Coordination Minority Games in Delay Tolerant Networks

Habib B.A. Sidi*, Wissam Chahin*, Rachid El-Azouzi*†, and Francesco De Pellegrini‡

***Abstract*—In this paper we introduce a novel framework for the distributed control of DTNs. The mechanism that we propose tackles a crucial aspect of such systems: in order to support message replication the devices acting as relays need to sacrifice part of their batteries. The aim is thus to provide a reward mechanism able to induce activation of relays in a coordinated fashion. The proposed scheme functions in non-cooperative fashion, and requires minimal message exchange to operate. In particular, relays choose among two strategies: either to participate to message relaying, or not to participate in order to save energy. The base for our mechanism design is to define the relays' utility function according to a minority game; in fact, relays compete to be in the population minority with respect to activation. By tuning the activation level, the system can hence control and optimize the DTN operating point in a distributed manner. To this respect, we characterize extensively the possible equilibria of this game. Finally, a stochastic learning algorithm is proposed which can provably drive the system to the equilibrium solution without requiring perfect state information at relay nodes. We provide extensive numerical results to validate the proposed scheme.**

***Keywords*—Minority game, energy efficiency, delay tolerant networks, Nash equilibrium, learning algorithms, mechanism design**

I. Introduction

Delay Tolerant Networks (DTNs) have gained much interest in the research community in recent years [4, 14]. They have been identified as a promising mean to transport data in intermittently connected systems, i.e., where persistent connectivity cannot be assumed [3]. Also, DTNs can be employed in order to offload traffic from telecommunication networks with respect of specific classes of data when no stringent delivery requirements exist, e.g., video podcasting.

To date, the problem that attracted the most attention from the research community is how to efficiently route messages towards the intended destination(s). This problem is usually solved by disseminating multiple copies of the message in the network. Replication of the original message by the so called epidemic routing protocol ensures that at least some copy will reach the destination node with high probability minimizing the delay to reach the intended destinations. In turn, the standard optimization problem becomes how to maximize the delivery probability under constraints on the resources spent to forward it to destination. As a consequence, several papers have further extended the possible optimizations to the use of activation and/or forwarding control at relays [10].

However, any such optimization is made under the implicit assumption that relays are willing to cooperate with the source node. But, due to limited energy or memory capacity, not always relays can be active and participate to message routing. More precisely, the core issue is whether owners of relay devices, e.g., either smartphones or tables, are willing to have battery depleted to sustain DTNs communications. From the forwarding standpoint, in turn, massive de-activation of relays becomes a core threat which hinders any possible attempt to optimize network performance.

In different contexts, user participation to network operations is granted by means of appropriate incentive mechanisms. The customary example is the usage of credit exchange in peer to peer systems in order to discourage free riders. In the case of DTNs, nevertheless, there is an additional technical issue to be solved. In fact, the credit exchange mechanism cannot be based on end-to-end communications between nodes due to the fact that feedback messages in DTNs may incur into large delays.

In this work we solve precisely the design of a credit-based mechanism for relay participation in DTNs. In particular, it is based on a mechanism design that attains a twofold objective. First, the decision to participate to relaying or not is taken autonomously by relays according to the incentive scheme. Second, since incentives engender a competition among relays that play strategies on their activation, they can be driven to attain a desired operating point for the DTN. Such an operating point, in turn, is precisely the solution of a joint optimization problem involving the number of active relays.

To this respect, we rely on a novel and specific utility structure rooted on the following trade off. The success of a tagged relay depends explicitly on the number of opponents met, namely, nodes adopting the same strategy. In fact, the bigger the number of relays participating to the message delivery, the higher the delivery probability for the message, but indeed the less the chance for the tagged relay to receive a reward from the system.

Rooting our approach in the theory of the Minority Game (MG) [8] we can avoid explicit coordination among the relay nodes. The MG tunes performance of competing relays and welfare of the DTN (number of message copies and message delivery probability). Hence, it configures as an appropriate tool to drive the network to a desired operating point. We thoroughly investigate the properties of our coordination game in which relays compete to be in the population minority.

Finally, the coordination scheme rules the number of active relays via the rewarding mechanism: as such the message source can tune the number of active relays so as to achieve a target performance figure, e.g., the probability of successful message delivery. Conversely, in order to spare network resources, the source can lower performance requirements and

*CERI/LIA, University of Avignon, 339, chemin des Meinajaries,Avignon, France. † University of California at Berkeley, California-94720. ‡CREATE-NET, via Alla Cascata 56 c, 38100 Trento, Italy.

thus reduce the number of relays actively routing the message to destination.

A. Background and contribution

The minority game studies how individuals of a population of heterogeneous agents may reach a form of coordination when sharing resources for which the utility decreases in the number of competitors. Upon introducing adaptation of strategies based on each one's expectation about the future, the game can describe a dynamical system with many interacting degrees of freedom where cooperation is implicitly induced among agents. The MG was first introduced in literature as a simplification of the El Farol Bar's attendance problem [7, 8]. In the El Farol bar problem [5] N users decide independently whether to go to the unique bar in Santa Fe that offers entertainment. However, the bar is small, and they enjoy only if at most Ψ of the possible N attendees are present, in which case they obtain a reward r at a cost $0 \leq c \leq r$ for going to the bar. Otherwise, they can stay home and watch stars with utility 0. Players have two actions: go if they expect the attendance to be less than Ψ people or stay at home and watch stars if they expect the bar will be overcrowded.

The extension of the game introduces a learning component based on the belief of future attendance that every player has: the only information available is the number of people who came to El Farol in past weeks [12],[11],[9].

All those works consider an odd number of interacting agents and do not suggest the exact analysis of equilibrium points as we suggest in this paper; a further key added value of our work is the application of a standard economic estimator, namely, the logit belief model, which provides a suitable convergence framework for our mechanism design. Finally, from the application standpoint, and to the best of our knowledge, it is the first time the concept of MG is applied to DTNs with the aim to derive a mechanism to induce coordination in a non-cooperative fashion.

Some proofs are not presented here due to lack of space, the reader can refer to the technical report [13] for details.

II. Network model

In this section, we present the overall architecture and the intuitions behind our design.

A. System architecture and reward mechanism

We consider a DTN with several source-destination pairs s_i and a large number of mobiles acting as relay nodes in the system. Each mobile is equipped with a wireless interface allowing communication with other mobiles in their proximity. Messages are generated at the source nodes and need to be delivered to the destination nodes; however, each such message is relevant for a time interval of length τ: this is also the horizon by which we intend to optimize network performance.

The network is assumed to be sparse: at any time instant, nodes are isolated with high probability.[1] Nevertheless, due to mobility patterns, communication opportunities arise whenever two nodes get within mutual communication range, i.e., a "contact" occurs. The time between subsequent contacts between any two nodes is assumed to follow a random distribution.[2]

Consider now a message generated at $t = 0$: each source node attempts to deliver the message to its destination; it does so eventually with several copies spread between the relays nodes. Each such message contains a time stamp reporting its age and can be deleted when it becomes irrelevant, e.g., after time τ. Due to lack of permanent connectivity, we exclude the use of feedback that allows the sources or other mobiles to know whether the message has been successfully delivered to its destination or not. For the same reason, the design of our activation mechanism should not require centralized coordination and any such scheme should indeed run fully distributed on board of the relay nodes.

Now assume that the system aims to achieve a target performance (e.g., delivery probability or end-to-end delay). Without loss of generality, we focus only on the probability of successful delivery and our results can be extended to any performance measures satisfying the monotonicity on the number of active user nodes in DTNs. However, based on this target and the parameters of DTNs (e.g., mobility, transmission range, density of nodes), the system can estimate the number of nodes which should participate, named Ψ, in order to guarantee this target level. This value can be defined as the minority threshold of our game. Now the question is how to stimulate Ψ user nodes to participate to delivery message in a distributed manner. A basic scheme to achieve this objective is as follows: we introduce a reward mechanism, in which each source-destination s in the system proposes a reward r^s for relays. For example this reward can be a number of credits that relays may use to send their own messages into the network. Furthermore, we assume that upon successful delivery of a message, the relay node receives a positive reward r^s if and only if it is the first one to deliver the message to the corresponding destination. Recall that the objective of the payment scheme is to provide incentive for the node to participate in forwarding. But, larger rewards engender more nodes to be active which yields a higher delivery probability at the expense of battery depletion and networks lifetime. This trade-off rises the following question for the source: How to define the reward in order to involve Ψ relay nodes as active nodes such in a way to attain a given performance level? This question will be investigated in the next subsection.

B. Network Game

In this subsection we detail the payoff for the relays. When a message is generated by a source node, the competition takes place during the message lifetime, i.e., with duration τ. Each mobile has two strategies: either to participate to forwarding, i.e., pure strategy *transmit* (T), or not to participate, i.e., pure strategy *silent* (S). Mixed strategies, i.e., probability distributions over the two possible actions, are also possible and will be described later on.

Each strategy adopted by a relay corresponds to a certain utility it receives. Clearly, the utility of the relay also depends

[1] This is also the case when disruption caused by mobility occurs at a fast pace compared to the typical operation time of protocols, e.g., the TPC/IP protocol suite.

[2] We don't consider any specific mobility pattern. Indeed several works on mechanism design in DTNs have made, for the sake of tractability, specific assumptions on nodes' mobility (e.g. Random Walk, Random Waypoint).

by the actions performed by N opponent mobiles. For each player in the game, it is worth playing a given action if the number of peer nodes that adopt the same strategy does not exceed a given fraction of the total population of N interacting nodes. Hence, the utility of player is designed in such a way that, upon successful delivery of message to the destination, an active mobiles may receive a positive expected reward conditional to the fact that the actives mobiles represent the minority and to the mechanism selected by network operator. Other nodes receive in this case the opposite as a non-positive expected reward. The customary way to interpret this non-positive reward is that of a regret for abstention.

Formally, let N be the total number of nodes involved in the competition. The probability that an active mobile relays the copy of the packet to the destination within time τ is denoted by $1 - Q_\tau$ where Q_τ is the probability for the tagged relay for not succeeding in message relaying to destination.

At time $t = 0$, each relay plays T or plays S: players who take the minority action win, whereas the majority loses. Now, let $N = N_T + N_S$, where N_T (resp. N_S) is the number of agents selecting strategy T (resp. S). A tagged relay playing strategy T is member of the minority if $N_T \leq \Psi$, otherwise it loses; silent agents win as $N_S \leq N - \Psi$. The probability of receiving a reward R, for an active relay is a function of inter-meeting rate, live time, reward mechanism used by sources and number of active relays. The total reward $R = \sum_s r^s P^s_{succ}(T, k, s)$ with $P_{succ}(T, k, s)$, the probability of an active node to receive a reward r^s from source s when k nodes are active. We denote by g the energy spent by a relay node when it remains active during $[0, \tau]$.

From the sources point of view, performance should be guaranteed above some target level: $D^s_{succ} \geq D^{th}_{succ}$, where D^s_{succ} is the probability of successful delivery of a message:

$$D^s_{succ}(N_T) \qquad = 1 - \prod_{k=1}^{N_T} Q_\tau \tag{1}$$

and D^{th}_{succ} is the performance threshold imposed by the source. The connection between the network performance and the game depends on the total reward R set by sources for successful delivery. Hence the value Ψ should obey the following relation

$$\sum_s r^s \cdot P^s_{succ}(T, \Psi, s) = g\tau$$

where $g \geq 0$ is a constant cost of activation per second for each relay. Note that Ψ is chosen such as to equalize the total energy cost spent by nodes for being active in $[0, \tau]$ and the expected reward obtained for a successful delivery. In the homogeneous case ($P^s_{succ} = P_{succ}\ \forall s$), in which the relay and sources have similar physical characteristic, e.g. transmission range, mobility patterns, energy capacities etc, the last relation becomes $n_s r \cdot P_{succ}(T, \Psi) = g\tau$ where n_s is the number of sources in the network.

We now state the assumption required for the function $P_{succ}(T, k, s)$:

Assumption A
The function $P^s_{succ}(T, k, s)$ is decreasing in k, i.e., number of active relays.

Now we can introduce two utility functions for our game, under the assumption that the population of sources is homogeneous: $P^s_{succ}(T, k, s) = P_{succ}(T, k)\ \forall s$:

Scenario 1: Zero-sum utility

$$U(T, N_T) = \sum_s r^s \cdot P^s_{succ}(T, N_T, s) - g\tau, U(S, N_S) = -U(T, N_T)$$

Scenario 2: Fixed regret utility

$$U(T, N_T) = \sum_s r^s \cdot P^s_{succ}(T, N_T, s) - g\tau, \quad U(S, N_S) = -\alpha,$$

where in the second case the utility of non-active nodes expresses the regret or satisfaction for not participating to message relaying. In particular, we assume $\alpha \geq 0$, and we define N^α_T such that $U(T, N^\alpha_T) = -\alpha$.

The formulation of **Scenario 1**, requires nodes to estimate P_{succ}. This can be calculated over time by interrogating neighboring nodes and averaging their success rate: this amounts to run a pairwise averaging protocol as in [6]. In case we want to avoid the use of gossip mechanisms, we can model regret of non-active nodes as a constant negative perceived utility, which corresponds to **Scenario 2**.

III. Characterization of equilibria

In this section we provide the exact characterization of the equilibria induced by the game: we distinguish pure Nash equilibria and mixed Nash equilibria.

A. Pure Nash Equilibrium

The Nash Equilibrium in pure strategy for our game is given by the relation :

$$U(S, N_T) \geq U(T, N_T + 1),\ U(S, N_T - 1) \leq U(T, N_T)$$

Thus, no player can improve its utility by unilaterally deviating from the equilibrium.

Proposition 1: Under assumption A, there exists a pure Nash Equilibrium for our game. Moreover
(i) for **scenario 1**, there exists a unique NE obtained when exactly Ψ among the total population of N nodes play T.
(ii) for **scenario 2**, there exists two Nash equilibria which are obtained when the total number of active relays is such that: $N_T \in \{N^\alpha_T, N^\alpha_T - 1\}$

Proof: Refer to [13] for details of the proof.

Remark 1: A crucial design issue is how to relate the parameters of the game to the performance of the DTN at the equilibrium. From (1), the number of active nodes required to attain D^{th}_{succ} needs to verify $N^{th}_T = \frac{\log(1-D^{th}_{succ})}{\log(Q_\tau)}$. Besides, from Proposition 1 it must be $\Psi = N^{th}_T$, thus

$$r^* = g\tau \frac{1}{n_s P_{succ}(T, N^{th}_T)}$$

Message reward r at the equilibrium is thus proportional to energy cost g through a positive constant.

B. Mixed Nash Equilibrium

Let's consider now that relay nodes maintain a probability distribution over the two actions. Compared to the pure strategy game, in the mixed strategy game every node can define the strategy by which it will be active only for a fraction of the time and stay silent the rest of the time. This kind of equilibrium is desirable for an homogeneous population of nodes with similar energy constraints.

In the mixed strategy game, node i can choose to play action T with probability p_i and play S with probability $(1-p_i)$. We let, $\mathbf{p} = (p_1, p_2, ..., p_N)$, the mixed strategy profile of our game. If $0 < p_i < 1,\ \forall i$ then $\mathbf{p}$ is a fully mixed strategy profile of the game. A standard companion notation that we use for $\mathbf{p}$ is $(p_i, \mathbf{p_{-i}})$: it denotes the strategy profile of the game when relay i uses strategy p_i and others use $\mathbf{p}_{-i} = (p_1, .., p_{i-1}, p_{i+1}, .., p_N)$. Let's denote by $V^i(\tilde{p}, \mathbf{p}_{-i})$ the utility of node i playing action T with probability $\tilde{p}$. We have the following definition of the mixed strategy Nash Equilibrium:

Definition 1: (i) A ***mixed strategy Nash Equilibrium*** specifies a mixed strategy $p_i^* \in [0,1]$ for each player i $(where\ i = 1 \dots N)$ such that :

$$V^i(p_1^*, .., p_{i-1}^*, p_i^*, p_{i+1}^*, .., p_N^*) \geq V^i(p_1^*, .., p_{i-1}^*, p_i, p_{i+1}^*, .., p_N^*),\ \forall p_i \quad (2)$$

(ii) We call a ***Fully mixed Nash Equilibrium*** a mixed strategy Nash equilibrium $\mathbf{p}$ with $p_i \notin \{0,1\}, \forall i$.

In the rest of the paper we will denote by the term **'mixer'** a relay who uses a mixed strategy $0 < p_i < 1$. The following proposition states that any mixed equilibrium $\mathbf{p}$ with $p_i \notin \{0,1\} \forall i$, is symmetric, i.e. $p_i = p\ \forall i$.

Proposition 2: Assume assumption A holds. At the equilibrium, all mixers must use the same probability p, i.e., $p_i = p_j\ \forall$ mixer i, j.

Proof: Assume that the set of mixers is not empty and let suppose that there are l relays that select pure strategy T and r pure strategy S. Without loss of generality let the strategy profile at the equilibrium : $\mathbf{p} = (p_1, \dots, p_{N-l-r}, 1, \dots, 1, 0, \dots, 0)$

Scenario 1: The utility for a mixer relay i writes

$$V^i(\tilde{p}, \mathbf{p}_{-i}) = (2\tilde{p}_i - 1)F(p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_N)$$

with $F(p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_N) =$

$$\prod_{j\neq i}^{N-l-r}(1-p_j)U(T, l+1) + \sum_{j\neq i}^{N-l-r} p_j \prod_{j'\notin\{i,j\}}^{N-l-r}(1-p_{j'}) \times U(T, l+2) + \sum_{j,j'\neq i}^{N-l-r} p_j p_{j'} \prod_{j''\notin\{i,j,j'\}}^{N-l-r}(1-p_{j''})U(T, l+3) + ... + \prod_{j\neq i}^{N-l-r} p_j U(T, N-r).$$

Note about this function that:

- F is strictly decreasing by any unilateral increase of p_j by node j. This comes from the fact that the utility function of an active node is decreasing with the number of active nodes (assumption A).
- For any two mixers $j \neq j'$, p_j and $p_{j'}$ are indifferently interchangeable variables in F.

At mixed equilibrium $\mathbf{p}$, $\frac{\partial V^i(\mathbf{p})}{\partial \tilde{p}_i} = 0\ \forall\ i \in \{1, \dots, N-l-r\}$. This implies that: $F(p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_N) = 0, \forall$ mixer i. Now suppose that there exists two mixers i and j, s.t. $p_i^* \neq p_j^*$. Without lost of generality assume that $p_i^* < p_j^*$, then

$$0 = F(p, .., p_{i-1}, p_{i+1}, .., p_j, .., p_N) > F(p_1, .., p_{i-1}, p_{i+1}, .., p_i, .., p_N) = F(p_1, .., p_{j-1}, p_{j+1}, .., p_N) > 0$$

which is absurd. Thus $p_i = p_j, \forall$ mixers i, j.

Scenario 2: Refer to [13] for details of the proof. ■

From proposition 2 we conclude that any fully mixed equilibrium $\mathbf{p}$ is symmetric, i.e. $p_i = p_j\ \forall i, j$. The following proposition characterizes the existence and uniqueness of a fully mixed Nash Equilibrium.

Proposition 3: Under assumption A, there exists a unique fully mixed Nash Equilibrium $\mathbf{p}^*$. Moreover, $\mathbf{p}^*$ is solution to:

- **Scenario 1** :

$$A(N, p^*) = \sum_{k=1}^{N} C_{k-1}^{N-1}(p^*)^{k-1}(1-p^*)^{N-k}U(T,k) = 0. \quad (3)$$

- **Scenario 2** :

$$A'(N, p^*) = \sum_{k=1}^{N} C_{k-1}^{N-1}(p^*)^{k-1}(1-p^*)^{N-k}[U(T,k)+\alpha] = 0.$$

Proof: Let p the symmetric mixed strategy adopted by every node in the game, $p_i = p,\ \forall i$.

Scenario 1: The utility of one relay i when the strategy profile (p_i, p_{-i}) is played is given by:

$$V^i(\tilde{p}_i, p_{-i}) = \tilde{p}_i \sum_{k=1}^{N} C_{k-1}^{N-1} p_{-i}^{k-1}(1-p_{-i})^{N-k}U(T,k) + (1-\tilde{p}_i)\sum_{k=0}^{N-1} C_k^{N-1} p_{-i}^k (1-p_{-i})^{N-k-1*}U(S, k+1)$$

$$= (2\tilde{p}_i - 1)\sum_{k=1}^{N} C_{k-1}^{N-1} p_{-i}^{k-1}(1-p_{-i})^{N-k}U(T,k)$$

Let $A(N, p_{-i}) = \sum_{k=1}^{N} C_{k-1}^{N-1} p_{-i}^{k-1}(1-p_{-i})^{N-k}U(T,k)$

if $A(N, p_{-i}) < 0$, $p_i = 0$ is the best response for player i and conversely, $p = 1$ is a best response when $A(N, p_{-i}) > 0$. A mixed strategy is obtained when $A(N, p_{-i}) = 0$. Also, we have

$$A(N, 0) = U(T, 1) > 0 > A(N, 1) = U(T, N)$$

thus there exists a mixed symmetric Nash Equilibrium which is unique since $A(N, p_{-i})$ is strictly decreasing with p. The mixed equilibrium is thus characterized by equation (3).

Scenario 2: The proof is similar to the one of scenario 1. For more details refer to [13].

The utility of one relay i when the strategy profile $(\tilde{p}_i, p_{-i})$ is played is given by:

$$V^i(\tilde{p}_i, p_{-i}) = \tilde{p}_i \sum_{k=1}^{N} C_{k-1}^{N-1} p_{-i}^{k-1} (1-p_{-i})^{N-k} U(T,k) - \alpha(1-\tilde{p}_i)$$

At the Nash equilibrium we have, $\forall$ player i, $\frac{\partial V^i(p^*)}{\partial p^*} = A'(N,p^*) = 0$ with

$$A'(N,p^*) = \sum_{k=1}^{N} C_{k-1}^{N-1} (p^*)^{k-1} (1-p^*)^{N-k} [U(T,k) + \alpha]$$

Since α is a fixed positive constant, $A'(N,p^*)$ has the same properties as $A(N,p^*)$ from the proof of scenario 1. Then we easily conclude that, p^* is unique and characterized by :

$$A'(N,p^*) = \sum_{k=1}^{N} C_{k-1}^{N-1} (p^*)^{k-1} (1-p^*)^{N-k} [U(T,k) + \alpha] = 0.$$

■

C. Equilibrium with mixers and non-mixers

We study here the existence of equilibrium when the population of agents is composed of pure strategy players: active or non-active, as well as mixers. In this case, a non-pure Nash equilibrium can be represented by the triplet (l, r, p^*), where $l, r \in \{0, 1, \ldots, N\}$ denote respectively the number of agents choosing pure strategy T or S, and $p^* \in (0,1)$ the probability with which the remaining $N - l - r$ mixers choose strategy T. Moreover, we denote by $v_T(l,r,p)$(resp. $v_S(l,r,p)$) the expected payoff to a player choosing T(resp. S). The expressions of $v_T(l,r,p)$ and $v_S(l,r,p)$ write as follow:

$$v_T(l,r,p) = \sum_{k=0}^{N-l-r} C_k^{N-l-r} p^k (1-p)^{N-l-r-k} U(T, l+k)$$

$$v_S(l,r,p) = -\sum_{k=0}^{N-l-r} C_k^{N-l-r} p^k (1-p)^{N-l-r-k} U(T, l+k)$$

Proposition 4: Using the previous notations, a strategy profile of type (l, r, p^*) is a Nash equilibrium with at least one mixer if and only if:

$$v_T(l+1, r, p^*) = v_S(l, r+1, p^*) \tag{4}$$

Proof: The condition (4) describes that a mixer is indifferent whether it chooses a pure strategy T or S. This is a necessary condition for the strategy profile (l, r, p^*) to be a Nash equilibrium.
In order to show sufficiency, we need to show that pure strategy players as well, cannot improve their expected utility through unilateral deviation from the equilibrium profile. Without loss of generality, suppose that there is at least one player using pure strategy T, we have

$$\begin{aligned} v_T(l,r,p^*) &\geq v_T(l+1,r,p^*) = v_S(l,r+1,p^*) \\ &\geq v_S(l-1,r+1,p^*) \\ &\geq p^* v_T(l,r,p^*) + (1-p^*) v_S(l-1,r+1,p^*) \end{aligned}$$

This last relation, states that an active user cannot improve its expected utility by unilaterally deviating from the strategy profile (l, r, p^*) using any strategy $p^* \in [0,1)$, given relation (4). As done for **Scenario 1**, in **Scenario 2**, we have, $v_S(l, r+1, p^*) = -\alpha$, let $v_S(l+1, r, p^*) = -\alpha$ then:

$$\begin{aligned} v_T(l,r,p^*) &\geq v_T(l+1,r,p^*) = -\alpha \geq v_S(l-1,r+1,p^*) \\ &\geq p^* v_T(l,r,p^*) + (1-p^*) v_S(l-1,r+1,p^*) \end{aligned}$$

moreover,

$$v_S(l+1, r-1, p^*) \leq v_T(l+1, r, p^*) = -\alpha = v_S(l, r, p^*).$$

This completes the proof. ■

Discussion on existence of (l, r, p^*) type equilibria
It is possible to isolate several cases where the relation (4) that characterizes a Nash Equilibrium of type (l, r, p^*), cannot be satisfied. We denote by, $p = 0^+$ (resp. $p = 1^-$) the mixed strategy infinitely close to 0 (resp. to 1), with which at least one mixer selects to be active. Since, $v_T(l, r, p^*)$ is strictly decreasing with l and p^*, we have, $v_T(l+1, r, p^*) = v_S(l, r+1, p^*)$

$$\iff \begin{cases} v_T(l+1, r, 0^+) > -v_T(l, r+1, 0^+) \\ v_T(l+1, r, 1^-) \leq -v_T(l, r+1, 1^-), \end{cases}$$

(1) If $l \geq \Psi$, then there is no Nash equilibrium of the desired type. Indeed, $l > \Psi$, then $v_T(l, r+1, 0^+) \leq 0$ and

$$v_T(l+1, r, 0^+) \leq 0 \leq -v_T(l, r+1, 0^+).$$

Then there is no possible Nash Equilibrium according to relation (4).

(2) If $l+r+1 > N-1$, then there is no Nash equilibrium. We already have $l < \Psi$, let $l + r + 1 = N$ then,

$$v_T(l+1, r, p) = C_1 \geq 0 \; \forall \, p \text{ and}$$

$$v_S(l, r+1, p) = C_2 > 0 \; \forall \, p.$$

Since v_T is decreasing with l, we have, $0 \leq C_1 < C_2$ which contradicts relation (4).

A Nash Equilibrium of type (l, r, p^*) exists then only for $l < \Psi$ and for $l+r \leq N-2$, thus there are exactly $\Psi(N-2) - \frac{\Psi(\Psi-1)}{2}$ Nash equilibria. In the following proposition we go further and decline some properties of the mixed strategy p^* at the equilibrium.

Proposition 5: The mixed strategy p^* at the equilibrium increases as r increase and reversely decreases as l increase.

Proof: Refer to [13] for details of the proof.

IV. Distributed reinforcement learning algorithm

In this section we introduce a distributed reinforcement learning algorithm: it permits to relays to adjust strategies they play over time in the framework of the DTN MG designed in section II. The analysis of convergence of the algorithm relies on a stochastic model that gives rise to an associated continuous time deterministic dynamic system. It can be proved that this process converges almost surely towards a stationary state which is characterized as ϵ-approximate Nash equilibrium.

In DTNs, limited computational power and low energy budget of relays requires adaptive and energy-efficient mechanisms letting relays adapt to operating conditions at low cost. The learning algorithm proposed here matches this reality of DTNs since, as we shall see, it has the following attractive features:

- It is genuinely distributed: strategy updating decision is local to relays;
- It depends uniquely on the realized payoffs: nodes utilize local observations to estimate their own payoffs;
- It uses simple behavioral rule in the form of logit rule.

We assume that each relay node i has a prior perception x_i of the payoff performance for each action (To be active, or not), and makes a decision based on this piece of information using a random choice rule. The payoff of the chosen action is then observed and is used to update the perception for that particular action. This procedure is repeated round after round, each round of duration τ generating a discrete time stochastic process which is the learning process.

For notation's sake, denote $A = \{T, S\}$ the set of pure strategies, and Δ_i is the set of mixed strategies for relay node i with $i \in \{1, ..., N\}$. Let $V^i(.)$ the payoff function for relay node i. The algorithm works in rounds of duration τ, at round k, each relay node i takes an action a_i^k according to a fully mixed strategy $p_i^k = \sigma_i(x_i^k) \in \Delta_i$. The fully mixed strategy is generated according to the vector $x_i^k = (x_{ia}^k)_{a \in A}$ which represents its perceptions about the payoffs of the available pure strategies. In particular, relay node i's fully mixed strategies are mapped from the perceptions based on the logit rule:

$$\sigma_{ia}(x_i) = \frac{e^{\beta x_{ia}}}{e^{\beta x_{iT}} + e^{\beta x_{iS}}} \tag{5}$$

where β is commonly called the temperature of the logit. The temperature has a smoothing effect: when $\beta \to 0$ it leads to the uniform choice of strategies, while for $\beta \to \infty$ the probability concentrates on the pure strategy with the largest perception. We assume throughout that σ_{ia} is strictly positive for all $a \in A$.

At round k, the perceptions x_{ia}^k will determine the fully mixed strategies $p_i^k = \sigma_i(x_i^k)$ that are used by each relay node i to choose at random action T (to be active) or S (to be silent). Then each relay node estimates his own payoff $\tilde{u}_i^k$, with no information about the actions or the payoffs of the other relay nodes, and uses this value ($\tilde{u}_i^k$) to update its perceptions as:

$$x_{ia}^{k+1} = \begin{cases} (1-\gamma^k)x_{ia}^k + \gamma^k \tilde{u}_i^k & \text{if } a_i^k = a \\ x_{ia}^k & \text{otherwise,} \end{cases} \tag{6}$$

Algorithm 1 Distributed reinforcement Learning Algorithm

1: **input:** $k = 1$, each relay node i chooses its action (T or S) according to distribution p_i and set its initial perception value $x_i^0 = 0$.
2: **while** $max(|x_{iT}^{k+1} - x_{iT}^k|, |x_{iS}^{k+1} - x_{iS}^k|) > \epsilon$ **do**
3: Each relay node i updates its fully mixed strategy profile at iteration k according to (5).
4: Relay node i selects its actions using its updated fully mixed strategy profile.
5: Relay node i estimates its payoff $\tilde{u}_i^k$.
6: Relay node i updates its perception value according to (7).
7: $k \leftarrow k + 1$
8: **end while**

where $\gamma^k \in (0, 1)$ is a sequence of averaging factors that satisfy $\sum_k \gamma^k = \infty$ and $\sum_k (\gamma^k)^2 < \infty$ (examples of such factor are $\gamma^k = \frac{1}{k}$ or $\gamma^k = \frac{1}{1+k \log k}$). A relay node only changes the perception of the strategy just used in the current round and keeps other perceptions unchanged. Algorithm (1) summarizes the learning process. The discrete time stochastic process expressed in (6) represents the evolution of relay node perceptions and can be written in the following equivalent form:

$$x_{ia}^{k+1} - x_{ia}^k = \gamma^k [w_{ia}^k - x_{ia}^k], \forall i \in \{1, .., N\}, a \in A \tag{7}$$

with

$$w_{ia}^k = \begin{cases} \tilde{u}_i^k & \text{if } a_i^k = a \\ x_{ia}^k & \text{otherwise.} \end{cases} \tag{8}$$

In what follows we will prove that this algorithm can attain a steady state for the coordination process among relay nodes. Also, the information it needs to operate is minimal.

A. Convergence of the Learning Process

Based on the theory of stochastic algorithms, the asymptotic behavior of (7) can be analyzed through the corresponding continuous dynamics [1]:

$$\frac{dx}{dt} = E(w|x) - x, \tag{9}$$

where $x = (x_{ia}, \forall i \in \{1, .., N\}, a \in A)$ and $w = (w_{ia}, \forall i \in \{1, .., N\}, a \in A)$.

Let us make equation (9) more explicit by defining the mapping from the perceptions x to the expected payoff of user i choosing action a as $G_{ia}(x) = E(V^i|x, a_i = a)$.

Proposition 6: The continuous dynamics (9) may be expressed as

$$\frac{dx_{ia}}{dt} = \sigma_{ia}(G_{ia}(x) - x_{ia}) \tag{10}$$

Proof: Using the definition of the vector w, the expected value $E(w|x)$ can be computed by conditioning on relay i's action:

$$\begin{aligned} E(w_{ia}|x_{ia}) &= p_{ia}U(a, p_{-i}) + (1 - p_{ia})x_{ia} \\ &= \sigma_{ia}G_{ia}(x) + (1 - \sigma_{ia})x_{ia} \end{aligned} \tag{11}$$

which with (9) yields (10). ■

This can be interpreted as follows: when the difference between the expected payoff and the perception value is large,

the perception value, from (7), will be updated with a large expected value $w_{ia}^k - x_{ia}^k$ and this difference will be reduced.

In the following theorem, we prove that the learning process admits a contraction structure with a proper choice of the temperature β .

Lemma 1: Under the logit decision rule (5), if the temperature satisfies $\beta < \frac{1}{n_s r}$, then the mapping from the perceptions to the expected payoffs $G(x) = [G_{ia}(x), \forall i \in \{1, .., N\}, a \in A)]$ is a maximum-norm contraction.

Proof: We give only a sketch of the proof in several points. For the full proof refer to [13].Let relay i action is to be active (action T). Then $G_{iT}(x) = \sum_{j=0}^{N} n_s r P_{succ}(T, j) C_j^N (\sigma_{iT}(x_i))^j (1 - \sigma_{iT}(x_i))^{N-j} - g\tau$

- We show that $|G_{iT}(x_i) - G_{iT}(\hat{x}_i)| \leq n_s r |\sigma_{iT}(x_i) - \hat{\sigma}_{iT}(\hat{x}_i)|$ for any two perceptions x_i and $\hat{x}_i$ of a relay node i.
- Then we prove that $\sigma_{iT}(x_i) - \hat{\sigma}_{iT}(\hat{x}_i) \leq \beta ||x - \hat{x}||_\infty$
- This allows to conclude that $|G_{iT}(x) - G_{iT}(\hat{x})| \leq \beta n_s r ||x - \hat{x}||_\infty$

Observing that since by the minority game rule $G_{iT}(\cdot) G_{iS}(\cdot) \leq 0$, then if $\beta < \frac{1}{n_s r}$, indeed $G(x)$ is a maximum-norm contraction. ■

Based on the property of contraction mapping, there exists a fixed point x^* such that $G(x^*) = x^*$. In the following theorem we show that the distributed learning algorithm also converges to the same limit point x^*.

Theorem 1: If $G(x)$ is a $||.||_\infty$-contraction, its unique fixed point x^* is a global at tractor for the adaptive dynamics (10), and the learning process (7) converges almost surely towards x^*. Moreover the limit point x^* is globally asymptotically stable.

Proof: Since $G(x)$ is a $||.||_\infty$-contraction, it admits a unique fixed point x^*. According to general results on stochastic algorithms the rest points of the continuous dynamic (10) are natural candidates to be limit point for the stochastic process (7). All together with ([1], corollary 6.6), we have the almost sure convergence of (7), given that we exhibit a strict Lyapunov function ϕ.

Now let $\phi(x) = ||x_{ia} - x^*||_\infty$, then $\phi(x^*) = 0, \phi(x) > 0, \forall x \neq x^*$. Let $i \in \{1, ..., N\}, a \in A$ be such that $\phi(x) = |x_{ia} - x_{ia}^*|$. If $x_{ia} \geq x_{ia}^*$, then $\phi(x) = x_{ia} - x_{ia}^*$. Since $G_{ia}(x)$ is a maximum norm contraction, there exist a Lipschitz constant ξ such that $G_{ia}(x) - G_{ia}(x^*) \leq \xi(x_{ia} - x_{ia}^*)$, and $G_{ia}(x^*) = x_{ia}^*$. All together combined with equation (10), we can write:$\forall x \neq x^*$

$$\begin{aligned} &\frac{d\phi(x)}{dt} = \frac{d(x_{ia} - x_{ia}^*)}{dt} = \frac{dx_{ia}}{dt} \\ &= \sigma_{ia}(G_{ia}(x) - x_{ia}) = \sigma_{ia}(G_{ia}(x) - G_{ia}(x^*) + x_{ia}^* - x_{ia}) \\ &\leq \sigma_{ia}\xi(x_{ia} - x_{ia}^*) + x_{ia}^* - x_{ia} - (1 - \sigma_{ia}\xi)\phi(x) < 0. \end{aligned}$$

and a similar argument for the case $x_{ia} \leq x_{ia}^*$ also shows that $\frac{d\phi(x)}{dt} < 0, \forall x \neq x^*$. Thus the function $\phi(x)$ is a strict Lyapunov function and x^* is globally asymptotically stable.■

B. Approximate fully mixed Nash Equilibrium

From lemma (1) and theorem (1), we have:

$$G_{ia}(x^*) = E(V^i | x^*, a_i = a) = x_{ia}^*.$$

This is a property of the equilibrium (x^*) of the distributed learning algorithm: its value x_{ia}^* is an accurate estimation of the expected payoff in the equilibrium. Moreover we show that the fully mixed strategy

$$p^* = (\sigma_{ia}^* = \frac{e^{\beta x_{ia}^*}}{e^{\beta x_{iT}^*} + e^{\beta x_{iS}^*}}, \forall a \in A, i \in \{1...N\})$$

is an approximate Nash equilibrium.

Proposition 7: Under the Logit decision rule (5), the fully mixed strategy $p^* = \sigma^*(x^*)$ at the equilibrium x^* is a ϵ-approximate Nash equilibrium for our game (proposition 3) with $\epsilon = -\frac{1}{\beta} \sum_{a \in A} \sigma_{ia}^* (ln(\sigma_{ia}^*) - 1)$.

Proof: A well-known characterization of the logit probabilities gives:

$$\begin{aligned} &\sigma_{ia}(x^*) = \arg \max_{\sigma_i = [\sigma_{iT}, \sigma_{iS}]} \sum_{a \in A} \sigma_{ia} E(V^i | x^*, a_i = a) \\ &- \frac{1}{\beta} \sum_{a \in A} \sigma_{ia}(ln(\sigma_{ia}) - 1) \\ &= \frac{e^{\beta E(V^i | x^*, a_i = a)}}{e^{\beta E(V^i | x^*, a_i = T)} + e^{\beta E(V^i | x^*, a_i = S)}} = \frac{e^{\beta x_{ia}^*}}{e^{\beta x_{iT}^*} + e^{\beta x_{iS}^*}}, \end{aligned}$$

and since ([2], pp.93) $\max_{\sigma_i} \sum_{a \in A} \sigma_{ia} E(V^i | x^*, a_i = a) - \frac{1}{\beta} \sum_{a \in A} \sigma_{ia}(ln(\sigma_{ia}) - 1) \leq \max_{\sigma_i} \sum_{a \in A} \sigma_{ia} E(V^i | x^*, a_i = a)$ then, we have:

$$\sum_{a \in A} \sigma_{ia}^* E(V^i | x^*, a_i = a) \geq \max_{\sigma_i} \sum_{a \in A} \sigma_{ia} E(V^i | x^*, a_i = a) - \epsilon$$

where $\epsilon = \max_{i \in \{1...N\}} \{-\frac{1}{\beta} \sum_{a \in A} \sigma_{ia}(ln(\sigma_{ia}) - 1)\}$.

Hence the fully mixed strategy $p^* = \sigma^*(x^*)$ in the equilibrium x^* is a ϵ-approximate Nash equilibrium. ■

V. Application and Numerical Results

In this section, we provide a numerical analysis of the performance achieved by DTN nodes following the distributed reinforcement learning mechanism proposed in section IV. For the rest of the paper, we will assume that relay nodes use the two hop routing scheme, and the inter-meeting rate between nodes follows an exponential distribution. Furthermore, we assume that upon successful delivery of a message, the relay node receives a positive reward R if and only if it is the first one to deliver the message to the corresponding destination. Under those assumptions, we can obtain the expressions of different quantities: in particular the probability that an active node relays a copy of a received packet to destination within time τ is $1 - Q_\tau$ where the expression of Q_τ is given by: $Q_\tau = (1 + \lambda\tau)e^{-\lambda\tau}$. Now, the probability of successful delivery of the message for an active node is:

$$P_{succ}(T, N_T) = \frac{1 - Q_\tau^{N_T}}{N_T}, \tag{12}$$

and this the probability that a relay is the first to deliver a given message to its destination (see [13]).

Fig. 1. Learning the fully mixed strategy: homogeneous case. $g = 6.6\times10^{-4}$

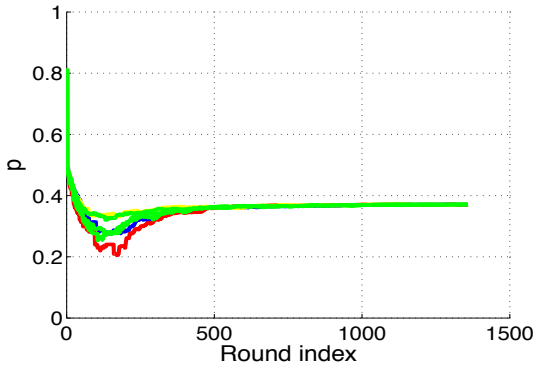
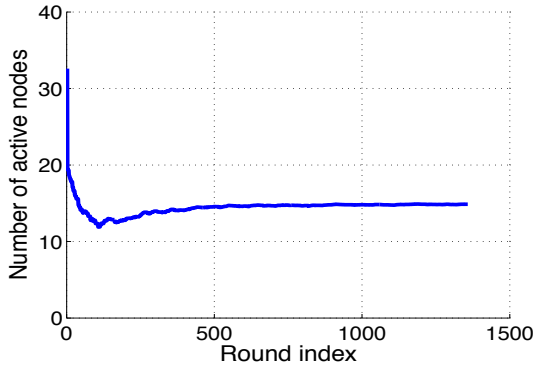


Fig. 2. Learning the fully mixed strategy: homogeneous case. $g = 6.6\times10^{-4}$



The numerical results presented here take into account the utility functions defined in Scenario 1. The parameters $\lambda = 0.03, \tau = 100$ are used through out the numerical analysis.

The performance of our learning algorithm in the homogeneous case is shown in Fig. 1. In this case we consider $g = 6.6 \times 10^{-4}, N = 40$. We set the sequence $\gamma^k = \frac{1}{k}$ for all iterations k, and the temperature $\beta \to \infty$, note that this choice of β is a good deal since it allows our algorithm to attain the Nash equilibrium.

In Fig. 1 we observe that the probability to be active for a node i ($p_i, \forall i \in \{1...N\}$) converges to the symmetric equilibrium ($p^* = 0.35$). Moreover, it is interesting to notice that the average number of active nodes at the equilibrium approaches the value of ($\Psi = 15$) where Ψ defines the comfort level of the minority game in pure strategy (Fig. 2). Such behavior is, in fact, a convergence to the strictly fully mixed Nash equilibrium discussed in proposition (2).

VI. Conclusion

Coordination of mobiles which are part of a DTN is a difficult task due to lack of permanent connectivity. Operations in DTNs, in fact, do not support the usage of timely feedback to enforce cooperative schemes which may be implemented on mobile nodes. Nevertheless, coordination is worth indeed in order to attain efficient usage of resources. Moreover, selfish behavior and activation control becomes core when owners of relay devices may need incentive to spend memory and battery. To this respect, our paper provides a novel mechanism designed using the theory of Minority Games (MGs). MGs are non-cooperative games which apply to contexts where the payoff of players decreases with the number of those who compete. We could design a reward mechanism for two hop routing protocols that runs fully distributed and with no need for any dedicated coordination protocol. I.e., the source controls how many nodes to activate in order to attain a target message delivery probability. It does so by setting the reward for nodes who deliver first and such in a way to avoid overprovisioning of activated relays. Finally, we developed a distributed stochastic learning algorithm able to converge to the optimal solution.

Future works will investigate how to extend the models and the properties of convergence of our algorithm to other types of networks such as cognitive radios and peer-to-peer networks.

VII. Acknowledgement

This work has been partially supported by the European Commission within the framework of the CONGAS project FP7-ICT-2011-8-317672, see www.congas-project.eu.

References

[1] M. Benaïm. Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 1–68. Springer, Berlin, 1999.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Mar. 2004.

[3] A. Chaintreau, P. Hui, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing*, 6(6):606–620, June 2007. (previously published in the Proceedings of IEEE INFOCOM 2006).

[4] R. El-Azouzi, F. De Pellegrini, H. B. Sidi, and V. Kamble. Evolutionary forwarding games in delay tolerant networks: Equilibria, mechanism design and stochastic approximation. *Computer Networks*, (0):–, 2012.

[5] H. Gintis. *Game Theory Evolving*. Princeton University Press, 2009.

[6] A. Guerrieri, I. Carreras, F. De Pellegrini, D. Miorandi, and A. Montresor. Distributed estimation of global parameters in delay-tolerant networks. *Computer Communications*, 33(13):1472–1482, 2010.

[7] Kets, W., Voorneveld, and M. Congestion, equilibrium and learning: The minority game. (2007-61), 2007.

[8] E. Moro. The Minority Game: an introductory guide. *eprint arXiv:cond-mat/0402651*, Feb. 2004.

[9] P. Mhnen and M. Petrova. Minority game for cognitive radios: Cooperating without cooperation. *Physical Communication*, 1(2):94 – 102, 2008.

[10] G. Neglia and X. Zhang. Optimal delay-power trade-off in sparse delay tolerant networks: a preliminary study. *in* Proc. of ACM SIGCOMM CHANTS 2006*, pp. 237–244*, 2006.

[11] Shang and L. Hui. Self-organized evolutionary minority game on networks. *2007 IEEE International Conference on Control and Automation*, 00:2186–2188, 2007.

[12] L. H. Shang. Self-organized evolutionary minority game on networks. *in International Conference of Control and Automation*, May 30- June 1, 2007.

[13] H. B. Sidi, W. Chahin, R. El-Azouzi, and F. De Pellegrini. Energy efficient minority game for delay tolerant networks. *Technical Report, url:http://arxiv.org/abs/1207.6760*, 2012.

[14] X. Zhang, G. Neglia, J. Kurose, and D. Towsley. Performance modeling of epidemic routing. Elsevier Computer Networks*, vol. 51, no. 10, pp.2867–2891*, 2007.