Adaptive Optimal Stochastic Control of Delay–Tolerant Networks

Eitan Altman, Francesco De Pellegrini, Daniele Miorandi, Giovanni Neglia

Adaptive Optimal Stochastic Control of Delay–Tolerant Networks *

Eitan Altman, Francesco De Pellegrini, Daniele Miorandi and Giovanni Neglia

September 2016

Abstract

Optimal stochastic control of delay tolerant networks is studied in this paper. First, the structure of optimal two-hop forwarding policies is derived. In order to be implemented, such policies require knowledge of certain global system parameters such as the number of mobiles or the rate of contacts between mobiles. But, such parameters could be unknown at system design time or may even change over time. In order to address this problem, adaptive policies are designed that combine estimation and control: based on stochastic approximation techniques, such policies are proved to achieve optimal performance in spite of lack of global information. Furthermore, the paper studies interactions that may occur in the presence of several DTNs which compete for the access to a gateway node. The latter problem is formulated as a cost-coupled stochastic game and a unique Nash equilibrium is found. Such equilibrium corresponds to the system configuration in which each DTN adopts the optimal forwarding policy determined for the single network problem.

1 Introduction

Delay–Tolerant Networks (DTNs) are sparse and/or highly mobile wireless ad hoc networks where no continuous connectivity guarantee can be assumed [2]. One central problem in DTNs is related to the routing of messages towards intended

*Published in IEEE Transactions on Mobile Computing. A shorter version appeared in the Proc. of IEEE INFOCOM, Rio de Janeiro (BR), Apr. 2009 [1]. E. Altman and G. Neglia are with University of Côte d'Azur, Inria, 2004 Route des Lucioles, Sophia-Antipolis (France), email: name.surname@sophia.inria.fr; D. Miorandi is with U-Hopper s.r.l., Via Antonio da Trento, 8, 38122 Trento (Italy), email: daniele.miorandi@u-hopper.com; this work was carried out while he was with CREATE-NET. F. De Pellegrini is with CREATE-NET, via alla Cascata 56/D, 38123, Povo, Trento (Italy), email: fdepellegrini@create-net.org.

destinations. Protocols developed in the mobile ad hoc networks field, indeed, would fail because a complete route to destination may not exist most of the time. A common technique for overcoming lack of connectivity is to disseminate multiple copies of the message in the network: this enhances the probability that at least one of them will reach the destination node within a given time delay [3]. This is referred to as epidemic-style forwarding [4], because, alike the spread of infectious diseases, each time a message–carrying node encounters a new node not having a copy thereof, the carrier may infect this new node by passing on a message copy; newly infected nodes, in turn, may behave similarly. The destination receives the message when it meets an infected node. Hereafter, two variants will be considered. In the first one, that referred to as *address-centric case*, nodes are assigned unique identifiers, and the source node generates messages destined to one specific node. In the second one, that is called *data-centric case* and is meant to model a content–centric DTN system [5], messages include metadata describing the content. Nodes have a filter based on which they discriminate whether a message is of interest or not, e.g., by matching with the message's metadata.

This paper addresses the case where mobile nodes have no *a priori* information on the encounter pattern. In order to develop all the theory thoroughly, the analysis is confined to the case when only the source of the message can generate new copies and hand them over to relays in radio range, while other infected nodes are allowed to forward it to the destination(s) only. The latter is referred to as two–hop forwarding [6, 7]. The main concern of this work is the optimal stochastic control of such routing protocol. The control variable is the probability of transmitting a message upon a suitable transmission opportunity, i.e., a contact with a node that is not the intended destination. The goal is to optimize the probability to deliver a message within a given time delay, while satisfying specific energy constraints.

The control of forwarding schemes in DTNs has received some attention in the literature. In [8], the authors propose an epidemic forwarding protocol based on the *susceptible-infected-removed* (SIR) model [9] and show that it is possible to increase the message delivery probability by tuning the parameters of the underlying SIR model. In [10] a detailed general framework is proposed in order to capture the relative performance of different self–limiting strategies. None of these two papers formalize specific optimisation problems. In [11] and its follow-up [12], the authors assume the presence of a set of special mobile nodes, the ferries, whose mobility can be controlled. Algorithms to design ferry routes are proposed in order to optimize network performance. Some works have been tackling the problem of estimation and consistency of the network state, e.g., how to produce reliable estimates of mobiles presence [13], or the number of mobiles [14], or even the intermeeting frequencies [15]. Compared to such works, the technique proposed in this paper performs such estimation *implicitly*: based on the theory of stochastic approximations applied to optimal forwarding control, no explicit estimates of the system state are needed but the number of released copies of the message. Works more similar to the one presented here are [16, 17, 18]. In [16] the authors consider buffer constraints and derive, based on some approximations, buffer scheduling policies in order to minimize the delivery time. The optimization goal in [17] can be considered a relaxed version of our problem, e.g., the weighted sum of delivery time and energy consumption. Adopting that optimization goal, the authors of [19] provide a rigorous proof of the fact that the dynamics corresponding to the closed loop optimal control acting over a fluid approximation is the fluid limit – for large number of relays – of the optimal Markov decision process.

Also, under a fluid model approximation, the work in [18] provides a general framework for the optimal control of the broad class of monotone relay strategies. Apart from the differences in the optimization functions, most of the above works do not address the problem of on-line estimation of optimal policies; an attempt is done in [10, 16] based on some heuristics for the estimation. Finally, game theoretical tools are applied to the study of competing classes of nodes in a DTN scenario.

The contributions developed in this paper can be summarized with the following three results:

- The analytical characterization of optimal control policies for two–hop routing in DTNs. Optimal policies are proved to have the threshold structure.

- A method – rooted in stochastic approximation theory – able to attain the optimal policy in the absence of knowledge of system-level parameters. Also, the scheme does not require explicit acknowledgement from the destination.

- The optimal control problem is extended to the case of several competing classes of mobile devices. The framework, in this case, is that of cost–coupled stochastic games [20]. The game is proved to have a unique Nash equilibrium, where the sources' best response coincides with the one determined for the single–class case.

The remainder of the paper is organized as follows. In Sec. 2, the problem is formalized and the system model used in the subsequent sections is presented. In Sec. 3 the structure of the optimal control policies is derived. Methods are presented in Sec. 4 to attain optimal control when part of the system's parameters are unknown. The multiclass case is introduced in Sec. 5. Numerical results are presented in Sec. 7. Sec. 8 concludes the paper.

2 System Model

A network of $(N+P)$ mobile nodes is considered, each equipped with some form of proximity wireless communications. For a given message, there exist P potential destinations. In particular, P takes value 1 in the address–centric case, whereas in the data–centric case it is equal to the number of nodes whose filter matches the message metadata. Let N be the number of potential relay nodes (including the source). The network is assumed to be sparse, so that, at any time instant, nodes are isolated with high probability. Communication opportunities arise whenever, due to mobility patterns, two nodes get within mutual communication range. Such events are referred to as "contacts."

The time between subsequent contacts of any pair of nodes is assumed to follow an exponential distribution with parameter $\lambda > 0$. The validity of this assumption for synthetic mobility models (including, e.g., Random Walk, Random Direction, Random Waypoint) has been discussed in [7]. There exist studies based on traces collected from real-life mobility [21] that argue that inter-contact times may follow a power-law distribution. But, it has been shown that these traces and many others exhibit an exponential tail after a cutoff point [22].

For the sake of tractability, the sequences of inter-contact times are assumed to be mutually independent. The degree to which real-world traces follow this assumption depends on a number of factors, related in particular to the correlation of the mobile devices' trajectories. Experimental evidence suggests that this assumption can model relatively well scenarios in which mobile trajectories are not correlated and the mobility pattern leads to a "fast mixing" of the IDs of nodes that meet (a good example is Random Waypoint in a sparse scenario with fast moving nodes). On the other hand, this assumption does not hold in case of (i) mobiles that tend to remain around a given point (e.g., Random Walk) (ii) nodes moving in clusters and/or mobility characterized by attraction points (points of interest).

The case of heterogeneous DTNs, i.e., with different intermeeting intensity, is out of the scope of the present work; nevertheless, the presented framework can be extended to handle such a case as well [23].

There can be multiple source-destination(s) sets, but the analysis is here limited to a single message, eventually with many copies, spreading in the network.[1] For simplicity, a message is assumed to be generated at time $t = 0$. The message transmitted is also assumed to be relevant before some deadline τ. This applies, e.g., to environmental information or data referring to events of transient nature (e.g., social events or meetings). The message contains a timestamp reporting its

[1]Results in sections 3 and 4 are valid when multiple messages are present in the network at the same time, provided that the bandwidth and the buffer are large enough to assure that the different propagation processes are not interfering.

generation time, so that it can be dropped when it becomes irrelevant.[2] Due to disconnected operations, it is assumed that no feedback mechanism allows the source or other mobiles to know whether the message has been successfully delivered to the destination within time τ. The focus of the work is on a set of relaying strategies that can be defined as probabilistic two-hop routing strategies. At each encounter between the source and a mobile that does not have the message, the message is relayed with some probability taking values in $U = [u_{\min}, u_{\max}] \subseteq [0, 1]$. It is worth observing that $0 < u_{\max} \leq 1$ is equivalent to a thinning operation of the meeting point process, i.e., assuming $\lambda' = u_{\max}\lambda$; $u_{min} > 0$ is required in certain rewarding mechanisms used in order to incentivize relays to take part to the forwarding process [24, 25].

If a mobile that is not the source has the message and it is in contact with another mobile, then it transfers the message if and only if the other mobile is the destination node. It is worth remarking that, by using the two-hop routing, control is operated only by the source and the main energy cost to deliver a message is met by the source of the message and not by relay nodes.

A discrete time model is adopted in the rest of the work, i.e., time is slotted with time slot duration Δ, where $\tau = K\Delta$. The n–th slot corresponds to the interval $[n\Delta, (n+1)\Delta)$. In this discrete time setting, a mobile receiving a copy during a tagged time slot can forward it starting from the following one. Moreover the forwarding probability during $[n\Delta, (n+1)\Delta)$ is a constant and it is denoted by u_n.

Let X_n be the number of mobiles, not including the potential destinations, that have a copy of the message at time $n\Delta$ (i.e., the beginning of the n–th slot), $X_0 = 1$. Under the assumptions above, X_n is a Markov chain with possible states $1, \cdots, N$. The transition rates depend on the forwarding probability used by the source at each time slot, so that a natural way to optimize the performance of the system is to control this parameter.

The problem addressed in this paper is to *maximize the probability to deliver the message to the destination.*[3] by time $K\Delta$, *under a constraint on the expected number of message copies injected in the system* By posing a constraint on the number of copies, the proposed model poses a constraint on the energy expended at the source and at relays. Actually, if constant per-contact energy is required in order to forward a message, such a constraint can express the network energy consumption due to both transmission and reception of a message.

The aim is to determine optimal time-dependent forwarding policies the source

[2]Time elapsed since generation can be traced summing up the time elapsed at each node with no need for global synchronization.

[3]In the data–centric case, one randomly chosen potential destination node is considered.

can implement.

More formally a forwarding policy (control policy) is defined as a function $\mu : \{0, 1, \ldots, K-1\} \to U$. The control used at time n is denoted by u_n.

In what follows a key role will be played by two types of forwarding policies, *static* and *threshold* policies, defined as follows:

Definition 2.1. *A policy μ is a* static *policy if μ is a constant function, i.e., $u_n = p \in U$, for $n = 0, 1, 2, \ldots, K-1$. A policy μ is a* threshold *policy if there exists $h \in \{0, 1, 2, \ldots, K-1\}$ (the threshold) such that $u_n = u_{\max}$ if $n < h$ and $u_n = u_{\min}$ if $n > h$.*

Static and threshold policies are different from the implementation standpoint. In fact, with static policies, at each communication opportunity, message forwarding is done with constant probability p. Conversely, with threshold policies, each time a mobile has a forwarding opportunity, it checks the time t elapsed since the message generation time and it forwards the message with some probability $u(t)$.

It is worth observing that static and threshold policies depend on few parameters only, i.e., the control p for static policies, and the threshold h and the corresponding value u_h for dynamic policies.

The following section provides the characterization of optimal static and threshold policies. In Table 1 the notation used throughout the paper is reported for the sake of reading.

3 Characterization of Optimal Policies

3.1 Preliminaries

Let $F_D(n)$ denote the probability that a message generated at time 0 is received before $n \cdot \Delta$ under control policy μ (by the intended destination in the address–centric case and by one randomly chosen potential destination in the data–centric case). The objective is to derive policies that maximize $F_D(K)$, while satisfying the following constraint on the expected number of message copies: $\mathrm{E}[X_K] \leq \Psi$. Let X_0 be the number of nodes with a copy of the message at time 0.

Before proceeding with the main statements of this section, closed forms for the dynamics of $\mathrm{E}[X_n]$ and $F_D(n)$ are derived.

Let $\zeta_r(j)$ be the indicator that the j–th relay among the $N - X_0$ relays that do not have the message at time 0, receives the message in slot $[r\Delta, (r+1)\Delta)$, for $0 \leq r \leq n-1$.

Random variables $\zeta_r(j)$ are stochastically increasing in the control actions u_k, and so are their sums (see [26] for definition and properties of usual stochastic

Symbol	*Meaning*
P	number of destinations
N	number of relay nodes (including the source)
λ	intermeeting intensity
Δ	time slot
$\tau = K \cdot \Delta$	timeout value
X_n	number of nodes having a copy of the message at time $n\Delta$
Ψ	maximum expected number of message copies
$F_D(n)$	probability that the message is delivered by time $n\Delta$
$\mu(\cdot)$	control policy
u_n	value taken by the control variable (i.e., forwarding probability) at time $n\Delta$
p	value taken by the control variable under static control
h	time threshold
θ	$= \sum_{k=0}^{K-1} u_k$
β	θ value for the optimal policy
$\zeta_r(j)$	indicator that the j–th mobile relay, among the $N - X_0$ ones that do not have the message at time 0, receives it during the r-the slot
$\overline{X}_m$	estimate of $\mathrm{E}[X_K]$ at the m-th round of the stochastic approximation algorithm
$\Pi_H(u)$	projection over H of the value u
$\{\cdot\}^{(i)}$	superscript indicates that the quantity refers to the i-th class of mobile nodes
$Y_n^{(i)}$	number of class i infected nodes that can transmit to the destination during the n-th time slot
S_n	total number of infected nodes that can transmit to the destination during the n-th time slot
$S_n^{(-i)}$	total number of infected nodes that can transmit to the destination during the n-th time slot but class i-th ones

Table 1: Notation used throughout the paper.

order). Also, variables $\zeta_r(j)$ are independent and identically distributed Bernoulli random variables. We observe that the expected value $\mathrm{E}\left[\zeta_r(j)\right] = (1-\exp(-\lambda u_r \Delta))\exp(-\lambda\Delta\sum_{h=0}^{r-1} u_h)$ is obtained by combining the independent increment of the intermeeting process with a thinning argument, because the source node implements independent random decisions in each slot.

Hence, $\sum_{r=0}^{n-1} \zeta_r(j)$ is the indicator that the j–th relay among the $N-X_0$ relays that do not have the message at time 0 receives the message within slot $n-1$. It holds

$$\Pr\left\{\sum_{r=0}^{n-1} \zeta_r(j) = 1\right\} = 1 - \exp\Big(-\lambda\Delta\sum_{h=0}^{n-1} u_r\Big)$$

The dynamics of the number of message copies obeys to the equation

$$X_n = X_0 + \sum_{j=1}^{N-X_0}\sum_{r=0}^{n-1} \zeta_r(j). \tag{1}$$

As observed above, X_k is stochastically increasing in the control action u_k as well: if policies μ' and μ differ only at index k, $0 \le k \le K$, where $u'_k > u_k$, it follows that

$$X'_n >_{st} X_n, \ \forall n > k \tag{2}$$

This formalizes the intuition that, under the monotonicity conditions of our theoretical framework, the higher the forwarding probability, the higher the number of infected nodes. A stronger characterization for the monotonicity in the control action is provided later in this section.

For the expected number of message copies, it holds

$$\begin{aligned}\mathrm{E}\left[X_n\right] &= X_0 + \mathrm{E}\left[\sum_{j=1}^{N-X_0} \zeta_{0,n}(j)\right]\\ &= X_0 + (N - X_0)\left(1 - \exp(-\lambda\Delta\sum_{k=0}^{n-1} u_k)\right)\end{aligned} \tag{3}$$

To characterize $F_D(n)$, some further notation is introduced. Hereafter let X denote the stochastic process corresponding to the number of nodes with a copy of the message until time n, i.e. $X = (X_0, X_1, \dots, X_{n-1})$ and let $x = (x_0, x_1, \dots, x_{n-1})$ represent its sample path x. Let $\chi = \{0, 1, \dots, N\}$ so that $X \in \chi^n$ and $X_h \in \chi$ for all $h = 0, 1, \dots, n-1$. Also, let $P_X(x)$, $x \in \chi^n$, be the probability distribution of X. Also, $G(n) = 1 - F_D(n)$ describes the complementary cumulative distribution function of the delay, which writes

$$
\begin{aligned}
& G(n) = 1 - F_D(n) = \Pr\big\{\text{no delivery by } n\Delta\big\} \\
& = \sum_{x \in \chi^n} P_X(x) \Pr\big\{\{\text{no delivery by } n\Delta\} | X = x\big\} \\
& = \sum_{x \in \chi^n} P_X(x) \Pr\big\{ \cap_{r=0}^{n-1} \ \{\text{no delivery in slot } r\} | X = x\big\} \\
& = \sum_{x \in \chi^n} P_X(x) \prod_{r=0}^{n-1} e^{-\lambda x_r \Delta} = \mathrm{E}\left[e^{-\lambda\Delta \sum_{r=0}^{n-1} X_r}\right]
\end{aligned}
$$

Now, it is possible to observe that

$$
\begin{aligned}
\sum_{r=0}^{n-1} X_r \quad & = nX_0 + \sum_{j=0}^{N-X_0} \sum_{r=0}^{n-2} (n-r-1)\zeta_r(j) \\
& = nX_0 + \sum_{j=0}^{N-X_0} V(j)
\end{aligned} \tag{4}
$$

where the auxiliary random variables

$$
V(j) := \sum_{r=0}^{n-2} (n-r-1)\zeta_r(j) \tag{5}
$$

In (5), index j identifies the j-th relay among those which do not have the message at time 0; also, slot $n-2$ is the last slot when j can receive the message in order to have a message copy at time $(n-1)\Delta$, i.e., at the beginning of slot $n-1$.

Hence, $V(j)$ is the number of slots during which j has been holding the message within the interval $[r\Delta, (n-1)\Delta)$. In fact, $(n-1) - r$ is the number of slots elapsed till the end of the interval once j received the message in slot r. Note that $V(j)$ and $V(i)$ are i.i.d for $j \neq i$ because the intermeeting intervals of source and relays i and j are i.i.d. as well.[4] Also, control u_{n-1} does not appear in (5): in fact, by model assumption, new nodes infected in slot $n-1$ are able to deliver the message to the destination only starting from the following time slot, i.e., beyond time $n\Delta$.

[4]Observe that independence assumption is needed just for intermeetings where one of the nodes is the source or a destination

In particular, let $p_V(v) = \Pr\{V = v\}$ (index j is omitted for simplicity's sake), it holds

$$p_V(v) = \begin{cases} e^{-\lambda\Delta \sum_{r=0}^{n-2} u_r}, & v = 0 \\ (1 - e^{-\lambda\Delta u_{n-v}})e^{-\lambda\Delta \sum_{r=0}^{n-v-2} u_r}, & v = 1, \ldots, n-1 \end{cases}$$

Hence, the following expression is obtained

$$\begin{aligned} G(n) &= \mathrm{E}\left[e^{-\lambda\Delta\left(nX_0 + \lambda\Delta \sum_{j=0}^{N-X_0} V(j)\right)}\right] \\ &= e^{-\lambda\Delta n X_0} \mathrm{E}\left[e^{-\lambda\Delta\, V(1)}\right]^{N-X_0} \end{aligned} \tag{6}$$

From (6), it follows that X_n has stronger monotonicity properties than those expressed by (2).

Corollary 3.1. *Let $k \in \{0, 1, \ldots, K-2\}$ and let $\mu = (u_0, \ldots, u_{K-1})$ be a given policy. A new policy $\mu' = (u'_0, \ldots, u'_{K-1})$ is defined as: $u'_n = u_n$ for $n \neq k$ and $u'_k > u_k$. Let X'_n and $F'_D(\cdot)$ be the number of message copies and the CDF of message delay when μ' is employed, respectively. It then holds: $F_D(K) < F'_D(K)$, $E[X_K] < E[X'_K]$.*

Proof. The statement concerning $\mathrm{E}[X_K]$ follows by inspection of (3).

The statement concerninig $F_D(K)$ requires some further insight. First observe that from (6), and considering $0 \le r \le K-2$, it follows:

$$\begin{aligned} &\frac{d}{du_r} p_V(v) = -\lambda\Delta p_V(v), \quad \text{if } 0 \le v < K - r - 1 \\ &\frac{d}{du_r} p_V(v) = \frac{\lambda\Delta\, e^{-\lambda\Delta u_r}}{1 - e^{-\lambda\Delta u_r}} e^{-\lambda\Delta \sum_{h=0}^{n-r-2} u_h}, \text{ if } v = K - r - 1 \\ &\frac{d}{du_r} p_V(v) = 0, \quad \text{if } K - r - 1 < v \le K - 1 \end{aligned} \tag{7}$$

Given policy $\mu = (u_0, \ldots, u_{K-1})$, consider the new policy $\mu' = (u'_0, \ldots, u'_{K-1})$, defined as: $u'_n = u_n$ for $n \neq k$ and $u'_k > u_k$. Denote the corresponding r.v. defined in (5) as V' and V, respectively and let $p_{V'}(v) = p_V(v) - \delta p(v)$. From (7), it follows $\delta p(v) > 0$ for $0 \le v < K - k - 1$, $\delta p(v) = 0$ for $K - k - 1 < v \le K - 1$, and, finally:

$$\delta p(K - k - 1) = - \sum_{h=0}^{K-k-2} \delta p(h)$$

Now, it is possible to write

$$\mathrm{E}\left[e^{-\lambda\Delta\, V'}\right] - \mathrm{E}\left[e^{-\lambda\Delta\, V}\right] = \sum_{h=0}^{K-k-1} \frac{1}{z^h}\big(p_{V'}(v) - p_V(v)\big)$$
$$= \sum_{h=0}^{K-k-1} \frac{1}{z^h}\delta p(h) = \frac{1}{z^{K-k-1}} \sum_{h=0}^{K-k-2} \delta p(h) - \sum_{h=0}^{K-k-2} \frac{1}{z^h}\delta p(h)$$
$$= \sum_{h=0}^{K-k-2} \delta p(h)\Big(\frac{1}{z^{K-k-1}} - \frac{1}{z^h}\Big) < 0 \tag{8}$$

where $z := e^{\lambda\Delta} > 1$ for the ease of notation. The statement follows from the strict monotonicity of $G(n)$ in $\mathrm{E}\left[e^{-\lambda\Delta\, V}\right]$ as in (6). $\square$

Also, as a direct consequence of the previous proposition, the following holds:

Corollary 3.2. *If an optimal policy exists, either it is the static policy μ_{max} with $\mu_{max}(n) = u_{max},\ \forall n$, or it saturates the constraint, i.e., $E[X_K] = \Psi$.*

Proof. Consider a policy μ, that is different from μ_{max} (i.e., k exists s.t. $\mu(k) < u_{max}$) and does not saturate the constraint ($\mathrm{E}[X_K] < \Psi$). As the expected number of infected nodes is a increasing continuous function of the forwarding probabilities, from μ a new policy μ' can be obtained, by increasing the forwarding probability in k, while satisfying the constraint $\mathrm{E}[X'_K] \leq \Psi$. The new policy has better performance, since $F'_D(K) > F_D(K)$, which is a contradiction. $\square$

It is immediate to observe that the set of admissible policies is empty if and only if the policy $\mu_{min}(n) = u_{min}$ for all n, does not satisfy the constraint.

3.2 Threshold structure of the optimal policy

Many Markov Decision Processes (MDPs) have optimal policies with threshold structure. There has been much work on characterizing properties of the cost and/or transition probabilities of an MDP that imply the threshold structure in the optimal policies. Some examples are [27, 28]. In absence of constraints, MDPs are known to possess pure optimal policies, so it is not surprising that unconstrained MDPs often have pure threshold optimal policies.

The control problem proposed in this work falls into the category of constrained MDP where one criterion is maximized while keeping the other one below a threshold. For such problems (where there is a single constraint) it is known from from [29] that pure optimal policies need not exist, but that Markov (and in some cases stationary) policies do exist which require randomization in at most one state.

In the rest of this section, the optimal policy in the constrained case is a threshold policy, where randomization is needed at most once, i.e, at the time that coincides with the threshold. Thus the optimal policy uses one pure action below the threshold, another one above it, and a randomization when at the threshold. Interestingly, there are many MDPs that have optimal pure threshold policies but for which, when adding constraints, the optimal policy is quite different than the one obtained here. Examples and conditions to obtain policies that have different structures are given in [30].

The first main result is described by the following theorem:

Theorem 3.1. *There exists an optimal threshold policy. A non-threshold policy is not optimal.*

Proof. The existence of an optimal policy follows from elementary properties of Markov decision processes [31]. By contradiction, consider a non threshold policy μ that satisfies the constraint ($\mathrm{E}[X_K] \leq \Psi$) and construct an alternate policy μ' which is better off. Since μ is non threshold, there exists some time $k \leq K-1$ and some $\epsilon > 0$ such that $u_{k-1} < u_{\max} - \epsilon$ and $u_k > u_{\min} + \epsilon$.

Let μ' be the policy obtained from μ by setting $u'_{k-1} = u_{k-1} + \epsilon$ and $u'_k = u_k - \epsilon$ (the other components are the same as those of μ). Let X'_n be the state process under μ'. Also, $F'_D(\cdot)$ is defined correspondingly.

First, observe that by construction the new policy is admissible since $\sum_{h=0}^{n-1} u'_h = \sum_{h=0}^{n-1} u_h$ from (3) $\mathrm{E}[X'_K] = \mathrm{E}[X_K] \leq \Psi$. Then, from (6) it is sufficient to show that $\mathrm{E}[e^{-\lambda\Delta V'}] < \mathrm{E}[e^{-\lambda\Delta V}]$.

In order to do so, the following observation is useful: from the expression of $p_V(v)$, it holds $p_{V'}(v) = p_{V'}(v)$ for $v \neq\in \{k-1, k\}$. Also, it holds

$$\begin{aligned}
p_{V'}(K-k-1) &= \left(1 - e^{-\lambda\Delta(u_{k-1}+\epsilon)}\right) e^{-\lambda\Delta\sum_{h=0}^{k-2} u_h} \\
&> \left(1 - e^{-\lambda\Delta u_{k-1}}\right) e^{-\lambda\Delta\sum_{h=0}^{k-2} u_h} = p_V(K-k-1) \\
p_{V'}(K-k) &= \left(1 - e^{-\lambda\Delta(u_k-\epsilon)}\right) e^{-\lambda\Delta(u_{k-1}+\epsilon)} e^{-\lambda\Delta\sum_{h=0}^{k-2} u_h} \\
&< \left(1 - e^{-\lambda\Delta u_k}\right) e^{-\lambda\Delta\sum_{h=0}^{k-1} u_h} = p_V(K-k)
\end{aligned} \tag{9}$$

From the normalization condition it follows

$$p_{V'}(K-k-1) = p_V(K-k-1) - \delta,\; p_{V'}(K-k) = p_{V'}(K-k) + \delta$$

Finally, it is possible to write

$$\mathrm{E}\left[e^{-\lambda\Delta V'}\right] - \mathrm{E}\left[e^{-\lambda\Delta V}\right] = \sum_{h=0}^{K-1} \frac{1}{z^h}\big(p_{V'}(v) - p_V(v)\big) =$$
$$= -\delta \cdot \left(\frac{1}{z^{n-k-1}} - \frac{1}{z^{n-k}}\right) < 0$$

It is hence proved that μ' is better off μ, which concludes the proof. □

It is now possible to determine the optimal threshold policy. Due to Corollary 3.2, the optimal policy is either the static policy μ_{max} or the constraint has to be saturated. In the second case, by using the expression derived above for the average number of copies, the optimal policy satisfies

$$N - (N - X_0)\exp(-\lambda\Delta \sum_k u_k) = \Psi.$$

Hence the following optimality condition is obtained:

$$\sum_{k=0}^{K-1} u_k = -\frac{1}{\lambda\Delta}\log\left(\frac{N-\Psi}{N-X_0}\right) =: \beta \tag{10}$$

This directly yields the threshold h^* of the optimal policy, by considering that $u_n = u_{max}$ for $n < h^*$ and $u_n = u_{min}$ for $n > h^*$ while satisfying Eq. (10). Then

$$\begin{aligned} h^* = \quad & \max\{l \in \mathbb{N} :\ 0 \le l \le K-1, \text{s.t.} \\ & v(l) = l \cdot u_{\max} + (K-1-l) \cdot u_{\min} \le \beta\}, \end{aligned} \tag{11}$$

and $u_{h^*} = \beta - v(h^*)$. In the particular case of $u_{min} = 0$, this reduces to $h^* = \lceil\beta\rceil$ and $v(h^*) = \beta - \lfloor\beta\rfloor$. If $u_{\min} = 0$ and $u_{\max} = 1$ then the optimal policy chooses $u_k = 1$ for all $k < \beta$ and $u_k = 0$ for all $k \ge \beta + 1$. At the remaining time, $k = \lceil\beta\rceil$, it uses $u_k = \beta - \lfloor\beta\rfloor$.

The same reasoning can be applied to determine the best static policy. In particular it is μ_{max}, if μ_{max} satisfies the constraint (and in such case the best static policy is also the optimal one), otherwise Eq. (10) holds, and by imposing $u_n = p^*$ for all n, then $p^* = \beta/(K-1)$.

It is worth remarking that the results on existence, structure and form of the optimal forwarding policy are the same for the address–centric and for the data–centric cases.

3.3 Comments

The result obtained on the optimality of a threshold-type policy is worth some comments. Intuitively, one would expect that an optimal policy for our problem depends on the (Markovian) state X_k. The above result shows that the knowledge of such variable is not necessary to derive an optimal policy. Indeed, the optimal policy falls in the class of *open loop* policies, i.e., which do not depend on the underlying state. The key point here is that, under the assumptions of the proposed model, a *strict monotonicity* property holds for the dynamics of the number of copies.

Indeed, the results described above state that (i) the delivery probability is strictly monotone in the number of copies in the network (ii) the evolution of the number of copies in the network (the state X_k) is strictly monotone in the control. These properties do not depend on whether the policy is open vs. closed loop. Also, (3) tells us that by using *open loop* policies – given the knowledge of some system level parameters such as λ and N – it is possible to fully control the evolution of $\mathrm{E}[X_k]$, i.e., of the average number of copies in the network. As the constraint is on $\mathrm{E}[X_k]$, the optimal policy results of the open loop type: hence, for any optimal closed loop policy it is possible to derive an equivalent open loop one.

It is possible to compare the threshold–type policy with the well-known spray–and–wait scheme. In a version of such a scheme, indeed, the source node distributes a copy of the message to the first L nodes encountered. This can be seen as a controlled two–hop routing, where the control is a *closed loop* policy, i.e., one that depends on the state of the system (the number of copies X_k). By assuming $u_{min} = 0$ it is possible to compare spray–and–wait with a threshold–type policy: the main difference is that the latter does not satisfy a bound on the number of copies made on the single realization, but, rather, on the *expected* number of copies.

4 Stochastic Approximations for Adaptive Optimization

From the results in the previous section and, in particular, from (10), it can be seen that the design of optimal policies requires the knowledge of global parameters such N and λ. In a real setting, such parameters may be unknown or can change over time. In this section methods are introduced in order to achieve optimal control policies in the absence of information on such parameters.

The approach is based on stochastic approximation theory [32]. This framework generalizes Newton's method to determine the root of a real-valued function when only noisy observations of such function are available.

Recall the two frameworks of optimization considered throughout the paper:

- Static control: find $p^* \in [u_{\min}, u_{\max}]$ such that the policy $u_n = p^*$ has the best performance among all static policies.

- Dynamic control: find $h^* \in \{0, 1, \cdots, K-1\}$ and $\mu(h^*)$ characterizing the optimal policy.

It is possible to approach on-line estimation of optimal static and dynamic control in a unified way. For a generic policy, let $\theta = \sum_{k=0}^{K-1} u_k$ denote the sum of the controls used over the K time slots. θ is uniquely determined from policy μ, but it also identifies uniquely a static or a threshold policy. For the static policy it is indeed $u_n = p = \theta/K$, while for the threshold policy h the dependence on β is given by (11).

Note that if $\theta = \beta$, then the two policies are the optimal static and threshold policies determined in the previous section. The problem reduces therefore to the distributed estimation of β.

The stochastic approximation algorithm will estimate β looking for the unique solution of a certain function in θ in the interval $[\theta_{\min}, \theta_{\max}] = [K \cdot u_{\min},\ K \cdot u_{\max}]$.

The algorithm works over rounds. Each round corresponds to the delivery of a set of messages. During a given round, a policy is used. $\mu(m)$ indicates the policy adopted at round m and $\theta(m) = \sum_{k=0}^{K-2} u_k(m)$ the corresponding θ value. At the end of each round an estimate of $\mathrm{E}[X_K]$ can be evaluated by averaging the total number of copies made during the round for each different message. Let $\overline{X}(m)$ denote such average. $\overline{X}(m)$ is used to update θ, according to the following formula:

$$\theta(m+1) = \Pi_H\Big(\theta(m) + a_m(\Psi - \overline{X}(m))\Big), \tag{12}$$

where $\Pi_H(\cdot)$ is the projection in $[\theta_{\min}, \theta_{\max}]$:

$$\Pi_H(\theta) = \max\left[\theta_{\min}, \min\left(\theta, \theta_{\max}\right)\right].$$

As discussed above, the new policy $\mu(m+1)$ is univocally determined from $\theta(m+1)$. The length of a round should be taken in such a way to enable a stable estimate of the mean number of copies performed with the policy currently in use.

The following theorem shows the convergence property of the algorithm.

Theorem 4.1. *If the sequence $\{a_m\}$ is chosen such that $a_m \geq 0\ \forall m$, $\sum\limits_{m=0}^{+\infty} a_m = +\infty$ and $\sum\limits_{m=0}^{+\infty} a_m^2 < +\infty$, the sequence of policies μ_m converges to the optimal policy with probability one.*

Proof. On the basis of the considerations at the beginning of this section it is sufficient to prove that $\theta(m)$ converges with probability one to β. The proof is divided in two parts. First sequence $\theta(m)$ is showed to converge to some limit set of the following Ordinary Differential Equation (ODE)

$$\dot{\theta} = \Psi - E[X_K|\theta]. \tag{13}$$

The convergence is a consequence of Theorem 2.1 in [32] (page 127). In order to use such a result, it is sufficient to show that:

i) $\sup_m E[Z^2(m)] < +\infty$, where $Z(i) = \Psi - \overline{X}(i)$: this is satisfied since $|Z(m)| \leq N$ for all m;

ii) $\sum_{m=0}^{+\infty} a_m^2 < +\infty$: this follows from the assumptions on the sequence $\{a_m\}$;

iii) There exist a measurable function $\overline{g}(\cdot)$ and r.v. η_i such that $\mathrm{E}_m Z(m) = \mathrm{E}[Z(m)|\theta(0), Z(i), i < m] = \overline{g}(\theta) + \eta(m)$;

iv) $\overline{g}(\cdot)$ is continuous;

v) $\sum_{m=0}^{+\infty} |\eta(m)| a_m < +\infty$ w.p. 1.

For the case at hand, it is possible to write in closed form

$$\begin{aligned} \mathrm{E}_m Z(m) \quad &= \mathrm{E}[Z(m)|\theta(0), Z(i), i < m] = \mathrm{E}[Z(m)|\theta(m)] \\ &= (\Psi - N) + (N - X_0)\, e^{-\lambda\theta(m)\Delta} \end{aligned} \tag{14}$$

Hence, it follows that $\mathrm{E}_m Z(m) = \overline{g}(\theta(m))$ where function $\overline{g}(\theta) = (\Psi - N) + (N - X_0)\, e^{-\lambda\theta\Delta}$ is clearly continuous in θ and thus also measurable. Note that $\eta(m) \equiv 0$ for $\forall m$, i.e., an unbiased estimator of the number of copies is employed. This fact, together with the fact that θ is bounded with probability one, concludes the first part of the proof.

The second part of the proof shows that the solution of such ODE converges to β as time diverges. From Eq. (1):

$$E[\overline{X}(m)|\theta(m)] = E[X_K|\theta(m)] = N - (N - X_0)e^{-\lambda\theta(m)\Delta} \tag{15}$$

so that Eq. (13) writes

$$\dot{\theta} = \Psi - N + (N - X_0)\, e^{-\lambda\theta(m)\Delta}. \tag{16}$$

In order to show that (16) converges to β, it is sufficient to observe two facts. First, by inspection, $\theta^* = \beta$ is an equilibrium point of (16). Second, as $E[X_K|\theta]$ is strictly monotone in θ, the equilibrium point is unique. In order to demonstrate the

stability of the estimator, the standard Lyapunov function $V(\theta) = (\theta - \theta^*)^2$ can be employed.
Then, it follows:

$$\dot{V}(\theta) = 2(\theta - \theta^*) \cdot \dot{\theta} = 2 \left[\theta + \frac{1}{\lambda\Delta} \log \left(\frac{N - \Psi}{N - X_0} \right) \right] \cdot \\ \cdot \left[\Psi - N + (N - X_0)\, e^{-\lambda\theta\Delta} \right] < 0 \text{ for } \theta \neq \theta^* \quad (17)$$

Thus, the asymptotic global stability of θ^* follows from the Lyapunov's theorem. □

Remark 4.1. *The proof is based on the fact that sequence $\theta(m)$ converges to some limit set of* (16). *Theorem 2.1 in [32] shows that $\theta(m)$ converges to $\theta(t_m)$, where $\theta(t)$ is the solution of Eq.* (13) *and $\{t_m\}_{m\geq 0}$ is the sequence defined as follows:*

$$t_0 = 0,\ t_m = t_{m-1} + a_m,\ \textit{for}\ m > 0 \quad (18)$$

Remark 4.2. *Given the expressions obtained in the previous section, Eq. (16) can be solved, leading to:*

$$\theta(t) = \frac{1}{\lambda\Delta} \ln \left\{ e^{\lambda\Delta[(\Psi - N)t + \theta(0)]} + \frac{N - X_0}{N - \Psi} \left[1 - e^{\lambda\Delta(\Psi - N)t} \right] \right\} \quad (19)$$

Remark 4.3. *The description of the algorithm above has been derived by assuming that the on-line estimation of the optimal control is obtained by using in Eq. (12) the estimation $\overline{X}(m)$ obtained from real message transmissions. This constraint can be overcome by using virtual messages. Indeed, the stochastic approximation technique requires the source to estimate the number of copies it would transmit during a time window of duration τ if it* had *a message to transmit. Then the source can simply register the contacts and "virtually" apply the policy. Multiple instances of the virtual relaying protocol (with different policies) can be run in parallel, speeding up the convergence to the optimal policy. If a real message has to be transmitted, the current policy estimation can be used.*

4.1 Optimal Choice of the Sequence $\{a_m\}$

The performance of the stochastic approximation algorithm (12) is known to depend heavily on the choice of the sequence $\{a_m\}$ [33]. By comparing Eq. (12) and Eq. (18), a trade-off can be observed. In fact, sequences $\{a_m\}$ vanishing slower guarantee a faster convergence to the ODE trajectory because the series $\sum a_m$ diverges faster and then t_m in Eq. (18) is larger. At the same time the corresponding estimation experiences increased noise effects due to the larger step size in the iterates equation (12).

A standard choice is $a_m = \frac{C}{m}$; the optimal value of C that guarantees the smallest asymptotic variance is $C = \left.\frac{\partial E[X(\tau)|\theta]}{\partial \theta}\right|_{\theta=\theta^*}$ ([32]). In general, however, C is unknown (as it depends on the unknown function $E[X(\tau)|\theta]$) and cannot be set a priori.

Another possible approach to improve the performance is to use techniques such as Polyak's averages [32, 34]. The idea is to use larger "jumps" to let the iterates converge faster, while using averages to smooth actual estimates. In Polyak's method, a sequence $a_m = O(m^{-1})$ can be adopted: in particular, one that satisfies the condition $a_m/a_{m+1} = 1 + o(a_m)$ and use the outcome as estimation of the optimal policy (i.e., as the control to be used on real messages)

$$\Theta(m) = \frac{1}{m}\sum_{k=1}^{m} \theta(k). \tag{20}$$

It is worth noting that such procedure does not add computational complexity, as the iterated Θ_m can be conveniently computed as

$$\Theta(m+1) = \Theta(m) + \frac{\theta(m+1) - \Theta(m)}{m+1}. \tag{21}$$

Observe that the use of virtual messages described before naturally decouple the iterations of the basic stochastic approximation equation (12) and the control used on the real messages, i.e., the averaged value of (21).[5]

In Section 7 Polyak's averaging techniques will be showed convenient in terms of convergence time.

4.2 Constant Step Approximations

In a real DTN implementation, one appealing feature of the solution proposed is that it can track changing conditions. This is attained by a specific stochastic approximation technique adopting constant step approximations, i.e., by assuming $a_m = \varepsilon$ for all m.

In this way, the system keeps on modifying its behaviour, in an open–ended fashion. Even for such case, results on convergence can be derived, even though in weaker form:

[5] If we would simply choose to plug in the averaged value into the primary control, the effect of coupling would vanish the advantage of averaging [32].

Theorem 4.2. *Consider iterates of the form:*

$$\theta^\varepsilon(m+1) = \Pi_H\Big(\theta^\varepsilon(m) + \varepsilon(\Psi - \overline{X}^\varepsilon(m))\Big). \tag{22}$$

For any $\delta > 0$, define by $N_\delta(\theta^*) = \{x \in \mathbb{R} : |x - \theta^*| < \delta\}$. As $\varepsilon \to 0$, sample paths $\theta^\varepsilon(m)$ converge in distribution to elements in $N_\delta(\theta^*)$. The fraction of time spent by the process in $N_\delta(\theta^*)$ during $[0, T]$ goes to 1 as time horizon T diverges.

Proof. The convergence is a consequence of [32][Thm.2.1, p. 248]: in order to apply that result, a set of sufficient conditions needs to be verified. The first such condition is verified by construction since the constraint set $[\theta_{\min}, \theta_{\max}]$ is a rectangle (condition A4.3.1 in [32][Thm.2.1, p. 248]).

In addition, once defined $Y^\varepsilon(i) = \Psi - \overline{X}^\varepsilon(i)$, the following conditions have to be verified:

i) $\{Y^\varepsilon(i)\}$ are uniformly integrable;

ii) There exist measurable functions $g_m^\varepsilon(\cdot)$ and r.v. η_i such that $\mathrm{E}_m Y^\varepsilon(m) = \mathrm{E}[Y^\varepsilon(m)|\theta(0), Y^\varepsilon(i), i < m] = g_m^\varepsilon(\theta) + \beta^\varepsilon(m)$;

iii) $\lim\limits_{m,n,\varepsilon}^{+\infty} \sum_{r=n}^{n+m-1} \mathrm{E}_m \beta^\varepsilon(m) = 0$;

iv) There exists a continuous function $g(\cdot)$ such that, for each $\theta \in [\theta_{\min}, \theta_{\max}]$, $\lim\limits_{m,n,\varepsilon}^{+\infty} \sum_{r=n}^{n+m-1} \big(g_m^\varepsilon(\theta) - g(\theta)\big) = 0$,

where the meaning of the limit appearing in iii) and iv) is that $m, n \to \infty$ and $\varepsilon \to 0$, in any way possible.

Indeed i) holds because $Y^\varepsilon(i)$ has finite support. As in Thm. 4.1, ii) derives from the closed form

$$\mathrm{E}_m Y^\varepsilon(m) = (\Psi - N) + (N - X_0)\, e^{-\lambda\theta^\varepsilon(m)\Delta} \tag{23}$$

so that $\overline{g}_m^\varepsilon(\theta) = (\Psi - N) + (N - X_0)\, e^{-\lambda\theta^\varepsilon(m)\Delta}$; also, as remarked in Thm. 4.1, $\beta^\varepsilon(m) \equiv 0$ for $\forall m$, i.e., an unbiased estimator of the number of copies is employed, thus iii) follows and iv) is verified by defining $g(\theta)$ as in Thm. 4.1. ☐

5 The Multiclass Case

In this Section the model is extended to the case of several competing DTNs. Such extension will draw on results from weakly coupled stochastic games [20].

5.1 Game model

Consider a network that contains M classes of mobiles. There are N_m mobile nodes in class m. In each class there is a source and a mobile of class i stores and forwards only messages originating from the source of that class. Nodes adopt two-hop routing. All sources generate messages for the same set of intended destinations. (This means that they generate messages to the same node in the address–centric case and messages with the same metadata in the data–centric case.) E.g., several competing applications may use different set of relays to diffuse a specific content such as advertisement messages. In particular, such a model is relevant is the case when each DTN is used to deliver a file to a remote host, and destinations represent a common set of access points acting as shared gateways.

Different classes interact in the dissemination process only when they attempt to transmit to the destination node.[6] However, collisions occur when two or more nodes from different classes attempt to transmit to the destination at the same time, e.g., when multiple DTNs leverage on the same gateway node.

Two cases are considered:

1. When at least two nodes from different classes attempt to transmit to the destination, the destination does not successfully decode any message; this is referred to as the collision model.

2. An arbitration procedure is coherently applied to all nodes, so that when many nodes have the possibility to transmit a message to the destination, one of them (drawn at random) succeeds; this is referred to as the arbitration model.

Concerning the traffic pattern, the model applies to the following two cases:

1. After a message is delivered or time τ has elapsed since its generation, the source can stay idle for a random amount of time after which a new message will be generated.

2. Sources synchronously generate messages with lifetime equal to τ. A message is generated every τ.

The problem falls into a certain category of stochastic games called cost–coupled stochastic games, introduced first in [20]. In such games, each player controls an independent Markov chain and knows only the state of that Markov chain. The interaction between the players is due to their utilities or costs which

[6]This is the case, e.g., when sources are scattered over a large area and collisions of sources transmitting messages to relays are negligible.

depend on the states and actions of all players. In the framework proposed hereafter, each source can infect relays of its own class only, independently from the other sources. In turn, coupling is due to possible collisions when transmitting to the destination. The potential occurrence of collisions affects the delivery probability and therefore the optimal strategy to be used.

Let $X_n^{(i)}$ be the number of relays of class i that are infected at time $n\Delta$. The following discrete time stochastic game is defined:

- *Players.* The M classes of mobiles.
- *Actions.* If at time $n\Delta$ class-i source encounters a mobile, it attempts transmission with probability $u_n^{(i)}$. $\mu^{(i)}$ is the time-dependent policy of class-i source. In this game theoretical framework $\mu^{(i)}$ denotes also the strategy of class-i, while $\mu^{(-i)}$ denotes the set of strategies adopted by the other classes.
- *Performance index.* The utility of each player/class is the probability of successful delivery, $F^{(i)}(K\Delta)$. Each class has also a constraint on the expected number of infected nodes, i.e., $E[X_K^{(i)}] \leq \Psi^{(i)}$.
- *Information.* Source i is assumed to know only $X_n^{(i)}$ and not know $X_n^{(j)}$ for $j \neq i$. The precise knowledge of $X_n^{(i)}$ is possible since source i knows exactly to how many mobiles it transmitted the packet to. Each source is assumed not to know whether the packet was delivered to the destination by time τ or not.

Let $Y_n^{(i)}$ be the number of infected nodes of class i that attempt to transmit to the destination during the n-th time slot ($0 \leq Y_n^{(i)} \leq X_n^{(i)}$). Also, $S_n^{(-i)} = \sum_{j\neq i} Y_n^{(j)}$ is the total number of infected nodes of classes different from class i that attempt a transmission to the destination during the n-th time slot.

5.2 Nash equilibrium

Proposition 5.1. *For both arbitration procedures considered, the complementary cumulative distribution function of the delivery delay at time $n\Delta$, $G^{(i)}(n)$, $n = 1, 2, \ldots$, is decreasing in the control action $u_r^{(i)}$, $r = 0, 1, \ldots, n-2$.*

Proof. Let event $A = \{$No class i node attempts delivery to the destination by $n\Delta\}$ and $B = A^C \cap \{$all class i nodes attempts to delivery to the destination by $n\Delta$ fail$\}$. Clearly

$$G^{(i)}(n) = \Pr\{A \cup B\} = G_1^{(i)}(n) + G_2^{(i)}(n)$$

First, $G_1^{(i)}(n)$ is strictly decreasing in the control action because it holds expression (6), i.e., $Y_r^{(i)} = 0$ for $r = 0, 1, \ldots, n-1$. It remains to be proved that $G_2^{(i)}(n)$ is non increasing.

At slot $r = 0, 1, \ldots, n-1$, all nodes $Y_r^{(i)}$ attempting to deliver to the destination will fail with probability

$$f(Y_n^{(i)}, S^{(-i)}(n)) = \begin{cases} \Pr\left\{S^{(-i)}(n) > 0\right\} & \text{collision model} \\ \dfrac{S^{(-i)}(n)}{Y_n + S^{(-i)}(n)} & \text{arbitration model} \end{cases} \tag{24}$$

It is immediate to observe that r.v. $f(Y_n^{(i)}, S^{(-i)}(n))$ is non increasing in the variable $Y_n^{(i)}$.

Hence, the probability of failure over all the slots is obtained by taking the expectation over the sample paths

$$G_2^{(i)}(n) = \mathrm{E}_{\substack{Y_n^{(i)}, S^{(-i)}(n) \\ n=0,1,\ldots,n-1}} \left[\sum_{h=0}^{n-1} f(Y_n^{(i)}, S^{(-i)}(n)) \right]$$

Using the same notation adopted in Thm. 3.1 policy μ' is such that $u_k' > u_k$ for a certain index $k \le n$: clearly, $Y_n' >_{st} Y_n$ (in fact, Y_n is a Poisson random variable with intensity $\lambda \Delta X_n' >_{st} \lambda \Delta X_n$).

Because $f(Y_n^{(i)}, S^{(-i)}(n))$ is non increasing in the variable $Y_n^{(i)}$, it is immediate that $f({Y'}_n^{(i)}, S^{(-i)}(n)) <_{st} f(Y_n^{(i)}, S^{(-i)}(n))$ for both the collision and the arbitration model, so that ${G'}_2^{(i)}(n) \le G_2^{(i)}(n)$, concluding the proof. $\square$

Theorem 5.1. *If for all M classes the complementary cumulative distribution function at time $(n+k)\Delta$, $G^{(i)}(n+k)$, is decreasing in the control action $u_{n-1}^{(i)}$ for $k = 1, 2, \ldots$, then the optimal threshold policy for the single-class case is also the best response to all the possible $\mu^{(-i)}$.*

Proof. The proof follows the same steps of that of Theorem 3.1: given a non-threshold policy $\mu^{(i)}$, it is possible to build in the same way a new policy $\hat{\mu}^{(i)}$, which can be showed, by means of Prop. 5.1, to achieve better performance than $\mu^{(i)}$. $\square$

From the theorem above the following result follows immediately:

Corollary 5.1. *The considered game has a unique Nash equilibrium. This Nash equilibrium is obtained when each class adopts its optimal single-class threshold policy.*

Proof. The optimal threshold policies are mutual best responses, so they are a Nash Equilibrium. Moreover any different set of strategies cannot be a Nash equilibrium, because at least one class can improve its performance by adopting the optimal single-class threshold policy. ☐

Remark: in general, in order to derive the objective function, node i should know the statistics of $X_n^{(j)}$, $j \neq i$ (but not its actual value). In the present case, however, as the best response does not depend on the statistics of the number of infected nodes in other classes, such knowledge is not required for deriving the best policy to be employed.

6 The Impact of Mobility Patterns

The results presented in the previous sections have been derived under the assumption that intermeeting times among pairs of nodes could be suitably described as sequences of independent and identically distributed exponentially distributed random variables. While it is known that the exponential distribution holds rather well for some synthetic mobility patterns such as random waypoint [7], this represents an idealization for most real-world situations. The objective hereafter is hence to discuss the applicability of the results derived so far in a more general setting.

The exact expression of the delay CDF has been obtained in Sec. 3 under the exponential assumption. But, in principle the optimality of threshold policies does not require such assumption. At the heart of the proof of Theorem 3.1, in fact, lies the monotonicity property of the number of message copies over time, as derived in Proposition 3.1. The framework can therefore account for non-exponential meeting times, provided a monotonicity assumption, as done in [35]. Such approach can be used to extend the results to situations in which the intermeeting time distribution is not exponential (but intermeeting times between any pair of nodes are still sequences of i.i.d. random variables). It is interesting to remark that the same monotonicity assumption is sufficient to apply the results derived in terms of stochastic approximation (Sec. 4) and for the multi-class case (Sec. 5). By relaxing the strict monotonicity, as done in [35], a threshold policy is still optimal under certain assumptions. Intuitively, in such cases a larger number of copies does not necessarily lead to a strictly higher probability of reaching the destination, as copies may be disseminated to nodes which never get in contact with the destination node. Such a case covers situations in which not all the pairs of nodes meet and/or meet with different intensities. However, in the absence of information on the global meeting pattern, our conjecture is that threshold policies can still be proved to be a solution of the optimization problem considered.

Concerning the i.i.d. assumption on intermeeting times, this is indeed rather idealistic with respect to real-world DTN deployments. In the real world, indeed (i) pairs of nodes may mutually meet at a different rate (ii) the sequence of intermeeting times between a given pair of nodes may present correlation (e.g., if node A meets node B at a given time, it is highly likely that they will meet again within a short time-frame). The kind of policies considered in this work are meant to model situations in which it is not possible (or: impractical) to keep track of the ID of the nodes encountered and of the last time two nodes met. It is possible to conjecture that —thanks to the monotonicity argument— the proposed framework provides an optimal policy (though it may not be the only one) also in the case of non i.i.d. intermeeting times, provided that no state information on either the ID of nodes meeting nor the time of last encounter is maintained, and no knowledge on the meeting process is available *a* priori.

Throughout the paper, it has been assumed that contacts are long enough so that all messages backlogged at a node can be transferred to the other one involved in the contact. In real-world settings, whereby duration of contacts is finite, this may not hold if the system gets congested (this depends on a number of application-dependent factors related to mobility, density of nodes, traffic pattern etc.). In general, it is possible to optimize the order by which messages are exchanged during a contact so as to maximize the delivery probability. As an example, one could send first the messages that are closest to timeout expiration, or use similar age-based scheduling and buffer management mechanisms [16]. Even in the case when node pairs meet with different intensity, this could be leveraged to introduce efficient scheduling policies. The presented results can be extended to the case of "blind" scheduling mechanisms (i.e., not based on age and not accounting for meeting intensity): again, in such cases, we conjecture that the optimality of threshold policies for forwarding is preserved.

Many delay-tolerant network deployments do not present a time-stationary meeting pattern. Mobility in many cases tends to follow a periodic pattern. This applies, e.g., to situations in which nodes are mobile phones and other personal computing devices, which are carried around by users in their daily routine. This raises issues related to (i) the optimality of threshold policies (ii) the convergence of the stochastic approximation method. As concerns the first point, intuitively, the respect of a monotonicity assumption, as discussed above for the non-exponential meeting time case, is required. This is expected to hold for most deployments. Concerning point (ii), this clearly depends on the timescale over which the stochastic properties of the meeting pattern change. Some experimental results using real-world traces are presented in [35]; the results therein suggest that convergence of stochastic approximation shall hold for a large share of DTN deployments.

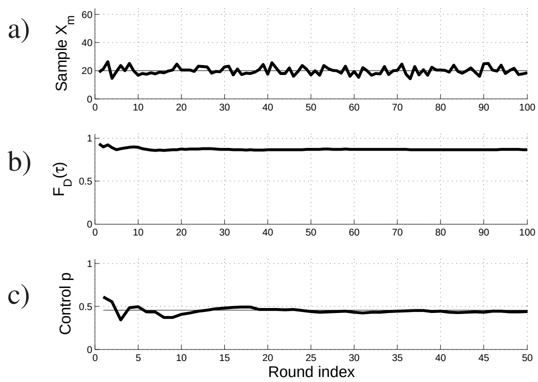


Figure 1: The dynamics of the stochastic approximation algorithm applied to the static forwarding policies.

7 Numerical Results

This section describes the numerical validation of the results derived on optimal control policies and on the convergence of the proposed stochastic approximation algorithm. The results of simulations have been obtained over a set of pre-recorder contact traces by emulating the message forwarding policies described before.

7.1 Stochastic Approximations

When considering stochastic approximation methods, one important performance indicator is the time needed for the algorithm to convergence and the accuracy in the identified optimal operating point.

The first set of tests are aimed therefore at evaluating the dynamics of the stochastic approximation process. A simplified setting was employed, in which any pair of nodes meets according to i.i.d. Poisson processes with intensity $1.0453 \cdot 10^{-5}$ s^{-1}; also, $N = 200$, $u_{\min} = 0$, $u_{\max} = 1$, $\Psi = 20$, $\Delta = 10$ s and $\tau = 20000$ s.

The source node performs at each round a sample measurement of $\overline{X}_m$, based on 30 different estimates of the number of infected nodes at time τ. At the end of each round, the control policy in use is updated according to (12). Unless otherwise specified, results in this section have been obtained considering $a_m = 1/(10 \cdot m)$ as scaling sequence.

Fig. 1 illustrates a specific run for the case when the source estimates the parameter p^* for the best static policy. The figure shows that the estimates $\overline{X}_m$ evaluated by the source are noisy, due to the limited number of samples per estimate.

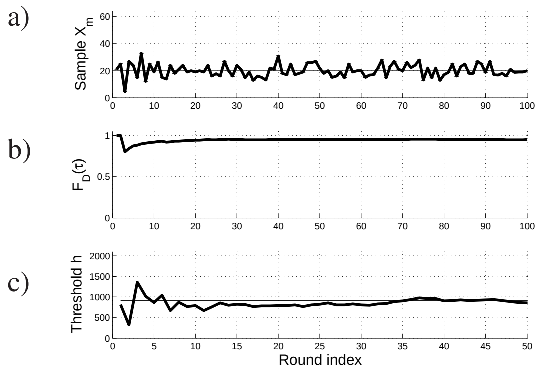


Figure 2: The dynamics of the stochastic approximation algorithm applied to the optimal forwarding policies.

Nevertheless, the convergence of the algorithm is apparent from the dynamics of the control p, i.e. the static forwarding probability, which stabilizes after about 20 rounds around the optimal value $p^* \approx 0.45$ (the horizontal line). For the sake of completeness, the time evolution of the message delivery probability is also reported (Fig. 1b)).

The same experiment was repeated for the dynamic control case. In this case, the source node tries to estimate the optimal threshold $h^* \approx 911$ slots. The dynamics of the estimated parameter is depicted in Fig. 2c). It is possible to observe that the convergence time in this case is similar to that measured in the case of static policies. This is due to the fact that in both cases the stochastic approximation algorithm estimates the same parameter β and even if the distribution of $\overline{X}_m$ (but not its expected value) is different for static and threshold policies, the sequence of estimates θ_m converges to the solution of the same ODE. Indeed, as mentioned in Sec. 4, the ODE trajectory provides more information than the "simple" asymptotic stability of the control variable: indeed the sample trajectories of the control estimates follow a shifted ODE dynamics with probability one. In particular, in Fig. 3 the dynamics of the parameter p has been depicted for the static case against a properly rescaled version of the ODE solution. Sample trajectories have been averaged over 10 runs of the algorithm. Also, in order to make the phenomenon visible, as described in [32], the dynamics are conveniently rescaled according to $t_m := \sum_{i=1}^{m} a_i$. It can be observed that, after an initial transient phase, the trajectory of the control mimics the original ODE; the maximum and minimum values of the trajectories have been superimposed to that graph for the sake of completeness. This pictorial representation confirms that the convergence speed of the algorithm is basically dictated by the dynamics of the related ODE solutions.

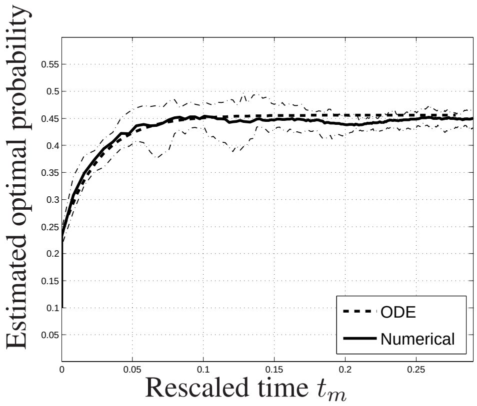


Figure 3: The convergence of the dynamics of the control variable against the reference ODE; at the time scale t_n and averaged over 10 sample trajectories in the case of static control. Thin dash-dotted lines delimit the maximum and minimum values attained by the estimate trajectories.

7.1.1 Polyak's averages

As mentioned in Sec. 4, a slowly decaying scaling sequence a_m obtains a fast convergence to the ODE dynamics and hence to the optimal control policy. The price to pay is a lower rejection to noise, with potentially larger oscillations. Here, the benefit of the Polyak-like averaging technique is showed, as a larger sequence is used, $a_m = 1/(10 \cdot m^{2/3})$, from which faster convergence but weaker noise rejection capabilities are expected.

In Fig. 4 the result of a run of the stochastic approximation procedure has been reported. The plain stochastic estimation of θ_m, based on the chosen a_m coefficients, is superimposed to the output, which has been obtained using the averaging procedure as in (20). It is possible to observe the smoothing performed by the Polyak averaging over the estimated optimal control values, both in the case of static control and in the case of threshold policies. Even though this is a particular case, this result shows, as anticipated in Sec. 4 that interesting tradeoff exist: it is possible to increase the speed of convergence of the algorithm by means of faster sequences, i.e., by approaching faster the tail of the ODE dynamics, while reducing at the same time the estimation noise by averaging.

7.1.2 Real-World Mobility Patterns

as described in Sec. 6, core aspect is the impact of real-world mobility patterns onto the convergence properties of the proposed stochastic approximation scheme. It is possible to perform numerical experiments using real world sample traces obtained

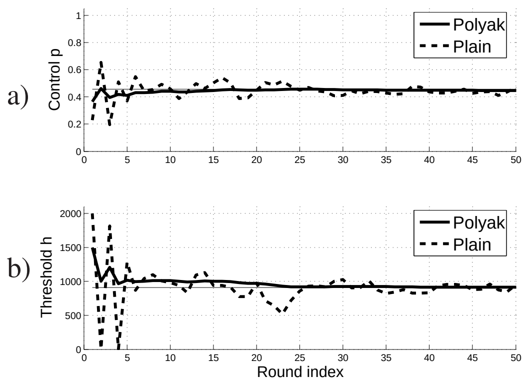


Figure 4: Algorithm employing Polyak's averages applied to a) static and b) threshold forwarding policies.

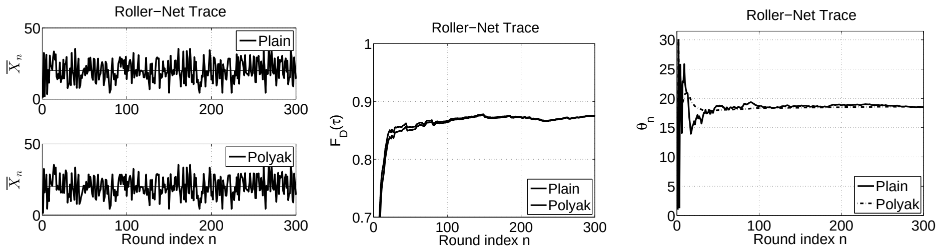


Figure 5: Rollernet trace: performance of the stochastic approximation algorithm a) per-round success probability (running average) b) average number of message copies c) dynamics of the control output of the algorithm.

by recording the encounter patterns of wireless devices. Those described hereafter are available online for experimentation.

Rollernet trace: the trace is described in [36] and it is available in CRAWDAD. This trace has been recorded on August 20, 2006 during a rollerblading tour which lasted three hours, composed of two sessions of 80 minutes; about 2500 people participated to the tour. The trace was collected using iMotes contact loggers provided to 62 volunteers; in practice, Bluetooth interfaces were used to sample the encounters that iMotes had with other devices, i.e., possibly different iMotes or cell phones in radio range. In the described experiments a subtrace involving contacts among iMotes, restricting the DTN network to 62 devices has been employed.

Office trace: the second dataset is the one used in [37]; this dataset has been obtained by monitoring the contact of 21 employees who volunteered to participate in the measurement campaign. The employees occupy different roles within the

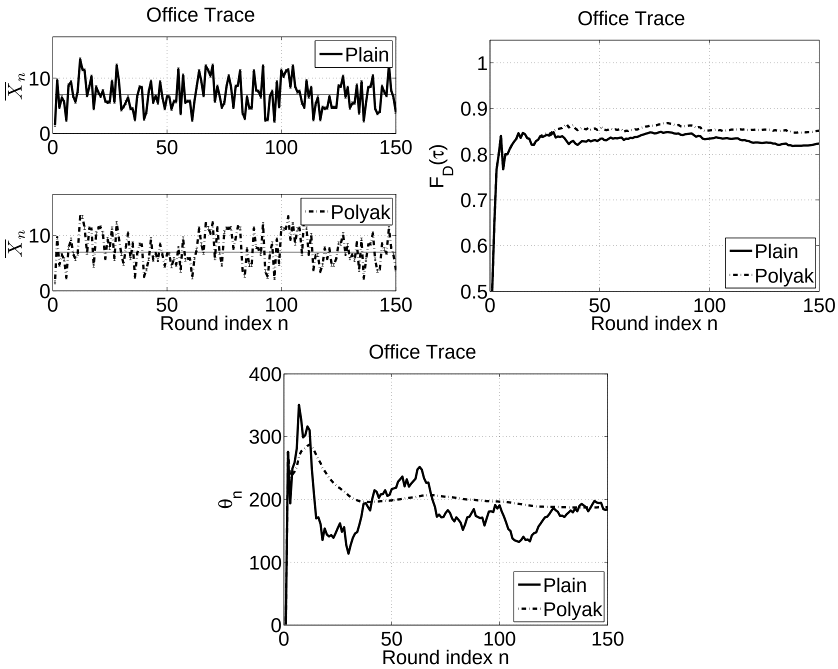


Figure 6: Office trace.

organization and their offices are distributed located on different floors of a research center, scattered over 4 floors in one single building. The experiment lasted 4 weeks; during such period the employees carried a mobile telephone running a Java application. The application was using Bluetooth connectivity in order to periodically trigger (every 60 seconds) Bluetooth node discovery. Upon detecting another device, its Bluetooth address, together with the current time-stamp was saved in the permanent storage of the device for later processing.
Campus trace: the last trace considered comes from the Student-Net project at the University of Toronto [38]. As in the case of the office trace, in the experimentation, students were equipped with Bluetooth–enabled hand–held devices, capable of generating inquiries and tracing any pairwise contact between users. The inquiry period was set so to preserve a 8-10 hours battery life-time. This resulted in a 16 seconds scan period. In particular, the trace used here involved 21 graduate students for a duration of two-and-a-half weeks.

Results: for real-world trace files, no optimality can be claimed with respect to the threshold policies proposed before. In fact, traces are not stationary-ergodic as assumed in Thm. 4.1.

Nevertheless, by design, the algorithm drives the system to match the constraint

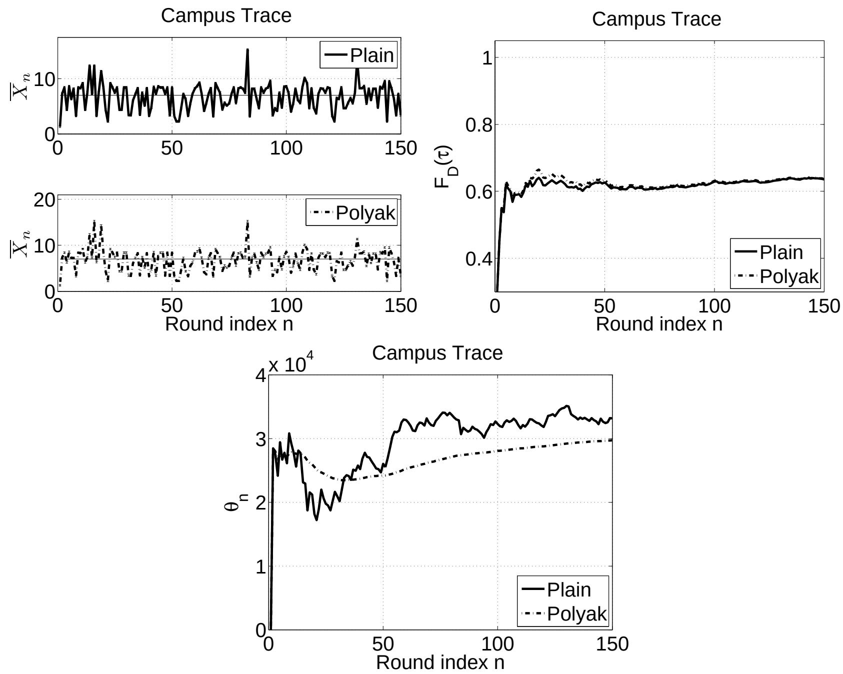


Figure 7: Campus trace.

on the average number of released copies. As showed in Fig. 5a and b, under non stationary-ergodic contact traces, the intermeeting contact process is subject to relevant fluctuations in the intermeeting intensities as described in [36]. In this setting, the chosen horizon is $\tau = 300$ s and $\Psi = 20$ copies, whereas the time slot considered is $\Delta = 10$ s.

Conversely, we tested Polyak averages, but they seem not to offer substantial advantage. This is due to the fact that a very fine grained adaptation step was adopted in order to reduce the bias in the estimation of the control; in turn this permits to comply with the constraint in the number of released copies.

The same analysis is repeated in the case of the *office* trace as depicted in Fig 6. In this case, $\Psi = 7$, $\tau = 5000$. The dynamics of the control variable θ_n is described in Fig. 6c. It is possible to observe again relatively large oscillations due to the non-stationary contact pattern of this trace. However, even in this case, the algorithm tends to stabilize at a working point that is quite far from the maximum possible threshold, but it is still able to track the constraint on the number of message copies. As before, the effect of Polyak averages is to produce a smoothed version of the control as seen in Fig. 6c. Overall, the effect of averaging is to slower the

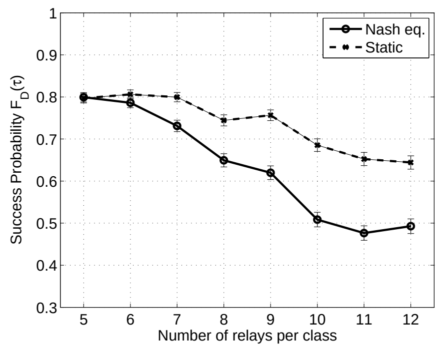


Figure 8: Performance at the Nash equilibrium compared to the best static strategy; $\tau = 200$ s, $\Psi = N - 1$.

convergence of the algorithm and to introduce a bias on the estimate of the control.

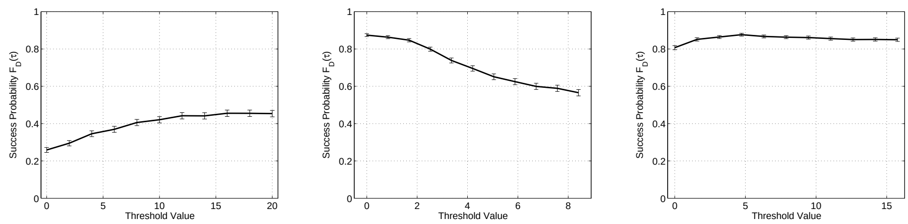
Finally, in the last run in Fig 7, the case of the *campus* trace is reported. The run is obtained for $\Psi = 7$, $\tau = 50000$. Standard stochastic approximations and Polyak's averages are much closer to each other than with the previous traces. This can be ascribed to the structure of the trace: in the case of the *office* trace, contacts are distributed uniformly among users. Hence, the dynamics of the number of copies generated is fast. In the *campus* trace, most contacts are concentrated over a small subset of users' ID pairs, while other nodes meet rarely. The resulting working point for the control threshold is very large: as expected since a large fraction of the considered interval of length τ is required in order to meet 7 nodes out of 20. The very slow dynamics of this trace makes the convergence slowdown introduced by the averaged version of the algorithm much less apparent compared to the previous cases.

The indication that we draw from the simulation experiments is that, even in case of non-stationary traces, the algorithm converges to an effective solution. In the case of fast time-varying traces the need to track the working point at a fast pace discourages the usage of Polyak averages.

7.2 Multiclass Case

In the game theoretical framework presented in Sec. 5, the result on the existence of a Nash equilibrium poses the question of the Pareto optimality of such equilibrium point. The answer is not straightforward since the success probability depends on the number of nodes involved, on the number of classes and on the underlying encounter process.

The multiclass case has been simulated considering the case of two classes, where $N_1 = N_2 = 5, 6, \ldots, 12$ nodes. Each pair of nodes meet with intensity $\lambda = 2.61 \cdot 10^{-2}$. Meetings among nodes follow i.i.d. Poisson processes. The one stage game has been iterated over 800 rounds in order to measure the impact of the different strategies in the collision model. First, the performance attained, namely the successful delivery probability by τ, has been measured in the case of the first interference model introduced in Sec. 5, whereby a collision between two nodes attempting transmission to the destination results in a failure for both nodes. The Nash equilibrium policy and the corresponding best static policy has been evaluated: results are reported in Fig. 8 using 95% confidence intervals. It can be noted that the social outcome can be improved if each class adopts the best static policy. It should be observed that this is not an equilibrium, because a class would find it more convenient to switch to its optimal threshold policy. But, it provides numerical evidence that the Nash equilibrium is not Pareto optimal.



(a) (b) (c)

Figure 9: Success probability as a function of the threshold value for three different cases of the communication range: (a) $R = 1$ m, (b) $R = 15$ m, (c) $R = 8$ m.

In order to better understand the dependency on the threshold value in the multi-class case, some further simulations have been performed. For this case, the number of nodes per class has been fixed at $N_1 = N_2 = 10$ and expected message copies at $\Psi = 9$. Mobility of nodes is identical to previous experiments, but with varying communication range in order to change the intensity of meetings. For each setting, the threshold at the Nash equilibrium is computed, θ^*, and the successful delivery probability is calculated as a function of the threshold value for $\theta \leq \theta^*$. Values larger than θ^* should not be considered as they violate the constraint on the expected number of copies.

As a result three regimes are identified. In the first one (sparse regime), the interference created by the other class is negligible, so that the success probability increases monotonically with the threshold value. As a direct consequence, the Nash equilibrium represents actually the optimal choice in terms of threshold value. This case is illustrated in Fig. 9a, where $R = 1$ m was used. In the second one (dense regime) interference is the driving factor. In this case the optimal choice is actually not to make copies at all (i.e., set the threshold to zero), and just rely on the ability of the source to directly meet the destination. This is shown in Fig. 9b for $R = 15$ m, where the success probability is monotonically decreasing in the threshold value used. In this case, clearly, the Nash equilibrium does not represent an efficient operating point for the system.

In the third regime (medium density), the two effects are present at the same time, so that there exists an optimal non-zero value of the threshold which is smaller than the value at the Nash equilibrium. This case is shown in Fig. 9c for $R = 8$ m.

8 Conclusions

In this paper a discrete time model for the control of mobile ad hoc DTNs was introduced. The closed form expressions for optimal static and threshold forwarding policies for two-hop routing have been derived. Such policies depend on network parameters, such as the number of nodes or nodes meeting rates. But, those parameters may be unknown a priori.

Using the theory of stochastic approximations, an algorithm has been designed that enables all nodes in the DTN to tune independently and optimally the forwarding policies, adapting them to the current operating conditions of the system. The algorithm does not require neither message exchanges, nor an explicit estimation of network parameters.

These features are appealing from the network design standpoint: indeed similar techniques can be applied to a wide set of problems in DTNs, a type of network where the estimation of global parameters is extremely challenging due to the absence of persistent connectivity.

The proposed model has been validated via numerical simulations, including both synthetic mobility traces as well as real-world ones. In particular, the scheme is able to adjust online the forwarding control such in a way to obtain a target number of generated message copies. Overall, this is an effective scheme able to drive the system to a desired operating point and applies to the case when the forwarding control is confined to run on board of source nodes.

Finally, the model has been applied to the case of competing DTNs. To this aim, a class of weakly coupled Markov games has been introduced where players are DTNs competing for resources. The coupling may occur because of interference when DTNs attempt to deliver messages to a common gateway node. In such games, a unique Nash equilibrium exists where each node applies the optimal policy determined for the single DTN case.

References

[1] E. Altman, G. Neglia, F. De Pellegrini, and D. Miorandi, “Decentralized stochastic control of delay tolerant networks,” in *Proc. of IEEE INFOCOM*, Rio de Janeiro, Brasil, 2009, pp. 1134–1142.

[2] S. Burleigh, L. Torgerson, K. Fall, V. Cerf, B. Durst, K. Scott, and H. Weiss, “Delay-tolerant networking: an approach to interplanetary Internet,” *IEEE Comm. Mag.*, vol. 41, no. 6, pp. 128–136, Jun. 2003.

[3] T. Spyropoulos, K. Psounis, and C. Raghavendra, “Efficient routing in intermittently connected mobile networks: the multi-copy case,” *ACM/IEEE Trans. on Networking*, vol. 16, pp. 77–90, Feb. 2008.

[4] A. Vahdat and D. Becker, “Epidemic routing for partially connected ad hoc networks,” Duke University, Tech. Rep. CS-2000-06, 2000.

[5] E. Nordström, P. Gunningberg, and C. Rohner, “Haggle: a data-centric network architecture for mobile devices,” in *Proc. of ACM MobiHoc S^3 Workshop*, New Orleans, Lousiana, USA, 2009, pp. 37–40.

[6] M. Grossglauser and D. Tse, “Mobility increases the capacity of ad hoc wireless networks,” *IEEE/ACM Trans. on Networking*, vol. 10, no. 4, pp. 477–486, Aug. 2002.

[7] R. Groenevelt, P. Nain, and G. Koole, “The message delay in mobile ad hoc networks,” *Performance Evaluation*, vol. 62, no. 1-4, pp. 210–228, October 2005.

[8] M. Musolesi and C. Mascolo, “Controlled epidemic-style dissemination middleware for mobile ad hoc networks,” in *Proc. of ACM Mobiquitous*, San Jose, California, July 2006, pp. 1–9.

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

[10] A. E. Fawal, J.-Y. L. Boudec, and K. Salamatian, “Performance analysis of self limiting epidemic forwarding,” EPFL, Tech. Rep. LCA-REPORT-2006-127, 2006.

[11] W. Zhao, M. Ammar, and E. Zegura, “Controlling the mobility of multiple data transport ferries in a delay-tolerant network,” in *Proc. of IEEE INFOCOM*, Miami USA, March 2005, pp. 1407–1418.

[12] M. M. B. Tariq, M. Ammar, and E. Zegura, “Message ferry route design for sparse ad hoc networks with mobile nodes,” in *Proc. of ACM MobiHoc*, Florence, Italy, May 22–25, 2006, pp. 37–48.

[13] K. Chen and H. Shen, “Dsearching: Distributed searching of mobile nodes in dtns with floating mobility information,” in *Proc. of IEEE INFOCOM*, April 2014, pp. 2283–2291.

[14] A. Guerrieri, I. Carreras, F. D. Pellegrini, D. Miorandi, and A. Montresor, “Distributed estimation of global parameters in delay–tolerant networks,” *Computer Communications*, vol. 33, no. 13, pp. 1472–1482, Aug. 2010.

[15] W. Gao, G. Cao, T. La Porta, and J. Han, “On exploiting transient social contact patterns for data forwarding in delay-tolerant networks,” *Mobile Computing, IEEE Transactions on*, vol. 12, no. 1, pp. 151–165, Jan 2013.

[16] A. Krifa, C. Barakat, and T. Spyropoulos, “Optimal buffer management policies for delay tolerant networks,” in *Proc. of IEEE SECON*, San Francisco, California, June 2008, pp. 260–268.

[17] G. Neglia and X. Zhang, “Optimal delay-power tradeoff in sparse delay tolerant networks: a preliminary study,” in *Proc. of ACM SIGCOMM CHANTS 2006*, Pisa, Italy, 15 September 2006, pp. 237–244.

[18] E. Altman, T. Başar, and F. De Pellegrini, “Optimal monotone forwarding policies in delay tolerant mobile ad-hoc networks,” in *Proc. of ACM InterPerf*, Athens, Greece, October 24 2008.

[19] C. Singh, E. Altman, A. Kumar, and R. Sundaresan, “Optimal forwarding in delay-tolerant networks with multiple destinations,” *IEEE/ACM Trans. Netw.*, vol. 21, no. 6, pp. 1812–1826, Dec. 2013. [Online]. Available: http://dx.doi.org/10.1109/TNET.2012.2233494

[20] E. Altman, K. Avrachenkov, N. Bonneau, M. Debbah, R. El-Azouzi, and D. S. Menasche, “Constrained cost-coupled stochastic games with independent state processes,” *Operations Research Letters*, vol. 36, pp. 160–164, 2008.

[21] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, “Impact of human mobility on opportunistic forwarding algorithms,” *IEEE Trans. on Mobile Computing*, vol. 6, no. 6, pp. 606–620, 2007.

[22] T. Karagiannis, J.-Y. L. Boudec, and M. Vojnović, "Power law and exponential decay of inter contact times between mobile devices," in *Proc. of ACM MobiCom*, September 9-14 2007, pp. 183–194.

[23] W. Chahin, R. El-Azouzi, F. De Pellegrini, and A. P. Azad, "Blind online optimal forwarding in heterogeneous delay tolerant networks," in *Proc. of IFIP Wireless Days*, Niagara Falls, Ontario, Canada, October 10-12 2011.

[24] R. El-Azouzi, F. De Pellegrini, and V. Kamble, "Evolutionary forwarding games in delay tolerant networks," in *Proc. of IEEE WiOpt*, Avignon, France, May 31 - June 4 2010.

[25] W. Chahin, H. Sidi, F. D. R. El Azouzi, and J. Walrand, "Incentive mechanisms based on minority games in heterogeneous delay tolerant networks," in *Proc. of ITC*, Shangai, China, 10-12 September 2013, *Best paper runner up*.

[26] M. Shaked and J. G. Shantikumar, *Stochastic Orders and Their Applications*. New York: Academic Press, 1994.

[27] E. Altman and G. Koole, "On submodular value functions and complex dynamic programming," *Stochastic Models*, vol. 14, no. 5, pp. 1051–1072, 1998.

[28] S. Stidham and R. Weber, "A survey of Markov decision models for control of networks of queues," *Queueing Systems*, vol. 13, no. 1-3, pp. 291–314, 1993.

[29] K. W. Ross, "Randomized and past-dependent policies for Markov decision processes with multiple constraints," *Operations Research*, vol. 37, no. 3, pp. 474–477, 1989.

[30] A. Hordijk and F. Spieksma, "Constrained admission control to a queueing system," *Advances in Applied Probability*, vol. 21, no. 2, pp. 409–431, 1989.

[31] E. Altman, *Constrained Markov Decision Processes*. Chapman and Hall/CRC, 1999.

[32] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, 2nd Edition, 2003.

[33] J. Maryak, "Some guidelines for using iterate averaging in stochastic approximation," in *Proc of IEEE CDC*, vol. 3, San Diego, California, 10-12 Dec. 1997, pp. 2287–2290.

[34] B. T. Polyak and A. Juditsky, "Acceleration of stochastic approximation by averaging," *SIAM Journal of Control and Optimization*, vol. 30, no. 4, pp. 838–855, Jul. 1992.

[35] F. De Pellegrini, D. Miorandi, and I. Carreras, "Optimal two-hop routing in delay–tolerant networks," in *Proc. of European Wireless (invited)*, Lucca, IT, 2010.

[36] P. U. Tournoux, J. Leguay, F. Benbadis, V. Conan, M. D. de Amorim, and J. Whitbeck, "The accordion phenomenon: Analysis, characterization, and impact on DTN routing," in *Proc. of IEEE INFOCOM*, Rio de Janeiro, Brasil, 2009.

[37] I. Carreras, F. De Pellegrini, D. Miorandi, D. Tacconi, and I. Chlamtac, "Why neighbourhood matters: Interests-driven opportunistic data diffusion schemes," in *Proc. of ACM CHANTS*, San Francisco, CA, US, September 2008.

[38] J. Su, A. Goel, and E. de Lara, "An empirical evaluation of the student-net delay tolerant network," in *Mobile and Ubiquitous Systems: Networking Services, 2006 Third Annual International Conference on*, July 2006, pp. 1–10.