INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE
INRIA
centre de recherche SOPHIA ANTIPOLIS - MÉDITERRANÉE

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Distributed Sub-gradient Methods for Delay Tolerant Networks

Giovanni Neglia (INRIA)

joint work with Riccardo Masiero (Univ. di Padova)

Grenoble June 20, 2011

Outline

- Motivations
- Method overview
- Our contribution
    - Mobility models
    - Asynchronous updates
    - Example of application in DTN
- Conclusions

General problem

In network applications ...

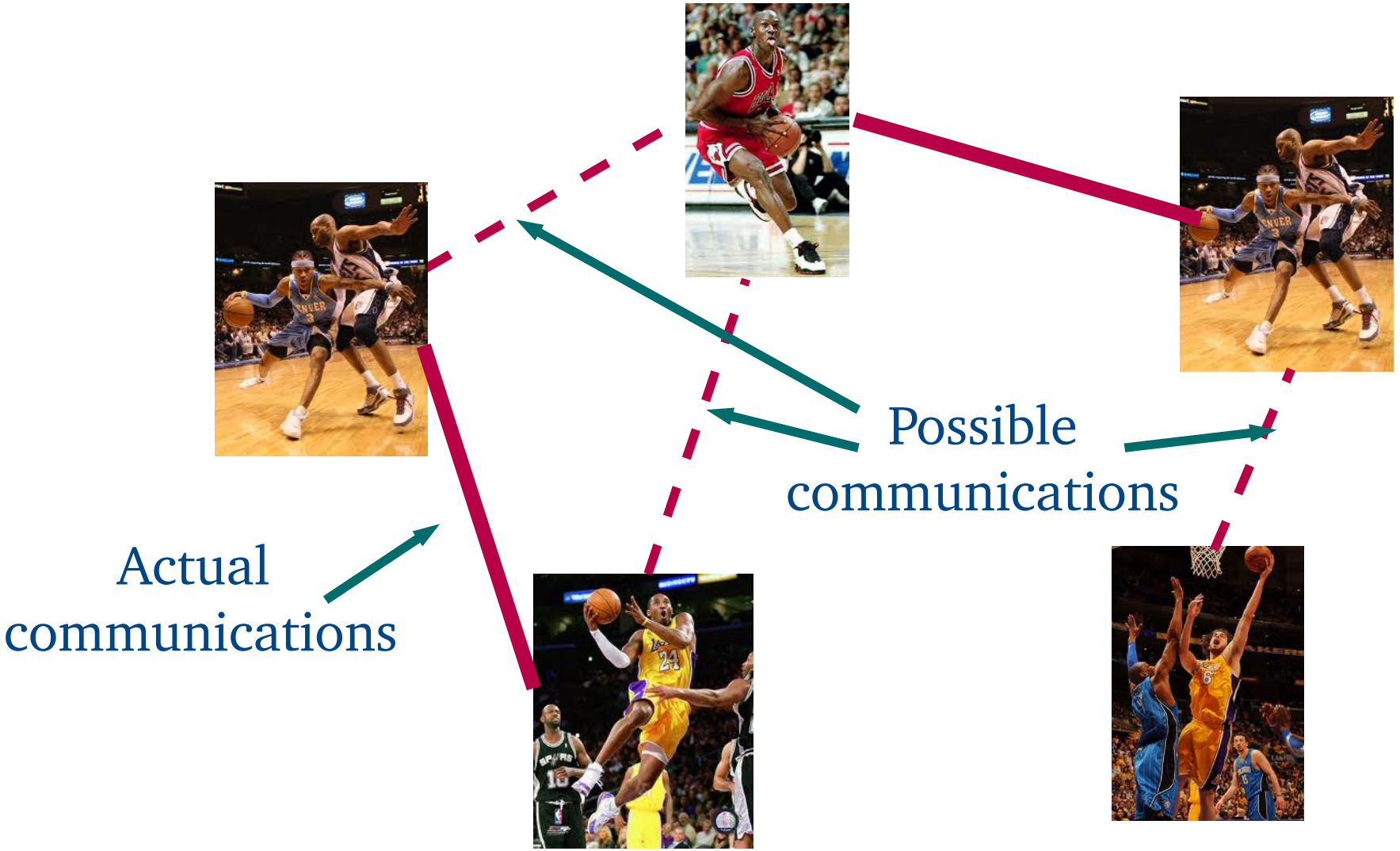
- node-wise:

Local beliefs about best actions for each player

- network-wise:

global performance

General problem



A particular case

- node-wise:
  convex local objective function

- network-wise:
  sum of objective functions

$$f_i(\boldsymbol{x}), \quad i=1,\ldots,M$$

$$f(\boldsymbol{x})=\sum_{i=1}^{M} f_i(\boldsymbol{x})$$

$$\boldsymbol{x}\in\mathbb{R}^N$$

parameter vector (actions to take), e.g. N=hM and the action of each node is identified by h parameters

Problem to solve

$$\text{determine } \boldsymbol{x}^*\in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\operatorname{argmin}}\, f(\boldsymbol{x})=\sum_{i=1}^{M} f_i(\boldsymbol{x})$$

Wrong approach

Local optimization of each player

$$\boldsymbol{x}^{i}(k+1)=\boldsymbol{x}^{i}(k)-\gamma(k)\boldsymbol{d}^{i}(k)$$

$\boldsymbol{d}^{i}(k)$: Gradient of $f_i(\boldsymbol{x})$ evaluated in $\boldsymbol{x}^{i}(k)$

$\gamma(k)$: Update weight sequence

Result

Each agent i would set its parameters on the basis of the minimum of its objective function $f_i(\boldsymbol{x})$

Distributed subgradient (Nedic, Asuman)

Iterative step

$$\boldsymbol{x}^{i}(k+1)=\sum_{j=1}^{M} a_{ij}(k)\boldsymbol{x}^{j}(k)-\gamma(k)\boldsymbol{d}^{i}(k)$$

average (gossip)

local optimization

$a_{ij}(k)>0$ iff node i and j have exchanged their estimates during slot k

$$\sum_{j=1}^{M} a_{ij}(k)=1$$

Distributed subgradient (Nedic, Asuman)

Iterative step

$$\boldsymbol{x}^{i}(k+1)=\sum_{j=1}^{M} a_{ij}(k)\boldsymbol{x}^{j}(k)-\gamma(k)\boldsymbol{d}^{i}(k)$$

average (gossip)

local optimization

In matrix form

$$\boldsymbol{X}(k+1)=\boldsymbol{A}(k)\boldsymbol{X}(k)-\gamma(k)\boldsymbol{D}(k)$$

$$\boldsymbol{X}(k)\stackrel{\text{def}}{=}[\boldsymbol{x}^{1}(k),\ldots,\boldsymbol{x}^{i}(k),\ldots,\boldsymbol{x}^{M}(k)]^{T}$$

$$\boldsymbol{D}(k)\stackrel{\text{def}}{=}[\boldsymbol{d}^{1}(k),\ldots,\boldsymbol{d}^{i}(k),\ldots,\boldsymbol{d}^{M}(k)]^{T}$$

Convergence/Optimality

Main assumptions about communications

1. Nodes communicate infinitely often, directly or indirectly

Time k

Time k+1

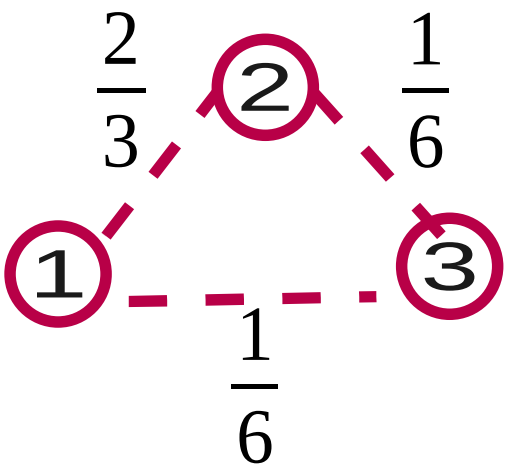
- 1-2 and 2-3 direct
- 1-3 indirect

2'. upper bound for the intercommunication interval between two nodes

OR

2''. matrices **A**(k) are i.i.d. random matrices

Toy example



in each slot only one link is activated with the corresponding probability

Local functions:

$$f_1(x) = f_2(x) = \frac{x(x-1)}{2}, \quad f_3(x) = 2x^2$$

Global function: $f(x) = \sum_{i=1}^{3} f_i(x) = 3x^2 - x$

$$x^* = \frac{1}{6}, \quad f(x^*) = -\frac{1}{12}$$

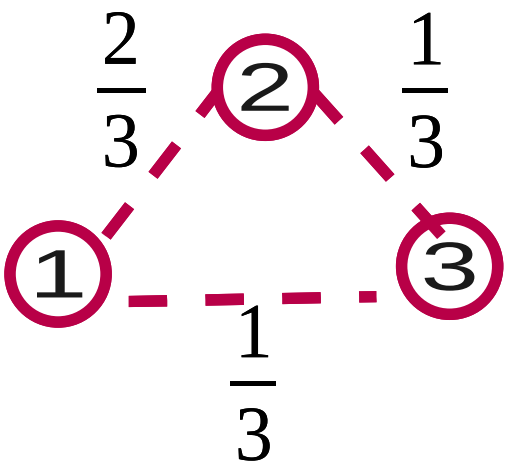
$\forall$ k, **A**(k) can be:

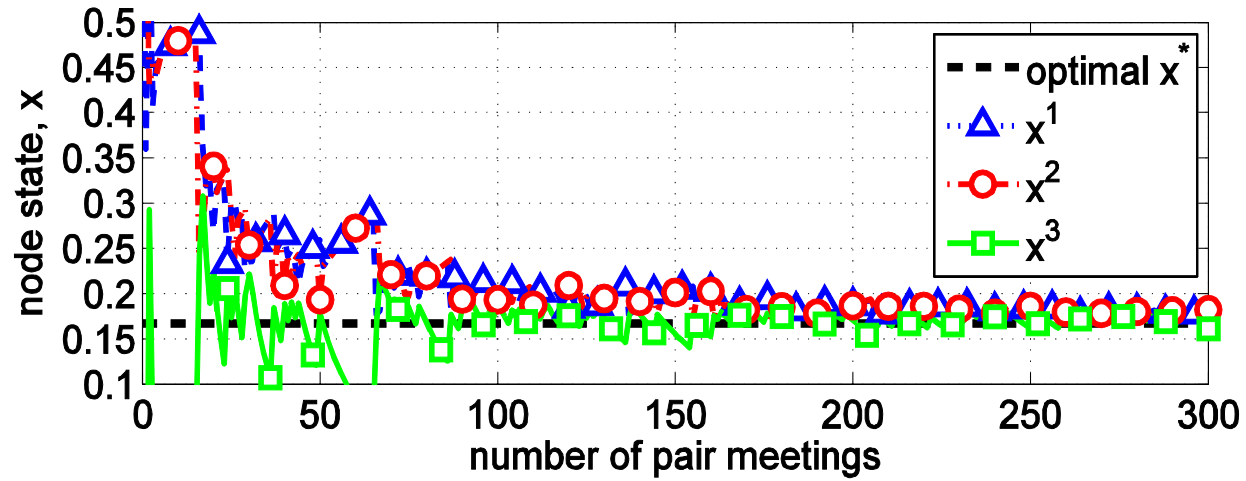
$$\begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

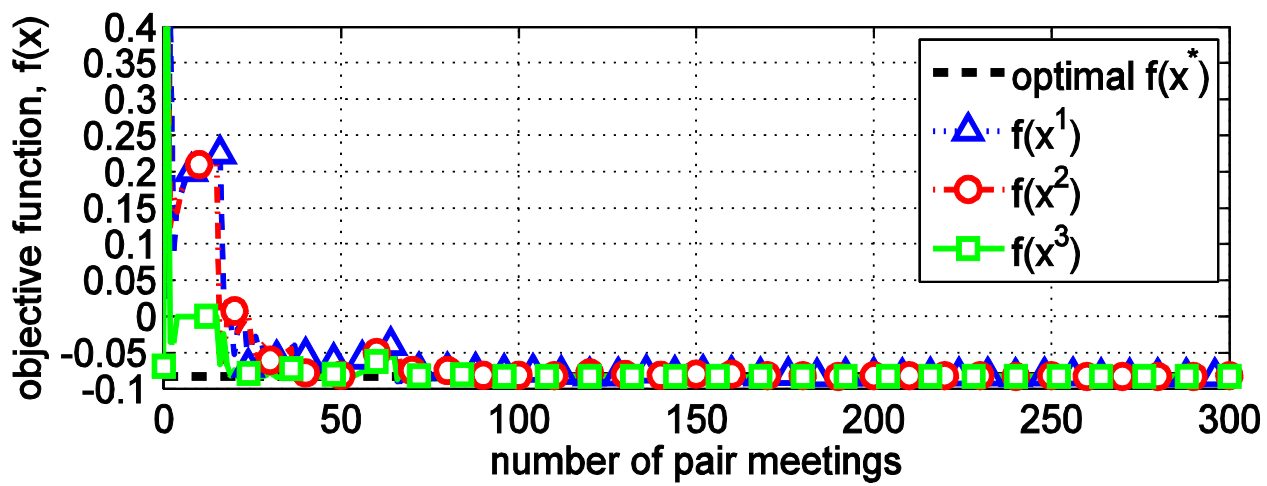
$$\begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

Toy example


2/3
1/3
1/3
1
2
3


node state, x
number of pair meetings
optimal x*
x1
x2
x3


objective function, f(x)
number of pair meetings
optimal f(x*)
f(x1)
f(x2)
f(x3)

Sketch of the proof

Define

$$\boldsymbol{A}_{(s)}^{(k)} \stackrel{\text{def}}{=} \boldsymbol{A}(k)\boldsymbol{A}(k-1)\cdots\boldsymbol{A}(s), \qquad \text{Backward product}$$

$$\boldsymbol{y}(k)^T \stackrel{\text{def}}{=} \frac{1}{M}\mathbf{1}^T\boldsymbol{X}(k), \qquad \text{Average opinion}$$

we have

$$\boldsymbol{y}(k+1)^T = \boldsymbol{y}(k)^T - \frac{\gamma(k)}{M}\mathbf{1}^T\boldsymbol{D}(k),$$

$$\boldsymbol{X}(k+1) = \boldsymbol{A}_{(1)}^{(k)}\boldsymbol{X}(1) - \sum_{s=2}^{k}\boldsymbol{A}_{(s)}^{(k)}\gamma(s-1)\boldsymbol{D}(s-1) - \gamma(k)\boldsymbol{D}(k)$$

Sketch of the proof

Proof

$$A_{(1)}^{(k)} \to J \stackrel{\text{def}}{=} \frac{1}{M} \mathbf{1}\mathbf{1}^T$$

Exponentially fast

$$\boldsymbol{x}^i(k) - \boldsymbol{x}^j(k) \to 0, \forall i, j$$

Gossip makes estimates converge

$$\boldsymbol{x}^i(k) \to \boldsymbol{y}(k), \forall i$$

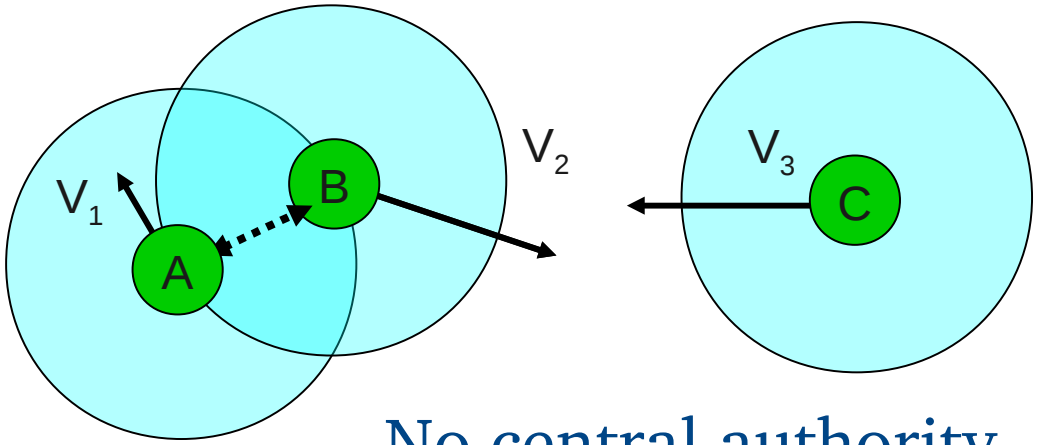
$$\boldsymbol{y}(k+1)^T = \boldsymbol{y}(k)^T - \frac{\gamma(k)}{M} \mathbf{1}^T \boldsymbol{D}(k),$$

It is the gradient of the global function

$$\boxed{\boldsymbol{y} \to \boldsymbol{x}^*, \forall i \quad \boldsymbol{x}^i \to \boldsymbol{x}^*}$$

Applications in DTNs

Delay Tolerant Networks
- mobile wireless networks
- no continuous connectivity



No central authority
Only local information
Network scenario unknown a priori

Online decentralized optimization appears to be particular important

Applications in DTNs

Opportunistic routing
selection of routing parameters to achieve optimal delivery time vs energy consumption trade off

Dynamic content injected by a base station
selection of nodes'update rates to minimize expected age in the network

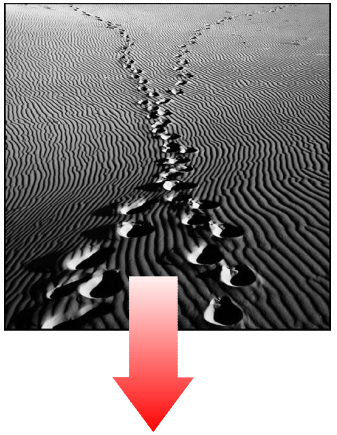
Applications in DTNs

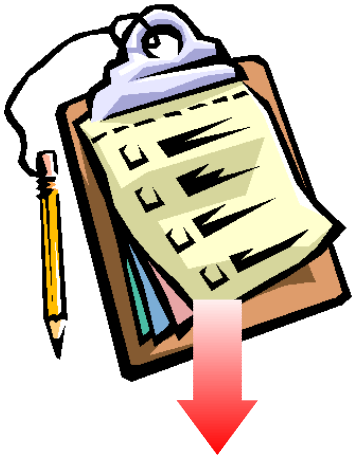
How to?

exchange of $\mathbf{x}^i$ → meetings

$\mathbf{A}(k)$ originate from the node meeting process

update of $\mathbf{x}^i$ → meetings or a priori





Asynchronous
easy to implement

Synchronous
difficult in a distributed setting

Applications in DTNs

How to? **A**(k) originate from the node meeting process

Formally ... contact matrix:

$$\mathbf{C}(k) = \begin{cases} c_{ij}(k) = 1 & \text{if node i has met node j} \\ c_{ij}(k) = 0 & \text{otherwise} \end{cases}$$

C(k) @ each meeting

$$a_{ij}(k) = a\, c_{ij}(k), \quad for\ i \neq j \quad with \quad 0 < a < 1 \ \wedge \ a_{ii}(k) = 1 - a$$

updates synchronized with meetings

Open issues

Mobility models under assumptions 2' and 2'' ?

2' → deterministic bounds

2'' → independent meetings

More general mobility models required

Do convergence and optimality still hold ?

Open issues

Synchronous operation

$$\boldsymbol{x}^{\boldsymbol{i}}(k+1)=\sum_{j=1}^{M} a_{ij}(k)\boldsymbol{x}^{\boldsymbol{j}}(k)-\gamma(k)\boldsymbol{d}^{\boldsymbol{i}}(k)$$

Asynchronous operation

$$\boldsymbol{x}^{\boldsymbol{i}}(k+1)=\sum_{j=1}^{M} a_{ij}(k)\boldsymbol{x}^{\boldsymbol{j}}(k)-\gamma^{i}(k)\boldsymbol{d}^{\boldsymbol{i}}(k)$$

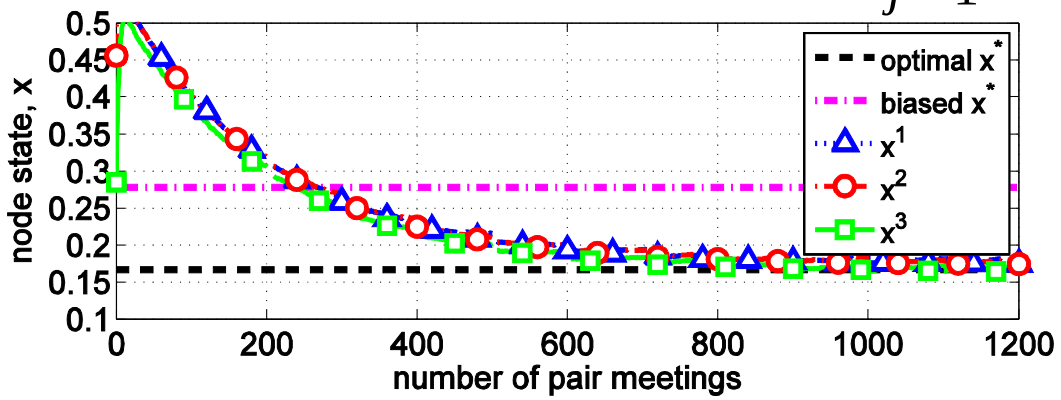
Open issues

Synchronous operation

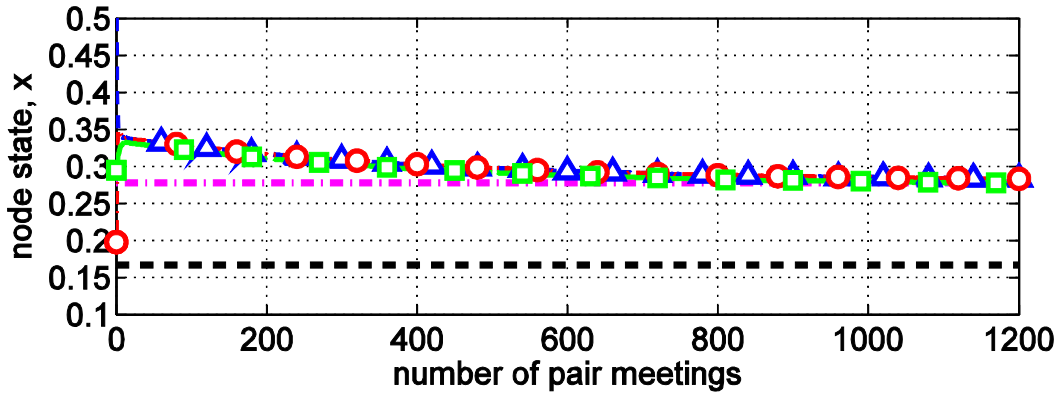
$$\boldsymbol{x}^{i}(k+1)=\sum_{j=1}^{M} a_{ij}(k)\boldsymbol{x}^{j}(k)-\gamma(k)\boldsymbol{d}^{i}(k)$$

Asynchronous operation

$$\boldsymbol{x}^{i}(k+1)=\sum_{j=1}^{M} a_{ij}(k)\boldsymbol{x}^{j}(k)-\gamma^{i}(k)\boldsymbol{d}^{i}(k)$$



synchronous



asynchronous

Why the bias ?
How to cope with it ?

More general mobility models

Assumption

It exists an irreducible, aperiodic and stationary Markov chain Φ with a finite or countable set of states S and a function $g: S \rightarrow C$, such that $\mathbf{C}(k) = g(\Phi_k)$, for each $\Phi_k \in S$. Moreover, $E[\mathbf{C}(k)]$ is an irreducible aperiodic matrix.

We prove that convergence holds also in this case under synchronous operation

$$\boldsymbol{A}_{(s)}^{(k)} \rightarrow \frac{1}{M} \mathbf{1}\mathbf{1}^T \stackrel{\text{def}}{=} \boldsymbol{J} \quad a.s., \forall\, k \geq s$$

$$\left\| \boldsymbol{A}_{(s)}^{(k)} - \boldsymbol{J} \right\|_\infty \leq C\beta^{k-s}$$

Key point of the proof

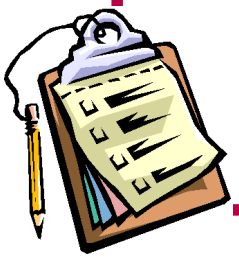
Then optimality follows as in the literature ...

Asynchronous Updates

Result

$$\lim_{k\to\infty} \left\| \boldsymbol{y}(k) - \boldsymbol{x}^{i}(k) \right\|_2 = 0 \quad (a.s.), \quad \forall i$$

Synchronous updates



$$0 \le \sum_{s=1}^{\infty} \gamma(s)\left[f(\boldsymbol{y}(s)) - f(\boldsymbol{x}^*)\right] < \infty \quad (a.s.)$$

and since $\sum \gamma(s) = \infty$

$$\liminf_{k\to\infty} f(\boldsymbol{y}(k)) = f(\boldsymbol{x}^*) \quad (a.s.) \quad \wedge \quad \lim_{k\to\infty} \boldsymbol{x}^{i}(k) = \boldsymbol{x}^*(a.s.)$$

$\Rightarrow$ **Optimality**

Asynchronous Updates

Asynchronous updates

$$\sum_{s=1}^{\infty}\sum_{i=1}^{M}\gamma_i(s)[f_i(\boldsymbol{y}(s))-f_i(\boldsymbol{x}^*)]<\infty \quad (a.s.)$$

Conjecture

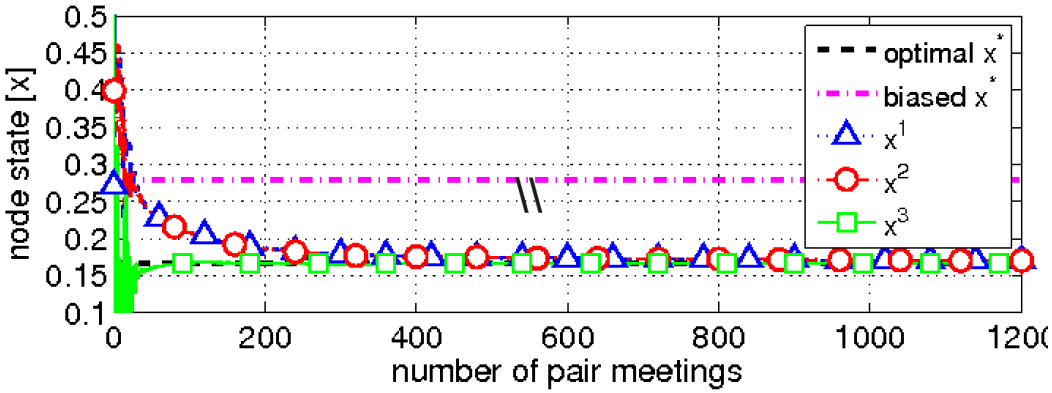
When updates are asynchronous, convergence results to optimality for sub-gradient methods hold if

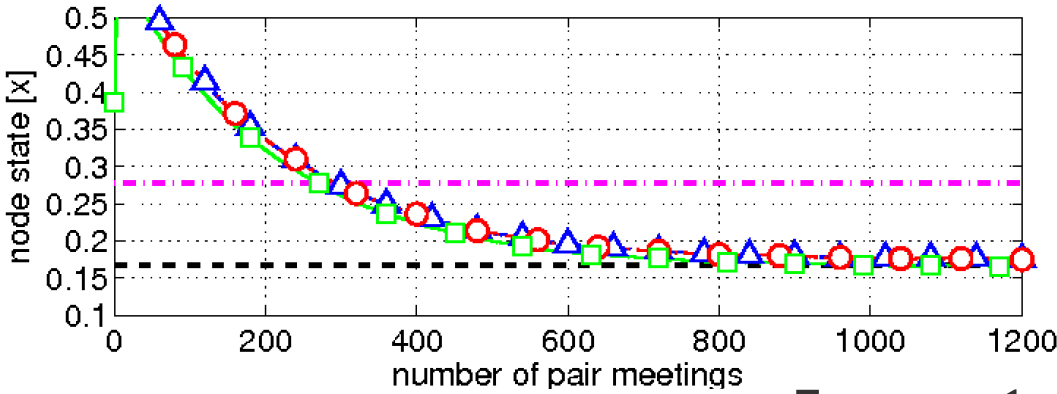
$$E[\gamma_i(k)] = E[\gamma_j(k)]$$

for each i and j

Asynchronous Updates

Conjecture's validity: examples





# of updates for i $u_i(k)$

probability i meets nodes p_i

meetings ~ Poisson

with $\gamma_i(k)=\frac{1}{u_i(k)}$

$$E[\gamma_i(k)]=E[\frac{1}{u_i(k)}]=\frac{p_i}{p_i k}=\frac{1}{k},\forall i$$

→ we reach optimality

with $\gamma_i(k)=\gamma \wedge p_1=p_2=\frac{5}{6}, p_3=\frac{1}{3}$

we optimize $\sum_i p_i f_i$ not $\sum_i f_i$

→ we observe bias

with $\gamma_i(k)=p_i^{-1}\gamma$

$$E[\gamma_i(k)]=E[p_i^{-1}\gamma]=p_i\frac{\gamma}{p_i}=\gamma,\forall i$$

→ we reach optimality

A Case Study [Ioannidis09,Altman09]

Application in DTNs

Scenario : dissemination of dynamic content in a large network

service Provider

periodic information updates

limited bandwidth

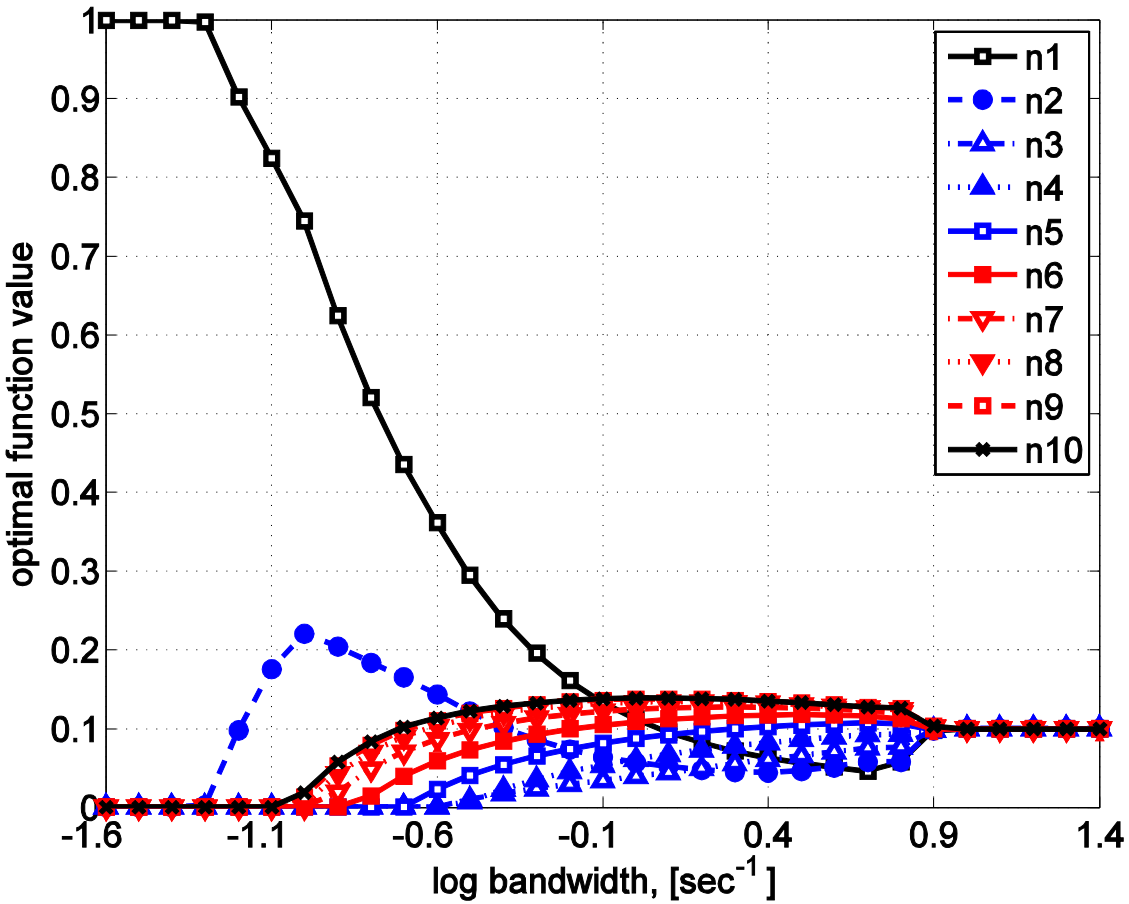
Taking advantage of opportunistic communications

Objective:

how to allocate bandwidth to nodes in order to reduce average information freshness

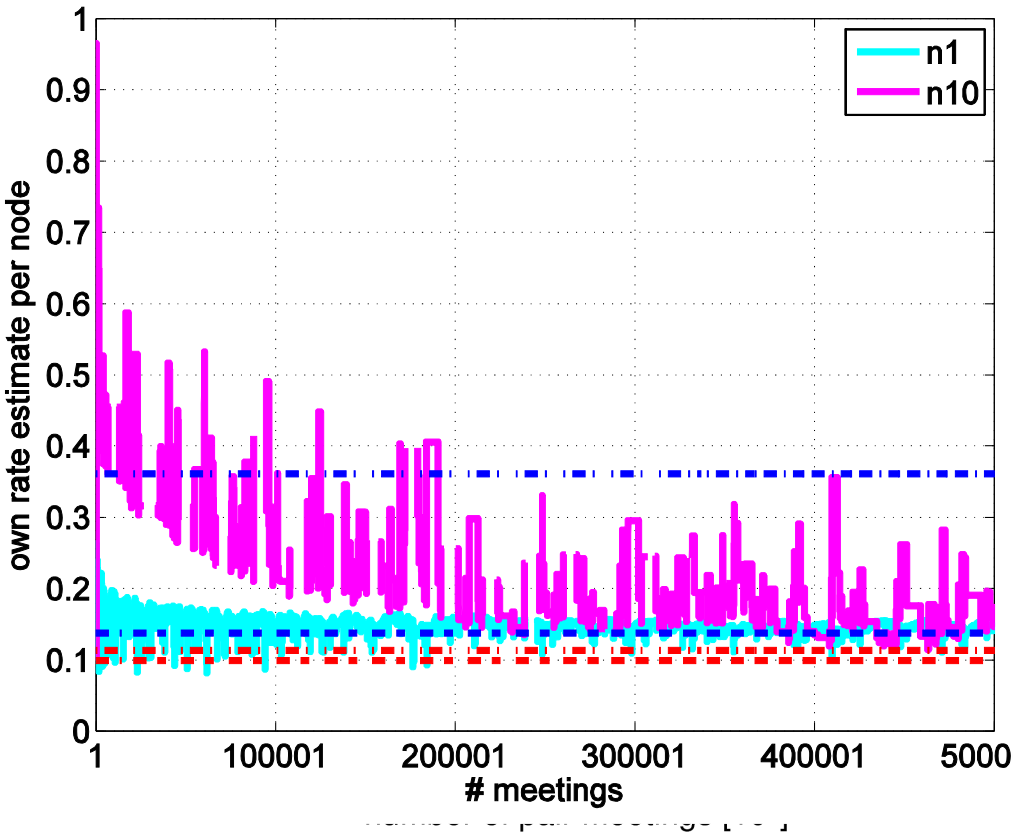
Centralized solution [Ioannidis09]

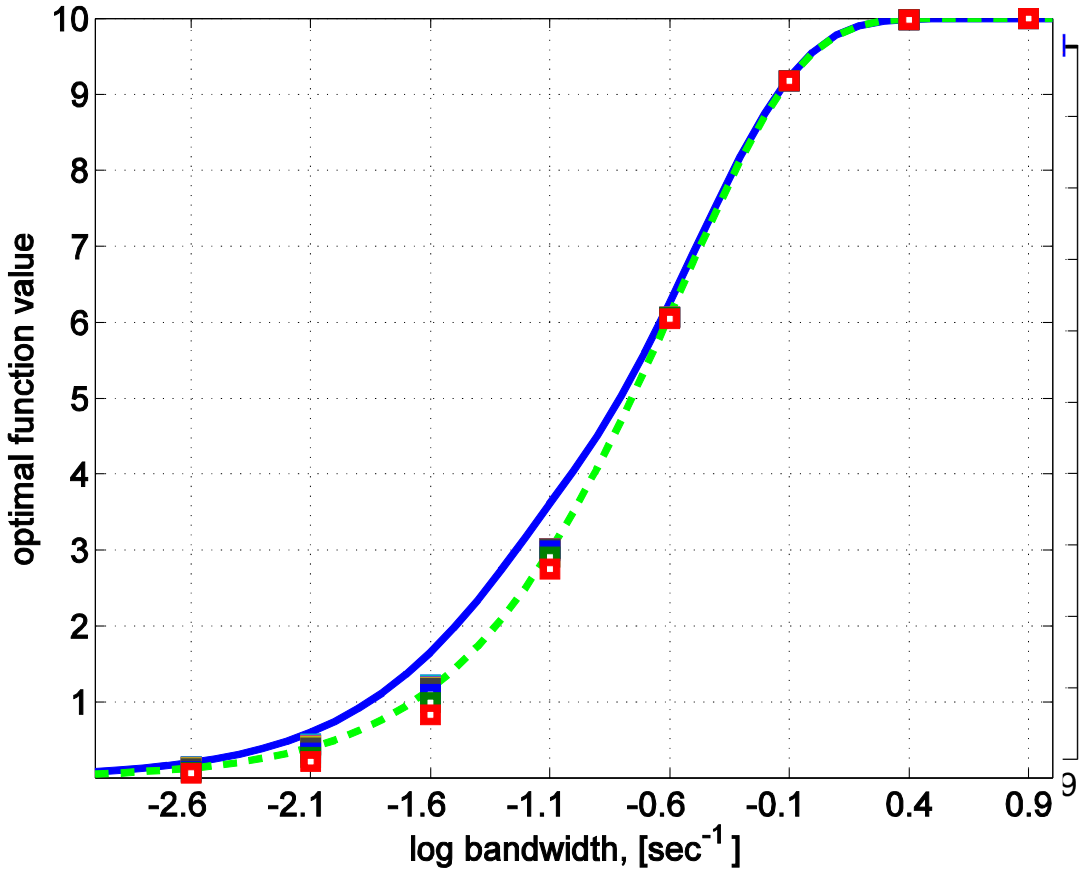
A convex optimization problem



10 nodes, ordered according to their contact rates (highest to lowest)

Our distributed approach





Conclusions

- Presentation of the Distributed Sub-gradient Method
- Application of the framework to DTNs
- Tackled issues:
    - Extension to more general mobility models
    - Asynchronous updates
- Case study:
    - Distributed Sub-gradient Method for dissemination of dynamic content