Incentive-based control of ad hoc networks: A performance study

Göbel, J.; Krzesinski, A.; Mandjes, M.

DOI
10.1016/j.comnet.2009.04.010

Publication date
2009

Document Version
Final published version

Published in
Computer Networks

Computer Networks 53 (2009) 2427–2443

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Incentive-based control of ad hoc networks: A performance study

Johannes Göbel [a], Anthony Krzesinski [b,*,1], Michel Mandjes [c,2]

[a] Department of Informatics, University of Hamburg, Vogt-Kölln-Str. 30, 22527 Hamburg, Germany
[b] Department of Mathematical Sciences, University of Stellenbosch, 7600 Stellenbosch, South Africa
[c] Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 22 November 2007
Received in revised form 2 April 2009
Accepted 28 April 2009
Available online 8 May 2009

Responsible Editor: M. van Steen

Keywords:
Autonomous motion
Congestion pricing
Credit incentives
Credit redistribution
Incentives for collaboration
Mobile ad-hoc networks
Mobility models
Radio interference

ABSTRACT

Ad hoc networks are self-configuring networks of mobile nodes, connected by wireless links. If a destination node is beyond the transmission range of an origin node, then the nodes must cooperate to provide a multi-hop route. Any node can act as a sender, receiver or transit node. It is clear that it is in a node's interest to be a sender or receiver, but it is less clear what the value is of forwarding traffic on behalf of other nodes. The nodes should therefore be given incentives to act as transit nodes, otherwise the network would fail to function. A way to do so is by introducing for each node a credit balance, where nodes use credits to pay for the costs of sending their own traffic, and earn credits by forwarding traffic from other nodes.

However, nodes that are located near the edge of the network will attract little transit traffic and earn few credits. In contrast, nodes located near the centroid of the network will attract transit traffic and earn credits. We investigate various ways of providing nodes near the edge of the network with preferential treatment in order to improve their credit balance and their throughputs.

We next focus on the situation where each node can move to improve its utility expressed in terms of either credit balance or throughput. Here radio interference plays an important role, as it defines an interesting trade-off: nodes may prefer to be close together in order to reduce the power needed to transmit data, but on the other hand proximity increases radio interference, and has therefore a negative effect on connectivity. Simulation experiments reveal that the positions of the nodes converge to non-trivial optimal positions on 2D and 3D surfaces.

1. Introduction

Ad hoc networks are self-configuring networks of mobile nodes, connected by wireless links. They enable infrastructure-free communication: no fixed equipment is needed, but instead each node acts as a router. An excellent survey on ad hoc networks, with special emphasis on Quality-of-Service aspects, is given in [1].

Any node can act as a sender, receiver or transit node. Although a node will benefit from being a sender or a receiver, it is less clear how nodes benefit from forwarding traffic on behalf of other nodes. However, if nodes do not act as transit nodes, the ad hoc network would fail to function. Therefore, a crucial question is: how can nodes be given incentives to act as transit nodes?

This question was studied in several papers. Crowcroft et al. [2] rely on earlier work on pricing and rate control [3,4] to devise such an incentive scheme. In their

* Corresponding author. Tel.: +27 21 808 4310; fax: +27 21 808 4416.
E-mail addresses: goebel@informatik.uni-hamburg.de (J. Göbel), aek1@cs.sun.ac.za (A. Krzesinski), mmandjes@science.uva.nl (M. Mandjes).

1 Supported by Grant numbers 2054027 and 2677 from the South African National Research Foundation, Nokia-Siemens Networks and Telkom SA Limited.

2 Also with: CWI, Kruislaan 413, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands and with EURANDOM, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

1389-1286/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2009.04.010

Nomenclature

$\mathscr{S}, \mathscr{D}$ the set of source, destination nodes
$O(r), D(r)$ the originating and the destination node of route r
$f_r(i)$ the node that node i forwards traffic to on route r
σ_{ri} the signal-to-interference ratio for a signal sent from node i to the next node $f_r(i)$ along route r
$\mathscr{R}^S(s), \mathscr{R}^D(s), \mathscr{R}^T(s)$ the set of routes that originate, terminate, transit at node s
$N_s(t)$ the number of calls originating from node s at time t
$x_s(t)$ the flow sent by a source node s at time t
$y_r(t), Y_{ri}(t)$ the net flow sent along route r at time t, the gross flow sent from node i to node $f_r(i)$ at time t
$c_j(t), \gamma_j(t)$ the bandwidth, power in use at node j at time t
$w_s(t), b_s(t)$ the willingness-to-pay for congestion costs incurred at node s at time t, the credit balance at node s at time t
$\mu_{rj}(t), \mu_r(t)$ the price at node j for carrying a unit of flow along route r at time t, the price at node $O(r)$ for carrying a unit of flow end-to-end along route r at time t
$\mu_j^P(t), \mu_j^B(t)$ the power, bandwidth congestion price at node j at time t
$\boldsymbol{z}_i$ the (x, y) location of node i
d the Euclidean distance $\|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2$

Constants
$p_i = 10^{-3}$ W/Mbps the power radiated per unit flow by a call at node i
$\Delta = 0.01$ s the update interval for prices, credit balances and flows
$\ell(d) = kd^{-u}$ the attenuation function where $k = 1.82 \times 10^{-14}$ and $u = 3.52$
$W = 3.84$ Mbps the chip rate of the spreading code
$N_0 = 10^{-13}$ the power of the thermal background noise
$L = 60$ bits the packet size
$\alpha_s = 0.3\ \mathrm{s}^{-1}$ the willingness-to-pay factor
$\beta = 0.01\ \mathrm{s}^{-1}$ the credit balance discount factor
$\eta = 1$ the code orthogonality factor
$\kappa = 0.05\ \mathrm{s}^{-1}$ the bandwidth and power congestion price factor
$C_j = 10.0$ Mbps the target bandwidth usage at node j
$\Gamma_j = 0.5$ W the target power usage at node j
$e^{\mathrm{rx}} = 10^{-3}$ W/Mbps the power consumed per unit flow when receiving
$e_{ij}^{\mathrm{tx}} = \max(10^{-2}, 10^{-4}d^2)$ the power consumed per unit flow when transmitting a distance d from node i to node j

framework, each node has a credit balance that determines how much the node can spend on transmission resources during the next time interval. For each node there are two resources: bandwidth and power, each with its own price. The price of the resource increases when the resource is scarce, and decreases when the resource is abundant. When a call arrives, given the current prices, it is connected along the least cost route to its destination. At the same time the node earns credits when acting as a source, destination or transit node. The interplay between the prices, rate allocations, and credit balances is such that global stability of the system is achieved. Importantly, such a scheme is *decentralized*: no central controller is needed, and the scheme therefore has favourable scalability properties.

Chen et al. [5] consider a setting in which there are several types of nodes: some wish to communicate ('clients'), while others do not, but are willing to serve as relay nodes. The clients cannot reach the base station directly, and can use the relay nodes to forward their traffic, in which case the relay nodes should be provided with an incentive to do so. Mechanisms are proposed to pay the relay nodes for the service they deliver. Every relay node is entitled to set its own price at any point in time: the lower the number of relay nodes the more substantial their market power. Interestingly, the authors consider a situation both with and without communication between the relay nodes. Thus [5] unlike [2] assumes an infrastructure where the relay nodes – in the case where they communicate amongst each other – form an oligopoly and set their prices such that their profits are maximised, which does not necessarily promote resource usage to the largest extent possible as in [2].

According to the credit distribution scheme proposed in [2], nodes acquire credits by providing resources to relay calls from other nodes. Nodes that spend their resources in providing a transit service for traffic originating elsewhere will in return receive a higher resource allocation when they need transit services themselves in the future. Such a protocol ensures *mutual cooperation*. However, one consequence of this mutual cooperation is that nodes located near the edge of the network which attract little transit traffic will earn few credits: these nodes will have a low data rate for originating calls. In contrast, nodes located near the centroid of the network will attract transit traffic and will earn a relatively large amount of credits: these nodes will have a large data rate for originating calls. In this paper, we will investigate various ways of modifying this concept of mutual cooperation by providing nodes near the edge of the network with preferential treatment in order to improve their credit balances and their throughputs.

One shortcoming of [2], admitted by its authors, was that it did not incorporate a model of radio interference. As a result, in the framework of [2] it *always* helps for nodes to be closer together, as this reduces the power needed to transmit data. The reality, however, is that proximity may have a negative effect on connectivity: the increased radio interference may lead to certain nodes in the network becoming disconnected. The paper [6], which can be regarded as a predecessor of [2], does incorporate radio interference, but it does so in a rather partial way.

It may be desirable to allow the nodes to move to different locations in order to improve their utilities expressed, for instance in terms of throughput, or, in the framework of [2], credit balance; we call this "autonomous motion". The idea is that the nodes evaluate the radio conditions in their immediate neighbourhood, and then decide in which direction to move. In the framework of [2], *without* radio interference modelled, one would expect that this would lead to the nodes "clumping together". In a more realistic setting *with* radio interference modelled, it is not *a priori* clear what will happen, as there is a trade-off between the nodes being close together (thus reducing the power needed to transmit data) and the nodes being further apart (thus reducing radio interference and improving connectivity). In this paper we investigate autonomous motion in a model that incorporates radio interference. Simulation experiments reveal that several non-trivial motion patterns appear.

The remainder of this paper is organized as follows: Section 2 presents a summary of the mobile ad hoc network model and describes the radio interference model, the bandwidth and power congestion prices, flow allocation and the bandwidth and power usage. Section 3 investigates several ways for redistributing credits from those nodes that are over-provisioned with an above-average amount of credits to those nodes that have an under-provision of credits. Section 4 presents a model that describes the motion of the nodes and presents insights concerning the effect of radio interference on the motion of the nodes. Conclusions are given in Section 5. An initial version of part of Section 4 has appeared in [7].

2. A summary of the model

This section presents a summary of the features of the mobile ad hoc network model which is based on [2]. We present the model in sufficient detail for the reader to follow the experimental results presented below: we refer to [2] for further background on the model.

2.1. Model basics

Consider an ad hoc network consisting of a set of mobile nodes located on a two-dimensional surface. An origin node transmits to a destination node. If the destination node is beyond the transmission range of the origin node, then the other nodes must cooperate to provide a multi-hop route. This raises the question as to why an intermediate node on the route would expend bandwidth and power in forwarding transit traffic without being compensated? The main idea behind the model of Crowcroft et al. [2] is that transit nodes are rewarded for forwarding traffic, thus creating incentives for collaboration, and the rewards are continuously adjusted such that the network converges towards an operating point which maximises the overall data rates.

The setup of [2] can be summarized as follows: there are two scarce resources: bandwidth and power. Each node has a target bandwidth usage and a target power usage. When a node acts as a transit or destination node, it obtains compensation in the form of *credits* for the congestion costs of the bandwidth and power resources consumed. This provides the basis for mutual cooperation: a node can use its credits (which it receives for making its own resources available to be used by other members of the community) to pay for the bandwidth and power congestion costs incurred when it sends its own traffic. New calls are connected on the least cost routes. The bandwidth and power congestion prices are updated at regular intervals, and are meant to reflect the level of congestion at any specific node along any specific route.

Our model extends the model of Crowcroft et al. [2] by including a representation of radio interference among the nodes, and by allowing the nodes to move in order to improve the values of their utilities.

In order to provide the reader with an insight as to how the system dynamics work, we first briefly sketch in Section 2.2 how the bandwidth prices and bandwidth usages mutually affect each other (similar arguments apply to the power prices and the power usages). A detailed description of the model is then given in Sections 2.3 through 2.9.

2.2. System dynamics

Consider an ad hoc network consisting of nodes 0, 1 and 2 located at coordinates $(25,50)$, $(50,30)$ and $(40,70)$ on a 100×100 square. We consider what happens within one simulated time unit during which three calls arrive. The first call arrives at node 2 at time 0.27, transmits to node 1 and departs at time 0.49; the second call arrives at node 0 at time 0.41, transmits to node 2 and is still present at time 1; the third call arrives at node 2 at time 0.59, transmits to node 0 and is still present at time 1. All calls are directly connected so there is no relay via a third node. The plots in Fig. 1 are read as follows.

- **Time 0.0 to 0.27.** No calls are present and no bandwidth is in use: the price of bandwidth at each node decreases from its initial value of 0.01 and the credit balance at each node is constant.
- **Time 0.27 to 0.41.** The first call is in service and uses bandwidth at nodes 1 and 2. Nodes 1 and 2 were idle so the price of bandwidth is low. The bandwidth in use at nodes 1 and 2 increases rapidly and initially exceeds the target bandwidth usage by a factor of approximately 6. The price of bandwidth at nodes 1 and 2 increases and the bandwidth in use at nodes 1 and 2 quickly converges towards their target bandwidth usages of 10 bandwidth units (see the remark below). The price of bandwidth at the idle node 0 continues to decrease. Node 1 is receiving traffic therefore its credit balance increases. Node 2 is transmitting traffic therefore its credit balance decreases.
- **Time 0.41 to 0.49.** The second call is in service. The bandwidth in use at nodes 0 and 2 increases rapidly. From now onwards nodes 0 and 2 are in use and the price of bandwidth at these nodes increases. The bandwidth in use at nodes 0 and 2 converges to 10 bandwidth units.

Node 2 receives more traffic that it sends (compare the bandwidth used by node 0 – this is the traffic received by

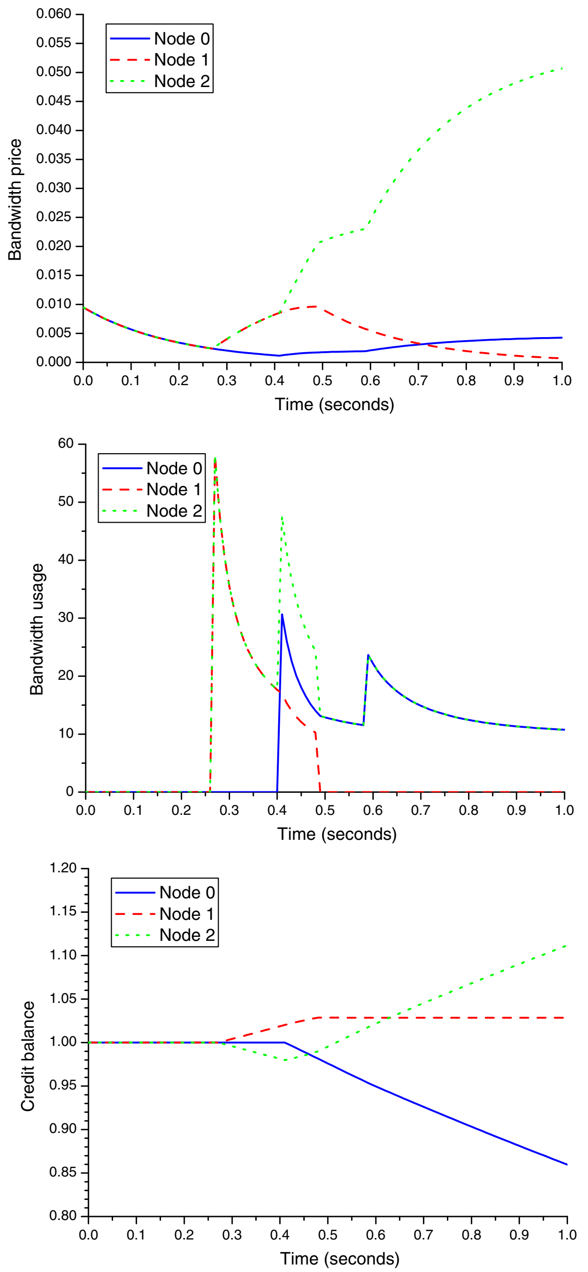


Fig. 1. The system dynamics of a 3-node network as assigned by the model derived from Crowcroft et al. [2].

node 2 – with the bandwidth used by node 1 – this is the traffic sent by node 2): the credit balance at node 2 therefore increases.

- **Time 0.49 to 1.0.** The first call terminates at time 0.49. From now onwards node 1 is idle: the bandwidth in use at nodes 1 and 2 decreases rapidly. Since node 1 is idle, the credit balance at node 1 is constant and the price of bandwidth at node 1 decreases.

The third call arrives at time 0.59 and transmits from node 2 to node 0. The bandwidth in use at nodes 0 and 2 increases rapidly The price of bandwidth at nodes 0 and 2 increases. The bandwidth in use at nodes 0 and 2 converges towards their target bandwidth usages of 10 bandwidth units.

Node 0 transmits at a higher rate than node 2. This is because when node 0 starts transmitting call 2 at time 0.41 it can use all of the bandwidth at node 0, whereas when node 2 starts transmitting call 3 at time 0.59 it can only use that part of the bandwidth at node 2 that is not being used to receive call 2. Node 2 thus receives more traffic than it sends. The credit balance at node 2 therefore increases. Likewise node 0 sends more traffic than it receives. The credit balance at node 0 therefore decreases.

The net shift of credit from node 0 to node 2 places a bound on how long node 0 can maintain the traffic rate of call 2 which significantly exceeds the traffic rate of call 3. To continue transmitting traffic in the long run, node 0 must either earn credits by making more of its bandwidth available to the community, or it must acquire credits by other means, for example by "buying" credits from node 2 and monetarily compensating node 2 for the credits acquired.

- **Remark.** The equations which determine the average bandwidth in use at each node are evaluated at discrete time intervals. The average bandwidth in use can differ from the target bandwidth usage. This effect is exaggerated when solving a small model with a few nodes and a few calls in service. Thus Fig. 1 shows that the average bandwidth in use exceeds the target bandwidth usage for short periods of time. This effect is less evident, though still present, when solving larger models where many calls are in progress. The variance of the flow at each node in a large model is lower so that transitions from idle (zero flow, low bandwidth price) to high utilization (large flow, the bandwidth price adjusts rapidly upwards) are less likely. The model ensures that the average bandwidth in use at each node will converge towards a value that does not exceed the target bandwidth usage.

In summary, the control mechanism is designed such that the nodes are given incentives to efficiently use their resources. The resource prices reflect the current level of congestion: when resources are scarce the prices increase, when resources are abundant the prices decrease.

Having introduced the main ideas behind the model, we now systematically explain its elements in more detail. The reader is referred to the glossary at the end of the paper for a complete list of the notation used.

2.3. The radio interference model

The original model [2] does not contain a description of radio interference. Our model takes radio interference into account as follows.

Node i can reach node j when the signal received by node j from node i is strong enough to be successfully decoded. Consider a call (the "tagged" call) in service at node $i \neq D(r)$ on route r where $D(r)$ denotes the destination node of route r. Let $f_r(i)$ denote the node that node i forwards traffic to on route r. The strength of the signal received at node $j = f_r(i)$ from the tagged call is $p_i y_r \ell(\boldsymbol{z}_i - \boldsymbol{z}_j)$ where p_i is the power radiated per unit flow by the tagged call at node i, y_r is the flow along route r, and the attenuation function is given by $\ell(\boldsymbol{z}_i - \boldsymbol{z}_j) = kd^{-u}$, where $u = 3.52$ is an attenuation factor, $k = 1.82 \times 10^{-14}$ is a scale factor and $d = \|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2$ is the Euclidean distance between $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$.

Let W denote the chip rate of the spreading code and let N_0 denote the power of the thermal background noise. The signal-to-interference ratio is σ_{ri} [10] – see also [8,9,11]

$$\sigma_{ri} = \frac{W}{y_r} \frac{p_i y_r \ell(\boldsymbol{z}_i - \boldsymbol{z}_j)}{N_0 + \eta(A + B)}, \tag{1}$$

where $0 < \eta \leqslant 1$ represents the effect of the radio interference (in terms of the orthogonality of the codes),

$$A = \sum_{k \neq i,j} p_k \ell(\boldsymbol{z}_k - \boldsymbol{z}_j) \sum_{\substack{r \in \mathscr{R}^S(k) \\ f_r(k) = j}} y_r$$

is the interfering signal at node j arising from calls originating at the neighbours k of node j where $\mathscr{R}^S(k)$ is the set of routes that originate at node k, and

$$B = \sum_{k \neq j} p_k \ell(\boldsymbol{z}_k - \boldsymbol{z}_j) \sum_{\substack{r \in \mathscr{R}^T(k) \\ f_r(k) = j}} y_r$$

is the interfering signal at node j arising from calls transiting the neighbours k of node j where $\mathscr{R}^T(k)$ is the set of routes that transit node k. Note that the node antennae are assumed to be directional, and the effectiveness of the orthogonal coding scheme is assumed to be represented by the term η in Eq. (1): node k will therefore contribute to the interfering signal at node j only if $f_r(k) = j$ so that nodes adjacent to node j whose transmissions are not directed at node j do not interfere with the tagged call at node j.

We next describe the relationship between the *effective* (net) transmission rate y_r on route r, and the *actual* (gross) transmission rate Y_{ri} between the nodes i and $j = f_r(i)$ on route r. If the effective rate, at some point in time is y_r, it requires a gross rate of $Y_{ri} = y_r / P(\sigma_{ri})$, where $P(\sigma_{ri})$ denotes the packet success probability. We compute $P(\sigma_{ri})$ as follows [10]. We first determine

$$\overline{P}(\sigma_{ri}) = (1 - 0.5e^{-\sigma_{ri}})^L,$$

where L is the packet size in bits. However, $\overline{P}(\sigma_{ri})$ cannot be the correct packet success probability when $y_r/\overline{P}(\sigma_{ri})$ exceeds W. Taking also into account the fact that probabilities cannot be larger than 1, this leads to the following expression for the packet loss probability

$$P(\sigma_{ri}) = \min(1, \max(y_r/W, \overline{P}(\sigma_{ri}))).$$

We assume that nodes i and j are within transmission range and can reach each other if the packet success probability $P(\sigma_{ri}) \geqslant 0.9$.

2.4. The bandwidth and power congestion prices

The nodes make decentralized decisions on route selection and flow allocation based on *congestion prices* announced by the nodes.

The congestion prices are determined [2] by two first-order linear differential equations (DEs): the first DE describes the price of power and the second DE describes the price of bandwidth. The solutions of the DEs yield prices that are constant when the resource is fully utilized, increase when the resource is over-utilized and decrease when the resource is under-utilized, in line with the principles explained in Section 2.2, thus facilitating the target resource usage. Other DEs could have been used, but we use the DEs presented in [2]. In view of the findings of [4], these DEs are to be preferred since they are designed to perform optimally in terms of maximizing the social welfare, under minimal regularity conditions on the utility functions involved, subject to the restrictions on the target resource usages imposed by the nodes.

We first consider the power congestion price $\mu_j^P(t)$. Let $\gamma_j(t)$ be the power in use at node j at time t and Γ_j be the target power usage at node j. If $\gamma_j(t)$ exceeds Γ_j it means that the current power consumption at node j exceeds the target power usage at node j and the price of power at node j should increase. Similarly, the price of power should decrease when $\gamma_j(t) < \Gamma_j$. This principle can be implemented through the DE

$$\frac{\mathrm{d}}{\mathrm{d}t} \mu_j^P(t) = \kappa \mu_j^P(t) \frac{\gamma_j(t) - \Gamma_j}{\Gamma_j} \tag{2}$$

with initial value $\mu_j^P(0) = 1$ where κ is a constant of dimension seconds^{-1}. Likewise, the bandwidth congestion price $\mu_j^B(t)$ satisfies the DE

$$\frac{\mathrm{d}}{\mathrm{d}t} \mu_j^B(t) = \kappa \mu_j^B(t) \frac{c_j(t) - C_j}{C_j} \tag{3}$$

with initial value $\mu_j^B(0) = 1$ where $c_j(t)$ is the bandwidth in use at node j at time t and C_j is the target bandwidth usage at node j. Expressions for evaluating $\gamma_j(t)$ and $c_j(t)$ are presented in Section 2.9.

The DE for the bandwidth congestion price can be approximately evaluated through

$$\mu_j^B(t + \Delta) \approx \mu_j^B(t) \left(1 + \kappa\Delta \left(\frac{c_j(t) - C_j}{C_j}\right)\right)$$

for some suitably small value of Δ, with a similar expression for $\mu_j^P(t + \Delta)$. The congestion prices are thus adjusted every Δ seconds so that the resource usage at each node converges towards being utilised as much as is possible without exceeding the target resource usage.

2.5. The route prices

Once the prices of the resources at the nodes are known, the price for sending a unit of flow along a particular route can be determined.

Let e_{ij}^{tx} denote the power used to transmit a unit flow from node i to node j in the absence of radio interference. We assume that e_{ij}^{tx} is an increasing function of the distance between the positions $\boldsymbol{z}_i, \boldsymbol{z}_j$ of nodes i and j, and also there is a certain minimum power needed irrespective of the positions [10]. We therefore choose

$$e_{ij}^{\mathrm{tx}} = \max(10^{-2}, 10^{-4} \|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2^2),$$

which is a non-zero function in the vicinity of the transmitting node i. If node i cannot reach node j then we set $e_{ij}^{tx} = \infty$. Crowcroft et al. [2] specify $e_{ij}^{tx} = 10^{-4}\|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2^{1/2}$. Let e^{rx} denote the power used to receive a unit flow.

The price $\mu_{rj}(t)$ that node j charges the originating node $O(r)$ of route r for processing a unit of flow along route r is computed as follows. If node j is the originating node of route r then bandwidth and power resources are used at node j for transmitting to node k where $k = f_r(j)$ so that

$$\mu_{rj}(t) = \frac{e_{jk}^{tx}}{P(\sigma_{rj})}\mu_j^P(t) + \frac{1}{P(\sigma_{rj})}\mu_j^B(t).$$

If node j is a transit node of route r then bandwidth and power resources are used at node j for receiving from the node preceding node j and for transmitting to node k where $k = f_r(j)$ so that

$$\mu_{rj}(t) = \left(e^{rx} + \frac{e_{jk}^{tx}}{P(\sigma_{rj})}\right)\mu_j^P(t) + \left(1 + \frac{1}{P(\sigma_{rj})}\right)\mu_j^B(t).$$

If node j is the destination node of route r then bandwidth and power resources are used at node j for receiving from the node preceding node j so that

$$\mu_{rj}(t) = e^{rx}\mu_j^P(t) + \mu_j^B(t).$$

The congestion prices of the resources consumed in transmission are inflated by a factor $1/P(\sigma_{rj})$ to model the resources consumed by the transmission of errored packets. It is assumed that only correct packets are received.

2.6. The credit balance

The nodes are given an incentive to provide resources for forwarding transit traffic and for receiving traffic: by doing so they earn credits which they in turn require to send their own data. In this subsection, we describe how the credit balance evolves as a function of time.

Each node s maintains a *credit balance* $b_s(t)$ with an initial value of 1. The credit balances are modelled as continuous variables. The credit balance of node s is adjusted according to the following three principles

(i) The node s spends credits $\sum_{r\in\mathscr{R}^S(s)} y_r(t)\mu_r(t)$ for the congestion costs (bandwidth, power) incurred in transmitting its own traffic through the source, transit and destination nodes on its outbound routes, where $\mu_r(t)$ is the sum of the prices charged by the nodes along route r

$$\mu_r(t) = \sum_{j\in r}\mu_{rj}(t). \tag{4}$$

(ii) The node s receives credits $\sum_{r:s\in r} y_r(t)\mu_{rs}(t)$ for the congestion costs (bandwidth, power) incurred in acting as a source node, transit node, or a destination node. Note that the route prices also include the costs incurred at the originating node. These costs do not cause credits to be transferred since the originating node pays itself, reflecting the fact that the resources at the originating node may be scarce.

(iii) For each node s, that part of the credit balance that differs from the average value of 1 is discounted using a factor $\beta = 0.01$ (say) where β has dimension seconds^{-1}. Thus over one second, the under-provisioned nodes that possess a credit balance of less than 1 receive 1% of the credits that they lack and the over-provisioned nodes that possess a credit balance larger than 1 surrender 1% of that part of their balance that exceeds 1.

Note that credit discounting allows under-provisioned nodes to increase their credit balance at the expense of over-provisioned nodes. In practice, credit discounting should be organized such that it does not undermine the incentive system which encourages over-provisioned nodes to make their resources available to the network. This can be achieved by monetarily charging under-provisioned nodes for creating credit, such charges being accumulated in a fund which is used to monetarily compensate over-provisioned nodes for destroying credit. This type of monetary compensation is discussed in Section 3.5.

We now present a DE for $b_s(t)$ which satisfies the three principles mentioned above

$$\frac{d}{dt}b_s(t) = -\beta(b_s(t) - 1) + \Omega_s(t), \tag{5}$$

where the reimbursement

$$\Omega_s(t) = \sum_{r:s\in r} y_r(t)\mu_{rs}(t) - \sum_{r\in\mathscr{R}^S(s)} y_r(t)\mu_r(t). \tag{6}$$

With this definition, the credit balance $b_s(t)$ is affected by credit discounting (the term $-\beta(b_s(t) - 1)$ which increases or decreases the credit balance at node s depending on whether the node s is under- or over-provisioned) and by the reimbursement $\Omega_s(t)$ which represents the credits earned by node s by forwarding traffic from other nodes as well as the credits spent by node s in transmitting its own traffic.

The DE (5) can be approximately evaluated through

$$b_s(t+\Delta) \approx b_s(t) - \beta\Delta(b_s(t) - 1) + \Delta\Omega_s(t). \tag{7}$$

In the remainder of this paper the term "reimbursement" refers to credit allocation via $\Omega_s(t)$ defined in Eq. (6) and the term "redistribution" refers to credit allocation via $\beta\Delta(b_s(t) - 1)$ as in Eq. (7). Note that the credits transferred via Eqs. (6) and (7) take place not necessarily in integral units.

If the credits earned are equal to the credits spent so that $\sum_s \Omega_s(t) = 0$, then the total credit balance is equal to the number of nodes N. This can readily be shown by induction over t. Recall that $b_s(0) = 1$ for all s so that $\sum_s b_s(0) = N$. Next suppose that $\sum_s b_s(t) = N$. Then

$$\begin{aligned}\sum_s b_s(t+\Delta) &= \sum_s b_s(t) - \beta\Delta\sum_s(b_s(t) - 1) + \Delta\sum_s\Omega_s(t)\\ &= \sum_s b_s(t) = N,\end{aligned}$$

which completes the proof.

Note that in [2], the term $\sum_{r\in\mathscr{R}^S(s)} y_r(t)\mu_r(t)$ in Eq. (6) which denotes the credits spent by node s in transmitting its own traffic along route r, is replaced by $\alpha_s b_s(t)$ where $0 < \alpha_s \leqslant 1$ so that the credit balance satisfies the DE

$$\frac{\mathrm{d}}{\mathrm{d}t} b_s(t) = -\beta(b_s(t) - 1) - \alpha_s b_s(t) + \sum_{r:s \in r} y_r(t)\mu_{rs}(t). \tag{8}$$

However, when radio interference (Section 2.3) and autonomous motion (Section 4) are modelled, Eq. (8) is no longer valid. For example, if at least one node has moved, then Eq. (8) describes the change incurred in the credits received namely $\sum_{r:s\in r} y_r(t)\mu_{rs}(t)$ yet Eq. (8) assumes that the credits spent remain what they were before node s moved, namely $\alpha_s b_s(t)$. The incorrect computation of the amount of credits spent has two consequences. First, the total credit balance in the network is no longer constant, and second, credit balance optimisation will maximize the credits received but not the credits spent.

2.7. The willingness-to-pay

Each node s determines its resource usage according to its *willingness-to-pay* $w_s(t)$ at time t for the congestion costs incurred in sending its traffic. The willingness-to-pay $w_s(t)$ has dimension seconds^{-1}.

As in [2], we assume that the willingness-to-pay is a fixed share of the total credits available so that $w_s(t) = \alpha_s b_s(t)$ where $0 < \alpha_s \leqslant 1/\Delta$. Observe that $\alpha_s > 1$ is feasible. The credit spent by node s for transmitting data over an interval of time Δ is thus $\alpha_s b_s(t)\Delta$. The maximum value $\alpha_s = 1/\Delta$ represents the extreme case where all credits available are spent during a time slice Δ.

2.8. Flow allocation

A call between an originating node and a destination node is connected on the least cost route connecting these two nodes. Once the route prices are known, the least cost route can be determined. This is done as follows.

A call originating at node s at time t is connected to a randomly selected destination node d along the least cost route $r^\star$ where

$$r^\star = \arg \min_{\substack{r \in \mathscr{R}^S(s) \\ D(r)=d}} \sum_{j \in r} \mu_{rj}(t). \tag{9}$$

The routes are assigned unique labels and are lexicographically sorted according to a collation sequence: ties in Eq. (9) are resolved lexicographically.

The total flow $x_s(t)$ generated by node s at time t, assuming that the willingness-to-pay $w_s(t)$ is equally distributed among the $N_s(t) > 0$ calls originating at node s at time t, is

$$x_s(t) = \sum_{r \in \mathscr{R}^S(s)} y_r(t) = \frac{1}{N_s(t)} \sum_{r^\star \in \mathscr{R}^S(s)} \frac{w_s(t)}{\mu_{r^\star}(t)}, \tag{10}$$

where $y_r(t) > 0$ only on those routes $r^\star$ that attain a minimum in Eq. (9).

Note that the route flows $y_r(t)$ at time t are no longer decision variables to be determined directly. Instead, as we see from Eq. (10), the total flows $x_s(t)$ are derived from the prices $\mu_{rj}(t)$ that themselves depend on the congestion price decision variables $\mu_j^P(t)$ and $\mu_j^B(t)$ which are obtained as solutions to Eqs. (2) and (3).

2.9. The bandwidth and power usage

Once the routes used by the calls in service are known, the sets of routes $\mathscr{R}^S(j), \mathscr{R}^D(j)$ and $\mathscr{R}^T(j)$ which originate, terminate and transit node j can be updated. The bandwidth $c_j(t)$ used at node j at time t is given by

$$c_j(t) = \sum_{r \in \mathscr{R}^S(j)} \frac{y_r(t)}{P(\sigma_{rj})} + \sum_{r \in \mathscr{R}^D(j)} y_r(t) + \sum_{r \in \mathscr{R}^T(j): j \in r} \left(1 + \frac{1}{P(\sigma_{rj})}\right) y_r(t). \tag{11}$$

Note that the transmitted flows are inflated by a factor $1/P(\sigma_{rj})$. It is assumed that no packets are lost when the flow is received. The factor $1 + 1/P(\sigma_{rj})$ (which is larger than 2) accounts for the fact that at any transit node along the route, the flow has to be both received and forwarded.

The power $\gamma_j(t)$ used at node j at time t is given by

$$\gamma_j(t) = \sum_{r \in \mathscr{R}^S(j)} \frac{e_{jk}^{\mathrm{tx}}}{P(\sigma_{rj})} y_r(t) + \sum_{r \in \mathscr{R}^D(j)} y_r(t) e^{\mathrm{rx}} + \sum_{r \in \mathscr{R}^T(j): j \in r} \left(e^{\mathrm{rx}} + \frac{e_{jk}^{\mathrm{tx}}}{P(\sigma_{rj})}\right) y_r(t), \tag{12}$$

where $k = f_r(j)$. As was the case for the bandwidth usage, the power usage relates to gross rates, and therefore the outgoing flows are inflated.

The numerical evaluation of Eq. (10) can yield flow values $y_r(t)$ which, when used in Eqs. (11) and (12), cause the total bandwidth $c_j(t)$ and the total power $\gamma_j(t)$ to differ from their target usage values of C_j and Γ_j, respectively. However, the model ensures that the total bandwidth in use and the total power in use at each node converge towards values that do not exceed the target values. The rate at which the resource usages converge towards their target values depends upon the time interval Δ between successive evaluations of the DEs which determine the congestion prices and the credit balances at the nodes. The interval Δ should be chosen sufficiently small such that the discrete approximations to the DE's (2), (3) and (5) provide accurate solutions. Following [2] we use $\Delta = 0.01$ seconds.

2.10. The simulation model

The simulator uses the equations described in Section 2.3 through 2.9 above. The simulation proceeds as a sequence of steps. In each step, for those nodes which currently originate calls, the simulator computes the node prices, which determine the route prices; the credit balances are updated and the willingness-to-spend is computed; the new flow allocations are computed; the resources in use are computed, and the next simulation step begins. Additional updates take place when calls enter or leave the simulated network.

We developed the simulator in Java using the DESMO-J simulation framework [12]. The simulation model computes Eq. (1) through (12) and is parametrised as follows:

– each node originates one call at a time so that $N_s(t) = 1$
– each call selects a destination node at random
  - the call is connected on the route with the lowest (at the instant of the call initiation) costs
  - the route is used for the duration of the call: calls are not rerouted, which implies low signalling costs and no route flap
– the call holding times and the call idle times are exponentially distributed, mean 0.5 s
– the prices are updated every $\Delta = 0.01$ s
– the simulation lasts for 100,000 s.

Note that the order in which Eq. (1) through (12) are presented corresponds to the order in which these quantities are computed every Δ seconds in the simulator.

2.11. Implementation issues

A protocol to implement the incentive-based control was presented in [13]. In this regard we observe that the parameter values required for the calculation of Eq. (1) through (12), apart from Eq. (4), are local in the sense that any quantity computed for node i is expressed in terms of the parameters of node i and its neighbours $j = f_r(i)$.

The calculation of Eq. (4) requires access to non-local information. However, Eq. (4) can be evaluated if a distance vector routing protocol is used to communicate at regular intervals to each node an estimate of the route costs from each of its neighbours to all destinations.

A practical implementation would also require the average call holding time $1/\mu$ to be of the order of several minutes so that the calculation of the congestion prices, flows and resources consumed would occur relatively infrequently at intervals of $O(\Delta/\mu)$ and would therefore place a small computational burden on the nodes.

3. Credit redistribution

According to the credit reimbursement scheme proposed in [2], nodes acquire credits by providing resources to relay calls from other nodes. The nodes then use their credits to pay for the cost of sending their own traffic. Such a protocol ensures mutual cooperation. One consequence of this mutual cooperation is that nodes at the edge of the network which attract little transit traffic will earn few credits from transporting transit calls. The data rates for traffic originating from such nodes will be low. Nodes at the centre of the network will attract a relatively large amount of transit traffic and will earn credits from transporting transit calls. The data rates for traffic originating from such nodes will be larger.

However, one or more nodes may need to increase their data rate(s) beyond that afforded by the resources that its current credit balance allows it to acquire. This might in particular hold for nodes located at the periphery of the network that attract only a limited number of transit calls and thus have a limited credit balance. Such nodes, though relatively isolated from the rest of the network, may have urgent data to transmit and they should be allowed to send their data at a rate such that their data transmissions complete in their allotted time.

Credit reimbursement as given by Eq. (6) is therefore augmented [2] by a mechanism for redistributing credits from those nodes that are over-provisioned with an above-average $b_s(t) > 1$ amount of credits to those nodes that are under-provisioned with a below-average $b_s(t) < 1$ amount of credits.

The following sections investigate various methods for improving the credit redistribution mechanism. We distinguish between the *global* redistribution of credits from all nodes with $b_s(t) > 1$ to all nodes with $b_s(t) < 1$, and the *local* redistribution of credits where an under-provisioned node with $b_s(t) < 1$ receives credits from one or more of its neighbouring nodes that are over-provisioned.

Note that any credit redistribution, be it global or local, requires a mechanism that compensates those nodes that surrender their credits for the benefit of others, and penalizes those nodes that replenish their credits at the expense of others. If this were not done then well-provisioned nodes, for example, would lack an incentive not to spent all their credits before their credits were removed by redistribution: see the discussions in Sections 2.6 and 3.5.

3.1. The original global credit redistribution model

In this section, we examine credit redistribution as presented in [2] but modified via Eq. (6) to take the effect of motion and radio interference into account.

Consider Eq. (7) which is used to compute the credit balance every Δ seconds. The term $\beta(b_s(t) - 1)$ in Eqs. (5) and (7) models the global redistribution of credits at node s. If $b_s(t)$ exceeds the average value of the node credit balance (the average value is 1), then the credit balance at node s will be decreased by an amount that is proportional to that part of the credit balance that exceeds 1. This is interpreted as a transfer (redistribution) of credits from node s to other nodes. Similarly, if $b_s(t)$ is less than 1, then the credit balance at node s will be increased, which we interpret as a transfer (redistribution) of credits to node s from other nodes. Credit redistribution thus allows a node to acquire/release credits in addition to those credits that were earned/spent via the reimbursement mechanism.

This approach towards global redistribution does scale even though an increase/decrease in the credit balance of one node corresponds to an decrease/increase in the credit balances of one or more nodes that are possibly located in a distant part of the network. No mechanism is required to transfer the redistributed credit. A node that has a surplus of credits will, over a period of time Δ, *destroy* a fraction β of that surplus. Likewise a node that has an under-supply of credits will, over a period of time Δ, *create* credits corresponding to a β fraction of its deficit.

Note that for $\beta > 0$, the credit balance at each node s is bounded below by $b_s(t) \geqslant \beta/(\alpha + \beta)$ where $\alpha = \alpha_s$ for all node indices s. This positive minimum credit balance is realised when a node s earns no transit revenue, in which case the reimbursement becomes

$$\Omega_s(t) = -\sum_{r \in \mathscr{R}^S(s)} y_r(t)\mu_r(t) = -\alpha b_s(t),$$

since $y_r(t) = \alpha b_s(t)/\mu_r(t)$. In general, when a node s earns transit revenue, we have $\Omega_s(t) \geqslant -\alpha b_s(t)$. If the reimbursement $\Omega_s(t)$ is equal to the credit balance redistributed $\beta(b_s(t)-1)$ then $\beta(b_s(t)-1) \geqslant -\alpha b_s(t)$ so that $b_s(t) \geqslant \beta/(\alpha+\beta)$.

Global redistribution thus ensures that a minimum amount of credits is available at each node, including those nodes at the edge of the network that are unable to attract transit traffic. Global redistribution also ensures that credits cannot be permanently accumulated by nodes that are well-off. This can be interpreted as the credits expiring at each node s in the sense that using resources at other nodes along routes $r \in \mathscr{R}^S(s)$ in return for services granted to the routes that transit node s cannot be postponed indefinitely. The total credit balance of all the nodes in the network remains equal to the number of nodes in the network even if nodes leave the network (in general the credit balance of a departing node is not equal to 1) and/or if nodes enter the network with an arbitrary initial credit endowment.

We now investigate three value-ranges for β.

- $\beta = 0$ implies that credits are allocated by reimbursement and no credit redistribution takes place. In this case Eq. (7) yields

$$b_s(t+\Delta) = b_s(t) + \Delta\Omega_s(t), \tag{13}$$

  and the nodes acquire credits only by providing resources to relay the calls from other nodes.
- $0 < \beta < 1/\Delta$ implies that credits are allocated both by reimbursement and by credit redistribution. Thus [2] suggests to discount that part of the credit balance that differs from 1 using $\beta = 0.01$, so that over one second, under-provisioned nodes receive 1% of the credits that they lack in comparison to an average node's credit balance (a balance of 1), see Eq. (7). Correspondingly, nodes that possess a credit balance larger than 1 surrender 1% of that part of their balance that exceeds 1.Increasing the value of β increases the amount of the credit balance discounted, so that the value $(1-\beta\Delta)b_s(t)$ of the credits earned in the past becomes less important. This decreases the relative importance of the reimbursement scheme over credit redistribution, since the advantageous (disadvantageous) treatment of nodes that in the past have allocated a high (low) share of their resources to other nodes' calls is further reduced due to the increasing amount of credits being transferred by redistribution.
- $\beta = 1/\Delta$ implies that credits are allocated primarily by redistribution and almost no reimbursement takes place. In this case Eq. (7) yields

$$b_s(t+\Delta) = 1 + \Delta\Omega_s(t)$$

  so that the credit allocation becomes "memoryless". The credit balance available for spending during the next period depends on what was earned/paid during the current period: any credit balance history preceding the current period is irrelevant. Note that for Δ sufficiently small, the credit balance $b_s(t+\Delta) = 1 + \Delta\Omega_s(t) \approx 1$ closely approximates the case where the credit balance at each node is reset to 1 at the start of each update period.

Fig. 2 shows the relationship between the value of β and the average credit balance and the data transmission rates of four nodes in the 10-node network presented in Fig. 3. The simulation executes for 100,000 time units during which more than 700,000 calls arrive to the network model. The congestion prices are adjusted every $\Delta = 0.01$ s.

As β increases, the credit balance at each node converges towards 1. This does not imply that the data rates of the originating traffic at all the nodes become identical. A node in a central position like node 8 is closer to an average destination node than a node at the periphery like node 1. This yields a lower cost per unit flow at node 8 due to the relatively lower cost of carrying calls on shorter routes. In addition, nodes with one or more nodes in close proximity (like node 7) benefit from an increased credit balance to a higher degree than the average under-provisioned node since the data rates of such short distance calls are disproportionately highly increased. Thus node 7 is able to relatively cheaply increase its (average) data rate of originating calls. However, increasing the value of β yields only a small improvement in the data rate at node 1.

The lower plot of Fig. 2 shows that $\beta \approx 0.1$ yields the highest average total data rate of 27.6 units/second. Thus throughput maximisation requires (slightly) increasing the flow allocation model's emphasis on reimbursement by reducing the value $(1-\beta\Delta)b_s(t)$ of the credits earned in the past.

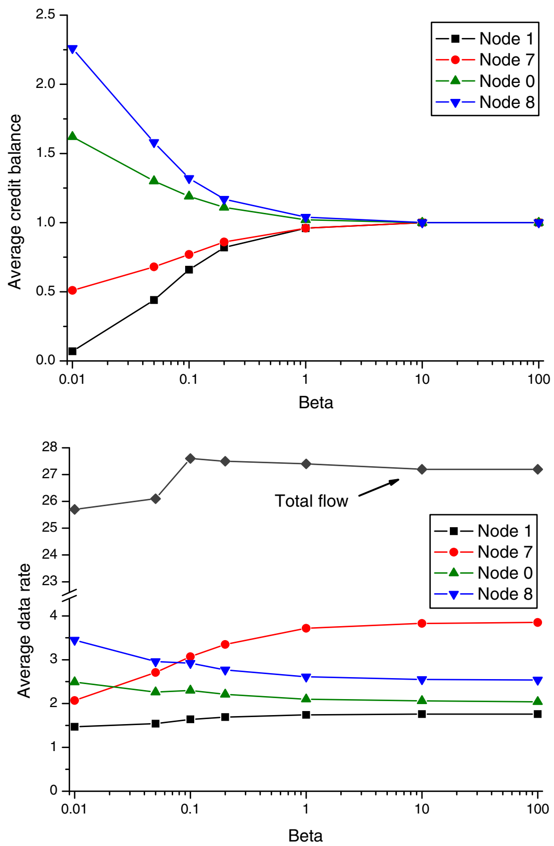


Fig. 2. Global redistribution [2] versus β.

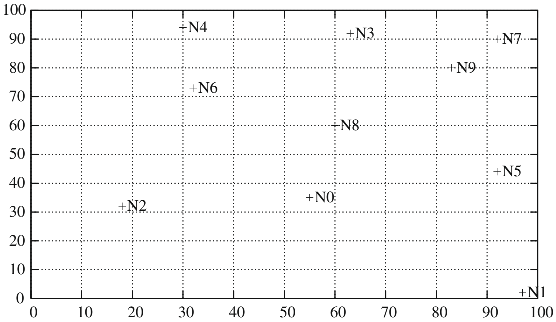


Fig. 3. The geographic topology of the 10-node network [2].

3.2. A modified global credit redistribution model

The above experiments significantly change the credit and the resource allocation (for large values of β) yet the data rate at node 1 is increased only slightly and remains significantly smaller than the data rates at the other nodes.

This raises the question as to whether it is possible to *either* further improve the data rate at node 1 by increasing the discount factor of node 1 while leaving the discount factors of the other nodes unchanged, *or* to at least achieve similar results in terms of the data rate at node 1 by less drastic changes to the credit and resource allocation. The following approaches were investigated.

Adjusting the discount factor at node 1. Here the discount factor β of all the nodes remains fixed at the default value 0.01 apart from node 1 where β is varied in the range $0.01 \leqslant \beta \leqslant 1/\Delta$. In this case, the nodes earning credits do not lose the amount gained as quickly as in the experiments presented in Section 3.1 above (these nodes now have a constant β-value of 0.01 so that credits are allocated primarily by reimbursement) yet the credit balance at the under-provisioned node 1 (this node now has a larger β-value) is replenished as rapidly as in the experiments above.

Note that this update policy violates the model's property that the total credit balance remains equal to the number of nodes in the network: the amounts by which discounting increases or decreases the credit balances are no longer necessarily zero-sum. The increase in the total credit balance can be regarded as "inflation": the increased amount of credits lowers the value of the credits.

Fig. 4 shows that in comparison to the original experiment, the central nodes 0 and 8 are now better off; under-provisioned peripheral nodes like node 7 no longer benefit from an increased β at the expense of central nodes. However, the data rate at node 1 remains more or less the same as before: node 1 spends its credits to pay its relay nodes (provided these credits are no longer quickly removed by a high β). Inflation increases the congestion prices since more credits are available in total, with the undesirable result that the increased amount of credits that node 1 possesses cannot be used to acquire additional resources.

Adjusting the target credit balance at node 1. In order to permanently increase the credits available to node 1, we

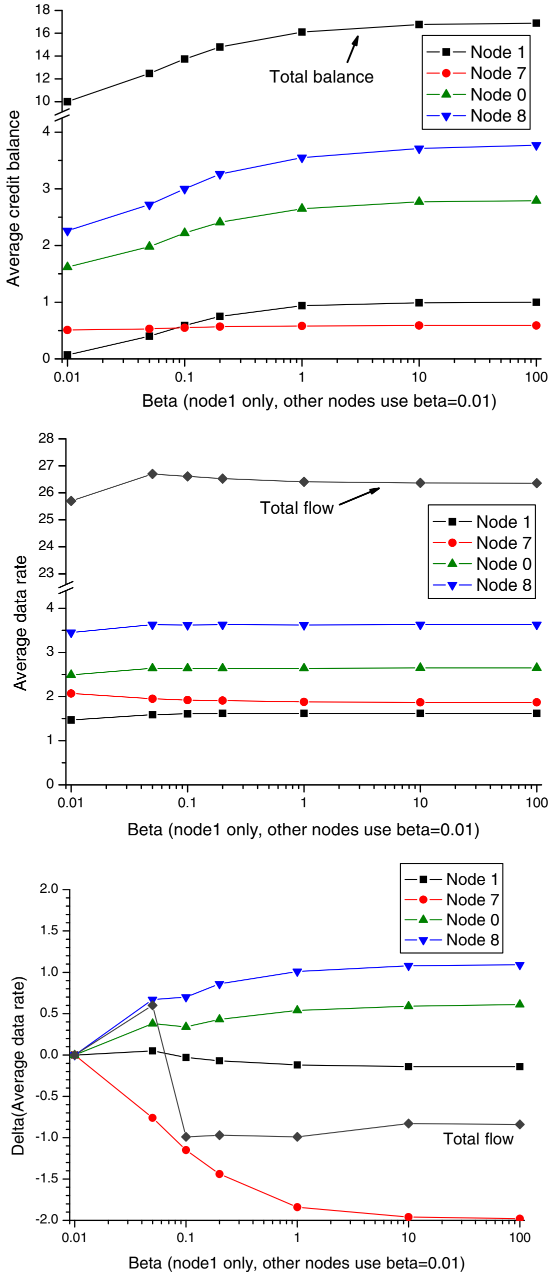


Fig. 4. Global redistribution versus β of node 1.

introduce a target credit balance k_s for each node s such that the target balance is not necessarily equal to 1. Eqs. (5) and (7) become

$$\frac{\mathrm{d}b_s(t)}{\mathrm{d}t} = -\beta(b_s(t) - k_s) + \Omega_s(t),$$

and

$$b_s(t+\Delta) = b_s(t) + \Delta\Omega_s(t) - \beta\Delta(b_s(t) - k_s).$$

This approach leaves the total credit balance $\sum_s k_s$ constant and so avoids inflation. Fig. 5 shows the results for

$$k_s = \begin{cases} 2 & s = 1, \\ 1 & \text{otherwise,} \end{cases}$$

which doubles the target credit balance at node 1.

Fig. 5 shows a limited improvement in the data rate at node 1, especially if β is set sufficiently large that the credit balance at node 1 reaches a value close to 2. The reasons for the modest improvement in the data rate at node 1 are the increased call traffic demand (yielding increased congestion prices, so that doubling the value of the willingness-to-pay does not result in twice the resources being allocated to a call), and the overall resource restrictions.

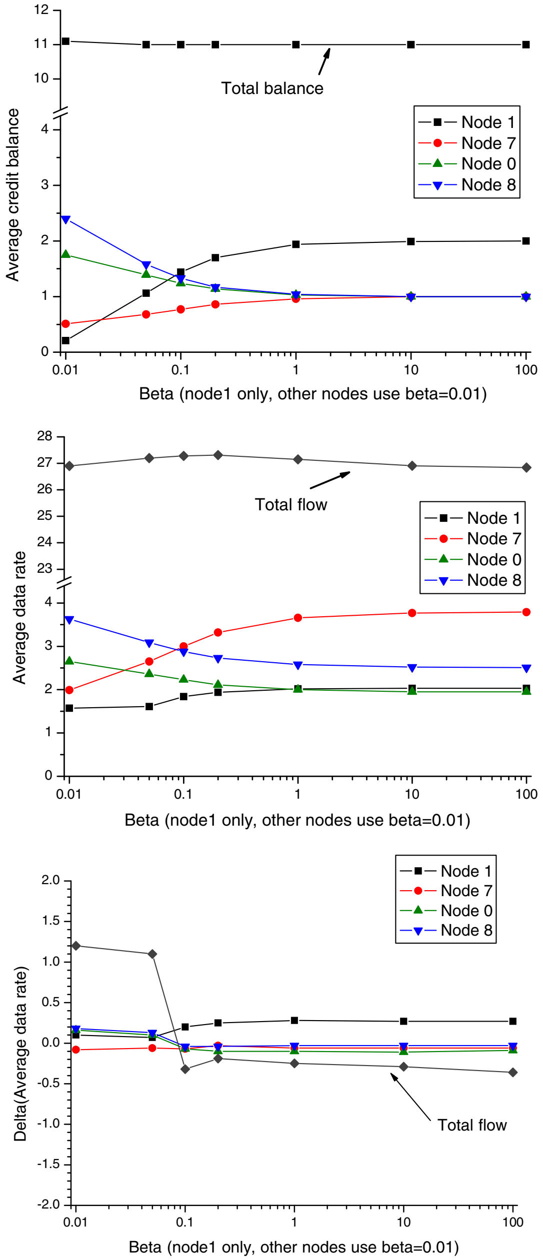


Fig. 5. Global redistribution with the target credit balance of node 1 set to 2.

Since all the nodes adjacent to node 1 are relatively far away, call traffic is limited by the power available – another experiment using $k_1 = 10$ yielded only a minor increase in resources allocated.

Further suggestions. The willingness-to-pay $w_s(t)$ is given by $w_s(t) = \alpha_s b_s(t)$. We have not investigated the effect of modifying the value of the α_s of a subset of the nodes. We do not expect that changing the value of α_s will yield significantly different results than in the last experiment (which yielded the highest data rate possible for node 1) since the central idea of both attempts is to linearly scale the willingness-to-pay at node 1 – whether this is done by doubling the credits available or by doubling the share of the (not doubled) amount of credits should yield similar results.

Finally, it is also possible to continuously transfer a fixed amount of credits to node 1. This approach was not investigated since constant credit transfers do not adapt to the current state of node 1 and of the network in general. Thus node 1 could potentially accumulate a large amount of credits if no call is in service for a while, and this would cause inflation if the nodes which contribute credit to node 1 are not identified.

3.3. Local credit redistribution

Local credit redistribution limits the under-provisioned nodes to acquiring credits from nodes within their transmission range only. For the purpose of credit balance updates, β in Eq. (7) is set to zero, again yielding Eq. (13) where global credit redistribution is disabled. However, once the credit balance of a node s drops below a given threshold $0 < b < 1$, the surplus credits (a node is deemed to have a surplus of credits if its credit balance exceeds 1) of an adjacent node (or more than one node, if need be) are transferred to node s until either node s reaches a credit balance of 1, or no adjacent node disposes of a surplus. For any node this yields a "social cluster" (defined by adjacency) of nodes that can provide credits as the need arises.

In a typical (connected) network, these social clusters are not disjoint, so that the credits transferred to an adjacent node might be further transferred to more distant nodes in the event that the first receiver is able to build up a surplus of credits. However, successive credit transfers eventually yielding a dispersion of credits throughout the network are unlikely. A node that is under-provisioned due to its disadvantageous location will typically, yet not necessarily, spend its (locally transferred) credits for call transmission and will soon become under-provisioned again; thus the node is unable to further transfer the credits received to other nodes. This mechanism of local redistribution scales since only adjacent nodes need communicate in the event of a shortage of credits.

Since nodes may be forced to surrender a significant share of their credit balance, the concept of mutual cooperation is relaxed. A node that is repeatedly forced to surrender a large portion of its surplus credits to its under-provisioned neighbours cannot attain a high data rate for its originated calls. We therefore impose a weaker definition of mutual cooperation which preserves the order in which the adjacent nodes are forced to release their surplus credits. Each attempt to acquire credits will commence at the last node that was not required to completely release its surplus. Therefore, any node will not have to surrender its entire surplus twice in succession unless all the other neighbours did the same. Moreover, the order in which all the nodes (that potentially compete for the surplus from the same nodes) scan their credit balance and initiate local credit redistribution if need be is randomised at each update interval Δ: thus no node has the advantage of always acquiring credits first.

Similar to Fig. 2, Fig. 6 shows the effect of different values of the credit threshold b on the average credit balance and the data transmission rate at the same four nodes. The figure shows that the effects of local and global credit redistribution are approximately the same. As b increases, the average credit balance converges towards 1, which improves the data rates of the previously under-provisioned nodes at the expense of the over-provisioned neighbours. Due to the influence of the topology (the distance to the neighbouring nodes, the average distance to all nodes) on the price a node has to pay for originating calls, the data rates do not converge.

The experiment indicates that the total flow is robust with respect to the choice of b: for $b \geqslant 0.5$ the total flow is almost constant, with the greatest total flow (27.4 units/second) obtained at $b = 0.7$.

Note that credit redistribution allows any node to (temporarily) increase its data rate. Such an approach is not restricted to nodes with small data rates: credits can be locally redistributed so that a well-provisioned node that has a pressing need to transmit as quickly as possible can be provided with even more credits. However, local credit redistribution is most likely to have a significant impact on data rates when applied to under-provisioned nodes since the willingness to pay of a well-provisioned node might already dominate any other demand for resources: the well-provisioned node may already be receiving (almost) all the resources available.

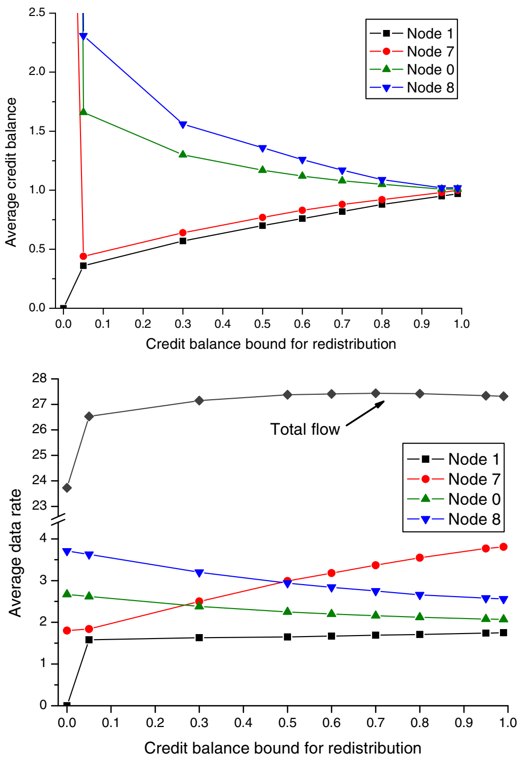


Fig. 6. Local redistribution versus the credit threshold b.

3.4. A comparison of global versus local redistribution

The experiments show that from both the viewpoints of enabling peripheral nodes to send call traffic as well as the total network data rate maximisation (subject to highest possible resource utilisation) it is beneficial to augment the reimbursement scheme with the additional mechanism of global or local credit redistribution.

Global redistribution as suggested by Ref. [2], using a value of $\beta = 0.01$, provides a reasonable trade-off between reimbursement whereby a node is compensated for making its resources available to the community, and redistribution which allows an under-provisioned node to transmit traffic.

Credit redistribution enables higher total data rates. Thus redistribution serves another purpose in addition to the purpose intended by the authors of [2] who proposed redistribution as a mechanism for keeping the total amount of credits equal to the number of nodes in the network in the event of nodes entering and/or leaving the network.

For the network model under investigation, global redistribution yields slightly better total data rates than local redistribution. Global redistribution should thus be preferred unless it is necessary to identify pairs of nodes involved in any credit transfer, for example to enable nodes to monetarily compensate each other for credits moved, which would imply centralised control in the case of global redistribution. Preliminary experiments with both global and local redistribution enabled yielded no significant improvement in comparison to conducting global redistribution only.

The inferior performance of local redistribution is due to the fact that an under-provisioned node must acquire credits from nodes with a credit surplus within its connectivity range only. This restriction might even cause local redistribution to fail in network topologies where under-provisioned nodes are locally clustered. Such nodes will not be able to locally acquire bandwidth while the surplus credit balance of nodes clustered at the centre (adjacent to other nodes that are able to build up a surplus only) cannot be acquired.

In the 10-node network topology, node 8 (that has the highest credit surplus) as well as nodes 5 and 0 are located within range of the peripheral node 1, while node 9 cannot be reached from node 1. Therefore node 5, although originally building up a higher surplus than node 9, has to transfer credits to node 1 and is thus worse-off than node 9. Therefore node 9 benefits from the under-provision of credits at node 1 when competing with node 5 for the same resources such as using node 8 as transit node. This can be explained as follows. Since node 5 has to transfer credits to node 1, its credit balance is diminished. Node 5 has a correspondingly low data rate and receives a correspondingly low resource allocation from the transit node 8. This yields a higher resource allocation to other nodes at node 8, which benefits node 9 and the data rate at node 9 improves. Regardless of what the data rate at node 9 was before, now that node 5 has fewer credits to spend, the data rate at node 9 increases.

Global redistribution avoids such a biased treatment (biased in the sense that local redistribution penalises well-provisioned nodes that are adjacent to under-provisioned nodes) since the share of the surplus credit removed is the same for all nodes able to obtain a surplus. Global redistribution will minimise perturbations to the competition among nodes with surplus credits.

3.5. Cooperative behaviour and compensation

The issue of trust is not addressed in this paper. In this regard, a security mechanism such as in [14] is needed to ensure that nodes cannot create credits without corresponding monetary payments which accrue to a fund. Likewise nodes that destroy credits are monetarily compensated from the fund.

The scheme is stable. Assume for example that all nodes increase their credit balances simultaneously. All nodes contribute to the fund in return for the creation of credits. Since the credit balances at all the nodes increase, no node will receive a significant better (or worse) service: each node's willingness to pay increases and so do the resource prices, so that no node can significantly improve its data rate. Over time, global redistribution will slowly remove the credits that every node possesses in excess of the target credit balance so that the credit balances converge towards the level that prevailed before each node increased its balance.

If monetary compensation is used, then when initially increasing their credit balances the nodes contribute to the fund in return for the creation of credits. Conversely, as credits are removed from the system due to global redistribution, the nodes are compensated from the fund, which yields zero-sum overall payments.

A node may attempt to create more credits (cheat) by increasing its credit balance $b_s(t)$ at some discrete time instants t. In addition, an under-provisioned node can accelerate global credit redistribution by continuously applying a node-specific discount factor β_j larger than β; likewise an over-provisioned node may increase its credit balance by using a discount factor smaller than β to delay redistribution. Note that such behaviour will cause the total credit balance to no longer be equal to the number of nodes.

Such an increase in the total credit balance can be regarded as inflation: the increased amount of credits lowers the value of the credit. However, if *all* the nodes increase their balance $b_s(t)$ by applying the same increased discount factor β, the prices will be higher, yet each node will receive the same amount of resources, which corresponds to inflation. Eventually the total amount of credits in the system (and thus the prices) decreases due to global redistribution and the total balance converges to the number of nodes. To receive preferential treatment when competing with other nodes for access to resources, a node must be the only cheat, or cheat by a significantly larger margin than all its competitors. Cheating must be repeated continuously for a node to receive a permanent advantage over other nodes. Local redistribution can be viewed as a form of cheating. In this case, inflation is avoided since credits are moved, not created, so that the total amount of credits in the network remains constant.

Note that both local and global credit redistribution require some form of compensation to provide an incentive for over-provisioned nodes to take part in the network. This compensation can take the form of monetary payments.

Monetary payments could be conducted as follows: credits transferred among the nodes in return for relaying calls require no monetary compensation. Recall that the target credit balance of a node is 1. Nodes which have a credit balance less than 1 and which are therefore under-supplied with credits can be charged a monetary amount when they create credits through the application of Eq. (5). The money raised in this way is used to compensate those nodes which have a credit balance greater than 1 and which are therefore over-supplied with credits: these nodes destroy credits through the application of Eq. (5). Since the creation/destruction of credits increases/decreases a node's expected service, such a monetary compensation presents an incentive for a node with a balance above 1 (such nodes provide more resources to the community than they remove from the community) to take part in the network. The same argument in favour of monetary compensation also applies when local redistribution is used to move credits from over-provisioned to under-provisioned nodes.

4. Autonomous motion

Some nodes may be placed so that they do not attract transit flows. These nodes earn revenue only from sending and receiving traffic and so they have little credit for sending their own traffic. One could argue that these nodes must be content with a low data rate, but on the other hand these nodes could move to more favourable positions where they could earn revenue and so improve their data rate. We call this "autonomous motion". Here radio interference plays an important role, as it defines an interesting trade-off: nodes may prefer to be close together in order to reduce the power needed to transmit data, but on the other hand proximity increases radio interference, and has therefore a negative effect on connectivity. In the remainder of the paper we will obtain insight into the resulting behaviour by means of simulation experiments which reveal that the positions of the nodes converge to non-trivial optimal positions on 2D and 3D surfaces.

4.1. The motion model

The location of the nodes determines (i) the battery power e_{ij}^{tx} required to transmit unit flows, and (ii) the signal-to-interference ratios σ_{ri}. These quantities in turn will have an impact on the congestion costs which determine the routes and the flows. In this initial study, the costs of the energy expended in the motion are not taken into account.

The autonomous motion of the nodes is modelled as follows: we assume that each node can estimate its own position. At intervals of 100Δ seconds, each node broadcasts its position and the magnitudes of its originating and transit flows to its neighbours. These data allow each node to compute, at intervals of 100Δ seconds, an approximate value of its utility function $u_{i,n}(t+\Delta)$ that would be realised if the node moved from its current position $(x_i(t), y_i(t))$ to any of N candidate positions

$$(x_i(t) + \dot{x}_i\Delta\cos(2\pi n/N), y_i(t) + \dot{y}_i\Delta\sin(2\pi n/N)),$$

where $(\dot{x}_i, \dot{y}_i)$ is the (fixed) velocity of node i, and $n \in \{0, 1, \ldots, N-1\}$. The best move for node i (the value of n that maximises the utility $u_{i,n}(t+\Delta)$, say $n^\star$) is determined. Node i will not move if $u_{i,n^\star}(t+\Delta) \leqslant u_i(t)$. First each node determines its optimal direction, and only then do they move (simultaneously). The accuracy with which node i can compute its utility function depends upon the extent to which its estimates of the positions and the flows of its neighbours are both accurate and up-to-date, and the computational resources available to node i.

Note that the nodes do not move to the candidate positions to measure the values of the utility; each node remains in its current position while it computes an estimate of its utility function at each of the candidate positions.

If a node is out of transmission range and cannot reach any other node, then such a disconnected node will move at random until it can reach a node whereupon it will move autonomously as described above.

4.2. Experimental results

We next present the results of a set of experiments that allow the nodes to move autonomously in order to maximise their utility functions, as described below. As mentioned earlier, there is an interesting trade-off between the nodes being close together (thus reducing the transmission power needed) and being far apart (thus reducing the radio interference). The experiments indicate how the nodes move in specific situations.

We consider two utility functions namely $u_i(t) = b_i(t)$ so that the utility at node i is equal to the credit balance at node i, and $u_i(t) = \sum_{r\in\mathscr{R}^S(i)} y_r(t)$ where $i = O(r)$ where $O(r)$ denotes the originating node of route r so that the utility at node i is equal to the total flow originating at node i.

*A 3-node network.*In this experiment three nodes are located on a 100 m × 100 m plane: the three nodes lie almost on a North–South straight line. Node 0 moves from North to South. Radio interference is included in the model.

Fig. 7 shows that nodes 1 and 2 display interesting (but expected) behaviours as they move in order to *maximise their credit balances*. First consider node 1. As node 0 approaches node 1, node 1 moves South, away from node 0. When node 0 is near node 1, node 1 moves West, away from node 0. When node 0 is distant from node 1, node 1 is immobile.

Next consider node 2. When node 0 is distant from node 2, node 2 is immobile. As node 0 approaches node 2, node 2 moves South, away from node 0. When node 0 is near node 2, node 2 moves East, away from node 0. When node 0 moves past node 2, node 2 moves North, away from node 0. When node 0 is distant from node 2, node 2 moves South, away from node 1.

A 21-node network. In this experiment 21 nodes are located on a 100 m × 100 m plane in three concentric rings surrounding a central node. Radio interference is included in the model. Fig. 8(upper) shows what happens when the nodes move autonomously to *maximise their credit balances.* The nodes located on the outer ring move to the edge of the network. The other nodes remain more or less in place. Fig. 8(lower) shows what happens when the nodes move autonomously to *maximise their throughputs*. The nodes move towards the centre of the network where they form an approximate ring, yielding an increase of some 25% in the total average throughput due to the shorter transmission distances even though the radio interference among the nodes has increased.

A 10-node network: a collective motion is induced. In this experiment 10 nodes are located at random on a 100 m × 100 m plane as in Fig. 3. Radio interference is included in the model. Node 1 moves at a constant velocity from its initial location close to the South-East edge of the network through the centroid of the network and reaches the North boundary of the network at the end of the simulation.

Fig. 9 shows what happens when the nodes move autonomously to *maximise their throughputs*. The nodes initially move towards the centre of the network. As node 1 approaches the centroid of the network, the remaining nodes move in a cluster to intercept node 1. Having intercepted node 1, the cluster tracks the motion of node 1 but remains at a distance from node 1: the nodes in the cluster align themselves along an approximately co-linear front, which configuration affords energy efficient multi-hop connectivity among the nodes in the cluster.

A 10-node network: motion on a 3D surface. Consider the surface

$$z = f(x,y) = \sum_{i=1}^{N} a_i e^{-((x-x_i)^2+(y-y_i)^2)/s_i},$$

which places a hill i centred on (x_i, y_i) where a_i denotes the scale (height) and s_i the shape (steepness) of the hill.

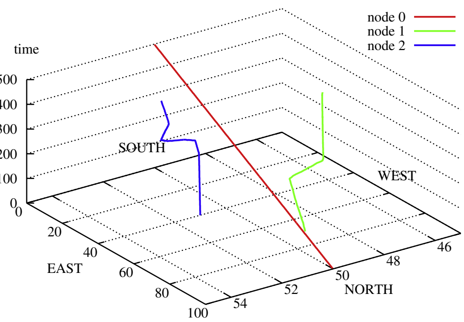


Fig. 7. A 3-node network: the nodes move autonomously in order to maximise their credit balances.

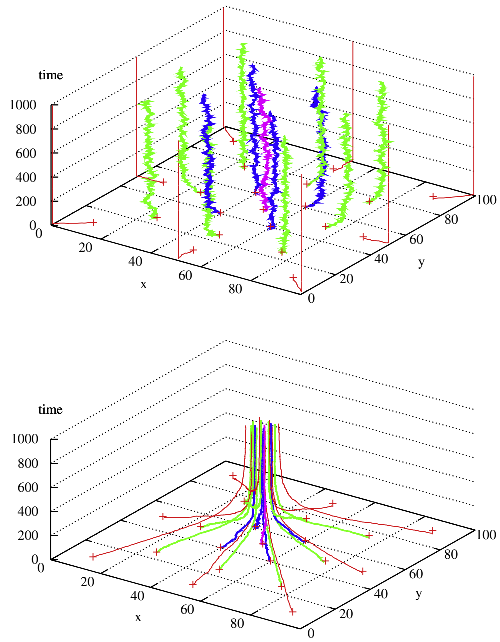


Fig. 8. A 21-node network: the nodes move autonomously in order to maximise their credit balances (upper); throughputs (lower).

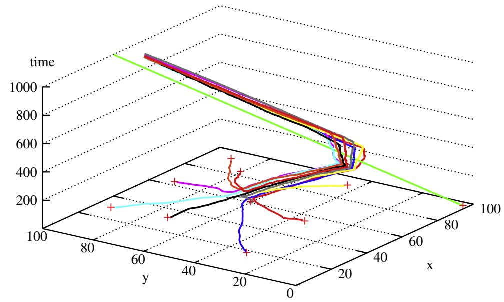


Fig. 9. A 10-node network: the nodes move autonomously in order to maximise their throughputs; a collective motion is induced by the deterministic motion of node 1.

In this experiment, we construct $N = 2$ hills located on a $100\ \text{m} \times 100\ \text{m}$ plane. The hills are centred on $(x_1, y_1) = (20, 80)$ and $(x_2, y_2) = (70, 30)$ and are of height $a_1 = 10$, $a_2 = 5$ and shape $s_1 = s_2 = 100$. Radio interference is included in the model. The nodes are placed at random on the (x, y)-plane as in Fig. 3 except for nodes 1 and 4 which are initially located behind hills and which are out of radio contact with respect to the other nodes.

At intervals of 1 s, an unconnected node determines the slope of the surface in the 8 cardinal compass directions and moves along a line orthogonal to the line of steepest ascent. An unconnected node therefore moves around the hill until it makes radio contact. Other heuristics can be used to compute the movements of disconnected nodes and may result in more rapid reconnection. The connected nodes move as described in Section 4.1.

Fig. 10 shows nodes 1 and 4, which are initially located behind hills and are unconnected, moving around their respective hills and establishing contact with the other nodes. The nodes then move autonomously *to maximise their throughputs* which results in a motion towards the centre of the network where the nodes form an approximate ring. As required, the autonomous motion results in an increased throughput. The total credit balance is equal to the number of nodes.

4.3. Interpretation

The pattern of movement which arises when the nodes move autonomously in order to maximise their credit balances is fundamentally different from the pattern which arises when the nodes move to maximise their throughputs. In the latter case, the nodes move closer together in order to reduce the power needed to transmit data, but

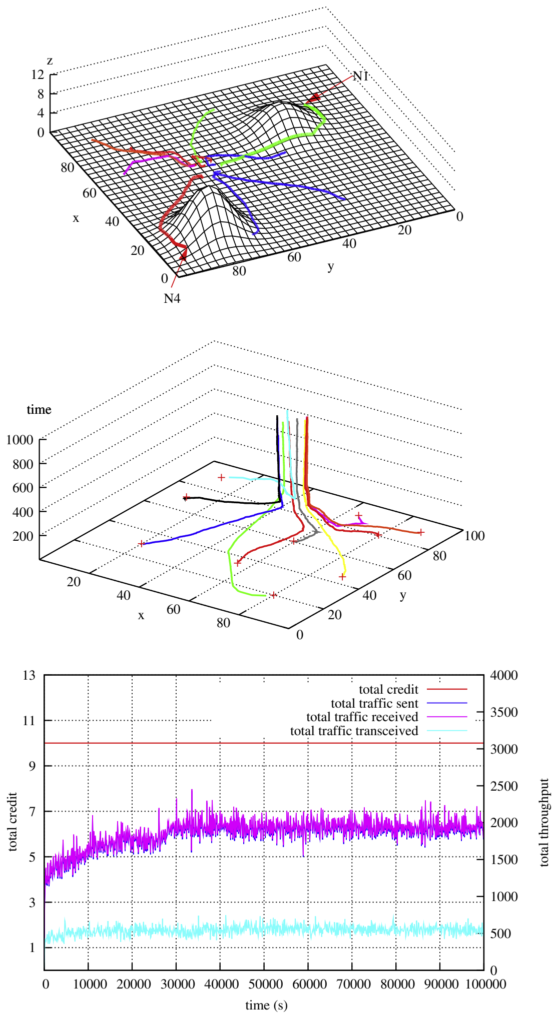


Fig. 10. A 10-node network where two unconnected nodes move autonomously to make radio contact: motion on the surface (upper); (x, y) motion as a function of time (middle); the total throughput and the total credit balance (lower).

on the other hand proximity increases radio interference, which has a negative effect on connectivity. Maximizing the credit balance induces a distribution of the nodes throughout the area: in particular, a node s at the periphery moves towards the edge of the network. To understand this behaviour, we examine the impact of such a motion on the payments made and payments received by a node.

First we observe that when node s moves to the edge of the network, the distance between node s and most (if not all) of the other nodes will increase. Node s will therefore need more power to transmit data, and the price of power at node s will increase. This in turn will lead to a reduction in the payments made by node s to other nodes, and an increase in the payments received by node s *from* other nodes.

The payments made by an originating node s to other nodes for originating calls are reduced. Consider a node s at the origin of route r. If node s moves towards the edge of the network, then the price of power at node s will increase. The price $\mu_r(t)$ of route r will increase and the data rate $y_r(t) = w_s(t)/\mu_r(t)$ of the originating call on route r will decrease since node s's willingness to pay $w_s(t)$ for data transmission is a constant share of its credit balance $b_s(t)$: thus $w_s(t) = \alpha b_s(t)$.

When computing the cost of an originating call on route r, since the increased costs of route r are incurred at node s only, the source node s is able to retain a relatively higher share of its willingness-to-pay $w_s(t)$ (the originating node pays for bandwidth and power congestion costs at the source, transit and destination; since node s is the source, only the congestion costs at the transit and the destination nodes are transferred to other nodes). The payments made by node s to the other nodes on route r decrease.

The payments received by a transit node s from other nodes for transit calls are increased. Consider a node s on route r where $s \neq O(r)$. If node s moves towards the edge of the network, then the price of power at node s will increase. When computing the cost of a call on route r, since the increased costs are incurred at node s only, a relatively higher share of the willingness-to-pay of the originating node $O(r)$ is received by the transit node s. The payments received by the other nodes on route r decrease.

If the direction of motion of node s that is seeking to maximize its credit balance is based on its short term income (as computed during the next time interval), then node s has an incentive to move to the edge of the network since payments made *to* other nodes will decrease and payments received *from* other nodes will increase. However, this is a short-sighted perspective. The movement towards the edge of the network will degrade the ability of node s to attract transit calls and hence its long-term income will decline.

Seeking maximum throughput does not degrade the ability of the nodes and the network to carry transit calls. We therefore conclude that autonomous motion targeting throughput optimization is better suited for organizing the spatial distribution of the mobile nodes. The motion model is one-step look-ahead: it takes only the current call occupancy into account and does not consider the effect of future calls. In contrast to the case of credit balance maximization, an optimal throughput-maximizing decision based on the next period has no adverse impact on a longer time scale.

5. Conclusion

This paper has considered an ad hoc network in which nodes are given incentives to collaborate. We rely on a framework that was developed in [2], which we augmented with a radio interference model. Each node has a credit balance; credits are spent when a node sends or receives traffic, and credits are earned when a node acts as source, transit or destination node.

Our first focus was on mutual cooperation issues: nodes that are located near the centre of the network are considerably better off, compared to nodes at the periphery, since nodes near the centre are more likely to serve as transit nodes, and therefore earn more credits. We proposed a number of remedies for this effect. The efficacy of these remedies was assessed in a detailed performance study.

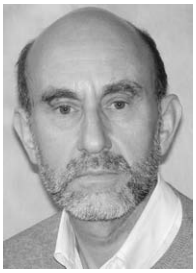
The second topic that we addressed concerned so-called "autonomous motion". Our model allows us to study the optimal compromise for the nodes between being close together (to reduce the power costs of transmission) and being further apart (to reduce radio interference). We investigated how the nodes would move if they were to optimize some utility function such as their credit balance, or their throughput. Through simulation experiments we investigated the evolution of the position of the nodes as a function of time. Some of the simulations take a 3D model of the terrain into account where simple heuristics are used to establish connectivity, starting from a disconnected situation. We observe that, depending on the choice of the utility function, the positions of the nodes converge to non-trivial optimal operating points.

References

[1] T. Bheemarjuna Reddy, I. Karthigeyan, B.S. Manoj, C. Siva Ram Murthy, Quality-of-service provisioning in ad hoc wireless networks: a survey of issues and solutions, Journal of Ad Hoc Networks 4 (1) (2006) 82–124.
[2] J. Crowcroft, R. Gibbens, F. Kelly, S. Östring, Modelling incentives for collaboration in mobile ad hoc networks, Performance Evaluation 57 (2004) 427–439.
[3] F. Kelly, Charging and rate control for elastic traffic, European Transactions on Telecommunications 8 (1997) 33–37.
[4] F. Kelly, A. Maulloo, D. Tan, Rate control in communication networks: shadow prices, proportional fairness and stability, Journal of the Operational Research Society 49 (1998) 237–252.
[5] K. Chen, Z. Yang, C. Wagener, K. Nahrstedt, Market models and pricing mechanisms in a multihop wireless hotspot network, in: Proceedings of the Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2005, pp. 73–84.
[6] J. Crowcroft, R. Gibbens, F. Kelly, S. Östring, Modelling incentives for collaboration in mobile ad hoc networks, in: Proceedings of the WiOpt, 2003.
[7] J. Göbel, A.E. Krzesinski, M. Mandjes, Analysis of an ad hoc network with autonomously moving nodes, in: Proceedings of the Australasian Telecommunication Networks and Applications Conference ATNAC 2007, Christchurch New Zealand (December 2007) pp. 41–46.
[8] O. Dousse, F. Baccelli, P. Thiran, Impact of interferences on connectivity in ad hoc networks, IEEE/ACM Transactions on Networking 13 (2) (2005) 425–436.
[9] H. Koskinen, Generalization of critical transmission range for connectivity to wireless multihop network models including

interference, in: Proceedings of the Communications and Computer Networks 2005, pp. 88–93.
[10] V. Siris, Resource control for elastic traffic in CDMA networks, in: Proceedings of the Mobicom 2002, Atlanta GA (September 2002).
[11] V. Siris, C. Courcoubetis, Resource control for loss-sensitive traffic in CDMA networks, in: Proceedings of the IEEE INFOCOM 2004, Hong Kong (March 2004).
[12] B. Page, W. Kreutzer, The Java simulation handbook – simulating discrete event systems with UML and Java, Chapter 10, Shaker, Aachen, 2005.
[13] J. Göbel, A.E. Krzesinski, Modelling Incentives and protocols for collaboration in mobile ad hoc networks, in: Proceedings of the 11th IEEE/ACM Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2008), Vancouver BC, Canada (October 2008), pp. 78–85.
[14] L. Buttyán, J.-P. Hubaux, Security and Cooperation in Wireless Networks: Thwarting Malicious and Selfish Behaviour in the Age of Ubiquitous Computing, Cambridge University Press, Cambridge, 2008.

Johannes Göbel received his diploma degree in business information technology from the University of Hamburg, Germany in 2006. He is presently a research staff member at the Department of Computer Science at the University of Hamburg. His Ph.D. currently in progress focuses on the decentralised optimisation of self-organising transport networks.



Anthony E. Krzesinski obtained the M.Sc. from the University of Cape Town and the Ph.D. from Cambridge University, England. He joined the Shell Research Laboratory in Amsterdam where he worked on the development of mathematical models to predict the performance of computer systems. He later joined the Department of Computer Science at University of Stellenbosch. In 1985 he was appointed Professor of Computer Science at the University of Stellenbosch. His research interests centre on the performance evaluation of telecommunication networks.

Michel Mandjes received his M.Sc. (in both mathematics and econometrics) and Ph.D. degrees from the Vrije Universiteit (VU), Amsterdam, the Netherlands. After having worked as a member of technical staff at KPN Research (Leidschendam, the Netherlands) and Bell Laboratories/Lucent Technologies (Murray Hill NJ, USA), as a part-time full professor at the University of Twente, and as department head at the Centre for Mathematics and Computer Science (CWI), Amsterdam, he currently holds a full professorship at the University of Amsterdam, the Netherlands. His research interests include performance analysis of communication networks, queueing theory, Gaussian traffic models, traffic management and control, and pricing in multi-service networks.