

Decentralized Stochastic Control of Delay Tolerant Networks

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

▶ **To cite this version:**

Eitan Altman, Giovanni Neglia, Francesco de Pellegrini, Daniele Miorandi. Decentralized Stochastic Control of Delay Tolerant Networks. [Research Report] RR-6654, INRIA. 2008. inria-00322443

HAL Id: inria-00322443
https://inria.hal.science/inria-00322443

Submitted on 17 Sep 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Contrôle Stochastique Distribué des Réseaux Tolérants aux Délais

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

N° 6654

August 2008

Thème COM



ISSN 0249-6399 ISRN INRIA/RR--6654--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
SOPHIA ANTIPOLIS - MÉDITERRANÉE

Contrôle Stochastique Distribué des Réseaux Tolérants aux Délais

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

Thème COM — Systèmes communicants
Équipes-Projets Maestro

Rapport de recherche n° 6654 — August 2008 — 24 pages

Résumé : Nous étudions dans le présent document des questions liées au contrôle stochastique optimal dans les réseaux tolérants aux délais. Nous dérivons d'abord la structure des politiques optimales de transmission à deux sauts. La mise en œuvre de ces politiques requiert la connaissance à priori de certains paramètres du système tels le nombre de mobiles ou le taux de contact entre ceux-ci. Toutefois, de tels paramètres pourraient ne pas être connus à la conception du système ou même changer avec le temps. Pour remédier à ce problème, nous élaborons des politiques d'adaptation combinant l'estimation et le contrôle et permettant d'obtenir des performances optimales malgré le manque d'information. Nous étudions ensuite l'interaction entre plusieurs classes rivales de mobiles, en la formulant comme un jeu stochastique avec couplage de coût. Nous montrons que ce jeu a un équilibre de Nash unique, où chaque classe adopte la politique de transmission trouvée optimale dans le problème avec une seule classe.

Mots-clés : Stochastic Control, Game Theory, Delay Tolerant Networks

* INRIA, 2004 Route des Lucioles, Sophia-Antipolis, France, email: name.surname@sophia.inria.fr
† CREATE-NET, Via alla Cascata 56/D, Povo, Trento, Italy, email: name.surname@create-net.org

Centre de recherche INRIA Sophia Antipolis – Méditerranée
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Decentralized Stochastic Control of Delay Tolerant Networks

Abstract: We study in this report optimal stochastic control issues in delay tolerant networks. We first derive the structure of optimal 2-hop forwarding policies. In order to be implemented, such policies require the knowledge of some system parameters such as the number of mobiles or the rate of contacts between mobiles, but these could be unknown at system design time or may change over time. To address this problem, we design adaptive policies combining estimation and control that achieve optimal performance in spite of the lack of information. We then study interactions that may occur in the presence of several competing classes of mobiles and formulate this as a cost-coupled stochastic game. We show that this game has a unique Nash equilibrium where each class adopts the optimal forwarding policy determined for the single class problem.

Key-words: Stochastic Control, Game Theory, Delay Tolerant Networks

1 Introduction

Delay–Tolerant Networks (DTNs) are sparse and/or highly mobile wireless ad hoc networks where no continuous connectivity guarantee can be assumed [1, 2]. One central problem in DTNs is related to the routing of packets towards the intended destination. Protocols developed in the mobile ad hoc networks field, indeed, fail since a complete route to destination may not exist most of the time. One common technique for overcoming such problem is to disseminate multiple copies of the message in the network, enhancing the probability that at least one of them will reach, within a suitable time-frame, the destination node [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 meets an infected node.

In this report we consider the zero knowledge scenario [5, 6], where mobile nodes have no *a priori* information on the encounter pattern. Moreover we constrain the analysis to the case when the source of the message can copy it, while the other infected nodes can only forward it to the destination. This is referred to as 2-hop forwarding [7]. We investigate the problem of optimal stochastic control of such routing protocol. The control variable is the probability of transmitting a message upon a suitable transmission opportunity (i.e., contact). The goal is to optimize the probability to deliver a message, while satisfying specific energy constraints. The main contributions of our work summarizes as follows :

- We introduce a discrete–time framework to model message diffusion in DTNs ; within such framework, we characterize analytically the structure of optimal policies for routing control using sample path techniques. In particular, threshold policies are proved optimal.
- We introduce methods for handling the control problem in the case where some parameters of the system are unknown. The described solutions are based on stochastic approximation theory. Convergence to the optimal control policies, under suitable conditions, is analytically derived.
- We extend the problem of optimal control to the case of several competing classes of mobile terminals. The framework, in this case, is that of cost–coupled stochastic games [8, 9]. We prove that the game has a unique Nash equilibrium where each class adopts the optimal forwarding policy determined for the single class problem.

The results obtained are validated numerically through extensive simulation studies.

The control of forwarding schemes has been addressed in DTNs literature before. In [10], the authors propose an epidemic forwarding protocol based on the *susceptible-infected-removed* (SIR) model [11] and show that it is possible to increase the message delivery probability by tuning the parameters of the underlying SIR model. In [12] a detailed general framework is proposed in order to capture the relative performances of different self-limiting strategies. None of these two papers formalize a specific optimization problem. In [6] and its follow-up [5], 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. Works more similar to

ours are [13, 14, 15]. In [13] 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 [14] can be considered a relaxed version of our problem (e.g., the weighted sum of delivery time and energy consumption), also in this case the optimal policy is a threshold one. Also, under a fluid model approximation, the work in [15] 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 online estimation of optimal policies ; an attempt is done in [12, 13] based on some heuristics for the estimation.

Finally, to the best of our knowledge, this is the first formulation of a game with competing nodes in a DTN scenario.

The remainder of the report is organized as follows. The system model is introduced in Sec. 2. The structure of optimal control policies is derived in Sec. 3. Methods for optimization in the presence of unknown system's parameters are presented in Sec. 4. The multiclass case is introduced in Sec. 5. Numerical results are presented in Sec. 6. Sec. 7 concludes the report pointing out possible research directions.

2 System Model

Consider a network of $N + 1$ mobile nodes, each equipped with some form of proximity wireless communications. The network is assumed to be extremely 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. We refer to such events 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 model for synthetic mobility models (including, e.g., Random Walk, Random Direction, Random Waypoint) has been discussed in [16]. There exist studies based on traces collected from real-life mobility [17] that argue that inter-contact times may follow a power-law distribution. Recently, the authors of [18] have shown that these traces and many others exhibit exponential tails after a cutoff point. For this reason, we choose to stick with the exponential meeting time assumption, which makes our analysis tractable.

There can be multiple source-destination pairs, but we assume that at a given time there is a single message, eventually with many copies, spreading in the network[1]. For simplicity we consider a message originated at time $t = 0$. We assume that the message that is transmitted is relevant during some time τ. This applies, e.g., to environmental information or data referring to events of transient nature (e.g., happenings). The message contains a time stamp reporting its generation time, so that it can be deleted at all when it becomes irrelevant. We do not assume any feedback that allows the source or other mobiles to know whether the message has been successfully delivered to the destination within the time τ.

[1] Results in sections 3 and 4 are valid even for multiple messages at the same time, but we assume that the bandwidth and the buffer are large enough to assure that the different propagation processes are independent.

We focus on a set of relaying strategies that can be defined as probabilistic 2-hops 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 $U = [u_{\min}, u_{\max}]$. 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.

We adopt a discrete time model, considering a time slot duration Δ. The n–th slot corresponds to interval $[n\Delta, (n+1)\Delta)$ and the number of slots is equal to $K = \lfloor \tau/\Delta \rfloor$. In this discrete time model, we assume that a mobile that receives a copy during a time slot can forward it starting from the following time slot. 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 destination, that have a copy of the message at time $n\Delta$ (i.e. at the beginning of the n–th slot), $X_0 = 1$. Under the assumptions above, X_n is a Markov chains with possible states $1, 2, \cdots N-1$. The transition rates depend on the forwarding probability used by the source in each time slot, so a natural way to optimize performance system is to control such forwarding probabilities.

The problem we address in this paper is to *maximize the probability to deliver the packet* by the K-th time slot, *under a constraint on the expected number of infected nodes*. The number of infected nodes is related to the total energy consumption. In particular they are simply proportional if we assume that most of the energy is consumed for transmission and a constant per-contact energy expenditure in order to forward a message. We want to determine optimal time-dependant forwarding policies the source can implement. More formally we define a forwarding policy (control policy) as a function $\mu : \{0, 1, 2, \cdots K-1\} \to U$.

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. $\mu(n) = p \in U$, for $n = 0, 1, 2, K-1$. A policy μ is a* threshold *policy, if there exist $h \in \{0, 1, 2, \cdots K-1\}$ (the threshold) such that*

$$\mu(n) = \begin{cases} u_{\max}, & \text{if} \quad n < h \\ u_{\min}, & \text{if} \quad n > h \end{cases} \tag{1}$$

Observe that a threshold policy is identified by two parameters, the threshold value h and corresponding control value $\mu(h)$.

Static and threshold policies are different from the implementation standpoint. In fact, with static policies, at each communication opportunity, message forwarding is done with a 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 $u(t)$, i.e. they require a dynamic approach[2].

It is worth noticing that static and threshold policies are defined based on few parameters only, i.e., the control p for static policies, and the threshold h and the corresponding value $\mu(h)$ for dynamic policies, which leads to a simple implementation.

[2]Incidentally, time t can be traced just summing up the time elapsed at each node with no need for nodes' synchronization.

Symbol	*Meaning*
$N+1$	number of nodes
λ	intermeeting intensity
τ	timeout value
K	$\lfloor \tau/\Delta \rfloor$
Δ	time slot
X_n	number of nodes having a copy of the message at time $n\Delta$
Ψ	maximum expected number of infected nodes
$F_D(n)$	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
$\zeta_{n,m}(i)$	indicator that the i–th mobile, among the $N - X_n$ ones that do not have the message at time $n\Delta$, receives it during the next m slots
$Q_{n,m}$	probability that a mobile does not receive the message during time slots $n, n+1, ..., n+m-1$
$\gamma_n(s)$	$= \mathrm{E}[\exp(-s\xi_{0,n}^{(1)}]$
$X_n^*(s)$	Laplace-Stieltjes transform of X_n
$\overline{X}_m$	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
$\{\cdot\}^{(i)}$	superscrit 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

TAB. 1 – Notation used throughout the report

In the following section we characterize optimal static and threshold policies. Then in Sec. 4 we show how the source can *learn* online the optimal policy. In Table 1 the notation used throughout the report is described.

3 Characterization of Optimal Policies

We define $F_D(n)$ the probability that a message generated at time 0 is received before $n\Delta$, i.e. $F_D(\cdot)$ is the CDF of the message delay (considering the messages not delivered by τ as delivered at ∞).

We want to derive policies that maximize $F_D(K)$, while satisfying the following constraint on the expected number of infected nodes : $\mathrm{E}\{X_K\} \leq \Psi$.

Let us first characterize the evolution of X_n. Let $\zeta_{n,m}(j)$ be the indicator that the j–th mobile among the $N - X_n$ mobiles that do not have the message at time $n\Delta$, receives the message during $(n\Delta, (n+m)\Delta]$. Then we have

$$X_{n+m} = X_n + \sum_{j=1}^{N-X_n} \zeta_{n,m}(j). \tag{2}$$

Variables $\zeta_{n,m}(j)$ are i.i.d. Bernoulli random variables with expected value :

$$\mathrm{E}\left[\zeta_{n,m}(j)\right] = 1 - \exp(-\lambda\Delta \sum_{k=n}^{m-1} u_k) = 1 - Q_{n,m}, \tag{3}$$

where $Q_{n,m}$ is then the probability that a mobile does not receive the message in time slots $n, n+1, ..., n+m-1$. We observe that $\zeta_{n,m}(j)$ are stochastically increasing in the control actions u_k (see Appendix A for definition and results). More formally, given a policy μ, consider the policy μ' such that $\mu'(n) = \mu(n)$ for $n \neq k$ and $\mu'(k) > \mu(k)$, and denote as $\zeta'_{n,m}(j)$, X'_n and $F'_D(\dot{})$ respectively the indicator variables. the number of infected nodes and the delivery probability function of μ', then

$$\zeta'_{n,m}(j) >_{st} \zeta_{n,m}(j) \; \forall n < k \text{ and } m > k.$$

Moreover being that the number of infected nodes X_n (X'_n) can be obtained as sum of the indicator variables $\zeta_{0,n}(j)$ ($\zeta'_{0,n}(j)$) (Eq. (2)), it holds

$$X'_n >_{st} X_n, \; \forall n > k.$$

This formalizes the intuition that the higher the forwarding probability the higher the number of infected nodes (the same conclusion can be reached through a simple sample path reasoning).

From the previous equations we can easily derive the expected value of X_n, that will be used in the next section :

$$\mathrm{E}[X_n] = X_0 + (N - X_0)\,(1 - Q_{0,n}) \tag{4}$$

Using the Laplace Stieltjes Transform of X_n, $X^*_n(s) := \mathrm{E}[\exp(-sX_n)]$, we can derive the following useful formula for $F_D(n)$:

$$F_D(n) = 1 - \prod_{i=0}^{n-1} X^*_i(\lambda\Delta). \tag{5}$$

In order to prove (5), let us define $G(n) = 1 - F_D(n\Delta)$, then it follows

$$\begin{aligned} G(n+1) &= G(n)\ \Pr\{\text{no delivery in the n-th slot}|X_n\} \\ &= G(n)\ \mathrm{E}\Big[\Pr\big\{\text{no delivery in the n-th slot}|X_n\big\}\Big] \\ &= G(n)\ \mathrm{E}\Big[\exp(-\lambda\Delta X_n)\Big] = G(n) X^*_n(\lambda\Delta) \\ &= \prod_{i=0}^{n} X^*_i(\lambda\Delta) \end{aligned} \tag{6}$$

From Eq. (5), and above considerations on stochastic orderings, it follows that the delivery probability and the final number of infected nodes are increasing in the control actions u_k. Formally,

Proposition 3.1 *Given two policies μ and μ', defined as above, it holds : $F_D(K) < F'_D(K)$, $E[X_K] < E[X'_K]$.*

Proof. Being that $\exp(-sx)$ is decreasing in x for each $s > 0$ and $X_n =_{st} X'_n$ for $n < k$ and $X_n <_{st} X'_n$ for $n \geq k$, then, for each $s > 0$, $X^*_n(s) = X'^*_n(s)$ for $n < k$ and $X^*_n(s) < X'^*_n(s)$ for $n \geq k$. It follows then that $F_D(K) < F'_D(K)$. The inequality for the expected values follows immediately from $X'_K >_{st} X_K$. □

As a consequence of this proposition, the following holds :

Corollary 3.1 *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. Let us consider a policy μ, that is different from μ_{max} (i.e. $\exists k$ s.t. $\mu(k) < u_{max}$) and does not saturate the constraint ($\mathrm{E}[X_K] < \Psi$). Being that the expected number of infected nodes is a continuous function of the forwarding probabilities (this is evident from Eq. (4) and Eq. (3)), we can obtain from μ a new policy μ', by increasing the forwarding probability in k, while satisfying the constraint $\mathrm{E}[X'_K] \leq \Psi$. The new policy has better performance, being that $F'_D(K) > F_D(K)$. □

We observe that the set of admissible policies could be empty. It can be verified that this happens if and only if the policy $\mu_{min}(n) = u_{min}$ for all n, does not satisfy the constraint.

In what follows we consider that admissible policies exist and we are going to characterize policy optimality. To this purpose it is useful to derive an explicit formula for the Laplace Stieltjes transform. Let us introduce

$$\begin{aligned} \gamma_n(s) &:= \mathrm{E}[\exp(-s\zeta_{0,n}(1))] = (1 - Q_{0,n}) \exp(-s) + Q_{0,n} \\ &= e^{-s} - (1 - e^{-s}) \exp\left(-\lambda\Delta \sum_{k=0}^{n} u_k\right) \end{aligned} \tag{7}$$

Then $X^*_n(s)$ can be expressed as a function of $\gamma_n(s)$ as follows :

$$\begin{aligned} X^*_n(s) &= \mathrm{E}[e^{-sX_n}] = \mathrm{E}\left[\exp\left(-s\left(X_0 + \sum_{i=1}^{N-X_0} \zeta_{0,n}(i)\right)\right)\right] \\ &= e^{-sX_0} \left(\mathrm{E}\left[\exp(-\zeta_{0,n}(1))\right]\right)^{N-X_0} = \\ &= e^{-sX_0} \gamma_n(s)^{N-X_0} \end{aligned} \tag{8}$$

We can now introduce the main result of this section.

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 (see for example [20]). We need simply to prove that a non threshold policy cannot be optimal.

Let us consider a non threshold policy μ that satisfies the constraint ($X_K \leq \Psi$), then there exists some time $k < K$ and some $\epsilon > 0$ such that $u_k < u_{\max} - \epsilon$ and $u_{k+1} > u_{\min} + \epsilon$.

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

We notice that $\gamma'_n(s) = \gamma_n(s)$ for $n \neq k$ and $\gamma'_k(s) = \gamma_k(s)\exp(-\lambda\Delta\epsilon) < \gamma_k(s)$. Then from Eq. (8), it follows that ${X'}^*_n(s) = X^*_n(s)$ for $n \neq k$, while ${X'}^*_k(s) < X^*_k(s)$, which in turn brings $F'_D(n\Delta) > F_D(n\Delta)$ for $n \geq k$. Moreover ${X'}^*_K(s) = X^*_K(s)$ implies that $\mathrm{E}[X'_K] = \mathrm{E}[X_K] \leq \Psi$, then the new policy satisfies the constraint and improves the delivery probability. Hence a non threshold policy μ cannot be optimal. □

Let us now determine the optimal threshold policy. Due to Corollary 3.1, the optimal policy is μ_{max} if it satisfies the constraint. Otherwise, the constraint has to be saturated and we can obtain the threshold value from Eq. (4), imposing $\mathrm{E}[X_K] = \Psi$:

$$Q_{0,K} = \frac{N - \Psi}{N - X_0}.$$

Hence

$$\sum_{k=0}^{K-1} u_k = -\frac{1}{\lambda\Delta}\log\left(\frac{N-\Psi}{N-X_0}\right) =: \beta \tag{9}$$

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. 9. Then $h^* = \max\{h \in \mathbb{N} : v(h) = h \cdot u_{\max} + (n-h)\cdot u_{\min} \leq \beta\}$, and $u_{h^*} = \beta - v(h^*)$. In the particular case of $u_m in = 0$, this reduces to $h^* = \lceil\beta\rceil$ and $v(h^*) = \beta - \lfloor\beta\rfloor$.

In fact, denote $y = \lfloor\beta\rfloor$. 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 $k > \beta + 1$. At the remaining time, $k = y + 1$, it uses $u_k = \beta - y$.

In the general case when $u_{\min} \geq 0$, the optimal threshold computes $h^* = \max\{h \in \mathbb{N} : v(h) = h \cdot u_{\max} + (K-h)\cdot u_{\min} \leq \beta\}$: again, the optimal policy chooses $u_k = u_{\max}$ for all $k < h^*$ and $u_k = u_{\min}$ for all $k > h^*$; also, $u_{h^*} = \beta - v(h^*)$.

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. (9) holds, and imposing $u_n = p^*$ for all n, we obtain $p^* = \beta/K$.

4 Stochastic Approximations for Adaptive Optimization

In this section we introduce methods for achieving the optimal control policies in the case where some parameters (i.e., N and λ) are unknown. We show that simple iterative algorithms may be implemented

at each node, allowing them to discover the optimal policy in spite of the lack of information on such parameters[3].

Our approach is based on stochastic approximation theory [21]. 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 which we use :

[3]Note that the estimation of N and λ is per se non-trivial in the lack of persistent connectivity.

– Static control : find the constant $p^* \in [u_{\min}, u_{\max}]$ such that the policy $\mu = p^*$ has the best performance among all static policies.
– Dynamic control : find the threshold $h^* \in \{0, 1, \cdots, K-1\}$ and $\mu(h^*)$ characterizing the optimal policy.

We can approach online estimation of optimal static and dynamic control in the same way. Let us denote $\theta = \sum_{k=0}^{K-1} u_k$, the sum of the controls used over the K time slots. θ is univocally determined from the policy μ, but it also identifies univocally a static or a threshold policy. For the static policy is $\mu(n) = p = \theta/K$, while for the threshold policy it is $h = \max\{h \in \mathbb{N} : v(h) = h \cdot u_{\max} + (n-h) \cdot u_{\min} \leq \theta\}$, and $\mu(h) = \theta - v(h)$. Note that if $\theta = \beta$, then the two policies are the optimal static and threshold policies determined in the previous section. Then in both cases our policy estimation problem comes down to estimate β. Again mobiles do not know quantities such as λ, N, etc., so that they can not compute β a priori using Eq. (9). 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 in rounds. Each round corresponds to the delivery of a set of messages. During a given round, a policy is used. Let us denote by μ_m the policy adopted at round m and $\theta_m \sum_{k=0}^{K-1} \mu_m(k)$ 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 :

$$\theta_{m+1} = \Pi_H\Big(\theta_m + a_m(\Psi - \overline{X}_m)\Big), \tag{10}$$

where

$$\Pi_H(\theta) = \begin{cases} \theta_{\max} & \text{if } \theta \geq \theta_{\max} \\ \theta & \text{if } \theta_{\min} \leq u \leq \theta_{\max} \\ \theta_{\min} & \text{if } \theta \leq \theta_{\min} \end{cases}$$

As discussed above, the new policy μ_{m+1} is univocally determined from θ_{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 $a_m \geq 0\ \forall m$, $\sum_{m=0}^{+\infty} a_m = +\infty$ and $\sum_{m=0}^{+\infty} a_m^2 < +\infty$, the sequence of policies μ_m converges to the optimal policy with probability one.*

Proof. On the basis of the considerations at the begin of this section we only need to prove that θ_m converges with probability one to β. The proof is divided in two parts. First we show that the sequence θ_m converges to some limit set of the following Ordinary Differential Equation (ODE)

$$\dot{\theta} = \Psi - E[X_K|\theta]. \tag{11}$$

For this reason the Eq. (10) is said to be the stochastic approximation of Eq. (11). The convergence is a consequence of Theorem 2.1 in [21] (page 127).

In the second part we show that the solution of such ODE converges to β as time diverges.

We observe that from Eq. (4) and Eq. (3)

$$E[\overline{X}_m|\theta_m] = E[X_K|\theta_m] = N - (N - X_0)e^{-\lambda\Delta\theta_m} \tag{12}$$

so that Eq. (11) writes

$$\dot{\theta} = \Psi - N + (N - X_0)\, e^{-\lambda\Delta\theta_m}. \tag{13}$$

Notice that, the application of Theorem 2.1, as reported in [21], requires to show that some conditions holds. The first two of them follow easily from the problem settings :

i) $\sup_m E[(Z_m)^2] < +\infty$, where $Z_i = \Psi - \overline{X}_i$: this is automatically 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 $\{a_n\}$;

The remaining two of them are general measure-related conditions that write as follows :

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.

In our case, we can write directly $\mathrm{E}_m Z_m = \mathrm{E}[Z_m|\theta_0, Z_i, i < m] = \mathrm{E}[Z_m|\theta_m] = (\Psi - N) + (N - X_0)\, e^{-\lambda\Delta\theta_m}$. Hence, it follows that $\mathrm{E}_m Z_m = \overline{g}(\theta_m)$ where function $\overline{g}(\theta) = (\Psi - N) + (N - X_0)\, e^{-\lambda\Delta\theta}$ is clearly continuous and thus measurable. Notice that it also follows $\eta_m \equiv 0$ for $\forall m$.

We now need to show that the ODE (13) converges as time diverges to the asymptotically global fixed point given by β.

First, it is easy to check that $\theta^* = \beta$ is an equilibrium point of (13).

Second, as $E[X_K|\theta]$ is strictly monotonic in θ, the equilibrium point is unique. In order to demonstrate the stability of the estimator, we use the Lyapunov function $V(\theta) = (\theta - \theta^*)^2$.
Then, we have :

$$\dot{V}(\theta) = 2(\theta - \theta^*) \cdot \dot{\theta} = 2\left[\theta + \frac{1}{\lambda\Delta}\log\left(\frac{N - \Psi}{N - X_0}\right)\right] \cdot \\ \cdot \left[\Psi - N + (N - X_0)\, e^{-\lambda\Delta\theta}\right] < 0 \text{ for } \theta \neq \theta^* \tag{14}$$

Asymptotic global stability follows in both cases from Lyapunov's theorem. □

Remark 4.1 *It is worth to dig a bit more in depth the result exposed in Theorem 2.1 [21]. Basically, it states that the sample paths of the estimates θ_m converge w.p.1 to the solution $\theta(t)$ of Eq.* (11)*. Also, such convergence is apparent at the "natural" time scale of the algorithm, i.e., over the time sequence $\{t_n\}_{n\geq 0}$ defined as follows :*

$$t_n = \begin{cases} 0, & n = 0 \\ t_{n-1} + a_n, & n > 0 \end{cases} \tag{15}$$

The convergence, in particular, is guaranteed for the piecewise constant interpolation $\theta(t; n)$, defined for $t \geq 0$ as :

$$\theta(t; n) = \theta_k \quad \textit{for } t_k - t_n \leq t < t_{k+1} - t_n \tag{16}$$

More precisely, Theorem 2.1 in [21][page 127] ensures the convergence, for almost all sample paths, of (16) to the solution of the ODE (11).

Remark 4.2 *After some cumbersome derivation, the closed form solution of Eq.* (13) *is :*

$$\theta(t) = \frac{1}{\lambda\Delta}\log\Big(\frac{N-\Psi}{N-X_0}\frac{1}{1+\big(\frac{N-\Psi}{N-X_0}e^{\lambda\Delta\theta(0)}-1\big)e^{-\lambda\rho t}}\Big) \tag{17}$$

where $\rho = \Delta(N-\Psi)$.

In Section 6 we will provide numerical evidence of the convergence of the "tail" of the iterates to the ODE dynamics.

In the description of the algorithm above we have suggested that the online estimation of the optimal control is obtained by using in Eq. (10) the estimation $\overline{X}_m$ obtained from real message transmission. However, in the case of two-hop routing, we may circumvent this constraint by using a sort of "virtual messages" : indeed, the stochastic approximation technique works also if the source simply keeps track of the number of mobiles it *would* infect during the 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 keeping track of the nodes it would have infected if it had a message. If a real message has to be transmitted, the current policy estimation can be used.

4.1 Choice of the Sequence $\{a_n\}$

The performance of the stochastic approximation algorithm (10) is known to depend heavily on the choice of the sequence $\{a_n\}$ [22]. By comparing Eq. (10) and Eq. (15), we can observe that a trade-off arises, peculiar for stochastic approximation algorithms in the form (10). In a informal manner, we can describe such effect as follows. First, we notice that the series $\{a_n\}$ diverges : the slower $\{a_n\}$ vanishes, the faster t_n in Eq. (15) diverges and the faster the trajectory of the estimates get closer to the tail of solution of the ODE. But, there is a competing effect : the slower $\{a_n\}$ vanishes, the larger is the asymptotic variance of the estimation [21] ; this is due to the fact that larger $\{a_n\}$ sequences have weaker filtering capabilities in the iterates equation (10).

This trade-off has been studied in literature. For example, a standard choice is $a_n = \frac{C}{n}$; the optimal value of C that guarantees the smallest asymptotic variance is [21] $C = \frac{\partial E[X(\tau)|\theta]}{\partial\theta}\Big|_{\theta=\theta^*}$. 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 of (10) is to use techniques such as Polyak's averages [21, 23]. The idea is to use larger "jumps" to let the iterates converge faster, while using averages to smooth actual estimates.

In Polyak's method, we may use a sequence $a_n = O(n^{-1})$, and in particular one that satisfies the condition $a_n/a_{n+1} = 1 + o(a_n)$ and use as estimation of the optimal policy (i.e., as control to be used on real messages)

$$\Theta_n = \frac{1}{n}\sum_{k=1}^{n}\theta_k. \tag{18}$$

In Section 6 we will show that using Polyak's averaging techniques may lead to advantages in terms of convergence time to the optimal control.

4.2 Constant Step Approximations

In a real DTN implementation, we may be interested in tracking changing conditions. This can be done through stochastic approximation techniques by considering constant step approximations, i.e., iterates of the form :

$$\theta^a_{m+1} = \Pi_H\Big(\theta^a_m + a(\Psi - \overline{X}^a_m)\Big). \tag{19}$$

In this way, the system does not "get stuck" at a given θ but keeps on modifying its behaviour, in an open–ended fashion. Following the approach in [21][Chap.8] it is possible to prove that the iterations in (19) converge in distribution as $n \to +\infty$.

In particular for small enough step size a, the limit process is, with arbitrary high probability, concentrated in an arbitrary small neighbourhood of the optimal control θ^*. This is important in ensuring that the approximation obtained is close to the optimal control policy. Formally, we get the following result :

Theorem 4.2 *For any $\delta > 0$, define by $N_\delta(\theta^*)$ the set $\{x \in \mathbb{R} : |x - \theta^*| < \delta\}$. As $a \to 0$, almost all sample paths of the algorithm (19) converge with arbitrary high probability to elements in $N_\delta(\theta^*)$.*

Proof. The first step is to verify that the conditions required in [21][Thm.2.1, p. 248] hold ; this is true since they are a subset of the conditions already verified in the proof of Theorem 4.1, plus a mild measure-theoretical requirement that holds here since $\mathrm{E}_m Z_m = (\Psi - N) + (N - X_0)\, e^{-\lambda\Delta\theta^a_m} = \overline{g}(\theta^a_m)$ and because $\overline{g}(\cdot)$ does not depend on a and m (compare [21][Condition 1.5, p. 245]).

The referred theorem ensures that with high probability, as $a \to 0$, almost all sample paths of (19) will eventually converge to an arbitrary small neighbourhood of some limit set of the ODE (11). Then we can use the second part of the proof of Theorem 4.1 for the definition of the limit set of the ODE (11). □

5 The Multiclass Case

In this Section we model the decentralized stochastic control problem in the presence of several competing DTNs as a weakly coupled stochastic game, introduced in [8, 9].

5.1 The 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 destination. Here we assume that message transmission time is equal to a time slot duration and meetings occur at the begin of a time slot. The transmission technique uses receiver based codes, and an arbitration procedure can avoid collisions among the members of the same class, so that collisions occur if and only if two or more nodes from different classes are trying to deliver their messages to the destination at the same time. We also study the case when the arbitration

procedure is coherently applied from all nodes, so that when many nodes have the possibility to transmit a message to the destination, one of them is successful.

We consider two different traffic generation models. In both case each source has a single relevant message at a given time instant. In the first traffic generation model sources synchronously generate messages with lifetime equal to τ. In the second one, 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. Hence sources operate asynchronously.

As in the previous section, it may not be desirable for a source to transmit a copy of its message at each opportunity it has since this consumes expensive network resources such as energy, hence the source can decide to forward the message with a given probability. Due to interactions among different mobile classes, a problem of non-cooperative control of those probabilities arises.

Our problem falls into a category of stochastic games that was recently introduced in [8, 9], in which each player control 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 depend on the states and actions of all players. Indeed in our framework each source can infect mobile of its own class independently from the other sources and the only coupling derives from collisions when transmitting to the destination. The possibility of having collisions affects the delivery probability.

A different problem is a classless model where a relay node can be infected by all the available source nodes. In this case the state needs in general to specify which messages are carried by each node. Nevertheless if we consider the synchronous traffic generation model and performance metrics only depending on the delivery of the first message among the competing ones, the problem can be addressed in the same framework ([24]).

5.2 A Weakly Coupled Markov Game Formulation

Let $X_n^{(i)}$ be the number of mobiles of class i that are infected at time $n\Delta$. We consider the following discrete time stochastic game.

- **The players** the M classes of mobiles, act independently.
- **The 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-dependant policy of class-i source. In this game theoretical framework we refer to $\mu^{(i)}$ also as the strategy of class-i, while $\mu^{(-i)}$ denotes the set of strategies adopted by the other classes.
- **The 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 $j \neq i$. But it knows its statistics. The precise knowledge of $X_n(i)$ is possible since the source i knows exactly to how many mobiles it transmitted the packet for relay. Note that it is not assumed to know if the packet was delivered to the destination.

Let us define $Y_n^{(i)}$ as the number of infected nodes of class i that can transmit to the destination during the n- th time slot ($0 \leq Y_n^{(i)} \leq X_n^{(i)}$), $S_n^{(-i)} = \sum_{j \neq i} Y_n^{(j)}$ and $S_n = \sum_j Y_n^{(j)} = Y_n^{(i)} + S_n^{(-i)}$.

A recurrence law analogous to Eq. (6) can be derived for the CDF of the delivery time of messages of each class. For example for class i :

$$\begin{aligned}
&1-F_D((n+1)\Delta) = G(n+1) = \\
G^{(i)}(n+1) &= G^{(i)}(n) \\
&\Pr\{\text{class } i \text{ message not delivered in n-th slot}\} \\
&= G^{(i)}(n)\Pr\left\{\{S_n=0\}\cup\{S_n^{(-i)}>0\}\right\} \\
&= G^{(i)}(n)\left(\Pr\{S_n=0\}+\right. \\
&\left.+\left(1-\Pr\{S_n^{(-i)}=0\}\right)\right) \\
&= G^{(i)}(n)\left(\prod_j X_n^{(j)*}(\lambda\Delta)+\right. \\
&\left.+\left(1-\prod_{j\neq i} X_n^{(j)*}(\lambda\Delta)\right)\right).
\end{aligned}$$

For the case of a cross-class arbitration procedure, then one needs to take into account the possibility that a node of class i succeeds even in presence of other nodes. In a fair arbitration scheme this will happen with probability $Y_n^{(i)}/(Y_n^{(i)}+Y_n^{(-i)})$. We can then derive the following expression for $G^{(i)}(n)$:

$$\begin{aligned}
G^{(i)}(n+1) &= G^{(i)}(n) \\
&\quad \Pr\{\text{class } i \text{ message not delivered in n-th slot}\} \\
&= G^{(i)}(n)\Pr\{\{S_n=0\}\cup\ \{\{S_n>0\} \\
&\quad \cap\{\text{class } i \text{ nodes lose arbitration}\}\}\} \\
&= G^{(i)}(n)\left(\Pr\{S_n=0\}+\right. \\
&\quad \left.(1-\Pr\{S_n=0\})\ \mathrm{E}\left[\frac{S_n^{(-i)}}{S_n}|\ S_n>0\right]\right).
\end{aligned}$$

We observe that $G^{(i)}(n+1)$ depends on the vectors of control actions $(u_k^{(1)}, u_k^{(2)}, \ldots u_k^{(M)})$, for $k \leq n-1$. Before stating our main results we introduce the following observation.

Proposition 5.1 *For both the arbitration procedures, $G_{n+1}^{(i)}$ is decreasing in the control action $u_{n-1}^{(i)}$.*

Proof. From Eq. (2)

$$X_n^{(i)} = X_{n-1}^{(i)} + \sum_{j=1}^{N^{(i)}-X_{n-1}^{(i)}} \zeta_{n-1,1}^{(i)}(j),$$

where $\zeta_{n-1,1}^{(i)}(j)$, $i=1,2,\ldots N^{(i)}$ are i.i.d. Bernoulli random variables with expected value $1-\exp(-\lambda u_{n-1}^{(i)}\Delta)$. It follows that if $\hat{u}_{n-1}^{(i)} > \tilde{u}_{n-1}^{(i)}$ then $\hat{X}_n^{(i)} >_{st}$

$\tilde{X}_n^{(i)}$, and consequently $\hat{X}_n^{(i)*} < \tilde{X}_n^{(i)*}$, because $\exp(-\lambda\Delta x)$ is a decreasing function of x. We can then conclude that $G_n^{(i)}$ in Eq. (20) is decreasing in $u_n^{(i)}$.

Moreover $\hat{X}_n^{(i)} >_{st} \tilde{X}_n^{(i)}$ implies also that $\hat{Y}_n^{(i)} >_{st} \tilde{Y}_n^{(i)}$, because

$$Y_n^{(i)} = \sum_{j=1}^{X_n^{(i)}} \eta^{(i)}(j),$$

where $\eta^{(i)}(j)$, $j = 1, 2, \ldots X_n^{(i)}$ are indicators showing which infected mobile can transmit to the destination in the n-th time slot and are i.i.d. Bernoulli random variables with expected value $1-\exp(-\lambda\Delta)$. Intuitively if the number of infected nodes is higher, also the number of infected nodes in the transmission range of the destination is higher.

Let $\hat{S}_n = \hat{Y}_n^{(i)} + S_n^{(-i)}$ and $\tilde{S}_n = \tilde{Y}_n^{(i)} + S_n^{(-i)}$, it follows that : $\hat{S}_n >_{st} \tilde{S}_n$. Then $\Pr\{\hat{S}_n = 0\} < \Pr\{\tilde{S}_n = 0\}$. About the second addend multiplying expected value in Eq. (20), first we observe that :

$$(1-\Pr\{S_n = 0\})\mathrm{E}\left[\frac{S_n^{(-i)}}{S_n} |\ S_n > 0\right] =$$
$$\Pr\{S_n^{(-i)} > 0\}\mathrm{E}\left[\frac{S_n^{(-i)}}{S_n} |\ S_n^{(-i)} > 0\right],$$

because $S_n^{(-i)}/S_n = 0$ when $S_n^{(-i)} = 0$. And being $S_n^{(-i)}/S_n = 0$ a decreasing function of $Y_n^{(i)}$, then we have

$$\mathrm{E}\left[\frac{S_n^{(-i)}}{\hat{S}_n} |\ S_n^{(-i)} > 0\right] < \mathrm{E}\left[\frac{S_n^{(-i)}}{\tilde{S}_n} |\ S_n^{(-i)} > 0\right].$$

This concludes the proof that also $G^{(i)}(n+1)$ in Eq. (20) is decreasing in the control action $u_{n-1}^{(i)}$. □

Theorem 5.1 *If $\forall n$ $G^{(i)}(n+1)$ is decreasing in the control action $u_{n-1}^{(i)}$, then the optimal threshold policy for the singleclass 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 $\mu^{(i)}$, we build in the same way a new policy $\hat{\mu}^{(i)}$. In fact equations (3), (4) and (8) hold also for each specific class i and the hypothesis on $G_n^{(i)}$ permits to conclude that $\hat{\mu}^{(i)}$ has better performance. □

Remarks. We observe that the result above applies to both the arbitration schemes and the traffic generation models considered. In fact Finally the different traffic models, for a given class i, only have an effect of the probability distributions of $X_n^{(-i)}$ and $Y_n^{(-i)}$, but they not change the best response strategy for class i.

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 singleclass threshold policy.*

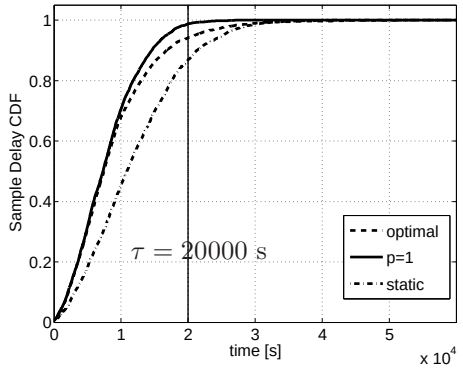


FIG. 1 – Delay CDF in the case of a) optimal control policy (dashed line) b) static control (dot-dashed line) and c) $p = 1$ (dotted line).

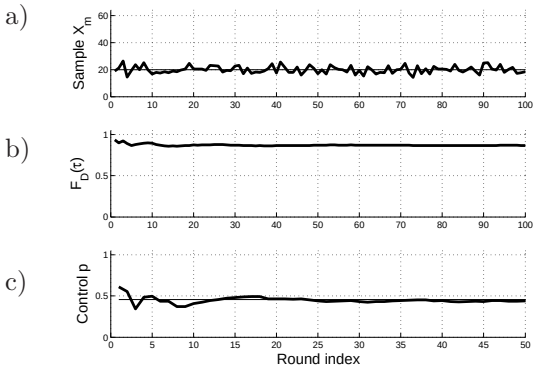


FIG. 2 – The dynamics of the stochastic approximation algorithm applied to the static forwarding policies.

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

6 Numerical Results

Numerical results have been obtained simulating the discrete-time system with Matlab.

The intensity λ of the pairwise meeting process has been selected considering a standard Random Waypoint (RWP) mobility scenario.In fact it is known [16], that for the RWP $\lambda = \frac{8wRv}{\pi L2}$, where L is the playground size, R the communication range, $w = 1.3683$ is a constant and v is the scalar speed of nodes. Here, we have chosen $L = 5000$ m, $N = 200$, $R = 15$ m and $v = 5$ m/s. The corresponding value is $\lambda = 1.0453 \times 10^{-5}\ s^{-1}$. For the timeslot we have chosen $\Delta = 10$ s.

6.1 Discrete control policies

In the first set of experiments, we simulated the discrete control policies in order to evaluate their relative performances. In Fig. 1 we reported the comparison of the optimal control policy and the static control policy. For the considered setting, where $u_{\min} = 0$ and $\tau = 20000$ s, we obtain $h^* = 911$ for the optimal threshold policy, and $p^* = 0.46$ for the static policy. It can be noticed that the static policy attains a much lower success probability, whereas, as expected, the delay CDFs under the optimal control and under the policy $\mu(n) = 1$ coincide at times smaller than $h^*\Delta$.

6.2 Stochastic Approximation

In the following we describe the application of the stochastic approximation algorithm described in Sec. 4 and we show that it is able to discover the optimal control policy for the two hops relay protocol. The setting is similar to what described above, but in this case several rounds are performed (see Sec. 4). Basically, the source performs for 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 the round, a novel policy is generated and is employed in the following run. Unless otherwise specified, results in this section have been obtained with $a_m = 1/(10 \cdot m)$, $\tau = 20000$ and $\Psi = 20$.

Fig 2 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. Nevertheless, the convergence of the algorithm is evident from the dynamics of the control p, i.e. the static forwarding probability, which stabilizes after about 20 rounds around the optimal value p^* (the horizontal line). For the sake of completeness, we also reported the running value of the delay CDF, measured at time τ, obtained during the run of the algorithm (Fig 2b)).

We repeated the same experiment in the case of the optimal threshold policies. In this case, the source tries to estimate the optimal threshold h^*, and the dynamics of the estimated parameter is depicted in Fig 3c). We observe that the convergence time is similar to that measured in the case of the 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, we have observed that the sequence of estimates converges to the solution of the same ODE.

Despite the dynamics showed so far are interesting, we have overlooked a lot of information that is carried by the ODE technique for stochastic approximations. In fact, as mentioned in Sec. 4, the ODE trajectory provides more information than the "simple" asymptotic stability of the control variable. In fact, the ODE method leverages a stronger property, i.e., the sample trajectories of the control estimates follow a shifted ODE dynamics with probability one. In particular, we depicted in Fig 4 the dynamics of the controlled variable against a (properly) shifted version of the reference ODE of the control for the static case. We averaged the trajectory over 10 runs of the algorithm. Also, in order to make the phenomenon visible, as described in [21], the dynamics are conveniently rescaled according to $t_n := \sum_{i=1}^{n-1} a_n$, i.e., at the time scale of the

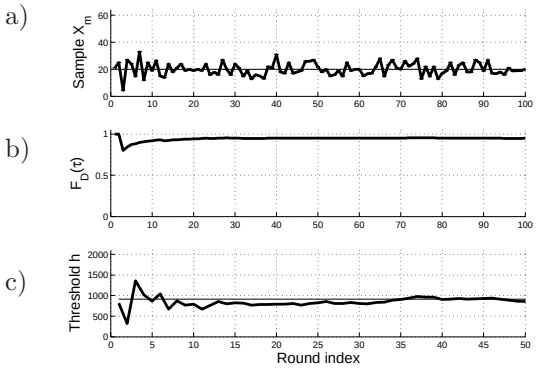


FIG. 3 – The dynamics of the stochastic approximation algorithm applied to the optimal forwarding policies.

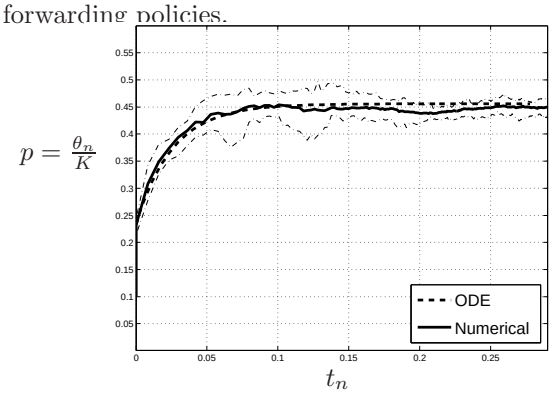


FIG. 4 – The convergence of the dynamics of the control variable against the reference ODE ; at the time scale t_n and averaged over 10 sample trajectories in the case of static control. Thin dash-dotted lines delimit the maximum and minimum values attained by the estimate trajectories.

control. It can be observed that, after an initial transient phase, the trajectory of the control mimics the original ODE ; we superimposed the maximum and minimum values of the trajectories 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.

6.2.1 Polyak's averages

As mentioned in Sec. 4, a slowly decaying a_n obtains a fast convergence to the ODE dynamics, i.e., the optimal control value. The price to pay is a lower rejection to noise, with larger oscillations. Here, we show the benefit of the Polyak-like averaging technique, as we choose a larger sequence, $a_n = 1/(10 \cdot n^{2/3})$, from which we expect faster convergence but a more noisy estimate.

Again, in Fig. 5 we reported the results of the stochastic approximation procedure : we superimposed the plain stochastic estimation of θ_n, based on the chosen a_n coefficients, and the output, obtained using the control from (18). We note the smoothing performed by the Polyak averaging over the estimated

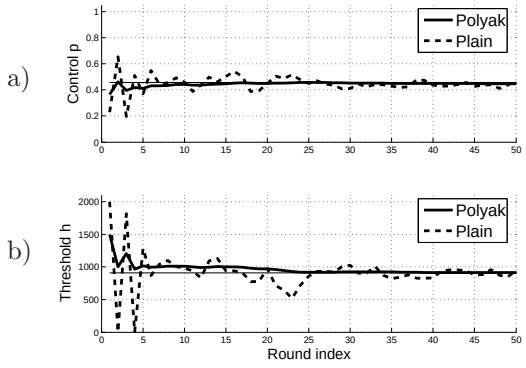


FIG. 5 – Algorithm employing Polyak's averages applied to a) static and b) threshold forwarding policies.

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 tradeoffs exist : it is possible to increase the speed of convergence of the algorithm by means of faster sequences, i.e. approaching faster the tail of the ODE dynamics, while reducing at the same time the estimation noise by averaging.

6.2.2 Nash Equilibrium

In the game theoretical framework, the result on the existence of a Nash equilibrium poses the question whether such equilibrium is Pareto optimal. 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.

For such reason, we resorted to numerical simulations in order to get better insight. In particular, we considered increasing number of nodes for a two-player game where each DTN has $N_1 = N_2 = 5, 6, 7$ nodes, and we rescaled the reference playground side to $L = 100$ m. Also, $\tau = 200$ s in this experiments. We repeated game rounds in order to measure the impact of the different strategies under the collision model. As depicted in Fig. 6, at the Nash equilibrium, the success probability is smaller than the one experienced in isolation by single players using the optimal threshold policy. This was expected, due to the effect of collisions. But, as shown in Fig. 6, if each class adopts the best static policy, the social outcome can be improved. We observe that this is not an equilbrium, because a class would find more convenient to switch to its optimal threshold policy, but it provides numerical evidence that the Nash equilibrium is not Pareto optimal.

7 Conclusions

In this report we introduced a discrete time model for the control of mobile ad hoc DTNs. We provided closed form expressions for optimal static and threshold forwarding policies for two-hops routing. Based on such results, we provided an algorithm based on the theory of stochastic approximations; the

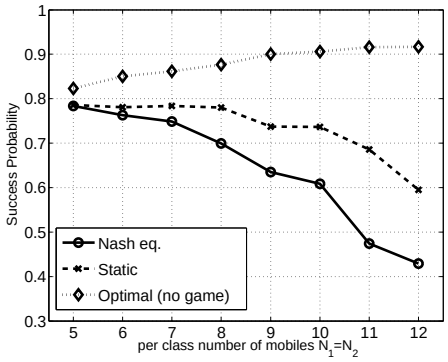


FIG. 6 – Performance at the Nash equilibrium compared to the ideal case and to a static strategy ; $\tau = 200$ s, $\Psi = N - 1$.

algorithm, in particular, enables all nodes in the DTN to tune independently and optimally the parameters of static and dynamic optimal forwarding policies, adapting to the current operating conditions of the system. It does not require actually message exchanges to operate such tuning and, more important, it guarantees optimality but it does not require to estimate global parameters of the DTN, such as the number of nodes or the intermeeting intensities. We believe that these features are very appealing ; similar techniques promise application 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.

Finally, the discrete model has been applied to the case of competing DTNs : we studied a class of weakly coupled Markov games where players are DTNs, and the coupling occurs because of interference at a common destination node. Based on the previous discrete model, we proved that a unique Nash equilibrium, exists for the entire class of games based on threshold policies.

A Stochastic order

In this appendix we provide definition and properties of usual stochastic order. Given two real valued random variables x and y, x is said to be smaller than y in *usual stochastic order*, if

$$\text{Prob}\{x > t\} \leq \text{Prob}\{y > t\} \; \forall t \in \mathbb{R},$$

and this is written as

$$x \leq_{st} y.$$

Definition and properties of usual stochastic order can be found in [19].

In this paper we rather use strict stochastic order relations. We say that x is *strictly stochastically smaller* than y, $x <_{st} y$, if

$$\begin{aligned} &1) \quad \text{Prob}\{x > t\} \leq \text{Prob}\{y > t\} \;\; \forall t \in \mathbb{R}, \\ &2) \quad \exists\, t^* \in \mathbb{R} \;\mid\; \text{Prob}\{x > t^*\} < \text{Prob}\{y > t^*\}. \end{aligned}$$

We prove below the two main properties of strict stochastic order that we employ in this report.

Proposition A.1 *Given two real-valued random variables x and y and $\Phi : \mathbb{R} \to \mathbb{R}$ a strictly increasing function, if $x <_{st} y$ then $\Phi(x) <_{st} \Phi(y)$.*

Proof. Let us define $\Phi^{-1}(u) \triangleq \inf\{z : \Phi(z) > u\}$.

$$\text{Prob}\{\Phi(x) > u\} = \text{Prob}\{x > \Phi^{-1}(u)\} \leq \text{Prob}\{y > \Phi^{-1}(u)\} = \text{Prob}\{\Phi(y) > u\}.$$

If t^* is the value for which $\text{Prob}\{x > t^*\} < \text{Prob}\{y > t^*\}$, define $u^* = \lim_{t \to t^{*+}} \Phi(t)$ (this is needed to deal with eventual discontinuity of Φ in t^*). It holds :

$$\text{Prob}\{\Phi(x) > u^*\} = \text{Prob}\{x > t^*\} > \text{Prob}\{y > t^*\} = \text{Prob}\{\Phi(y) > u^*\}.$$

□

Proposition A.2 *Given two non-negative real-valued random variables x and y, if $x <_{st} y$ then $E[x] < E[y]$.*

Proof.

$$\text{E}[x] = \int_0^\infty \text{Prob}\{x > t\}\text{d}t < \int_0^\infty \text{Prob}\{y > t\}\text{d}t = \text{E}[y],$$

where strict inequality follows from right continuity of the complementary distribution function. □

Références

[1] 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.

[2] L. Pelusi, A. Passarella, and M. Conti, "Opportunistic networking : data forwarding in disconnected mobile ad hoc networks," *IEEE Communications Magazine*, vol. 44, no. 11, pp. 134–141, November 2006.

[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] 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.

[6] 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 13–17 2005.

[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] 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.

[9] ——, "Constrained stochastic games in wireless networks," in *IEEE Globecom General Symposium*, Washington D.C., 2007.

[10] M. Musolesi and C. Mascolo, "Controlled Epidemic-style Dissemination Middleware for Mobile Ad Hoc Networks," in *Proc. of ACM Mobiquitous*, July 2006.

[11] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Comput. Netw.*, vol. 51, no. 10, pp. 2867–2891, 2007.

[12] 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.

[13] A. Krifa, C. Barakat, and T. Spyropoulos, "Optimal buffer management policies for delay tolerant networks," in *Proc. of IEEE SECON*, 2008.

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

[15] E. Altman, T. Başar, and F. D. Pellegrini, "Optimal monotone forwarding policies in delay tolerant mobile ad-hoc networks," *Proc. of ACM InterPerf*, October 24 2008.

[16] R. Groenevelt and P. Nain, "Message delay in MANETs," in *Proc. of ACM SIGMETRICS*, Banff, Canada, June 6, 2005, pp. 412–413.

[17] 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.

[18] T. Karagiannis, J.-Y. L. Boudec, and M. Vojnović, "Power law and exponential decay of inter contact times between mobile devices," in *Proc. of MobiCom '07*. ACM, 2007, pp. 183–194.

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

[20] E. Altman, *Constrained MArkov Decision processes*. Chapman and Hall/CRC, 1999.

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

[22] J. Maryak, "Some guidelines for using iterate averaging in stochastic approximation," *Decision and Control, 1997., Proceedings of the 36th IEEE Conference on*, vol. 3, pp. 2287–2290 vol.3, Dec 1997.

[23] B. T. Polyak and A. Juditsky, "Acceleration of stochastic approximation by averaging," *SIAM Journal of Control and Optimization*, vol. 30, pp. 838–855, 1992.

[24] E. Altman, G. Neglia, F. De Pellegrini, and D. Miorandi, "Decentralized stochastic control of delay tolerant networks, INRIA Research Report 2008-Number Pending," August 2008. [Online]. Available : http://www-sop.inria.fr/maestro/personnel/Giovanni.Neglia/publications.htm

Table des matières

1 Introduction **3**

2 System Model **4**

3 Characterization of Optimal Policies **6**

4 Stochastic Approximations for Adaptive Optimization **9**
4.1 Choice of the Sequence $\{a_n\}$ 12
4.2 Constant Step Approximations 13

5 The Multiclass Case **13**
5.1 The model 13
5.2 A Weakly Coupled Markov Game Formulation 14

6 Numerical Results **17**
6.1 Discrete control policies 18
6.2 Stochastic Approximation 18
6.2.1 Polyak's averages 19
6.2.2 Nash Equilibrium 20

7 Conclusions **20**

A Stochastic order **21**



Centre de recherche INRIA Sophia Antipolis – Méditerranée
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399