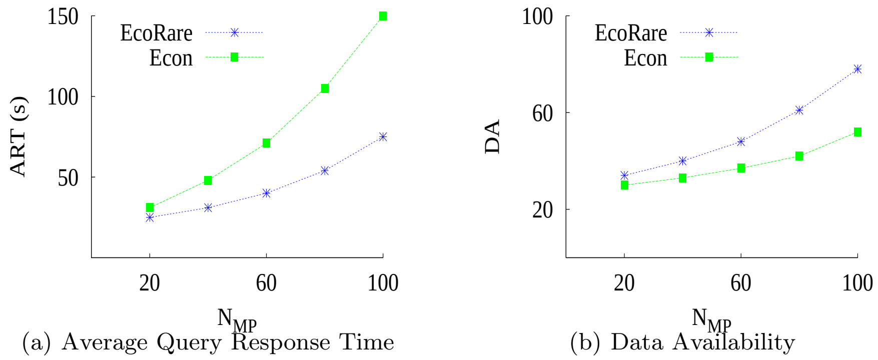
Figure 4 depicts the performance of EcoRare using default values of the parameters in Table 2. As more queries are answered, the energy of MPs decreases and increasing number of MPs run out of energy, hence query paths to data items become longer or inaccessible. Hence, ART increases and DA decreases for both EcoRare and Econ. With increasing number of queries, network congestion and overloading of some of the data-providers also increases ART.



(a) Average Query Response Time (b) Data Availability

**Fig. 4.** Performance of EcoRare

The performance of EcoRare degrades at a considerably slower rate than that of Econ because EcoRare's selling mechanism (via the *sell_price*) results in the creation of multiple copies of rare items, thereby ensuring that rare items are hosted at some MP. In contrast, since Econ does not consider any selling mechanism, rare items become inaccessible when their providers run out of energy.



(a) Average Query Response Time (b) Data Availability

**Fig. 5.** Effect of varying the number of MPs

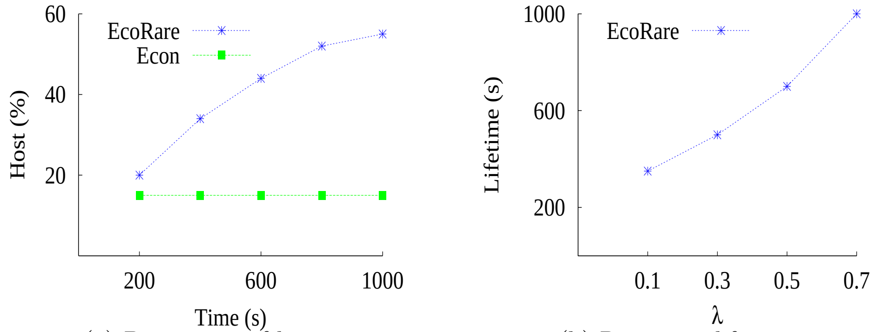
### Scalability of EcoRare

Figure 5 depicts the results when the number $N_{MP}$ of MPs is varied, while keeping the number of queries proportional to $N_{MP}$. In each case, about 15% of the MPs were data-providers, while the remainder were free-riders. With increasing number of MPs, the number of hops to reach queried data items increases due to larger network size, hence ART increases for both EcoRare and Econ. DA also increases for both approaches since larger number of MPs implies more query paths to reach data items. EcoRare outperforms Econ due to its rare item selling mechanism, as explained for Figure 4. Increasing number of MPs implies higher number of free-riders, hence there are more opportunities for rare data-providers to sell their items to free-riders, which explains the increasing difference in performance between the two approaches. Thus, EcoRare exhibits good scalability.

### MP participation and rare item lifetimes

Figure 6a depicts the percentage of host MPs in the M-P2P network over time. An MP is regarded as a host MP if it hosts a data item, which is accessed at least once (either via the *use_price* or the *sell_price* mechanism) during a given time period. In EcoRare, the selling mechanism entices the effective conversion of free-riders to host MPs. The percentage of host MPs plateaus over time due to the majority of the free-riders having been already converted to host-MPs and owing to some of the free-riders running out of energy. In contrast for Econ, the percentage of hosts remains relatively constant over time since only the data-providers host data due to the absence of a selling mechanism. In essence, EcoRare's selling mechanism increases MP participation levels upto the point where the majority of the MPs are providing service to the network.

Figure 6b depicts the results when the rarity score $\lambda$ of data items is varied. For this experiment, we normalized the values of $\lambda$ of different data items such that $0 < \lambda \leq 1$ for each item $d$. We selected five data items corresponding to each value of $\lambda$ and computed the average of their lifetimes. Higher value of $\lambda$ implies higher rarity score of data item $d$. Observe that lifetimes of items increase with increasing value of $\lambda$ because rare items are higher-priced, hence they are more likely to be hosted by free-riders for maximizing revenues. We do not present the results of Econ for this experiment as it does not consider item rarity.




(a) Percentage of hosts (b) Rare item lifetimes

**Fig. 6.** MP participation and rare item lifetimes

## 6 Conclusion

We have proposed EcoRare, a novel economic incentive scheme for improving the availability of **rare data** in M-P2P networks. EcoRare combats free-riding and facilitates the creation of multiple copies of rare items since its selling mechanism allows a given data to have multiple owners. Our performance study demonstrates that EcoRare indeed improves query response times and availability of rare items in M-P2P networks.

## References

1. J. Broch, D.A. Maltz, D.B. Johnson, Y.C. Hu, and J. Jetcheva. A performance comparison of multi-hop wireless ad hoc network routing protocol. *Proc. MOBICOM*, 1998.
2. L. Buttyan and J.P. Hubaux. Stimulating cooperation in self-organizing mobile ad hoc networks. *ACM/Kluwer Mobile Networks and Applications*, 8(5), 2003.
3. R. Chakravorty, S. Agarwal, S. Banerjee, and I. Pratt. MoB: a mobile bazaar for wide-area wireless services. *Proc. MobiCom*, 2005.
4. J. Crowcroft, R. Gibbens, F. Kelly, and S. Ostring. Modelling incentives for collaboration in mobile ad hoc networks. *Proc. WiOpt*, 2003.
5. P. Daras, D. Palaka, V. Giagourta, and D. Bechtsis. A novel peer-to-peer payment protocol. *Proc. IEEE EUROCON*, 1, 2003.
6. E. Elrufaie and D. Turner. Bidding in P2P content distribution networks using the lightweight currency paradigm. *Proc. ITCC*, 2004.
7. D.F. Ferguson, Y. Yemini, and C. Nikolaou. Microeconomic algorithms for load balancing in distributed computer systems. *Proc. ICDCS*, pages 491–499, 1988.

8. A. Garyfalos and K.C. Almeroth. Coupon based incentive systems and the implications of equilibrium theory. *Proc. IEEE International Conference on E-Commerce Technology proceedings*, 2004.
9. P. Golle, K.L. Brown, and I. Mironov. Incentives for sharing in peer-to-peer networks. *Proc. Electronic Commerce*, 2001.
10. R. Guy, P. Reiher, D. Ratner, M. Gunter, W. Ma, and G. Popek. Rumor: Mobile data access through optimistic peer-to-peer replication. *Proc. ER Workshops*, 1998.
11. M. Ham and G. Agha. ARA: A robust audit to prevent free-riding in P2P networks. *Proc. P2P*, pages 125–132, 2005.
12. T. Hara and S.K. Madria. Data replication for improving data accessibility in ad hoc networks. *IEEE Transactions on Mobile Computing*, 5(11), 2006.
13. http://www.microsoft.com/presspass/presskits/zune/default.mspx.
14. S. Kamvar, M. Schlosser, and H. Garcia-Molina. Incentives for combatting free-riding on P2P networks. *Proc. Euro-Par*, 2003.
15. J. F. Kurose and R. Simha. A microeconomic approach to optimal resource allocation in distributed computer systems. *IEEE Trans. Computers*, 38(5):705–717, 1989.
16. N. Liebau, V. Darlagiannis, O. Heckmann, and R. Steinmetz. Asymmetric incentives in peer-to-peer systems. *Proc. AMCIS*, 2005.
17. Jinshan Liu and Valerie Issarny. Service allocation in selfish mobile ad hoc networks using vickrey auction. *Proc. Current Trends in Database Technology - EDBT Workshops revised papers, LNCS 3268*, 2004.
18. L. Ramaswamy and L. Liu. Free riding: A new challenge to P2P file sharing systems. *Proc. HICSS*, page 220, 2003.
19. D. Ratner, P.L. Reiher, G.J. Popek, and G.H. Kuenning. Replication requirements in mobile environments. *Mobile Networks and Applications*, 6(6), 2001.
20. O. Ratsimor, T. Finin, A. Joshi, and Y. Yesha. eNcentive: A framework for intelligent marketing in mobile Peer-to-Peer environments. *Proc. ICEC*, 2003.
21. B. Richard, D. Nioclais, and D. Chalon. Clique: A transparent, peer-to-peer replicated file system. *Proc. MDM*, 2003.
22. V. Srinivasan, P. Nuggehalli, C.F. Chiasserini, and R. R. Rao. Cooperation in wireless ad hoc networks. *Proc. INFOCOM*, 2003.
23. T. Straub and A. Heinemann. An anonymous bonus point system for mobile commerce based on word-of-mouth recommendation. *Proc. ACM SAC*, 2004.
24. D.A. Turner and K.W. Ross. A lightweight currency paradigm for the P2P resource market. *Proc. Electronic Commerce Research*, 2004.
25. O. Wolfson, B. Xu, and A.P. Sistla. An economic model for resource exchange in mobile Peer-to-Peer networks. *Proc. SSDBM*, 2004.
26. B. Xu, O. Wolfson, and N. Rishe. Benefit and pricing of spatio-temporal information in Mobile Peer-to-Peer networks. *Proc. HICSS-39*, 2006.
27. Yuan Xue, Baochun Li, and Klara Nahrstedt. Optimal resource allocation in wireless ad hoc networks: A price-based approach. *IEEE Transactions on Mobile Computing*, 2005.
28. S. Zhong, J. Chen, and Y.R. Yang. Sprite: A simple, cheat-proof, credit-based system for mobile ad-hoc networks. *Proc. IEEE INFOCOM*, 2003.