THIRD ERCIM WORKSHOP ON EMOBILITY

Marc Brogle, Geert Heijenk,
Torsten Braun, Dimitri Konstantas (Eds.)

University of Twente, Enschede, The Netherlands, May 27 and 28, 2009

Book orders:
Secretariat CTIT
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands
Phone +31 53 489 8031

Published: May 2009
University of Twente, Enschede, The Netherlands

Print: Ipskamp Drukkers, Enschede, The Netherlands

CTIT Workshop Proceedings WP 09-03
Centre for Telematics and Information Technology
University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

ISBN 978-90-365-2846-7
ISSN 1574-0846

Preface

ERCIM, the European Research Consortium for Informatics and Mathematics, aims to foster collaborative work within the European research community and to increase co-operation with European industry. In the ERCIM eMobility workshop, current progress and future developments in the area of eMobility should be discussed and the existing gap between theory and application closed.

This volume contains all accepted papers of the third eMobility workshop, which has been held in Twente, The Netherlands, on May 27 and 28, 2009. Papers from different areas have been selected for this workshop. The contributions discuss several topics of the ERCIM eMobility working group including, performance optimization in wireless and cellular networks, real world applications and case studies of WSNs and WLANs, service improvements in wireless (mesh) and fixed networks, and security and V2V communication in mobile networks.

At this point, we want to thank all authors of the submitted papers and the members of the international program committee for their contribution to the success of the event and the high quality program. The proceedings are divided into two sections, full papers and short papers / extended abstracts. While the latter present work in progress and ongoing research, the full papers have been carefully selected in a peer review process. The reviewers evaluated these papers and sent the authors the comments on their work.

The invited talk held by Fernando Boavida on "Ubiquitous Content Networking - The CONTENT NoE Approach" opened the NoE special session. The speaker provided some answers to the question about wether in a world full of content providers, where users can easily have Access to all kinds of content, there are still any challenges in ubiquitous content networking.
The other invited talk of the NoE special session held by Aiko Pras about "Emanics - Managing the future Internet" presented Emanics, a FP6 Network of Excellence (NoE) focussing on management of future Internet infrastructures and services. The speaker pointed out the three main topics of this NoE: 1) integrating research groups within Europe, 2) disseminating knowledge to the rest of the world and 3) performing research.

For the Vehicle-to-X Communications special session, the invited talk held by Thomas Michael Bohnert on "Vehicle-to-Business: Use Cases, Requirements, Architecture and Technology" presented the concept of Vehicle-to-Business. It re-visited existing business models and discussed novel models from a large-scale enterprise software and platform provider's point of view. Insights in what regards the requirements on a communication platform were provided and a communication and service delivery platform was presented.
The other invited talk of the Vehicle-to-X Communications special session held by Geert Heijenk about "Connect & Drive: On the use of Cooperative Adaptive Cruise Control to increase Traffic Stability and Efficiency" presented a research project carried out by a Dutch consortium of universities and industry. The

project aims at designing and evaluating a cooperative adaptive cruise control system, with the goal of increasing traffic efficiency and safety, and decreasing vehicle emission.

The University of Twente is an entrepreneurial research university. It was founded in 1961 and offers education and research in areas ranging from public policy studies and applied physics to biomedical technology. The UT is the Netherlands' only campus university. The next ERCIM eMobility workshop is scheduled for 2010. We hope that a lot of our participants and many new colleagues will take this opportunity to continue exchanging their knowledge and experiences devoted to the development and use of eMobility.

Torsten Braun
Marc Brogle
Geert Heijenk
Dimitri Konstantas

General chairs

Torsten Braun, University of Bern, Switzerland
Dimitri Konstantas, University of Geneva, Switzerland

TPC chairs

Marc Brogle, University of Bern, Switzerland
Geert Heijenk, University of Twente, The Netherlands

Technical program committee

Francisco Barcelo-Arroyo, Universitat Politecnica de Catalunya, ES
Robert Bestak, Czech Technical University in Prague, CZ
Gregory O'Hare, University College Dublin, IRL
Jean-Marie Jacquet, University of Namur, BE
Yevgeni Koucheryavy, Tampere University of Technology, FI
Saverio Mascolo, Politecnico di Bari, IT
Edmundo Monteiro, University of Coimbra, PT
Antonio M. Peinado, University of Granada, ES
Vasilios Siris, FORTH-ICS, GR
Dirk Stähle, University of Wuerzburg, DE
Do van Thanh, NTNU, Trondheim, NO
Mari Carmen Aguayo Torres, University of Malaga, ES
Vassilis Tsoussidis, Democritus University of Thrace, GR

Additional reviewers

Ageliki Tsioliaridou
Angel M. Gomez Garcia
Bharat Bhargava
David Palma
Dragan Milic
Gerald Wagenknecht
Giorgos Papastergiou
Guevara Noubir
Hans van den Berg
Ibrahim Khalil
Jordi Forne
Markus Anwander
Martina Zitterbart
Miki Yamamoto
Paulo Mendes
Peter Langendoerfer
Philipp Hurni
Rolf Kraemer
Sem Borst
Stylianos Dimitriou
Victoria Sanchez

Table of Contents

I Full papers

II Short papers and extended abstracts

Part I

Full papers

Increasing network lifetime by battery-aware master selection in radio networks

Maurits de Graaf[1] and Jan-Kees C.W. van Ommeren[2]

[1] Thales Division Land & Joint Systems,
Bestevaer 46, 1271 ZA Huizen, Netherlands
maurits.degraaf@nl.thalesgroup.com
[2] Faculty of Electrical Engineering, Mathematics and Computer Science,
University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
j.c.w.vanommeren@ewi.utwente.nl

Abstract. Mobile wireless communication systems often need to maximize their network lifetime (defined as the time until the first node runs out of energy). In the broadcast network lifetime problem, all nodes are sending broadcast traffic, and one asks for an assignment of transmit powers to nodes, and for sets of relay nodes so that the network lifetime is maximized. The selection of a relay set consisting of a single node (the 'master'), can be regarded as a special case of this problem. We provide a mean value analysis of algorithms controlling the selection of a master node with the objective of maximizing the network lifetime. The results show that already for small networks simple algorithms can extend the average network lifetime considerably.

Keywords network lifetime, ad hoc networks, average case analysis, random graphs

AMS classification 90B18

1 Introduction

Mobile wireless networks are often battery powered which makes it important to maximize the network lifetime: batteries are (relatively) heavy, large, and sometimes difficult to replace. Here, the network lifetime is defined as the time until the first node runs out of energy. The broadcast network lifetime problem asks for settings of transmit powers and (node-dependent) sets of relay nodes, that maximize the network lifetime, while all nodes originate broadcast traffic.

Literature in this area considers the lifetime maximization in mobile ad-hoc networks (MANETs). Often, the complexity is reduced by assuming transmissions originate from a single source (Kang and Poovendran

[1], Pow and Goh [2] and Park and Sahni [3]). The related problem of minimizing the total energy consumption for broadcast traffic has also been widely studied, because it provides a crude upper bound to the lifetime of the network. Liang [4] and Cagalj et al. [5] have proven independently that minimizing the total transmitted power is NP-hard. Another way to reduce the complexity of the general problem is to allow for transmissions from multiple sources but ask for a fixed (i.e., a node independent) *set* of relay nodes to maximize the network lifetime. This leads to lower bounds for the general network lifetime problem.

The contribution of this paper is a mean value analysis of a special case of this problem, where we ask for a *single* relay node (the master). We describe four algorithms controlling the selection of the master, while taking into account remaining battery capacity and transmit powers, and allowing transmission from multiple sources. For these algorithms, we provide a framework for calculation of the probability distribution and expectation of the network lifetime.

The results provide insight in the lifetime that can be gained by master selection which is directly relevant for some specific (military) VHF/UHF radio networks. For example IEEE 802.11 in infrastructure mode (where the access point has the master role). A more general interest lies in applications to Wireless Personal Area Networks (WPANs), and sensor networks. Here one could envisage a distinction between very simple devices (clients), and more powerful devices (eligible masters). Implementing the described master selection imposes little memory requirements while providing limited relaying capabilities. From a theoretical viewpoint this analysis provides a stepping stone for further generalizations: fixed relay sets of arbitrary size (leading to hierarchical trees) and dynamic master selection over time. Work on these extensions, involving a.o. an implementation of a dynamic master selection algorithm, is currently in progress. In the practical setting the transmit powers are dynamically adapted based on the RSSI (Relative Signal Strength Indications).

2 General model and notation

We only consider potential master nodes in a network. For a set $V \subseteq \mathbb{R}^d$ of potential master nodes, a power assignment is a function $p : V \to \mathbb{R}$. Following the notation of [6], to each ordered pair (u, v) of transceivers we assign a transmit power threshold, denoted by $c(u, v)$, with the following meaning: a signal transmitted by transceiver u can be received by v only when the transmit power is at least $c(u, v)$. We assume that $c(u, v)$ are

known, and that these are symmetric, i.e., $c(u,v) = c(v,u)$ for all pairs $\{u,v\} \in V$. For a node $m \in V$, let p_m denote the power assignment $p_m : V \to \mathbb{R}$ defined as:

$$p_m(v) = \begin{cases} c(v,m) & \text{for } v \neq m, \\ \max_{v \in V} c(v,m) & \text{for } v = m. \end{cases} \tag{1}$$

Note that with power assignment p_m the resulting graph has m as a master. Each vertex is equipped with battery supply b_v, which is reduced by amount $\lambda p_m(v)$ for each message transmission by v with transmit power $p_m(v)$. Similarly, b_v is reduced by amount $\mu r(v)$ for each message reception by v.

Let $T_1, T_2, T_3, \ldots$ denote the time periods under consideration. Let node i transmit $a_i(T_j)$ times during time period T_j. (Note, that as we focus on the network lifetime, we assume there is enough space between transmissions, so that collisions do not occur.) We assume that the $a_i(T)$ are constant for all time periods T_i, $(i = 1, \ldots,)$, and define $a_i = a_i(T)$. We call a series of transmissions were each node i transmits a_i times a *round*. Suppose node m is master. Based on these assumptions, we obtain after one round:

$$b_v = \begin{cases} b_m - \lambda p_m(m) \sum_{v \in V} a_v - \mu r(m) \sum_{v \neq m} a_v & \text{for } v = m, \\ b_v - \lambda a_v p_m(v) - \mu r(v) \sum_{v \in V} a_v & \text{for } v \neq m. \end{cases}$$

Note that this notion of rounds allows us to disregard the order in which the transmissions take place. Suppose that a master m is chosen which is kept for the whole lifetime of the network. The lifetime $L(m)$, expressed in the number of rounds, when node m is master can now be found as:

$$L(m) = \min_{v \in V} \{\rho_m, \rho_v\}, \tag{2}$$

where $\rho_m = b_m / (\lambda p_m(m) \sum_{v \in V} a_v + \mu r(m) \sum_{v \neq m} a_v)$ indicates that the lifetime is determined by the master node.

The expression $\rho_v = b_v / (\lambda a_v p_m(v) + \mu r(v) \sum_{v \in V} a_v)$ indicates that the lifetime may be determined by nodes that are 'far' from the master, and have too low battery capacity to reach the master, or have high reception powers.

In the general formulation, this paper is concerned with the following problem: given a graph $G = (V, E, c, b, a)$, $c : E \to \mathbb{R}$ denotes the transmit power thresholds, and $b : V \to \mathbb{R}$ denotes the initial battery levels $b_v, v \in V$, and the relative frequencies $a_1, \ldots, a_n$, one asks for a master m, maximizing the network lifetime: $L(m) = \max_{v \in V} L(v)$.

In our simplified analysis, we assume $\mu = 0$ (receive power is negligible), $\lambda = 1$ (by scaling), $V \subseteq \mathbb{R}^d$, E corresponds to a complete graph, $c(u, v) = \|u - v\|^2$, and relative message transmission frequencies $a_i = 1$ for $i = 1, \ldots, n$. In this case the only variables are the node locations and the initial battery levels: $G = (V, b)$. Note, however, that the methods used in this paper extend to other power attenuation laws. Moreover, if receive power is approximately equal to transmit power (i.e., if $\mu \approx \lambda$ and $r(m) \approx p_m(m)$) then the relative performance of various selection algorithms will be the same as for the analysis below.

3 Master selection algorithms

For a graph $G = (V, b)$, and a given master m, we say that b satisfies condition (*) if

$$\frac{b_v}{b_m} \geq \frac{1}{n} \text{ for all } v \in V. \qquad \text{condition (*)}$$

It immediately follows that

Proposition 1 *Suppose $G = (V, b)$ with vertex m as a master satisfies condition (*). Then the lifetime $L(m)$ is given by:*

$$L(m) = \frac{b_m}{np_m(m)} \tag{3}$$

Proof. Suppose condition (*) is satisfied. The lifetime $L(m)$ is determined by (2). By condition (*) and the fact that power is symmetric it follows for all $v \in V$ $p_m(m) \geq p_m(v)$ so that $b_v/p_m(v) \geq b_m/(np_m(m))$. So the minimization in (2) is obtained as in (3).

Note that condition (*) is satisfied for all masters $m \in \{1, \ldots, n\}$ if the ratio between the minimal and maximal element of b is at least $1/n$. This is particularly true if $b_i \cong U(1/n, 1)$. Note also that if condition (*) is not satisfied, then (3) provides an upper bound to the network lifetime. This upper bound can be far away from the network lifetime as there can be nodes v for which the ratio $b_v/p_m(v)$ could be very low.

Assuming condition (*) is satisfied, in view of (3) we define the message lifetime as the total number of messages that is transmitted during the lifetime. So:

$$M(m) = \frac{b_m}{p_m(m)} \tag{4}$$

Below we perform a mean value analysis of the message lifetime of the following algorithms.

– **Random Master Selection (RND).** Select a master node $m \in V$ at random. We include this for reference purposes.
– **Central Master Selection (CEN).** Select a master node m which is central in the sense that it minimizes the maximum power (distance) to reach the other nodes in the network.
– **Maximum Battery Master Selection (BAT).** Select a master node m in such a way that b_m is maximal among $b_1, \ldots, b_n$.
– **Optimal Master Selection (OPT).** Select a master node m in such a way that $M(m) \geq M(v)$, for all $v \in V$, as defined in (4).

4 Analysis of master selection algorithms

First, we present a common approach for the algorithms RND, BAT and CEN to find the expected lifetime of the network. For OPT, we need a more sophisticated analysis. In Section 4.2 we approximate the general d-dimensional case via the one-dimensional model. Sections 4.3 and 4.4 focus on uniform distributions in one- and two dimensions, respectively.

4.1 The one-dimensional case: a general approach

We consider the following scenario: nodes $Y_1, \ldots, Y_n$ are randomly distributed on $[0, 1]$. Let $Y_{(i)}$ denote the i-th order statistic of the random sample $Y_1, \ldots, Y_n$. That is, $Y_{(1)}$ denotes the smallest of these Y_i, $Y_{(2)}$ the next Y_i in order of magnitude and $Y_{(n)}$ the largest Y_i. So, $Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)}$.

Let $R = (Y_{(n)} - Y_{(1)})/2$, the radius of the shortest interval containing the nodes. With $a = Y_{(1)}$, there are nodes at point a and point $a + 2R$ and the other $n - 2$ nodes are located on $(a, a + 2R)$. Let X denote the distance from one of the $n - 2$ nodes in between the endpoints, to the midpoint $a + R$ ($a + R$ need not be an element of $\{Y_1, \ldots, Y_n\}$).

Denote the distance of node i to the midpoint by X_i and its battery capacity by B_i; the distance to the midpoint will be called location. Note that location and battery capacity are independent. Assume that B_i is $U(c, 1)$ distributed with $0 \leq c \leq \frac{1}{4}$, so $\mathrm{P}(B < b) = (b - c)/(1 - c)$ for $b \in (c, 1)$. For $c \geq \frac{1}{n}$ condition (*) is satisfied.

Under condition (*) the message lifetime of the network only depends on how long the master node works. For a master node at position $X \in [0, R]$ and battery capacity B, the lifetime $M = B/(X + R)^2$. Note that the distributions of battery capacity and location depend on the algorithm. We will use the notation M_{alg}, X_{alg} and B_{alg} to denote the

dependence of the message lifetime, distribution of master location, and master battery capacity on the algorithm. For the algorithms RND, BAT and CEN the battery capacity and the position of the master are independent. Therefore the expectation of $E[M_{\text{rnd,bat,cen}}]$ can be expressed as:

$$E[M_{\text{alg}}] = E[B_{\text{alg}}]E[1/(X_{\text{alg}} + R)^2], \text{where alg} = \text{bat}, \text{cen}, \text{rnd} \quad (5)$$

We easily find, $E[B_{\text{rnd}}] = E[B_{\text{cen}}] = (c+1)/2$. As B_{bat} is the maximum of n independent random variables, we get $E[B_{\text{bat}}] = (n+c)/(n+1)$.

For the OPT algorithm, the battery capacity and the position of the master are no longer independent. Therefore we focus on the distribution function of the lifetime. To simplify the analysis we assume $R = 1/2$. In [8] we present a general method to calculate $P(M_{\text{opt}} < t)$ without this assumption. However, this does not lead to insightful exact results. Define M_i to be the lifetime of the network if node i would be the master. Under $R = 1/2$, the lifetimes M_i are independent. Assuming that the nodes at the endpoints have index $i = 1$ and $i = n$ we obtain that

$$\begin{aligned} P(M_{\text{opt}} \leq t) &= P(\max\{M_i, i = 1, \ldots, n\} \leq t) \\ &= P(M_1 \leq t)P(M_n \leq t) \prod_{i=2}^{n-1} P(M_i \leq t), \end{aligned} \quad (6)$$

where the lifetime distribution at the boundary points M_1 and M_n is given by $\mathrm{P}(M_{\{1,n\}} \leq t) = \mathrm{P}(B \leq t) = t$ for $t \in [0,1]$. For the points $i = 2, \ldots, n-1$ we find that

$$\begin{aligned} \mathrm{P}(M_i \leq t) &= \int_0^{1/2} \mathrm{P}(B \leq (x+1/2)^2 t) f(x) dx \\ &= \begin{cases} 0 & \text{for } t < c, \\ \int_{\sqrt{\frac{c}{t}} - \frac{1}{2}}^{1} \frac{(x+1/2)^2 t - c}{1-c} f(x) dx & \text{for } c \leq t < 4c, \\ \int_0^{\frac{1}{2}} \frac{(x+1/2)^2 t - c}{1-c} f(x) dx & \text{for } 4c \leq t < 1, \\ \int_0^{\frac{1}{\sqrt{t}} - \frac{1}{2}} \frac{(x+1/2)^2 t - c}{1-c} f(x) dx + \int_{\frac{1}{\sqrt{t}} - \frac{1}{2}}^{\frac{1}{2}} f(x) dx & \text{for } 1 \leq t < 4, \\ 1 & \text{for } t \geq 4 \end{cases} \end{aligned} \quad (7)$$

where f denotes the density function of the location of an arbitrary point.

4.2 The *d*-dimensional case: reduction to one dimension

Via a simple construction the d-dimensional case can, in approximation, be reduced to the one-dimensional case. Consider the following scenario,

for some $d > 1$, nodes $Y_1, \ldots, Y_n$ are randomly distributed on the ball $B(0, 1/2) = \{x \subset \mathbb{R}^d | \ \|x\| \leq 1/2\}$ with unit diameter . Let R denote the radius of the smallest ball $B(0, R)$ containing $Y_1, \ldots, Y_n$, i.e., R is the maximum distance to the origin. Again, let X_i (the location) denote the distance of node i to the origin, and let B_i denote its battery capacity. Next, we make the simplifying assumption that the master node has to cover $B(0, R)$. With the squared power attenuation law, this means that for a master node $Y \in B(0, R)$ with corresponding distance X and battery capacity B, the lifetime $M = B/(X+R)^2$. In other words: for each master node Y, we assume there is always a node Y_j, $j = 1, \ldots, n$ which is diametrically opposite. As this in general not true, this overestimates the power assignment, leading to lower bounds for the network lifetime. This way a reduction to the one-dimensional case of Section 4.1 is obtained. Now (5), (6) and (7) can be applied, with minor modifications to account for the fact that now there is a only one endpoint.

4.3 One-dimensional case: uniform distribution

In this section, we specialize to the case where the distribution of the nodes is uniform on $U[0, 1] \subset \mathbb{R}^1$.

Theorem 1 *Let $Y_1, \ldots, Y_n$ be $U[0, 1]$ distributed with $B_i \cong U[c, 1]$. Then we have*

(a) $E[M_{\text{rnd}}] = \frac{(1+c)(n-1)^2}{(n-3)(n-2)}$,
(b) $E[M_{\text{bat}}] = \frac{2(n+c)(n-1)^2}{(n+1)(n-2)(n-3)}$,
(c) $E[M_{\text{cen}}] = 2(c+1)\phi_1(n)\frac{n(n-1)}{(n-3)}$,
(d) $E[M_{\text{opt}}] \geq 4 - \frac{1}{n+1}\left(\frac{7}{12}\right)^{n-2} - \int_{t=1}^{4}\left(2 - \frac{4}{3}\frac{1}{\sqrt{t}} - \frac{t}{12}\right)^{n-2} dt$

with

$$\phi_1(n) = \int_{y=0}^{1} \frac{(1-y)^{n-3}}{(y+1)^2} dy = \frac{1}{n-2} + O(n^{-2}). \tag{8}$$

Proof. Let $D = Y_{(n)} - Y_{(1)}$. Note that D has the following probability density function: $f_{\text{D}}(\ell) = n(n-1)\ell^{n-2}(1-\ell), \ell \in [0, 1]$. So, for $R = D/2$, we find,

$$f_{\text{R}}(\ell) = 2f_{\text{D}}(2\ell) = 2n(n-1)(2\ell)^{n-2}(1-2\ell). \tag{9}$$

Without loss of generality we assume $Y_{(1)} = 0$. First assume R is known, so there are nodes at 0 and at $2R$ and that the remaining $n-2$ nodes are uniformly on $[0, D]$. Also, the distance X from one of the remaining

$n-2$ nodes to the midpoint has a $U(0,R)$ distribution. This implies for the algorithms alg = RND and alg = BAT, using (9):

$$E[\frac{1}{(X_{\rm alg}+R)^2}]=\frac{n-2}{n}\int_{\ell=0}^{1/2}\int_{u=0}^{\ell}\frac{f_{\rm R}(\ell)}{\ell(u+\ell)^2}dud\ell+\frac{2}{n}\int_{\ell=0}^{1/2}\frac{f_{\rm R}(\ell)}{4\ell^2}d\ell$$
$$=\frac{2(n-1)^2}{(n-3)(n-2)}.$$

Where we condition on $R=\ell$. Clearly, with $B\cong U[c,1]$ we have: $E[B_{\rm rnd}]=(1+c)/2$ and $E[B_{\rm bat}]=(n+c)/(n+1)$, and (a) and (b) follow from (5). To see (c), we note that for the central algorithm, under the condition $R=\ell$, the master has probability density $f_{\rm cen}(\ell)=(n-2)(\ell-x)^{n-3}/\ell^{n-2}$. (This follows from the observation that now the master is the node that minimizes the distance to the center. So $P(X_{\rm cen}\leq u)=1-\left(\frac{\ell-u}{\ell}\right)^{(n-2)}$). So we obtain, using (9):

$$E[\frac{1}{(X_{\rm cen}+R)^2}]=(n-2)\int_{\ell=0}^{1/2}\int_{u=0}^{\ell}f_{\rm R}(\ell)\frac{(\ell-u)^{n-3}}{\ell^{n-2}(u+\ell)^2}dud\ell$$
$$=4\phi(n)\frac{n(n-1)}{(n-3)}.$$

As $E[B_{\rm cen}]=(c+1)/2$, we can again apply (5) to conclude (c). We observe that $(n-2)\phi_1(n)={}_2F_1(2,1,n-1,-1)$ where ${}_2F_1(2,1,n-1,-1)$ denotes Gauss's hypergeometric function (see [9], pp. 4-5). To see the righthandside of (8), we note that ${}_2F_1(2,1,n-1,-1)=1+O(n^{-1})$.

(d) For the optimal algorithm we only describe a lower bound. Clearly, $M_{\rm opt}$ with $B\cong U[c,1]$ is bounded from below by $M_{\rm opt}$ with $B\cong U[0,1]$. In its turn, this is bounded from below when we assume there are nodes at the boundary, i.e., $R=1/2$. Using $E[M_{\rm opt}]=\int_0^4(1-P(M_{\rm opt}\leq t))dt$, (6), and (7), we find with $c=0$ and $f(x)=1$, corresponding to the uniform distribution:

$$E[M_{\rm opt}]\geq 4-\frac{1}{n+1}\left(\frac{7}{12}\right)^{n-2}-\int_{t=1}^{4}\left(2-\frac{4}{3}\frac{1}{\sqrt{t}}-\frac{t}{12}\right)^{n-2}dt$$

Note that $0<\left(2-\frac{4}{3}\frac{1}{\sqrt{t}}-\frac{t}{12}\right)<1$, for $1\leq t<4$.

□

Corollary 1 *Let $Y_1,\ldots,Y_n$ be $U[0,1]$ distributed with B_i distributed according to $U[c,1]$. Then we have*

(a) $\lim_{n\to\infty}M_{\rm rnd}=1+c$

(b) $\lim_{n\to\infty} M_{\rm bat} = 2$
(c) $\lim_{n\to\infty} M_{\rm cen} = 2(1+c)$
(d) $\lim_{n\to\infty} M_{\rm opt} = 4$

Proof. Directly from Theorem 1. □

4.4 Two-dimensional case: uniform distribution

In this section, we specialize to the case that the n nodes are uniformly distributed on a disk with unit diameter in $\mathbb{R}^2$. We use the approach of Section 4.2 to derive lower bounds for the message lifetime.

Theorem 2 *Let $Y_1, \ldots, Y_n$ be uniformly distributed on a disk with unit diameter in $\mathbb{R}^2$, with $B_i \cong U[c,1]$. Then we have, with 'log' denoting the natural logarithm,*

(a) $E[M_{\rm rnd}] \geq \frac{(1+c)(1+4(n-1)(2\log(2)-1))}{2(n-1)}$,
(b) $E[M_{\rm bat}] \geq \frac{(n+c)(1+4(n-1)(2\log(2)-1))}{(n^2-1)}$,
(c) $E[M_{\rm cen}] \geq 4n(c+1)\phi_2(n)$,
(d) $E[M_{\rm opt}] \geq 4 - \frac{1}{n+1}\left(\frac{17}{24}\right)^{n-1} - \int_{t=1}^{4}\left(\frac{-48+64\sqrt{t}+t^2}{24t}\right)^{n-2} dt$

with

$$\phi_2(n) = \int_{u=0}^{1} u(1-u)^{n-2}(1+u)^{n-4}du.$$

Proof. Since $Y_1, \ldots, Y_n$ are uniformly distributed, the distribution of the locations X_i is given by $\mathrm{P}(X_i \leq y) = 4y^2$ for $y \in [0, 1/2]$ and $i = 1, \ldots, n$. Let $n = \mathrm{argmax}\{X_i | i = 1, \ldots, n\}$, so X_n has maximum distance to the center, say $X_n = R$. Now R has distribution function $\mathrm{P}(R \leq y) = (2y)^{2n}$, so

$$f_{\rm R}(y) = 4n(2y)^{2n-1}. \tag{10}$$

Given R, $X_1, \ldots, X_{n-1}$ are uniformly distributed on the disk with radius R. For the conditional location distribution of node i, it then follows that $\mathrm{P}(X_i \leq y|R) = (y/R)^2$ for $y \in [0, R]$ and $i = 1, \ldots, n-1$. This implies for the algorithms alg = RND and alg = BAT, using (5),

$$E[\frac{1}{(X_{\rm alg}+R)^2}] = \frac{1+4(n-1)(2\log(2)-1)}{n-1}.$$

Now (a) and (b) follow from $E[B_{\rm rnd}]$ and $E[B_{\rm bat}]$. For (c) we find noting that $f_{\rm cen} = 2(n-1)\frac{y}{\ell^2}(1-\frac{y^2}{\ell^2})^{n-2}$ and with $f_R(\ell)$ as in (10)

$$\begin{aligned} E[\frac{1}{(X_{\rm cen}+R)^2}] &= \int_0^{1/2}\int_0^{\ell} \frac{1}{(y+\ell)^2} f_{\rm cen}(y) f_R(\ell) dy d\ell \\ &= 8n\phi_2(n), \end{aligned}$$

where a change of variables $y = \ell u$ is applied.

(d) Analogous to the one-dimensional case we focus on a lower bound for $E[M_{\text{opt}}]$. Evaluating (7) with $R = 1/2$, $c = 0$ and $f(x) = 8x$ yields the bound presented in the theorem.

Corollary 2 *For the expected lifetime of the system the following limits hold:*

(a) $\lim_{n\to\infty} M_{\text{rnd}} = 2(1+c)(2\log(2)-1)$
(b) $\lim_{n\to\infty} M_{\text{bat}} = 4(1+c)(2\log(2)-1)$
(c) $\lim_{n\to\infty} M_{\text{cen}} = 2(1+c)$
(d) $\lim_{n\to\infty} M_{\text{opt}} = 4$

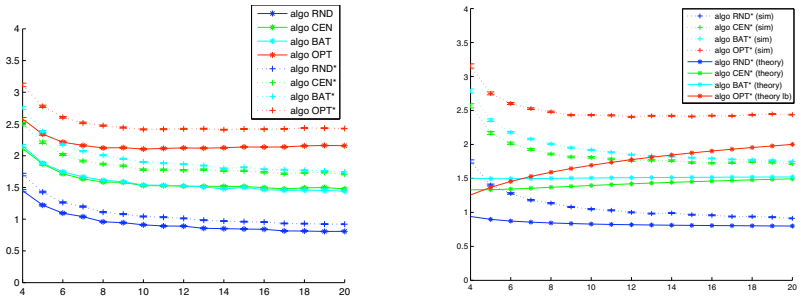
Proof. (a), (b) and (d) follow directly from Theorem 2. To see (c) note that $\int_0^1 u(1-u^2)du \geq \phi_2(n) \geq \int_0^1 u(1-u^2)^{n-4}du$ so $\frac{1}{2(n-1)} \geq \phi_2(n) \geq \frac{1}{2(n-3)}$. □

5 Simulation results

We present simulation results for the two-dimensional case. The network lifetime was evaluated for number of potential masters n, ranging from 4 to 20. The nodes were uniformly distributed in a disk of unit diameter. Battery levels were drawn from a $U(0,1)$- distribution. For each algorithm, the average network lifetime was evaluated over 10000 simulations. Confidence intervals of one standard deviation were calculated. Note that for $B_i \cong U(0,1), i = 1,\ldots,n$, condition (*) of Section 3 does not hold. To investigate the impact of this, in Figure 1(a) two simulated curves are drawn, for each algorithm. The uninterrupted line (indicated with 'algo ALG') shows the evaluation of the network lifetime according to $nL(m)$, with $L(m)$ as in (2), where m depends on the algorithm. The dotted line (indicated with 'algo ALG*') shows the message lifetime $M(m)$ according to (4). As for this parameter choice, condition (*) does not hold, the message lifetime $M(m)$ provides an upper bound to the actual message lifetime. The figure shows that this upper bound is 10-20 % higher than the actual lifetime. This is caused by the fact that slave nodes may run out of energy before the master does. However, for increasing values of n the quality of the approximation improves. The simulations for OPT also shows two effects: for small n, the length of the interval is small, increasing the network lifetime. When n increases, the interval size grows, but also the battery capacity does.

Figure 1(b) show how the simulations compare to the theory. Here 'algo ALG* (sim)' displays the message lifetime $M(m)$ according to (4). The curves corresponding to 'algo ALG* (theory)' show the lower bounds corresponding to Theorem 2. The difference between theory and simulation results is explained by the fact that we the theoretical analysis was based on a worst-case situation: the situation where the master and the node furthest away from the master are diametrically opposite. In simulations, this is not always the case. For the case of 'algo OPT* (theory)', the result provided in Theorem 2 (d) additionally assumed that $R = 1/2$.

Note that if condition (*) holds, then the the results of Theorem 2 provide a lower bound to the actual network lifetime $nL(m)$, with $L(m)$ as in (2). If condition (*) does not hold (as in the simulations), then the network lifetime $nL(m)$ is bounded from above by $M(m)$, which is bounded from below by lower bounds of Theorem 2. As a result, these provide only an indication of $nL(m)$.



(a) impact condition (*).

(b) simulations of $M(m)$ vs bounds of Theorem 2.

Fig. 1. Simulation results for the two-dimensional case.

6 Conclusions and future work

In this paper we describe and provide preliminary quantitative insight in four algorithms controlling the selection of master radios while aiming at maximizing the network lifetime. The algorithms take into account remaining battery capacity and transmit powers. We provide a framework

for calculation of the probability distribution of the network lifetime for the various algorithms. The results show that already for small networks (i.e., networks with a low number of potential masters) simple algorithms can extend the average network lifetime considerably. For large networks this effect is even stronger. For $n \to \infty$, OPT has the best performance as expected, then CEN which slightly outperforms BAT (especially when $B_i \cong U(c, 1)$ with $c > 0$), the performance of RND is (as expected) the worst of the analyzed algorithms. Future research includes extensions of the presented model to: (1) more general power laws and other distributions of locations; (2) variations of the master over time (dynamic selection) [7]; (3) extension of the model to a fixed set of relay nodes (instead of only one). In order to validate the results in practice, an implementation of this algorithm in a sensor network is currently in progress.

References

1. Kang, I., Poovendran, R.: Maximizing Network Lifetime of Broadcasting over Wireless Stationary Ad Hoc networks, Mobile Networks and Applications, 10, 879–89, (2005)
2. Pow, C.P., Goh, L.W.: On the construction of energy-efficient maximum residual battery capacity broadcast trees in static ad-hoc wireless networks, Computer Communications, 29, 93–103 (2005)
3. Park, J., Sahni, S.: Maximum lifetime broadcasting in wireless networks. In: 3rd ACS/IEEE International Conference on Computer Systems and Applications, 2005:1-8 (2005)
4. Liang, W.: Constructing minimum-energy broadcast trees in wireless ad hoc networks, In: Proceedings of the International Symposium on Mobile Ad Hoc Networking and Com-puting (MobiHoc), pp. 112–122 (2002)
5. Cagalj, M., Hubaux, J., Enz, C.: Minimum-energy broadcast in all-wireless networks, NP-completeness and distribution issues. In: Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM, pp 172–182 (2002)
6. Lloyd, E., Liu, R., Marathe, M., Ramanathan, R., Ravi, S.: Algorithmic Aspects of Topology Control problems for ad-hoc networks, Mobile Networks and applications, 10, Issue 1-2 , 19–34 (2005)
7. Graaf, M. de, Dynamic master selection in radio networks, patent pending, 2008.
8. Graaf, M. de, J.C.W. van Ommeren, Increasing network lifetime by battery-aware master selection in radio networks, Technical Report, University of Twente, February 2009.
9. W.N. Bailey, Generalized Hypergeometric Series, Cambridge, England: Univerity Press, 1935.

On the Efficient Policing of HTTP Traffic in WLANs

G.J. Hoekstra[1,2] and F.J.M. Panken[3]

[1]Innovation Research & Technology, Thales Nederland B.V., Huizen, The Netherlands
[2]Centre for Mathematics and Computer Science (CWI), Amsterdam, The Netherlands
[3]ICTRegie, The Hague, The Netherlands
hoekstra@cwi.nl

Abstract. A wireless communication medium typically includes two links: an uplink for transmitting information from the terminal to the access point and a downlink for transmitting information from the access point to the terminal. Separate traffic contracts are commonly defined for each link individually. This holds for both fixed and most wireless situations, since often different frequency bands are used to separate the links in the upstream and the downstream directions. However, in shared wireless access media such as e.g., Wireless Local Area Networks (WLAN), the uplink and downlink use the same frequency bands and compete with one another for airtime when transmitting packets. For this situation, separating the traffic contract over the two directions leads to an inefficient usage of the medium. This paper introduces and evaluates a new method for traffic policing that uses shared communication media more effectively by exchanging the contract parameters between the uplink and the downlink directions dynamically, when desired. The novel solution takes the commonly used policing mechanism as a base and shows how to alter this mechanism to use it more effectively for policing traffic contracts in shared media situations.

1 Introduction

Wireless LANs based on the IEEE802.11 standard have been applied successfully at home premises and public places to provide cheap internet access. Providing commercial services over WLAN by exploiting residential broadband access networks [3] has grown from the FON/Google initiatives towards the services provided currently, e.g. Wireless Internet Zone (WLAN) service from Norwegian-based operator Telenor. In order to make subscriptions for WLAN networks economically viable, Quality of Service (QoS) solutions are of crucial importance. For realizing QoS in telecommunication networks in general, end-users establish a traffic contract with the network (often for each flow independently), based on the quality of experience desired at the application level. This traffic contract specifies an envelope that describes the intended network traffic flow needed to realize a desired quality of service, most commonly expressed in delay, jitter,

and/or packet loss requirements. Most common parameters to parameterize a traffic flow include peak data rate, average sustained data rate, and burst size. Checking traffic rates and comparing them with the traffic contracts is referred to as *traffic policing*. If the received rate exceeds the contract, packets could be marked or dropped, depending on the local policy. The traffic contract is often the result of admission control; see [1] for an overall survey on admission control in wireless networks, [4] for a WLAN specific overview. In more recent work [10] an admission control scheme is presented that takes into account the dynamically changing channel conditions in WLAN to achieve multi-service QoS guarantees. Enterprises and neighboring operators often *shape* their traffic to meet traffic contracts, preventing that their traffic is dropped as a result of the policing mechanisms. Traditionally, only traffic that flows into the network was subjected to a predefined service level. The rationale for this is that if a network is properly dimensioned and the sum of all service level agreements does not exceed the network capacity, traffic will smoothly propagate through the network and hence checking inflow traffic for compliancy with the agreements is sufficient. Traffic contracts for Internet access and for interconnecting enterprises often dictate rates in both the upstream and the downstream directions. Today, these contracts are often dominated by the physical limitations of the interfaces (ISDN-30, T0,T1, etc.) but increased flexible transmission techniques (such as e.g., WLAN, Ethernet) introduced in the access and metro networks are changing this gradually.

The token bucket was first mentioned in [13] and is often implemented to police traffic contracts. It can be specified by three parameters: the token rate (TR), the bucket limit, C, and the leak rate l. Every $\frac{1}{TR}$ seconds a token is added to a bucket with limit C. Tokens are discarded if the bucket is full. When a packet of s bytes arrives, s tokens are removed from the bucket and the packet can pass. If the bucket contains fewer than s tokens, no tokens are removed from the bucket and the packet is considered to be non-compliant. The algorithm allows bursts of up to C bytes, but over the long run the output of conformant packets is limited to the constant rate TR. Non-compliant packets can be dropped, queued or marked, depending on the policy. A commonly used policy is instructing network elements to drop marked packets in the case of overload situations. Facilitating temporary storage of packets allows the integration of policing and shaping in a single device. As storing requires scarce storage capacity for each traffic contract, this is mainly used in situations where traffic is limited (e.g., in terminals) or for a bulk traffic contract, e.g., between operators.

Traffic contract parameters are negotiated on a relatively long time scale (in terms of weeks or months) and today's traffic policing techniques cannot adaptively change contract limits of the uplink and/or the downlink. For wired systems, the subdivision of separate traffic contracts in the upstream and the downstream directions makes sense since each direction has a separate physical connection. In shared media as e.g., WLAN, where the sending of information in the upstream direction competes with the sending in the downstream direction, policing in the two directions independently in general degrades the efficiency

of a shared communication medium. As an example, consider the situation that the network has more information to send to client A over the downlink than the downlink service contract of this customer allows. At the same time, assume that client A transmits less than the agreed uplink traffic contract allows. Applying the conventional traffic policing techniques to the uplink and downlink directions independently may queue or drop surplus downlink traffic, even though there may be room in the uplink contract that uses the very same medium for transportation. By considering the total service contract in the two directions simultaneously, variations in the demand for the downlink can be compensated with temporarily sending less information in the upstream direction, and vice versa. This cannot be realized by applying the traditional token bucket policing mechanisms, where each direction has its own bucket, operated independently.

The policing of a traffic contract goes hand in hand with the dimensioning of the network resources reserved for all traffic contracts. Ross discusses in [11] various models for computing the reservation level of a trunk that is shared by various classes of traffic which may differ in their demand and/or mean holding times. These models operate on relatively long time scales and provide insight in blocking probabilities of (state-dependent) trunk reservation schemas. Borst and Mitra describe in [2] the virtual partitioning schema, introducing dynamics to these reservation schemes by giving lower priority to misbehaving classes which violate their predetermined capacity allocation. Their outcome operates on long time scales and provides fairness and robustness in the admission of new session initiations, giving lower priority to misbehaving classes. This paper minimizes the misbehaving packets in shared access media and hence may serve as a base for the virtual partitioning and for other resource reservation schemas.

This paper introduces and evaluates a new method for traffic policing that uses shared communication media more effectively by exchanging policing tokens for the uplink and the downlink directions dynamically. Hence, the traffic contract is treated as a whole and is dynamically distributed over the upstream and downstream directions, when needed. The paper explains how the token bucket mechanisms commonly used for policing purposes can be extended to realize this. Extra attention is devoted to prevent oscillation effects and the situation that one of the directions dominates the total contract. The performance of the proposed token bucket mechanism is evaluated by extensive simulations for a wide range of parameter combinations. The results demonstrate the gains that can be achieved by applying this new mechanism and provide insight in the side conditions when the solution can be applied best.

2 Mechanism for congestion avoidance in shared media

Policing in the upstream and the downstream directions simultaneously can in principle be realized by two independent policing entities that read from a common reservoir filled with tokens. The rate at which the reservoir is filled represents the total traffic contract, for both directions. However, special precautions are needed to prevent that one of the directions dominates the other

and hence one direction may suffer from starvation. Starvation of one direction eventually results in disconnection (or total uselessness) of a bilateral connection. Prioritizing one direction to solve this for data applications is tempting. After all, packets for the downstream direction may have travelled from far and have nearly reached the final destination and hence giving them priority over upstream packets may seem a good idea at first glance. However, in the case of data applications, the upstream direction consists of mainly acknowledgements of received packets which are small in size by nature and dropping them will disturb the self-clocking of TCP, as described by Jacobson [6], and causes it to lower its throughput and resend information that eventually will need to pass in the downstream access link. For UDP-based real-time applications, the connections in both directions often contribute equally to the quality experienced by the person using the real-time application.

The aim is to design a simple mechanism that improves the overall throughput of the shared wireless medium by regulating the traffic contract in both the upstream and downstream directions simultaneously, and at the same time prevents starvation of one or two of the directions. An important but desired side condition for the solution that has its roots in financial considerations is that it should minimize changes required on existing telecommunication products that implement the token bucket mechanism.

2.1 Mechanism to police in upstream and downstream directions concurrently

To understand the *concurrent up and down policing mechanism*, consider two regular token bucket policing mechanisms: one in the upstream and one in the downstream direction, named U and D, respectively. Without loosing generality, the remainder assumes that U and D may also be used for spacing and hence contain a queue to temporarily store packets. The mechanisms U and D regulate traffic of a terminal named T. The mechanism can also be used to regulate a group of terminals (or an interface) but for didactical reasons the description is limited to the case where only a sole terminal feeds mechanisms U and D in two directions. Several of these policing mechanisms may work in parallel to police various groups of terminals simultaneously but since each of these mechanisms can work autonomously, describing a policing mechanism $(U + D)$ for a single terminal is sufficient.

The parameters of the policing mechanisms require only updates when conformance is checked, i.e., upon the arrival of a packet in either the upstream or the downstream direction. 1 illustrates the state diagram. The index $i(i = 0, 1 mod 2)$ indicates the direction (e.g., $i = 0$ for the upstream direction and $i = 1$ for the downstream direction). Let $L(i)$ denote the bucket level of direction i. Policing mechanisms U and D must support two bucket limits, namely a ceil bucket limit $C(i)$ (refered to as bucket limit) and a floor bucket limit $F(i)$ (refered to as floor limit).

When a packet arrives for transmission in direction i, first the token level $L(i)$ of policing mechanism for the corresponding direction is checked for sufficient

tokens. If the token level stores sufficient tokens to pass, the level is reduced according to the size of the packet just as regular leaky bucket systems, commonly used today. If, however, the token level of U is less than the size of the arriving packet, the conclusion that the packet is non-conformant could be prevented by using tokens from the bucket of the other direction (i.e., $i + 1 mod 2$). This realizes the shared token consumption. One important side condition is introduced to prevent one direction to dominate over the other, namely the level of the bucket used for borrowing tokens (i.e., the policing in the other direction, or $L(i+1)$) should not drop under the floor threshold $F(i+1)$. A packet is marked as non-conformant if both the bucket level $L(i)$ is lower than zero and the bucket level of the other direction $L(i+1)$ drops below the threshold value $F(i+1)$, if reduced with the size of the packet. Both threshold values can be parameterized to optimize performance. The ceil bucket level denotes the maximum level of the reservoir, similar as in traditional token bucket systems. To realize the sharing reservoir, the waste of tokens as a result of overflowing during low traffic activities in direction i can be added to $L(i+1)$. Note that by selecting for both directions $F(i) = C(i)$, the token bucket mechanism as used today can be obtained.

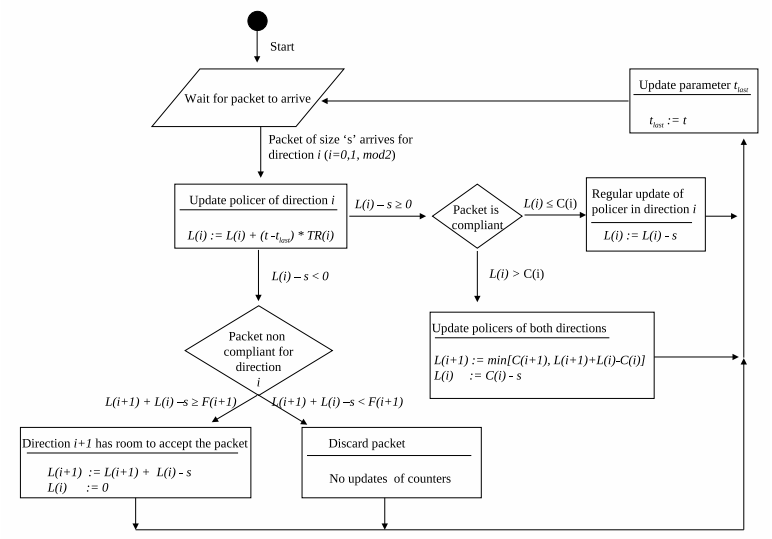


Fig. 1: State diagram for policing in two directions concurrently and preventing starvation in one of the directions. Here the parameters $L(i), C(i), F(i), TR(i)$ denote respectively the bucket level, the ceil bucket limit, the floor bucket limit, and the token rate in direction i (i = 0,1, mod 2).

2.2 Example

Let $i = 0$ denote the upstream direction and $i = 1$ the downstream direction. Figure 2 depicts the token level of the U and D token bucket mechanisms over time, the upstream bucket U located at the top and the downstream bucket D at the bottom. At time $T = t_1$ packet with size $s1$ arrives at the access point for transmission in the upstream direction. Since the value $s1$ can be subtracted from the token level of the upstream reservoir, there is no need for any interaction with the downstream bucket. At time t_2, a packet arrives at the downstream bucket. As the packet arrives after the ceil level of the downstream bucket has reached, an amount of $a2$ tokens will not be used by the downstream bucket and will be transferred to the upstream level. In this case, the upstream bucket level benefits from the low activity in the downstream direction. At times $T = t_4$, $T = t_5$, and $T = t_6$, packets arrive in the downstream direction. After subtraction of the packet sizes $s4$, $s5$ and $s6$ a negative bucket level value of respectively $a4$, $a5$ and $a6$ would occur at the downstream policer D. To prevent marking these packets as non-compliant, the bucket level of the upstream bucket level is reduced with these values. This is allowed as these decreases do not realize a bucket level beneath the threshold value $F(0)$, the floor limit of policer U. The

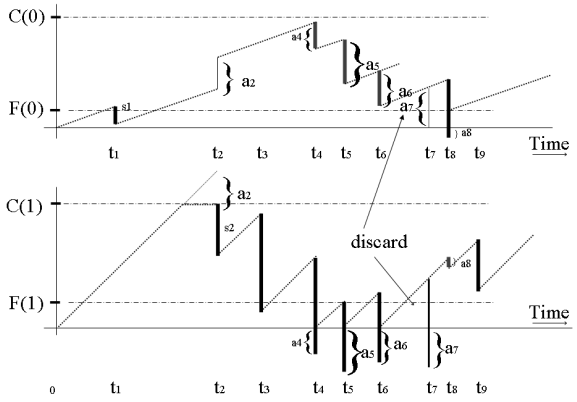


Fig. 2: Example of how the mechanism that polices traffic in the upstream and downstream directions concurrently. Index $i = 0$ denotes the upstream direction, with corresponding $C(0)$ and $F(0)$ as the ceil and floor limits. Index $i = 1$ denotes the downstream direction.

packet that arrives in the downstream direction at time $T = t_7$ is considered non-compliant, since subtracting the value a7 (the shortage after subtracting the value $s7$ from the downstream bucket level at $T = t_7$) from the upstream bucket level will result in lowering the upstream bucket level below the floor

threshold level $F(0)$. The packet arriving at $T = t_8$ in the upstream direction is compliant again, as at $T = t_8$ the downstream bucket level is more than value $a8$ above the floor value $F(1)$.

3 Performance Evaluation

By flexibly borrowing and assigning tokens among both directions, intuitively the shared medium is used better than when separating the contract over the two directions separately. The analysis of this chapter shows the effects of using the newly developed policing mechanism quantitatively, comparing the solution with the traditional way of policing. The performance evaluation is performed both analytically and by simulation. The analytical study shows the correctness of the intuition that the new mechanism uses the shared medium more effectively under certain conditions, whereas the simulation results provide insight when the improvement occurs in practice and quantifies the improvement.

3.1 Analytical Evaluation

A WLAN system that operates in mode Distributed Coordination Function (DCF) that is commonly implemented randomizes packets and distributes the transmit capacity fairly over the active stations. Following the reasoning from e.g. [7,5,9], the flow-level analysis of the DCF mode WLAN system can be modeled as a processor sharing queueing model where the rate at which files are served depends on the number of files that are simultaneously active. The Processor Sharing (PS) queueing model is used to compare the performance of traditional token bucket system with the proposed solution where two directions are policed concurrently. Consider a WLAN DCF system with a net capacity C Mbit/s. To model the flow-level behavior, we consider a classical M/G/1 PS-model, with flow-arrival rate λ, and where the service time B is generally distributed with mean β. In this model, incoming flows immediately enter the system, thereby receiving a fair share of the available capacity, C. Then the occupation rate is $\rho := \lambda\beta$, and the expected sojourn time $E[S]$ is known to be $E[S] = \frac{\beta}{C-\rho}$. Note here that the sojourn time is *insensitive* to the service-time distribution, i.e., depends on the service-time distribution only through its mean β. Previous work on the PS sharing properties of FTP and HTTP applications over WLAN shows that near-insentivity with respect to the file size distribution holds under the condition of having (1) sufficient buffer space at the access point and (2) small TCP window sizes [9,5]. In practice, the net WLAN capacity differs significantly from the gross rate. This is primarily due to the large protocol overhead of the TCP and WLAN MAC and are accounted for in the approach presented in these papers to determine the effective capacity, that we define as C, at the application layer. Similar to [9] our model assumes exponentially distributed think times between the successive file transfers in the context of retrieving a web page. Although the justification of exponentially distributed think times may be questioned, our modeling approach is solely *aimed at proofing the intuition* that

under these conditions policing traffic in two directions concurrently uses the medium more effectively than the traditional policing mechanism.

The traditional token bucket mechanism that polices transmissions in the up-and downstream directions separately is modeled by two distinct processor sharing models that each have a fixed share of the total effective processing capacity, denoted by C_{up} and C_{down}, respectively. The sum of the fixed shares of the two directions equals the total available WLAN processing time ($C = C_{up} + C_{down}$) at the application layer. The expected sojourn time of flows in the up-stream and downstream directions then yield by applying the formulas for these systems individually:

$$E[S_{up,down}] = \frac{\beta_{up,down}}{C_{up,down} - \rho_{up,down}}, where \quad \rho_{up,down} = \lambda_{up,down}\beta_{up,down},$$

When assuming the stability constraint ($0 < \rho_{up} + \rho_{down} < 1$) to these separated models, the mean sojourn time of the flows in the two distinct directions satisfies:

$$\frac{E[S_{up}] + E[S_{down}]}{2} = \frac{\beta_{up}}{2\,(C_{up} - \rho_{up})} + \frac{\beta_{down}}{2\,(C_{down} - \rho_{down})}$$

and

$$\frac{E[S_{up}] + E[S_{down}]}{2} > \frac{\beta_{up} + \beta_{down}}{2\,(C_{up} - \rho_{up}) + 2\,(C_{down} - \rho_{down})} \tag{1}$$

Following the same mathematical modeling principles as above, the mechanism that polices traffic in the up- and downstream directions concurrently is modeled as a single M/G/1 PS-model where jobs compete with one another for processing capacity $C = C_{down} + C_{up}$. For simplicity the complete mutual sharing situation is considered, i.e., floor limit F=0. File requests in the upstream and downstream directions still arrive according to a Poisson process with intensity λ_{up} and λ_{down} and request a processing time $E[B_{up}]$ and $E[B_{down}]$, respectively. Consequently, the expected sojourn time of these individual flows in a shared WLAN system is:

$$E[S_{up,down}] = \frac{\beta_{up,down}}{C - \rho}, where \quad \rho = \lambda_{up,down}\beta_{up,down} + \lambda_{down}\beta_{down},$$

Consequently, the overall mean waiting time of these jobs yields:

$$\frac{E[S_{up}] + E[S_{down}]}{2} = \frac{\beta_{up} + \beta_{down}}{2\,(C_{up} - C_{down}) - 2\,(\rho_{up} - \rho_{down})} \tag{2}$$

Comparing Equations (2) and (3) concludes that policing in two directions concurrently results in overall lower mean waiting times, confirming the intuition that (under the assumption of Poisson arrivals) this policing mechanism uses the shared access medium more effectively.

3.2 Simulation study

The analysis of the previous section shows that under the assumption of Poisson arrivals service times, the described mechanism is superior in performance. Simulation studies were performed to gain insight in the absolute performance gains of the proposed policing mechanism. In general, two telecommunication traffic types can distinguished, namely elastic (TCP-type, e.g., FTP, HTTP data) and real-time (e.g., voice / video kind of applications that use the non reliable UDP layer). For real-time traffic, the performance gained by using the new policing mechanism can best be captured by monitoring the user experience (i.e., mean opinion scores) and is beyond the scope of this paper. Elastic traffic, however, has build-in congestion avoidance functionality that effectively shares the transmission capacity equally over all active flows, an inherent property of the PS-queue. Blocking TCP packets as a result of non-compliancy of the traffic contract inevitably results in retransmissions and hence increases the duration of the associated service as experienced by the end-user. Hence, the length of the service time serves as a good base to quantify the affect of the decision to either block or let a packet pass.

Browsing application details Simulation results were obtained from a browsing application that is parameterized according to the web pages of a total of nine European telecommunication companies and research institutes, shown in Table 1. The application is configured to generate pages with a uniformly distributed body size of [13375, 61998] bytes, containing a uniformly distributed number of inline images in the range of [18, 51], where each image has a uniformly distributed size of [1761, 6854] bytes, resulting in an average page size is 131270 bytes. Referring to [8], consecutive clicks to retrieve a web page that result in the retrieval of the web page are can best be drawn from a lognormal distribution function with an average of 36.8s and a standard deviation of 3180.96s.

Table 1: Webpage measurements

	Page body (bytes)	images per page (images)	AV. Img size (bytes)	Total Img (bytes)
http://www.swisscom.com/	61998	18	6854	123370
http://www.francetelecom.com/	55429	51	1761	89799
http://www.npt.no/	57882	18	2891	52030
http://www.ismb.it/	40492	29	2061	59762
http://www.tu-berlin.de/	21765	23	6341	145837
http://www.motorola.com/	23576	20	1827	36532
http://www.bell-labs.com/	24538	29	4481	129938
http://www.tid.es/	13375	22	4335	95375
http://www.eurasip.org/Proceedings/Ext/IST05/	16457	25	5331	133278
Average	35057	26	3987	96213
Minimum	13375	18	1761	36532
Maximum	61998	51	6854	145837

Figure 3 depicts the network architecture used to obtain simulation results. There is a single web server that is connected to an IP network that should represent the Internet. Various IEEE 802.11b stations are connected to a WLAN

access point which executes the proposed upstream and downstream policing mechanism and connects the users via the IP network with the web application server that is assumed to be present somewhere on the Internet. Measurements on a significant number of TCP sessions on an Internet backbone of US-based operator Sprint reveal [12] that TCP sessions that traverse the Internet may experience a Round Trip Time (RTT) ranging from 10ms to 10seconds with a median of the minimum RTT of around 200ms. In accordance to this, the Internet IP network connecting the routers in our simulated network introduces a single-trip delay to each packet with a minimum of 100ms and a mild maximum of 200ms. Accordingly, the delay of consecutive packets are in our case assumed mutually independent and uniformly distributed between 100 and 200 ms: U[100, 200]. The network nodes (user stations and web server) are configured with a TCP stack that matches the implementation from the Windows XP operating system, which applies the Selective ACK (SACK) mechanism to reduce the impact of TCP retransmissions and match reality with contemporary circumstances.

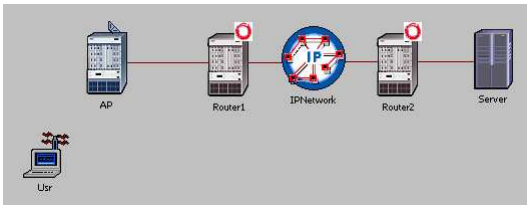


Fig. 3: Simulation setting used to obtain simulation results.

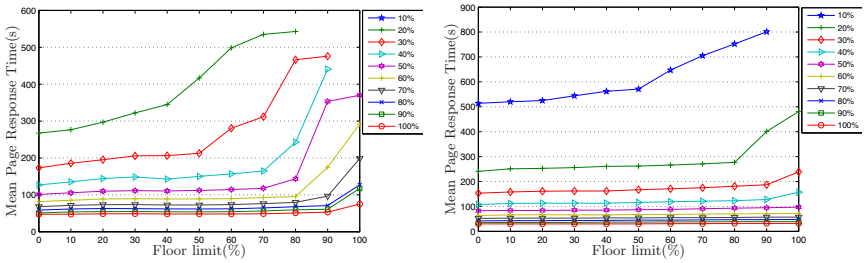
Quantitative comparison through simulation Various splitting ratios of a traffic contract can be envisioned, resulting in comparing many simulation data. Two splitting ratios were selected to compare our mechanism for policing traffic with traditional policing mechanism quantitatively, named *50/50* and *advance knowledge*. In the advance knowledge situation, the browsing application is first simulated without any traffic regulation restrictions. This provides clear insight in the desired network resources in both the upstream and downstream if there would not be any constraints. These values are subsequently used as a base for the traffic contract parameters for the upstream and downstream leaky bucket traffic policing parameters. The 50/50 simulation divides the traffic contract in two parts that are equal in size: 50% for the upstream and 50% for the downstream direction.

The browsing application was simulated under various bucket parameters and traffic contracts for both the 50/50 and the advanced knowledge situations. To understand the effect of sharing the traffic contract among various traffic streams

simulations were performed for two different user populations that share a single contract, namely 1 and 5 users. Each user downloads a web page according to the same statistics while the token rate was chosen proportionally with the user population. The page response times of the web pages were measured during consecutive simulations while the token rate was reduced in equal steps of 10% of the original contract, starting at 100% and stopping at 10%. The resulting response times may reach very high values that may not be appreciated by users in practice but serves the purpose of illustrating the differences between the various parametere settings. To study the effect on the tolerance of the policing mechanism, the simulations were repeated for various conditions of the token bucket limit, varying this limit in steps of 10s, from 10s to 60s and additionally for 90s, 120s, and for 60000s. For the simulations the bucket limit was expressed in bytes, obtained by multiplying the limit in seconds with the used token rate in bytes per second. The floor limits of the upstream and downstream buckets (F_u and F_d, respectively) that indicate to which extent the mutual borrowing of tokens is tolerated were chosen as a percentage of the bucket limit, varying from 0% to 100%. The floor limit of 0% corresponds with a complete sharing situation, meaning that the upstream and downstream buffers can always grab tokens from one another, if available. The 100% floor limit corresponds to the situation that the two buffers cannot exchange tokens and hence coincides with the traditional way of policing traffic. The quantitative comparison of the new mechanism was obtained by computing and comparing the mean page response times as function of the floor limits F for various settings of the token rates which are expressed as a percentage of the total desired contract, if there were no policing limitations.

3.3 Advanced knowledge

Figure 4a shows that for a small value of the bucket limit, C, the mean page response time increases when the floor limit increases from 0 to 100%. The figure also shows that the new policing mechanism performs significantly better for small percentages of the traffic contract. If the token rate is chosen too small, TCP connections time-out and pages cannot be retrieved for the case of traditional policing. This underlines the superior performance of concurrent policing over traditional policing (floor limit equal to 100% in the proposed solution) and explains why Figure 4a displays no data for very small service contracts (40% of the original data and less) and high floor limits. In addition, the proposed mechanism can gradually change its behavior from policing with complete sharing of the tokens all the way up to a strict separation of the different directions and token buckets. The newly developed policing mechanism clearly suffers less from this phenomenon and pages could even be retrieved for token rate set on 10% of the original traffic contract (naturally with high mean page response times and hence not shown in Figure 4a to maintain its readability). If the bucket limit C increases, the mean page response increases at a slower pace (see Figure 4b for a bucket limit of 30s). Figure 5 shows this aspect. By defining the relative gain as the difference of the mean page response time between the



(a) Bucket limit, C, equals *10 seconds*. (b) Bucket limit, C, equals *30 seconds*.

Fig. 4: Mean page response time of the browsing application for various levels of the traffic contract as function of the floor limit F. There is *one user* that makes use of the traffic contract.

newly (for floor limit set to 0%) and traditional policing (floor limit set to 100%) mechanisms divided by the mean response time of the new policing mechanism. Figure 5 shows that the relative gain decreases when the bucket size increases. This phenomenon is likely caused by defining the floor limit as a percentage of the bucket size which is expected to result in the situation that upstream token bucket hardly reaches the floor limit and hence the exchange of tokens does not take place. Figure 6 shows that the mean page response time as function of the floor limit while the traffic contract needs to be shared by five users remains flat, indicating that the gain in performance and airtime usage of the mechanism that polices traffic in two directions concurrently decreases when the contract needs to be shared among various users or TCP connections.

3.4 Results for 50/50 contracts

Comparing Figure 7 with Figure 4b shows that the newly developed policing mechanism improves the mean page response time for the advanced knowledge contract to exceed those from the 50/50 contract conditions, the same applies to the situation where the traffic contract has to be shared among 5 wireless users (not shown). Where in the case of advanced knowledge TCP sessions would timeout for smaller load values, this was not observed for any of the experiments using a 50/50 contract. In fact, the obtained mean response time for the 50/50 contract is far less sensitive to the floor limit. This behaviour may be explained by considering that the floor limit is based on the token rate that is now equal for both directions and facilitates apparantly the token exchange for almost the whole range of bucket limits. Figure 8 nicely illustrates this and suggests that the optimum gain for the 50/50 contract is reached when the bucket size exceeds the time between two the page inter click times (set to 36.8 seconds) and then declines slowly.

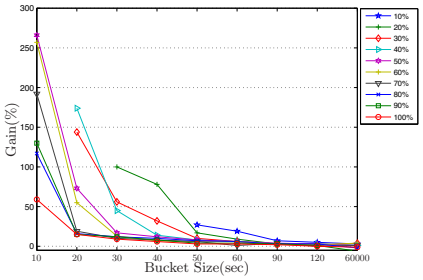


Fig. 5: The gain in response time as a function of the bucket limit C. The gain is expressed as the ratio of the mean page response time of the discussed policing solution divided by the mean page response time for the traditional leaky bucket policing mechanism with *one user*.

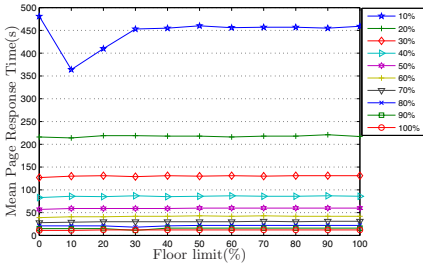


Fig. 6: Mean page response time of the browsing application for various levels of the traffic contract as function of the floor limit F. There are *five users* that make use of the traffic contract and the bucket limit, C, equals *30 seconds*.

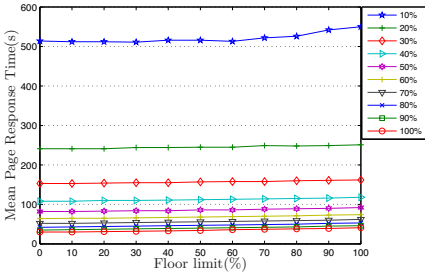


Fig. 7: Mean page response time of the browsing application for various levels of the traffic contract as function of the floor limit F. There is *one user* that makes use of the traffic contract and the bucket limit, C, equals *30 seconds*.

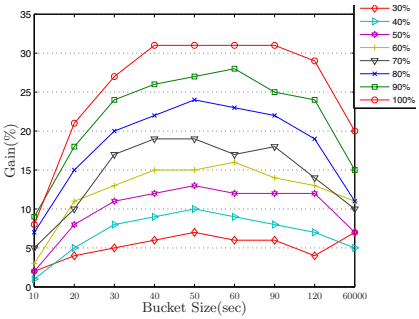


Fig. 8: The gain, expressed the ratio of the mean page response time of the discussed policing solution divided by the mean page response time for the traditional leaky bucket policing mechanism with *one user*.

4 Conclusions

In shared media, separating a traffic contract between end-user and network in the upstream and the downstream directions independently in general lowers the efficiency of the communication medium. Although the public and hotspot WLAN contracts are usually based on an *all you can eat* policy, this may change over time if the WLAN technology becomes more popular for usage outside the

home. The announced increase of WLAN in enterprises and the steep increase of dual model handsets that support WLAN are indicators that the WLAN traffic contract can follow the same path as public Internet access followed in the mid nineties: a cheap and unlimited contract in the beginning but a differentiated contract when popularity grows. Service level agreements as commonly seen in e.g., the xDSL access are then expected, inevitably resulting in various traffic contracts. The novel policing mechanism described in this paper that polices traffic in the upstream and downstream directions concurrently shows how the existing bucket mechanisms can be altered to use the shared WLAN medium more effectively. The more effective usage of the shared wireless media is, however, not the only result. Alternatively, the potential arises to create a more user friendly policing policy that regulates upstream traffic by not only taking into account the maximum agreed bitrate but also the consumption in the downstream direction. This provides a traffic contract with more flexibility for the end-users, where one day users may be more interested in consuming higher percentage of the upload capacity (e.g., gaining credits in P2P programs) whereas another day the download direction is used more extensively for downloading content. At the same time this flexibility also allows to police traffic to a fraction of the unpoliced volume while reducing TCP connection starvation.

The performance analysis performed in Section 3 shows that the novel policing mechanism that polices traffic in the upstream and downstream directions concurrently improves the performance of a shared medium when compared to applying the traditional policing solutions, especially for individual traffic contracts and for small traffic contract tolerances. In general, the positive effect in performance vanishes as the number of traffic flows increases that belong to the same contract. It then converges to the traditional approach of traffic policing mechanisms. In practice, these situations occur mainly in bulk contracts, where various users make use of a common access contract. In QoS enabled WLAN environments, the per-flow contract between end-user and WLAN operator is more common, which limits the number of simultaneous traffic flows per contract considerably. Exactly this circumstance provides a performance improvement when compared to applying traditional policing mechanisms.

5 Acknowledgments

The work reported in this paper was performed while the authors worked at Bell-Labs and was supported in part by the IST 6FP programme of the European Commission under contract OBAN, No 001889 and in part by the Netherlands Organisation for Scientific Research (NWO) under the Casimir project: Analysis of Distribution Strategies for Concurrent Access in Wireless Communication Networks. The authors kindly thank Sietse van der Gaast for his key contributions to the original idea on policing traffic in the upstream and downstream directions concurrently. The authors also thank Rob van der Mei and Michel Mandjes for their constructive discussions on the analytical performance analy-

sis and finally thank Matthew Andrews for his comments on earlier versions of this manuscript.

References

1. M.H. Ahmed. Call admission control in wireless networks: a comprehensive survey. *IEEE Communications Surveys and Tutorials*, 7(1):49–68, 2005.
2. S.C. Borst and D. Mitra. Virtual partitioning for robust resource sharing: computational techniques for heterogeneous traffic. *IEEE Journal on Selected Areas in Communications*, 16(5):668–678, 1998.
3. F.Panken, G.Hoekstra, D. Barankanira, C.Francis, R.Schwendener, O.Grondalen, and M.G.Jaatun. Extending 3G/WiMAX Networks and Services through Residential Access Capacity. *IEEE Communications Magazine*, 45(12):62–69, 2007.
4. D. Gao, J.Cai, and K.N. Ngan. Admission control in IEEE 802.11e wireless LANs. *IEEE Network*, 19(4):6–13, 2005.
5. G.J.Hoekstra and R.D. van der Mei. On the processor sharing of file transfers in Wireless LANs. In *VTC Spring*, page to appear, 2009.
6. V. Jacobson. Congestion avoidance and control. In *SIGCOMM '88: Symposium proceedings on Communications architectures and protocols*, pages 314–329, New York, NY, USA, 1988. ACM.
7. R. Litjens, F. Roijers, J.L. Van den Berg, R.J. Boucherie, and M.J. Fleuren. Performance analysis of wireless LANs: An integrated packet/flow level approach. In *Proceedings of the 18th International Teletraffic Congress - ITC18*, pages 931–940, Berlin, Germany, 2003.
8. Z. Liu, N. Niclausse, and C. Jalpa-Villanueva. Traffic model and performance evaluation of web servers. *Performance Evaluation*, 46(2-3):77–100, 2001.
9. D. Miorandi, A.A. Kherani, and E. Altman. A queueing model for HTTP traffic over IEEE 802.11 WLANs. *Computer Networks*, 50(1):63–79, 2006.
10. F.J.M. Panken and G.J. Hoekstra. Multi-service traffic profiles to realise and maintain qos guarantees in wireless lans. *Computer Communications*, 32(6):1022 – 1033, 2009.
11. K.W. Ross. Multirate loss models for broadband telecommunication networks. 1995.
12. S.Jaiswal, G.Iannaccone, C. Diot, J. Kurose, and D. Towsley. Inferring TCP connection characteristics through passive measurements. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, volume 3, pages 1582–1592 vol.3, 2004.
13. J.S. Turner. New directions in communications (or which way to the information age?). *IEEE Communications Magazine*, 40(5):50–57, 1986.

.

Analyzing the impact of relay station characteristics on uplink performance in cellular networks

D. C. Dimitrova[1], H. van den Berg,[1,2] G. Heijenk[1]

[1] University of Twente, Postbus 217, 7500 AE Enschede, The Netherlands,
{d.c.dimitrova,geert.heijenk}@ewi.utwente.nl
[2] TNO ICT, The Netherlands,
J.L.vandenBerg@ewi.utwente.nl

Abstract. Uplink users in cellular networks, such as UMTS/ HSPA, located at the edge of the cell generally suffer from poor channel conditions. Deploying intermediate relay nodes is seen as a promising approach towards extending cell coverage. This paper focuses on the role of packet scheduling in cellular networks with relay nodes. In particular, two uplink scheduling schemes deploying the relay functionality in different ways are compared in performance to a reference scenario where relaying is not used. We derive expressions which characterize for each of the two relay-enabled schedulers the service area of a relay station as a function of the relay location and transmit power. The results show that the service area is significantly influenced by the type of scheduling. Examining for both schedulers the impact on the effective data rates of mobile stations shows that there is an optimal combination of relay's position and transmit power which maximizes the service provided to all mobiles.

1 Introduction

Nowadays, a healthily functioning society is hardly imaginable without widely accessible and well operating communication networks. The tendency is towards wireless technologies providing high data rates and wide coverage. Pervasive wireless coverage, both indoors and outdoors, is hindered by the construction landscape of the area. An elegant way to improve performance in poor coverage area is to position a relay station (RS). A relay breaks up a direct communication path into two indirect paths. Given a well chosen position of the RS, indirect paths can provide better channel conditions than direct paths. In addition, relaying can be incorporated on both downlink, from base station (BS) to mobile station (MS), and uplink, from MS to BS. The application of relaying for coverage improvement has been studied for a broad range of wireless technologies, including WiMAX, see [10, 2], and cellular networks, see [9, 8, 12].

We propose a broader use of relaying, namely, to enhance offered service in general. In our proposal a mobile station, even if located in an area with coverage, can choose to use a relay if this improves its effective data rates. This paper is based on the EUL (Enhanced Uplink) technology which is described in the 3GPP Release 6 of the UMTS (Universal Mobile Telecommunications System) standard, see [1]. However, our approach has the potential to extend towards other cellular systems such as LTE technology, which is currently under standardization. Being the latest cellular technology

with high implementation percentage in Europe and North America and growing on other continents, UMTS/EUL is a good candidate to show the benefits of relaying for operators and mobile users.

The idea of relaying is rather attractive, however its implementation is not trivial. Placing a relay station within the cell poses new deployment decisions such as where to set the relay or how many do we need. In order to select the optimal solution we need to investigate how these decisions influence performance of the mobile stations. For example, [12] elaborates on the density or relays for the downlink in beyond-3G networks. Most importantly, the area over which a relay can improve performance, i.e. service area, has to be established.

An additional factor which influences the use of relays is the scheduling scheme at the base station. In order to understand why scheduling is relevant, we need to be familiar with certain features of EUL. In EUL the key resource is the maximum acceptable power received at the base station. The available channel resource is shared among all active users as the channel access is organized by the base station (BS) via time frames with fixed length, 2 or 10 ms, termed TTI (Transmission Time Interval). Furthermore, depending on location, a MS's transmit capacity may be insufficient to utilize the available resource, in which case diversity of scheduling schemes become an attractive choice. Several studies, for example [7, 3, 4], dedicated to scheduling in the EUL, show that the choice of scheme has a major impact on performance.

Our goal in this paper is to evaluate the performance of MSs under the combined influence of crucial relay characteristics and the type of scheduling at the BS. In particular, we are interested in finding optimal relay location and transmit power, which maximize the overal performance. Two relay-enabled scheduling schemes are proposed and investigated. For each scheme the mobile stations' performance at different locations is evaluated in terms of effective data rates. In addition we derive explicit analytical expressions to determine the relay's service area, which are subsequently supported by simulations.

Our work is related to [6] which discusses similar relay configurations for HSDPA. Although insightful, [6] does not provide any analytical derivations. On the contrary, [11] provides a very detailed cross-level analysis of relaying but does not consider the impact of particular scheduling schemes and other system characteristics. In summary, the most prominent contributions of our research are: discussion on the uplink of a cellular system; analytically defining the boundaries of a service area; and combined assessment based on both scheduling and RS specifics.

The paper continues as follows. First, in Section 2, we briefly discuss the relaying concept and describe the scheduling schemes considered in this paper. The model description and analysis appear in Section 3 and Secion 4, respectively. Section 5 presents our findings of the performance evaluation. Finally, Section 6 summarizes our work.

2 Relay-enabled Round Robin Scheduling Schemes

We would first discuss how relaying can be applied in a cellular system and will introduce relevant notation. Next, the scheduling schemes considered in this paper are described.

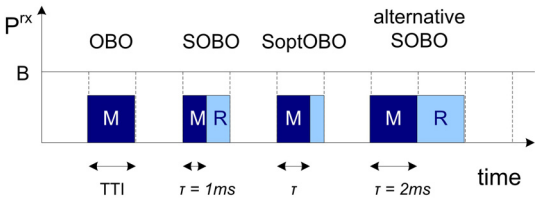


Fig. 1. Scheduling schemes: OBO, SOBO, SoptOBO, SidOBO and ExOBO

Relaying artificially breaks up a long communication path into two shorter ones. As result signal degradation decreases and received powers at the BS improve. These gains are partly lost to increased transmission time from the additional forwarding of data at the relay node. Thus, whether a relay can improve performance does not have a simple answer. In order to choose between an indirect path (MS-RS-BS) and a direct path (MS-BS) we need to evaluate the data rates realized on both. In a relay-enabled system the data of a MS which selects for the relay travels over two sub-paths, MS-RS and RS-BS, and the effective data rate on the indirect path depends on the data rates realized on the two sub-paths.

Each (sub-)path is characterized by a set of transmission parameters: the distance between the communicating devices d_{zz}, the path loss L_{zz}, the transmit power P_{zz}^{tx}, the duration of a transmission opportunity τ_{zz} and the instantaneous data rate r_{zz} during a transmission opportunity. The index zz refers to the specific (sub-)path, i.e. *ms* for the direct path from MS to BS, *mr* for the sub-path from MS to RS, and *rs* for the sub-path from RS to BS. The transmission times τ_{mr} and τ_{rs} are scheduler specific; their sum is denoted by $\tau = \tau_{mr} + \tau_{rs}$. Any further relations between the transmission parameters are discussed in Section 4. We now continue with introducing the scheduling schemes considered in our study.

Scheduling Schemes

The discussed schedulers belong to the Round Robin (RR) family where mobile users are served one-by-one (OBO), independently of their channel conditions. According to several studies, for example [7, 5], OBO is a rather inefficient in resource utilization if users with limited power capacity are served. Still we choose OBO since we expect relaying to increase the capability of a MS to fully use the available resource. In our study we consider two variants of a relay-aware OBO scheduler: *SOBO* and *SoptOBO*. A MS can select the direct or the relay path, depending on its location relative to the BS and the RS. In addition, plain *OBO* is considered as a reference scheme in which a MS always transmits directly to the base station independently of its location in the cell. All schemes assign a single TTI for the transmissions on direct and indirect paths, i.e. $\tau = 2ms$.

On the indirect path, a *shared OBO* (SOBO) scheme divides the TTI into two equal intervals of $1ms$ and MS and RS both receive one interval to transmit, i.e. $\tau_{mr} = \tau_{rs} = 1ms$. The benefits of working with fixed-length transmission times for direct and indirect paths exhibit during implementation. However, static subdivision for indirect paths is not the most efficient choice when the instantaneous rates on the sub-paths differ.

Apart from SOBO, various other schemes which apply fixed-length transmission times can be defined. For example, using a single TTI for the transmissions on both sub-paths, i.e. $\tau_{mr} = \tau_{rs} = 2ms$, meaning the BS reserves 2 TTIs for the service of a single MS. It can be shown that all schemes which, independently of precise values, use equal division of τ, i.e. $\tau_{mr} = \tau_{rs}$ have the same effective data rates. Therefore, we can claim that SOBO is representative for a group of schedulers.

In the *Optimized SOBO* (SoptOBO) the channel utilization for indirect paths is optimized by selecting the transmission times on the sub-paths such that both sub-paths MS-RS and RS-BS match in transmission capacity, i.e. $\tau_{mr} r_{mr} = \tau_{rs} r_{rs}$. Despite its maximized resource utilization SoptOBO is rather challenging for implementation since individually selecting transmission times for the sub-paths requires complex functionality in the base station.

3 Model

The model we consider consists of a single cell with EUL users (MSs) at random locations within the cell. A MS selects a direct or indirect path depending on which one provides higher effective data rate/ received power. Which parameter is used is a scenario specific choice and is explicitly indicated. All mobile stations are assumed to have the same maximum transmit power capacity $P^{tx}_{ms,max}$. The maximum transmit power of the RS is $P^{tx}_{rs,max}$. However, depending on location and the available budget B, MS/RS use either the maximum transmit power or a lower power, i.e. $P^{tx}_{zz} \leq P^{tx}_{zz,max}$[3]. The actually applied transmit power is chosen such that a MS maximizes its utilization of the budget B.

Given that we want to calculate the service area of a relay, we consider a cell with a single relay station. The service area is the aggregation of all locations within the cell from which a MS receives better service via the relay station. In addition, we are interested how the relay position and transmit power influence MS performance. Their impact is evaluated by both mathematical expressions and Monte Carlo simulations. We assume that a RS use the same frequency band for receiving and transmitting and that the change of mode is instantaneous.

At both BS and RS limited channel resource, budget B, is assumed. Given noise rise η and constant thermal noise N, the shared budget B at the BS, and RS, is derived: $B = \eta N$. Intentionally disregarding important factors such as inter- and intra-cell interference allows us to identify the effect relaying has on performance. However, we realize that such factors may have significant effects on the performance gains.

4 Analysis

In this section we concentrate on two distinctive aspects of relaying - the service area of the relay station and its impact on MS performance. Relaying has two opposite effects on mobile stations. On the one hand higher received powers are enabled (increasing the effective data rate), but on the other hand forwarding at the RS requires an additional

[3] Note that $P^{tx}_{ms,max} = P^{tx}_{mr,max}$.

transmission of the data (decreasing the effective data rates). For each scheduler we start with calculating the received powers from which subsequently data rates can be derived.

According to its definition the service area is the collection of locations from which the indirect path is preferred. Depending on the selection criteria we differentiate between service area based on received powers and based on effective data rates. For both cases, we derive generic expressions which allow us to calculate the service area as a function of the RS's position and transmit power.

4.1 Received Powers from Mobile Stations

According to signal propagation law, the received power P_{zz}^{rx} on any communication path is determined by the applied transmit power $P_{zz}^{tx} \leq P_{max}^{tx}$ and the path loss $L_{zz}(d)$. The maximum possible received power on an EUL path is limited by the available budget B at the receiver leading to:

$$P_{zz}^{rx} = min(\frac{P_{zz,max}^{tx}}{L_{zz}(d)}, B) \tag{1}$$

where the index zz=(ms,mr,rs) denotes the (sub-)path over which the transmission is done. The assumed path loss model is given by $L(d) = 123.2 + 10alog10(d)$ (in dB) with $a = 3.52$ the path loss exponent and d the distance in kilometer.

Note that in SOBO the received powers at the base and relay station are the same since the indirect path transmission is limited by the slower sub-path. In SoptOBO however, according the definition, generally these two received powers are different, as the unbalance in received powers is compensated for by difference in transmission time, see Section 2.

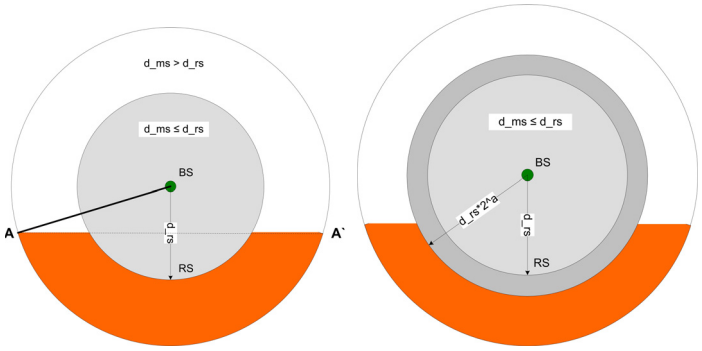
4.2 Effective Data Rates

The data rate achievable on a (sub-)path zz given a particular transmit power capacity is the *instantaneous rate* r_{zz}. Its dependency on other transmission parameters is given by:

$$r_{zz} = \frac{r_{chip}}{E_b/N_0} \cdot \frac{P_{zz}^{rx}}{N + (1-\omega)P_{zz}^{rx}} \tag{2}$$

In Equation (2) r_{chip} is the system chip rate and E_b/N_0 is the energy-per-bit to noise ratio. The parameter ω is used to account for reflected signals and the index $zz = (ms, mr, rs)$ refers to the (sub-)path. The maximum possible data rate a MS can realize is determined by the condition that the budget B can be filled, i.e. $P_{zz}^{rx} = B$.

The *effective rate* r_{eff} is the rate realized by a MS during the transmission opportunity τ. On the direct path, indexed ms, the effective rate is the same as the instantaneous, i.e. $r_{eff} = r_{ms}$, because the whole τ is used by the MS. On the indirect path however, due to data forwarding, the effective rate is lower than the instantaneous. r_{eff} depends on the part of τ used by the mobile, i.e. on τ_{mr}, and on the instantaneous rate that the MS can realize during τ_{mr}. In SOBO $\tau_{mr} = \tau/2$ while in SoptOBO $0 < \tau_{mr} < \tau$ depending



(a) Recevied power based SA (b) Effective rate based SA

Fig. 2. Service area (SA) of a single relay station based on: (a) received power leves; and (b) effective data rates.

on MS location such that $r_{mr}\tau_{mr} = r_{rs}\tau_{rs}$ and $\tau_{mr} + \tau rs = \tau$. Given the scheduler specific time assignment policy the expression of the effective rate becomes:

$$r_{eff} = \begin{cases} r_{ms} & \text{for OBO and direct path in SOBO\&SoptOBO} \\ min(r_{mr}, r_{rs}) * \frac{1}{2} & \text{for indirect path in SOBO} \\ \frac{r_{mr}\tau_{mr}}{\tau} & \text{for indirect path in SoptOBO} \end{cases} \quad (3)$$

4.3 Service area of a relay station

Our goal in this sub-section is to analytically determine the critical distance d_c at which a MSs changes its preference from direct to indirect path and which determines the boundaries of the service area. We will show that the shape and surface of the service area depend on whether received powers or effective data rates are used in the analysis. As we showed in Section 4.2, due to data forwarding the gains in effective rates are lower than what received powers 'promise'. Therefore, it is more appropriate to calculate service area based on effective rates. However, working with received powers provides us with a good reference base to illustrate our claim.

Service Area Based on Received Powers

In this scenario, a mobile station selects the transmission path - direct or indirect - which can offer higher received power at the base station. Recall that in SOBO the received power for the indirect path is determined by the sub-path with poorer channel conditions. The condition to select indirect transmission can be written as:

$$P_{ms}^{rx} < min(P_{mr}^{rx}, P_{rs}^{rx}) \quad (4)$$

From Equation (1), assuming the same transmit power for MS and RS, we can deduce that the received power is only dependent on the distance between the communicating stations. Therefore, the condition from Equation (4) can be rewritten as:

$$d_{ms} > max(d_{mr}, d_{rs}) \quad (5)$$

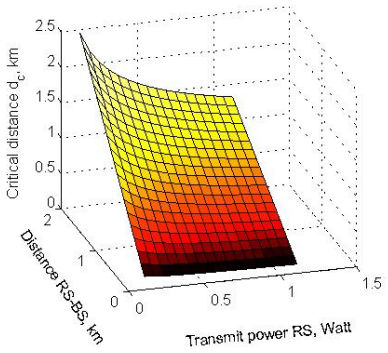


Fig. 3. Dependency of the critical distance on both distance of relay to the base station and the transmit power of the relay for a SOBO scheme.

We will now explain how Equation (5) can be transformed to a spatial limitations in the cell. A circle with radius d_{rs} around the base station can be drawn, see Figure 2(a), within which $d_{ms} \leq d_{rs}$ holds and a MS always selects the direct path; outside it - the indirect is preferred. The single relay station in our model is not sufficient to cover the whole disc for which $d_{ms} > d_{rs}$ is valid. However, by applying basic geometry rules for medians we can find the line AA', see Figure 2(a), all points of which have the same distance to the base station as well as to the relay. All MSs 'below' AA' have $d_{ms} > d_{mr}$ and can benefit from relaying. The intersection of the two inequalities defines the relay's service area.

Service Area Based on Effective Data Rates

In order to keep calculations tractable, yet providing useful insights, we consider the scenario when BS, RS and MS lay on a straight line. Both schemes - SOBO and SoptOBO - are discussed as we provide analytical expressions to find the critical distance. We begin with discussion on SOBO. A MS always selects the path, i.e. direct or indirect, which provides higher effective data rate. The critical distance d_c is then the distance d_{ms} for which direct and indirect paths realize the same effective rate. According to Equation (3), the condition can be formally expressed as:

$$r_{ms} = min(r_{mr}, r_{rs}) * \frac{1}{2} \tag{6}$$

Note that, given a straight line BS-RS-MS, when $d_{mr} > d_{rs}$ holds by default indirect transmission is chosen. Substituting with Equations (2) and (1) and solving for d_{ms}, we can derive a formal expression for d_c, namely:

$$d_c = d_{rs} * \left(2 \frac{P^{tx}_{ms,max}}{P^{tx}_{rs,max}} \right)^{1/a} \tag{7}$$

This general expression takes both characteristics of the relay - transmit power P^{tx}_{rs} and position d_{rs} - as parameters; a is the path loss exponent. The dependency is graphically presented in Figure 4.3. When RS have the same transmit capacity as of a mobile

the equations reduces to $d_c = d_{rs} * (2)^{1/a}$. The discussion continues with analysis for SoptOBO. From Equation (3) follows that the critical distance d_c is determined by the equality:

$$r_{ms} = \frac{r_{mr}\tau_{mr}}{\tau} \tag{8}$$

Per definition SoptOBO is optimized for the transfer of equal amount of data on both sub-paths. In combination with the limitations set on the transfer times, that knowledge allows us to express τ_{mr} in terms of rates, namely:

$$\begin{array}{rl} r_{mr}\tau_{mr} & = r_{rs}\tau_{rs} \\ \tau_{mr} + \tau_{rs} = \tau & \Rightarrow \tau_{mr} = \tau * \frac{r_{rs}}{r_{mr}+r_{rs}} \end{array} \tag{9}$$

After several substitutions and solving Equation (8) for d_{ms}, the critical distance can be given as a function of the positions of both, the relay and the mobile station. For a solution dependable only on relay characteristics we need to solve the system of equations:

$$\begin{aligned} d_c &= \left(d_{mr}^a + \frac{P_{ms}^{tx}}{P_{rs}^{tx}} . d_{rs}^a \right)^{1/a} \\ d_c &= d_{mr} + d_{rs} \end{aligned} \tag{10}$$

The system of equations (10) is in general difficult to solve explicitly and we therefore use numerical approaches. Again if we assume the relay to have the same transmit capacity as a mobile station Equation (10) simplifies to $d_c = (d_{mr}^a + d_{rs}^a)^{1/a}$.

5 Numerical results on relaying

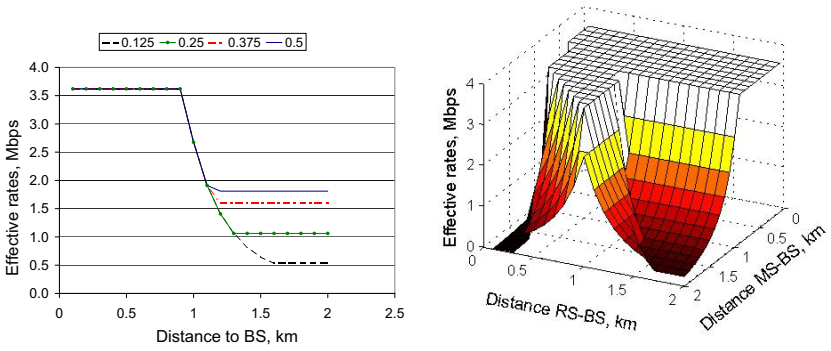
This section presents a quantitative evaluation of the impact a relay station has on the performance of mobiles. First, we discuss how key characteristics of the relay such as position and transmit power influence the effective data rates of the MSs, Sections 5.2 and 5.3 respectively. In both cases straight line BS-RS-MS is discussed. Since these are relay and not scheduler related characteristics we present the results only for SOBO. In the rest of the numerical results the relay station is located at 1 km from the BS and has transmit power $P_{rs}^{tx} = 0.125$ Watt, if not otherwise specified.

Subsequently, the service area of a relay station is presented in Section 5.4. We compare the results of Monte Carlo simulations for SOBO and SoptOBO. OBO does not use relaying and is excluded from the discussion. We generate 500 000 locations and for each compare whether direct or indirect path provides higher effective rate.

Finally, the three scheduling schemes - OBO, SOBO and SoptOBO - are compared in terms of effective data rates. The results are generated for 500 000 randomly taken locations of MSs by applying the analytical expressions of Section 4.2. On the hand of spatial graphs we illustrate how relaying can improve performance and what scheduling at the relay is more beneficial.

5.1 Parameter Settings

In the numerical experiments we apply system chip rate r_{chip} of 3840 kchips/s, thermal noise level N of -105.66 dBm and noise rise target η of 6 dB. From the given noise rise



(a) Tranmsit power of RS, d_{rs}=1.3 km (b) Position of RS

Fig. 4. Impact of RS characteristics on MS performance with SOBO scheduler: (a) RS transmit power; (b) RS position.

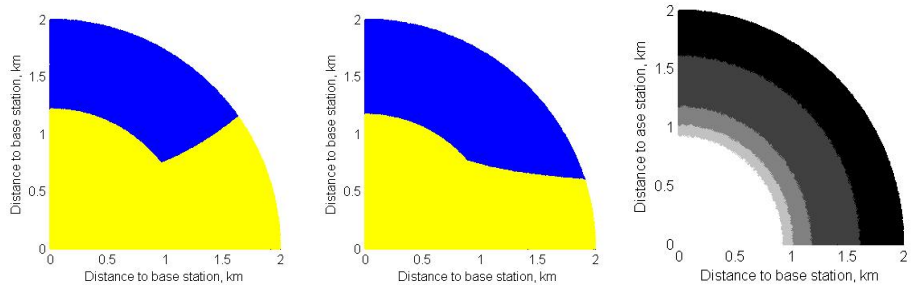
and thermal noise the available budget at both BS and RS can be calculated: $B = \eta * N$. Self-interference of 10% is considered, i.e. $\omega = 0.1$. All EUL flows are taken with E_b/N_0 target of 5 dB. By default, a single relay station is located at 1 km from the base station and has P_{rs}^{tx} either 0.125 Watt. Mobile stations have maximum transmit power $P_{\max}^{tx} = 0.125$ Watt.

5.2 Transmit Power of the Relay Station

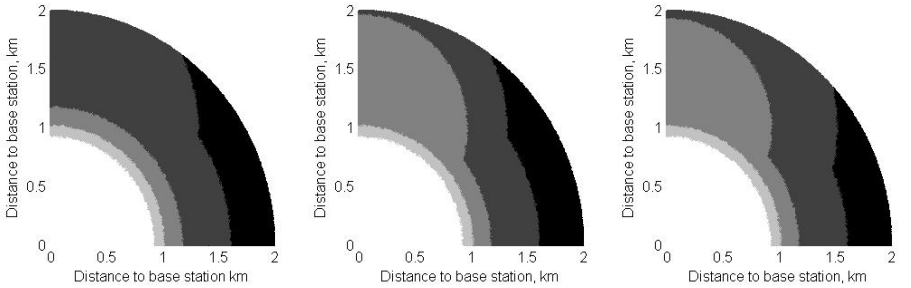
Figure 4(a) presents the results for four different RS transmit powers, namely (0.125 0.25 0.375 0.5) Watt, given the RS is located at 1.3 km from the base station. If the MSs transmit directly to the base station the graphs coincide and the differences start to appear when the relay is used. Interestingly, as the transmit power grows the distance at which indirect transmissions are preferred decreases. Higher RS transmit power leads to better budget utilization - better than a MS closer to the BS but with lower transmit capacity can achieve.

A MS on the direct path realizes the maximum possible effective data rate when it can fully utilize the available budget B, suggesting $P^{rx}(BS) = B$. Given a transmit power, from Equation (1) we can calculate the maximum distance for which $P^{rx}(BS) = B$ holds, namely $d_{ms,max}$. With the chosen parameter settings $d_{ms,max}$ equals 0.9 km, resulting, according to Equation (2), in 3.6 Mbps. As Figure 4(a) shows, 3.6 Mbps is the effective rate for all MSs at up to and including 0.9 km after which increasing the distance leads to decrease in the rate.

Given a cell radius of 2 km, the furthermost MS is located at 0.7 km from the relay. Thus all indirect transmissions are relay-limited, which explains the flat graphs of Figure 4(a) after relay is used. Under the condition of fully used budget B, i.e. the case of $P_{rs}^{tx} = 0.5$ Watt, the maximum possible effective rate on the indirect path is twice smaller compared to the direct path due to data forwarding, see Equation (3).



(a) Service area SOBO (b) Service area SoptOBO (c) Effective rate OBO



(d) Effective rate SOBO (e) Effective rate SOBO 0.25 Watt (f) Effective rate SoptOBO

Fig. 5. Spatial behaviour in the cell with a relay station located at 1 km and transmit power 0.125 Watt. Shape of service area for: (a) SOBO and (b) SoptOBO. Effective data rates for: (c) OBO; (d) SOBO; (e) SOBO with RS transmit power 0.25 Watt; and (f) SoptOBO.

5.3 Position of the Relay Station

The three dimensional graph in Figure 4(b) illustrates how the effective data rates change when the location of both mobile and relay station changes. It seems that a RS at 1 km has optimal performance. MSs close to the base station can use the total resource on their own and therefore show the same performance, i.e. flat area. The same holds for the relay station with the exception of remote MSs, e.g. at 1.5 km, whose own distance to the relay limits the performance. Moving the relay further away has a two-fold effect. On the one hand, increased distance to the base station leads to lower effective rates. On the other hand, the number of MSs which can benefit from relaying decreases. Furthermore, not only the number of directly sending MSs increases but their effective rates decrease due to longer distance to the BS.

5.4 Service Area

In this section the service area of a relay station at 1 km from the base station and with transmit power of 0.125 Watt is discussed. The service area is based on effective data rates. The results for SOBO and SoptOBO are presented in Figures 5(a) and 5(b)

respectively, where the dark surface corresponds to the service area, i.e. indirect transmissions, and the light surface - to all locations from which direct transmissions are preferred. The service area for only quarter of the cell is given since the area is symmetrical around the relay. Obviously, MSs from the other half of the cell always choose the direct path.

Comparing Figures 5(a) and 5(b) we can conclude that using SoptOBO extends the area mainly for locations far from the relay while the changes for close ones, i.e. at 1.2 km, are negligible. The bigger the difference in effective rates on the two sub-paths for SOBO, the bigger the improvements SoptOBO can offer. Since the asymmetry is larger for far self-limiting MSs they also gain the most with SoptOBO.

Note that the service area does not start right behind the relay which supports our conclusion from Section 4.3 that the actual service area is smaller than the improvements in received powers may suggest. Solving Equation (10) on the straight line BS-RS-MS for SOBO results in 1.22 km, what Figure 5(a) indicates as well.

Knowing the service area of a relay could assist us in choosing the number of relays to cover the whole outer part of the cell. However, it is more important to observe the effect of relaying on the MSs' performance.

5.5 Effective Data Rates

The spatial distribution of the effective rates for the three schedulers - OBO, SOBO and SoptOBO - is shown in Figures 5(c) to 5(f). In addition, an SOBO scheduler with higher transmit power of the relay is considered. Again a quarter cell is depicted as the brightness changes from high to low as the rates decrease. MSs located in the central white sector with sector around the BS with radius 0.9 km can fully utilize the available budget B. In all scenarios where relaying is used performance visibly improves.

Increasing the RS transmit power, see Figure 5(e), is only beneficial for RS-limited transmissions, which is most commonly the case of MSs close to the relay. Since the relay has the same available budget B as the BS, the same calculations for $d_{ms,max}$ hold, see Section 4.1. Thus explaining the circle with radius 0.9 km around the RS in Figure 5(e).

A relay configuration with SoptOBO scheme delivers better performance and larger service area than a SOBO scheduler offers, see Figure 5(f). Adapting the transmit times according to the MS location gives each mobile the opportunity to improve its effective rate. The improvement depends, as explained, on the instantaneous data rates on the two sub-paths. We do not expect significant changes if the RS transmit is increased with SoptOBO.

6 Conclusion

This paper discussed the benefits that relaying has to offer to mobile users by comparing the performance of two relay-enabled schemes to a reference scheduler with no relay. To analytically describe the service area of a relay we provided two independent approaches - based on received powers at the base station and based on effective data rates. These approaches are supported by simulations. The results show that the

shape and size of the service area depend on the relay characteristics but also on the scheduling schemes. We also evaluate how relay characteristics such as transmit power and location affect the effective rates of the mobiles. Interestingly, there is an optimal relay location, which maximizes the performance for all mobiles. Furthermore, we established that the maximum possible data rate is mostly limited not by the relay or the mobile but by the available resource at the base station.

It is our expectation that the changing number of active mobile users in a real system will influence the results. Therefore, we examine this in our current research. Another topic for further research is considering the impact of environment parameters such as inter-cell interference.

Acknowledgments We are greatful to Remco Litjens from TNO ICT, The Netherlands for the appropriate remarks on modelling and for his continuouse enthousiasm to be always of assistance.

References

1. 3GPP TS 25.309. FDD Enhanced Uplink; Overall Description.
2. Albert Bel and Gonzalo Seco-Granados Jose Lopez Vicario. The benefits of relay selection in WiMAX networks. In *ICT-MobileSummit '08*, 2008.
3. D. C. Dimitrova, H. van den Berg, G. Heijenk, and R. Litjens. Flow-level Performance Comparison of Packet Scheduling Schemes for UMTS EUL. volume 5031. WWIC '08, Tampere, Finland, 2008.
4. A. Mäder, D. Staehle, T. Liu, and H. Barth. Feasible load regions for different RRM strategies for the enhanced uplink in UMTS networks. In *Springer, Lecture Notes in Computer Science*, volume 4396, pages 213–228. EuroNGI Workshop, 2006.
5. S. Ramakrishna and J. M. Holtzman. A scheme for throughput maximization in a dual-class CDMA system. ICUPC '97, San Diego, USA, 1997.
6. Eike Reetz, Rainer Hockmann, and Ralf Tonjes. Performance Study on Cooperative Relaying Topologies in Beyond 3G Systems. In *ICT-MobileSummit '08*, 2008.
7. C. Rosa, J. Outes, T.B. Sorensen, J. Wigard, and P.E. Mogensen. Combined time and code division scheduling for enhanced uplink packet access in WCDMA. IEEE VTC '04 (Fall), Los Angeles, USA, 2004.
8. R. Schoenen, R. Halfmann, and B.H. Walke. An FDD Multihop Cellular Network for 3GPP-LTE. pages 1990–1994, 2008.
9. Martina Umlauft. Relay Devices in UMTS Networks - effects on Application Performance. In *Proceedings of the Fifth Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net 2006)*, 2006.
10. Josep Vidal, Olga Munoz, Adrian Agustin, Eduard Calvo, and Andreu Alcon. Enhancing 802.16 networks through infrastructure-based relays. In *ICT-MobileSummit '08*, 2008.
11. E. Weiss, S. Max, O. Klein, G. Hiertz, and B. Walke. Relay-based vs. Conventional Wireless Networks: Capacity and Spectrum efficiency. pages 1–5. Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on, 2007.
12. Marc Wener, Peter Moberg, and Per Skillermark. Cost assessment of radio access network deployment with relay nodes. ICT-MobileSummit 2008 Conference Proceedings, 2008.

Combined Coverage Area Reporting and Geographical Routing in Wireless Sensor-Actuator Networks for Cooperating with Unmanned Aerial Vehicles

Lodewijk van Hoesel, Aysegul Tuysuz-Erman and Paul Havinga

Department of Electrical Engineering, Computer Science and Mathematics,
University of Twente
P.O. box 217, NL-7500 AE Enschede, THE NETHERLANDS
{l.f.w.vanhoesel, a.tuysuz, p.j.m.havinga}@utwente.nl

Abstract. In wireless sensor network (WSN) applications with multiple gateways, it is key to route location dependent subscriptions efficiently at two levels in the system. At the gateway level, data sinks must not waste the energy of the WSN by injecting subscriptions that are not relevant for the nodes in their coverage area and at WSN level, energy-efficient delivery of subscriptions to target areas is required. In this paper, we propose a mechanism in which (1) the WSN provides an accurate and up-to-date coverage area description to gateways and (2) the wireless sensor network re-uses the collected coverage area information to enable efficient geographical routing of location dependent subscriptions and other messages. The latter has a focus on routing of messages injected from sink nodes to nodes in the region of interest. Our proposed mechanisms are evaluated in simulation.

Keywords: Wireless sensor networks, geographical routing, coverage area reporting, multi-sink networks

1 Introduction

The AWARE project (EU IST-2006-33579) considers self-deploying of wireless communication infrastructure with autonomous, unmanned aerial vehicles (UAVs) [1]. The AWARE platform targets to enable operation in sites which are difficult or impossible to access and which are without a pre-existent communication infrastructure. One of the focus application scenarios of the AWARE project is disaster management and civil security, in which wireless sensors collaboratively detect critical events (such as fire), or continuously monitor environmental conditions. In these applications, wireless sensors are the ears and eyes of the AWARE platform. They are added to the network on the fly and might be attached to mobile objects.

When wireless sensor networks (WSNs) contain multiple gateways, it is key to handle location dependent subscriptions efficiently to the set of gateways that service the particular region of interest. In the envisioned AWARE application scenarios, data sinks are interconnected via a powerful mobile ad-hoc network (MANET) and each communicates with a subset of the sensor network. Furthermore, data sinks

collaborate with other MANET enabled devices to extract contextual information from the sensor network by inserting subscriptions. These subscriptions inform the wireless sensors which information needs to be published and are only inserted into the (local) sensor network if relevant.

In this paper, we propose a mechanism in which (1) the wireless sensor network provides an accurate and up-to-date coverage area description to gateways and (2) the wireless sensor network re-uses the collected coverage area information to enable geographical routing of location dependent subscriptions and other messages. The latter has a focus on routing of messages injected from sink nodes to nodes in the region of interest (Figure 1).

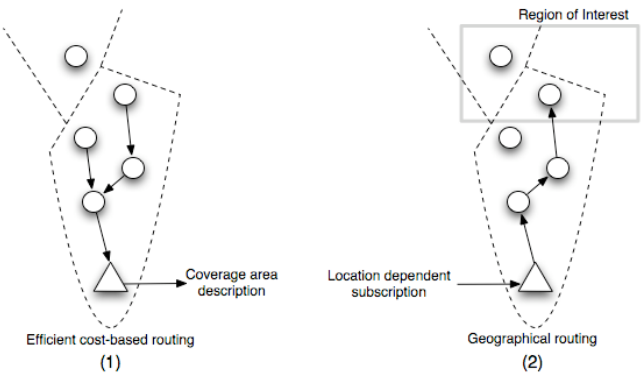


Figure 1: The WSN provides coverage area description to gateways (1) and reuse of collected information to route location dependent subscriptions (2)

2 The AWARE platform

The application scenarios considered in the AWARE project [1] motivate the research presented in this paper. The goal of the AWARE project is to develop a platform of self-deploying and self-organizing wireless sensor networks in collaboration with autonomous helicopters. The architecture of the AWARE platform comprises a number of heterogeneous sub-systems, which are described in relation to the global architecture in Figure 2. We have two key system layers of abstraction: the *sensor and dynamic networking layer*, and the *distributed services layer*.

The sensor and networking layer contains the sensor and the network protocols, which allow messages to be forwarded through multiple sensors taking into account the mobility of nodes and the dynamic change of topology. Assignment of each node to a sink in a reliable manner and handling the dynamics of the mobile sinks and sensors, and change of assignments are the concerns of this layer. The wireless sensor network can contain multiple mobile sinks e.g. attached to the helicopters, other vehicles, or humans. These sinks can communicate directly with each other via MANET links.

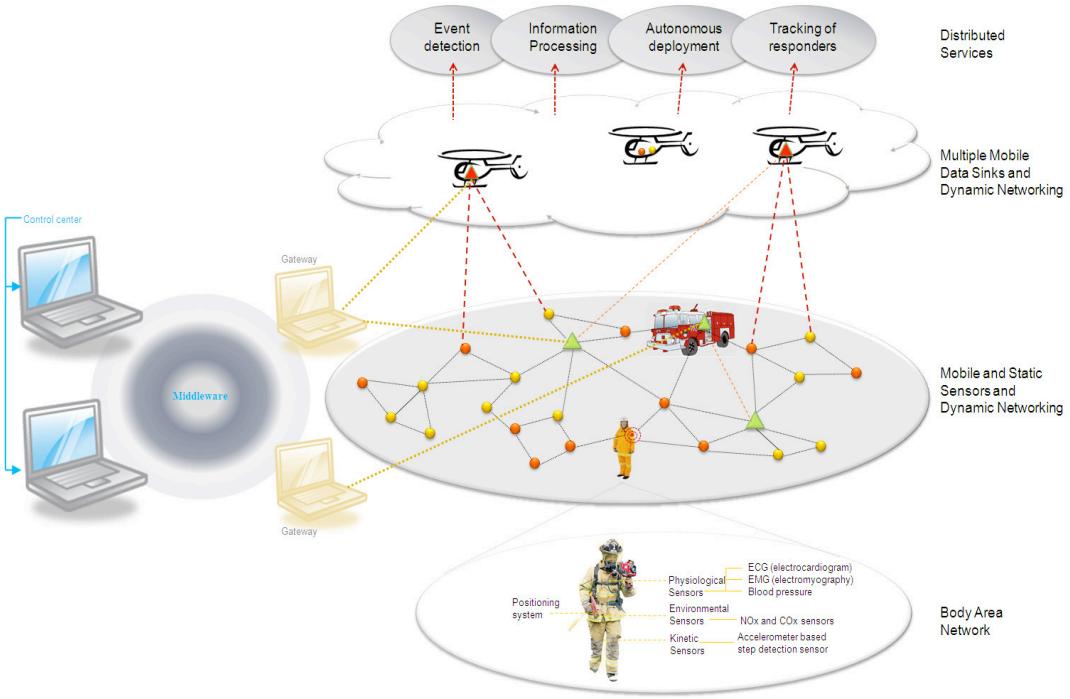


Figure 2: Overview of the AWARE platform architecture [1]

The distributed services layer contains different services to support mission critical management. We have identified four major services with the corresponding opportunities. The event detection supports reliable and timely detection of events. It is even capable of monitoring events in critical regions with mobile sensors. The information processing service deals with aspects of collecting and processing data. This service allows vast quantities of data to be easily and reliably accessed, aggregated, manipulated, filtered, disseminated, and used in a customized fashion by applications. The autonomous deployment supports detecting routing holes in the network and sends UAVs carrying sensors on-board to these regions to deploy additional nodes. It provides the ability of dynamically adapting the network to the requirements of the situation by increasing the coverage or repairing the connectivity of the network. Tracking of responders is also very important for safety-critical events. The body area network is used for this purpose. Readings from sensors on responders are collected/processed/integrated to provide a better insight into the user's state.

The coordination of the elements in the system is carried out by a control center. The middleware depicted in Figure 2 provides a publish/subscribe communication interface between all devices such as UAVs, responders and sensors in the system. Devices that produce data register themselves as data publishers. The middleware then creates the corresponding abstract data channel that takes care of taking this

information to other devices, which have registered themselves as subscribers to receive the data. Since the middleware tracks the data flow in the AWARE system, it can deliver the statistic data on system functionality to the control center to monitor the state of the system and its components. Also, the collected data can be archived in control centers for future information retrieval.

3 Related work

Nodes in a multi-hop wireless sensor network collaborate in forwarding packets to their destination(s) [2]. By the term routing protocol, we understand the mechanisms to select the ”best” node out of the set of nodes in radio range for forwarding the data to its final destination. In general, routing protocols try to optimize global performance (i.e. for example minimize network-wide energy consumption) by making local decisions on the best node to forward the data to. Another message routing strategy for wireless sensor networks is described in [3]: *geographical* routing. Instead of advertising an interest for data, or requesting to establish a route to a certain destination device, nodes use a routing technique based on node coordinates. Nodes are assumed to know their own position and the position of the sink node (i.e. the node where the data needs to be delivered). The idea is that nodes advertise data along with the coordinates where it must be delivered. Nodes closer to the sink node consider themselves candidates for relaying the message.

Face routing [4] routes packets along faces of planar network graphs by using simple *right hand rule* and proceeds along the line connecting the source and the sink. Although it guarantees to reach the destination, it does so with $O(n)$ messages, where n is the number of network nodes, and a simple flooding algorithm already reaches the destination with $O(n)$ messages. Also, it is not competitive with the shortest path algorithm in terms of cost depending on the number of hops between the source and the destination.

Adaptive Face Routing (AFR) [5] is the first algorithm competitive with the shortest path between the source and the destination. It basically enhances Face Routing [4] by the concept of an ellipse-bounding region restricting the searchable area. With a lower bound argument AFR was shown to be asymptotically optimal. On the other hand, AFR is not practicable due to its pure face routing concept. For practical purposes there have been attempts to combine greedy approaches (always send to the message to the neighbor closest to the destination) and face routing; for example *Greedy Perimeter Stateless Routing* (GPSR) [6], however, without competitive worst-case guarantees. There have been some other proposals for practical purposes to combine Greedy routing with face routing like the GOAFR and GOAFR+ algorithms by Kuhn et al. [7,8], which remain worst-case optimal.

In most of these protocols, the packets are sent from source to a destination position. For some other scenarios like AWARE scenario given in Section 2, it is also sufficient for some packets (e.g. subscriptions, etc.) to reach any destination currently located in a given area (i.e. geo-casting). Yu et al. in [9] proposes *Geographical and Energy-Aware Routing* (GEAR) algorithm, which shows how to broadcast a message to all the nodes in a target region. GEAR uses greedy forwarding to forward packets

to the nodes that are always progressively closer to the centroid of the target region, whilst trying to balance the energy consumption at the intermediate nodes. Once the message is delivered to the centroid of the target region, it then uses restricted flooding, namely *Recursive Geographic Forwarding*, to broadcast the message all remaining nodes in the given region.

There are some other protocols based on *window spanning infrastructure* (WSI) for routing to the specified message window (i.e. destination region). In this approach, the message first is forwarded towards the message window by an end-to-end routing protocol. Once the message reaches the window, an infrastructure within the message window is built along with the message propagation. The method in [10] uses a Greedy technique to find a routing path from message originator to a node N_c located at the center of the message's spatial window. This first part of the routing is similar with the approach used in GEAR. For the routing inside the window, the framework proposed in [10] uses two different approaches namely *WinFlood* and *WinDepth*. The *WinFlood* algorithm consists of a constrained parallel flooding, where a node broadcasts the message to its neighbors only if its own location is inside the message's spatial window. The alternative solution, *WinDepth*, is based on depth first search policy.

As we have seen from the related works given in this section, the first step of the window message processing techniques is generally based on Greedy approach which cannot guarantee that a routing path to a node in the message's spatial window will be found. The main difference between the protocols is observed in the second phase that is the routing inside the specified message window. However, our approach uses a different technique, based on *coverage area description*, in the first phase of the routing that is forwarding the packet from source to the given area. In the following sections, our approach is described in detail.

4 Distributed coverage area reporting

In this section, we describe how the sensor network is partitioned in the case of multiple data sinks and how the description of coverage area per sink is established.

4.1 Multi-sink partitioning of the sensor network

We assume that each of the nodes in the wireless sensor network has the ability to obtain an estimate of its position. This can be either by localization mechanisms [11-15], GPS or by other means (e.g. [16]). Whenever a node publishes information, it is augmented with the current position of the node.

Assume that several gateways are deployed in a certain area and that each of these gateways connects to one or more wireless sensors, which in their turn are part of a multi-hop network structure. In this setup, it is beneficial for e.g. bandwidth reasons to divide the sensor nodes between the gateways. Multi-hop routing of messages in the WSN is highly optimized for e.g. energy-efficiency (e.g. messages travel via shortest reliable paths) or latency (e.g. paths with congestion are avoided) [2,17]. The

efficiency of the network can be affected if messages need to be delivered at a particular gateway, while –from routing perspective- another gateway is more attractive. Therefore, another strategy of grouping nodes with gateways is to let the grouping be implicitly created by minimizing routing cost functions [18-20]. In that case, all topology constraints, such as connectivity, and load balancing are taken into consideration. Basically, the routing strategy of the wireless sensor network determines which node reports to which gateway. However, gateways have no prior knowledge on what area they cover and this information needs to be (dynamically) collected to efficiently deal with subscriptions that are valid only for particular regions. Note, that due to dynamics in the topology or node mobility, the set of nodes reporting to a particular gateway might change over time. This stipulates that a dynamic mechanism for collecting the coverage area is required. This mechanism can be passive or active, as we describe below.

A passive mechanism to obtain a coverage area description is to update the coverage area description whenever the gateway receives a sensor reading that is augmented with position information. Nodes that are not publishing data (e.g. no subscription has been injected into the WSN that matches their properties) would be excluded from the coverage area description. To overcome this problem, nodes can periodically publish their position information to the selected gateway, even if there is no relevant subscription active for them. A drawback of the passive mechanism is the amount of data that has to be transported within the wireless sensor network.

In this paper, we investigate a pro-active mechanism to establish a coverage area description. We let nodes (distributed) keep track of the local coverage area and apply a form of compression to the coverage area description: we describe the area with its *convex hull* i.e. a minimal and ordered set of coordinates that envelops the positions of the nodes that belong to a particular gateway. In such way the gateway can be efficiently informed of the service area while we reduce the amount of information each node needs to store and transmit/receive.

4.2 Establishing a coverage area description

In this section, we discuss our design for distributed coverage area reporting. Nodes determine the routing cost function to any of the gateways that can be reached within the (connected) multi-hop network. This requires gateway to announce themselves periodically through broadcast messages. We assume that the broadcast messages reach all sensor nodes in the connected network before the next broadcast period of the gateway, such that nodes can be sure that within one period all gateways can be discovered. Next, nodes select a gateway with minimum routing cost and send all their generated messages to this gateway. Meanwhile, nodes keep track of coordinates that are either (1) included in messages carrying sensor data, or (2) are explicitly transmitted. Using the received coordinate information, the nodes create a local version of the coverage area description, represented as a convex hull:

1. Nodes start with a convex hull with one coordinate, namely their own coordinate. This coordinate is either programmed during deployment or estimated using localization mechanisms.

2. When coordinates are received, the node checks if these need to be added to the local convex hull. If so, the node adds the coordinate to the local convex hull and (potentially) removes coordinates that are no longer on the convex hull. Nodes only store coordinates that describe the convex hull of their local coverage area and other coordinates are discarded.
3. To keep the local convex hull accurate, a time out mechanism is implemented to remove old coordinates from the local convex hull. The time out of a particular coordinate is reset, when a node receives a message containing the coordinate.

Periodically, the local convex hull is transmitted to neighboring nodes closer to the selected gateway. These nodes merge the received convex hull with their local convex hull. Optionally, the convex hull is reduced using some form of compressing before transmitting (in order to limit memory usage by the algorithm and energy consumption by reducing the size of transmitted/received coordinate list). Since most data will be augmented with position information in practice, explicit transmission of coordinates and local convex hulls would not be required to happen often. However, we do consider periodic transmission of local convex hulls to capture the area covered by none data producing sensor nodes.

With the above described algorithms, the WSN gateways are informed of the convex hull describing their coverage area. Next, this information can be used to optimize handling of position dependent information e.g. gateways can use the information whether a certain subscription is relevant for their coverage area. If not, the sink can decide not to insert the subscription in the WSN, which in the end saves energy and prolongs the lifetime of the wireless sensor network.

5 Geographical Routing using Local Convex Hulls

In the previous section, we discussed how nodes create local convex hull to facilitate coverage area reporting of partitions of the wireless sensor network. In this section, we discuss how this information can be reused to enable efficient geographical routing in the wireless sensor network, in particular the geographical routing of location dependent subscriptions that are injected at the data sink of the WSN and need to be executed in a particular region of the sensor deployment.

In fact, the local convex hull describes the area from which messages flow through the node towards a data sink. Our geographical routing exploits this information by using opposite routing paths i.e. a certain area can be reached by a node, if the area overlaps with the local convex hull description. We note that the reverse routing paths are not necessarily the cheapest paths in terms of routing costs. However, we assume that reverse routing paths are feasible to reach the particular region.

First, we have a closer look at the structure of location dependent subscriptions. We assume that these subscriptions consist of two parts: (1) a description of the area in which the subscription must be executed, and (2) a command sequence (e.g. sensor types, sample rates, critical thresholds, aggregate functions etc.). In this work, we are mainly concerned with the first part of the subscription.

We define $R = \{r_0, r_1, ..., r_n\}$ to be the coordinate set describing the region of interest extracted from the subscription, $L_i = \{l_0, l_1, ..., l_m\}$ the (compressed) local coverage area description of node i and p_i the (estimated) position of node i. The region of interest in the subscription R is in fact described as a closed polygon. We assume that the closed polygon is also a convex hull and that a subscription is generated per closed area. However, our assumptions about R are merely a choice to reduce the complexity of the routing functions described below.

5.1 Routing and executing decisions

The routing decisions in our proposed geographical routing protocol are straightforward. Upon receiving a location dependent subscription, sensor nodes or data sinks analyze the region of interest polygon in the subscription and carry out the following:

1. *Execute decision with forwarding* - The node checks if its (estimated) position is within the region of interest polygon in the subscription. If so, the device executes the subscription and propagates the subscription to neighboring nodes. Note that the execution of subscriptions might also be controlled with additional constraints in the subscription.
2. *Forward decision without execution* - The device checks if its local coverage area description geographically overlaps with the region of interest polygon. If so, the node decides to propagate the subscription e.g. using restricted flooding (Section 3).

The execute decision is in fact similar to the well-known point-in-polygon problem; the node checks if its position p_i falls within the polygon R. In [21] several algorithms are presented to efficiently determine if a point sits in a polygon. In general, these algorithms need complex geometrical operations, however, if we assume that R is a convex hull, the decision if p_i falls in R can be reduce to checking if p_i is geographically *left* to all line segments $r_0 \rightarrow r_1$, $r_1 \rightarrow r_2$, ..., $r_n \rightarrow r_0$. Then, per line segment in R, a node has to carry out three multiplications and two additions/subtractions. Namely, coordinate u is left of line segment $v \rightarrow w$ if

$$\det \begin{bmatrix} v_x & w_x & u_x \\ v_y & w_y & u_y \\ 1 & 1 & 1 \end{bmatrix} > 0 \qquad (1)$$

In fact, this function checks if the (oriented) area given by the vectors $v \rightarrow w$ and $w \rightarrow u$ is positive and hence u is left of the line segment $v \rightarrow w$.

The forward decision is more complex. The node checks if R geographically overlaps with L_i and if so, the node forwards the subscription. We distinguish two cases when the areas R and L_i (partly) overlap: one or more coordinates from the sets R or L_i fall within the polygon of the other set *or* one or more line segments from R intersects with line segment(s) from L_i. If neither of the cases is true, the node discards the subscription.

The same methodology as used in the execution decision can be to check the first case. If the case evaluates true, the areas overlap and the subscription can be

forwarded without evaluating the second case. In [22, 23] describe how the intersection point of lines can be efficiently checked using determinant calculations. Per set of line segment 6 multiplications and 9 additions are required. Whenever two line segments intersect, the check is aborted and the subscription forwarded.

6 Evaluation of routing accuracy

In this section, we evaluate the proposed combination of coverage area and geographical routing in terms of routing accuracy i.e. how well the proposed mechanisms deliver messages to the region of interest defined in a subscription. We define the following metrics:

- *Execution ratio* (ER) - The ratio of nodes that are within the region of interest and execute the subscription with the total number of nodes within the region of interest. This metric measures how well the routing is able to deliver the subscription to the region of interest.
- *False execution ratio* (FER) - The ratio of nodes that are outside the region of interest and execute the subscription with the total number of nodes outside the region of interest. Energy is wasted when subscriptions are executed outside the region of interest. The false execution ratio measures this effect.
- *False injection ratio* (FIR) - The ratio of data sinks that inject the subscription while none of the nodes in its partition executes the subscription with the total number of data sinks. Irrelevant subscriptions lead to higher energy expenditure in the WSN partition when injected. We measure this effect with the false injection ratio.

We implemented the coverage area reporting and geographical routing in a Matlab WSN model and study the accuracy of the routing under different conditions. Table 1 summarizes our simulation parameters. Nodes are randomly deployed in the deployment area. However, disconnected networks are discarded.

Table 1: Simulation parameters

Data sinks	5
Sensor nodes	95
Deployment area	400m x 400m
Transmission range	65m
Region of interest	square, 80m x 80m, center of deployment area
Runs per scenario	100

Table 2: Simulation results in ideal case (a), and simulation results with compression of local convex hull to 3 coordinates (b)

	(a)		(b)	
	Mean	**Std.dev.**	**Mean**	**Std.dev.**
ER	1.00	0.00	0.97	0.13
FER	0.00	0.00	0.00	0.00
FIR	0.08	0.13	0.11	0.14

First, we evaluate the proposed geographical routing without any disturbing factors (Table 2). In this scenario, our geographical routing scheme is able to deliver the subscription accurately to the region of interest as can been seen from the ER and FER results. However, some false injections exist. This is due to the fact that the convex hull of the coverage area can overlap with the region of interest without a node being present in the target area. This effect is due to our choice of describing the services area with a convex hull.

Next, we introduce compression of the local convex hull. In this scenario, each node reduces its local convex hull to at most three coordinates. This reduces execution time of the algorithms and message sizes nodes receive and transmit. We study the effect of the more energy-efficient operation on the accuracy metrics (Table 2). On average we see a 3% reduction in ER and a slight increase in FIR. We conclude that when compression is applied our algorithms are less successful in delivering the subscription to the target area.

In our scheme, we assume that nodes are able to estimate their position (Section 4.1). However, most localization schemes introduce errors and the accuracy of position estimates is also affected by node mobility. We model these errors as additive normal distributed errors to both x and y coordinates of nodes. As consequence, the service area descriptions do not match the reality exactly and the execute decision gets less accurate.

Table 3: Simulation results with additive normal distributed error with σ=5 and σ=10 in position estimates

	σ=5		σ=10	
	Mean	**Std.dev**	**Mean**	**Std.dev.**
ER	0.86	0.21	0.77	0.27
FER	0.00	0.01	0.01	0.01
FIR	0.07	0.12	0.09	0.12

Table 3 summarizes the effect of errors in position estimates on our geographical routing scheme. The results show that ER decreases with increasing position estimation errors. Additionally, we see that FER is indeed affected by inaccurate

position estimates. In the worst case (σ=10), 1% of the nodes outside the region of interest are executing the subscription on average.

7 Conclusion and future work

In this paper, we proposed a mechanism in which the wireless sensor network provides an accurate and up-to-date coverage area description to gateways. In our approach, nodes use their (cost-based) routing protocol to select a gateway to report to. Next, nodes keep track of all coordinates that flow through them towards the selected gateway and create actively a local coverage area description that is periodically forwarded a neighboring node along the route to the gateway. This ensures that coverage areas are up-to-date, even if nodes are e.g. mobile and that coverage area reports include nodes that are not publishing sensor data. As result, gateways are informed of the area they service. Additionally, we let nodes reuse the collected information to efficiently route location dependent subscriptions to a particular target region.

Simulation shows that the proposed routing is able to deliver the subscriptions accurately to the region of interest in the simulated scenarios. On average 97% of the nodes in the target area is reached, even if local convex hulls are extremely reduce to three coordinates (for energy and memory consumption reduction). However, inaccurate position estimates result in significant lower execution ratios (86% and 77% in the simulated cases) and introduce even executions of subscriptions outside the region of interest.

In our future work we intend to compare our geographical routing scheme with existing geographical routing protocols. Energy-efficiency is one of the key metrics to consider. Additionally, we are interested in a comparison with respect to the presented accuracy metrics.

References

1. Ollero, M. Bernard, M. La Civita, L.F.W. van Hoesel, P.J. Marron, J. Lepley and E. de Andres. AWARE: Platform for Autonomous self-deploying and operation of Wireless sensor-actuator networks cooperating with unmanned AeRial vehiclEs. IEEE International Workshop on Safety, Security and Rescue Robotics (SSRR 2007), Rome, pages 1-6, ISBN 978-1-4244-1569-4, September 2007.
2. I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. Elsevier Computer Networks, 38(4):393-422, 2002.
3. M. Zorzi and R.R. Rao. Geographic random forwarding (GERAF) for ad hoc and sensor networks: Multihop performance. IEEE Transactions on Mobile Computing, (Vol. 2 No. 4):337--348, 2003.
4. E. Kranakis, H. Singh, J. Urrutia. Compass Routing on Geometric Networks, in: Proceedings of 11th Canadian Conference on Computational Geometry (CCCG). Vancouver, August 1999, pp. 51–54.
5. F. Kuhn, R. Wattenhofer and A. Zollinger. Asymptotically Optimal Geometric Mobile Ad-hoc Routing, in: Proceedings of the International Workshop on Discrete Algorithms

and Methods for Mobile Computing and Communications (DIAL-M), Atlanta, Georgia, USA, September 2002.
6. B. Karp and H.T. Kung. GPSR: Greedy Perimeter Stateless Routing for Wireless Networks. In Proc. 6th Annual Int. Conf. on Mobile Computing and Networking (MobiCom), pages 243-254, 2000.
7. F. Kuhn, R. Wattenhofer and A. Zollinger. Worst-Case Optimal and Average-Case Efficient Geometric Ad-hoc Routing. Proceedings of ACM International Symposium on Mobile Ad hoc Networking and Computing (MobiHoc), 2003.
8. F. Kuhn, R. Wattenhofer, Y. Zhang and A. Zollinger. Geometric Ad-Hoc Routing: Of Theory and Practice, in: Proceedings of the 22nd ACM Symposium on the Principles of Distributed Computing (PODC), 2003.
9. Y. Yu, R. Govindan and D. Estrin. Geographical and energy aware routing: a recursive data dissemination protocol for wireless sensor networks. Technical Report UCLA/CSD-TR-01-0023, University of Southern California, 2001.
10. A. Coman, M.A. Nascimento and J. Sander. A framework for spatio-temporal query processing over wireless sensor networks, Proceedings of the 1st international workshop on Data management for sensor networks: in conjunction with VLDB 2004, Toronto, Canada, August 30-30, 2004.
11. A. Baggio and K. Langendoen. Monte-Carlo Localization for Mobile Wireless Sensor Networks. Elsevier's Ad Hoc Networks Journal, vol. 6, no. 5, July 2008.
12. J. Hightower and G. Borriello. SPOTON: An indoor 3D location sensing technology based on RF signal strength. Technical Report University of Washington, February 2000.
13. D. Niculescu and B. Nath. Ad hoc positioning system (APS). IEEE Global Telecommunications Conference (GLOBECOM '01), pp. (5)2926--2931, 2001.
14. T. He, C. Huang, B.M. Blum, J.A. Stankovic, T. Abdelza-her. Range-free localization schemes for large scale sensor networks. In MobiCom 2003, San Diego, CA, USA, September 2003.
15. B. Dil, S.O. Dulman, and P.J.M. Havinga. Range-Based Localization in Mobile Sensor Networks. In: Proceedings of Third European Workshop on Wireless Sensor Networks, 13-15 Feb 2006, Zurich, Switzerland. pp. 164-179. Lecture notes in computer science 3868. Springer Verlag, ISBN 3-540-32158-6, 2006.
16. C. Fischer, K. Muthukrishnan, M. Hazas, and H. Gellersen. Ultrasound-Aided Pedestrian Dead Reckoning for Indoor Navigation. In: Proceedings of the first ACM international workshop on Mobile entity localization and tracking in GPS-less environments, Co-located MOBICOM 2008, 15-19 September 2008, San Francisco, USA. pp. 31-36.
17. J.N. Al-Karaki and A.E. Kamal. Routing Techniques in Wireless Sensor Networks: A Survey. IEEE Wireless Communication Magazine, 11(6)6--28, December 2004.
18. J. Wu, S. Dulman, T. Nieberg and P. Havinga. EYES Source Routing Protocol for Wireless Sensor networks. In proceedings of: European Workshop on Wireless Sensor Networks (EWSN'04), January 2004.
19. C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F.Silva. Directed diffusion for wireless sensor networking. IEEE/ACM Trans. on Netw., 11(1):2--16, 2003.
20. P. Chatterjee and N. Das. A Distributed Algorithm for Load-Balanced Routing in Multihop Wireless Sensor Networks. In: Proceedings of 9th International Conference on Distributed Computing and Networking (ICDCN), 5-8 Jan 2008, India, pp. 332-338. Distributed Computing and Networking. Springer Verlag, ISBN 978-3-540-77443-3.
21. E. Haines. Point in Polygon Strategies. In Graphics Gems IV, ed. Paul Heckbert, Academic Press, p. 24-46, 1994.
22. Website: http://mathworld.wolfram.com/Line-LineIntersection.html. Accessed March 2009.
23. Website: http://tog.acm.org/GraphicsGems/gemsiii/insectc.c. Accessed March 2009.

Programming Wireless Sensor Networks Applications using SMEPP: A case study

Javier Barbarán, José A. Dianes, Manuel Díaz, Daniel Garrido,
Luis Llopis, Ana Reyna, Bartolomé Rubio

Department of Languages and Computing Science
University of Málaga, Spain
{barbaran,jdianes,mdr,dgarrido,luisll,reyna,tolo}@lcc.uma.es

Abstract. The SMEPP middleware is a secure and generic middleware, based on a new network centric abstract model for embedded peer-to-peer (EP2P) systems based on services with security as a main design issue. It has an adaptable architecture which can be used in different devices ranging from standard PCs to motes. This paper presents a case study that uses SMEPP for programming standard and small constrained devices such as motes. The middleware is implemented over a runtime infrastructure using a specific designed component model that controls the interactions between the different elements of the middleware. The development of SMEPP applications is carried out through a neutral-language API which can be used in different languages such as Java or nesC. The paper presents a real use of the middleware in an environmental monitoring application for nuclear power plants with wireless sensor networks showing the suitability of the middleware to develop this kind of applications.

Keywords. Middleware, Security, Services, P2P, Embedded

1 Introduction

Traditional middleware architectures have focused on achieving interoperability across heterogeneous platforms and software languages. Although the platforms have evolved from their creation incorporating new specific services and profiles (real time, embedded systems, telecommunications), but their architectures have remained to a great extent and stable.

The Secure Middleware for Embedded Peer-to-Peer systems (SMEPP) project [7] is a European project financed by the European Commission in the Sixth Framework Program. One of the main works carried out in this project has been the definition of a suitable architecture for secure EP2P systems. In the case of EP2P it is necessary to take into account other clearly different aspects such as the surveillance of application behaviour, the processing infrastructure and the underlying communication networks, in order to dynamically separate both the middleware and the applications, and to achieve the appropriate quality of service.

SMEPP supports a high-level, service-oriented model to program the interaction among peers, thus hiding low-level details that concern the supporting infrastructure.

Three key features of the model are the notion of group of peers, the notion of service offered by peers (or by groups), and the concern of security.

A key factor for the success of the middleware is reusability and adaptability [2]. The SMEPP architecture has been adapted to be used in motes. The result of this adaptation is SMEPP Light [3], the version of the middleware for constrained devices. SMEPP Light can run in these devices and it can fully interoperate with the rest of devices of the middleware such as PDAs or standard PCs.

The need for adaptation to different devices and domains makes it necessary to establish a component based software architecture as well as tailored software that achieves these requirements. SMCOM is a component model specially designed to be used in the implementation of the components of SMEPP that gives a modular architecture. This component model is supported by a runtime framework adapted to the different platforms of SMEPP.

Because of the special characteristics of SMEPP, there is a huge range of applications that can be easily solved using this middleware. But a particular field that can take benefit from the use of SMEPP and wireless sensor networks (WSNs), is the related to WSNs for energy power plants, and particularly for nuclear power plants due to the special security needs. Both SMEPP and SMEPP Light configurations are being tested in a real application where sensors are deployed outside and inside the plant and can measure different environmental conditions.

The structure of the paper is as follows: Section 2 introduces the architecture of SMEPP Light and the special needs in order to achieve the interoperability with SMEPP. Following it is shown SMEPP design focusing on the component model used and some details of the runtime framework. In Section 4, the main part of this paper is presented with the case study of developing a real application using SMEPP/SMEPP Light for environmental monitoring application for nuclear power plant. The paper finishes with some conclusions.

2 SMEPP Light Architecture

A main goal of SMEPP is to provide an efficient and flexible architecture that, at the same time, must be scalable and adaptable to the resource constrained devices that will be used in the SMEPP application domains. As result, the architecture of SMEPP for WSNs is SMEPP Light. Customizing a middleware is an error-prone task and requires deep knowledge of platform and middleware design. To solve this, SMEPP has a configuration tool which allows selecting an adequate and possibly reduced set of components of SMEPP for a concrete platform such as SMEPP Light.

The SMEPP Light architecture must support the sensors requirements that are: small memory, small computing power and battery powered. The conceptual architecture for this version of SMEPP is quite similar to the full one, but since SMEPP Light offers a subset of functionalities of SMEPP, also its architecture exploits a subset of the SMEPP components. Figure 1 shows the architecture of SMEPP Light taking into account the special characteristics of motes and TinyOS [9].

The architecture is divided into non-hierarchical layers. The higher layer is composed just by Service Model Support that is the SMEPP Light API. The implementa-

tion of the API primitives will be developed following the SMEPP Runtime Component Framework philosophy but using also the Runtime Component Framework provided by TinyOS, which will be visible by all layers.

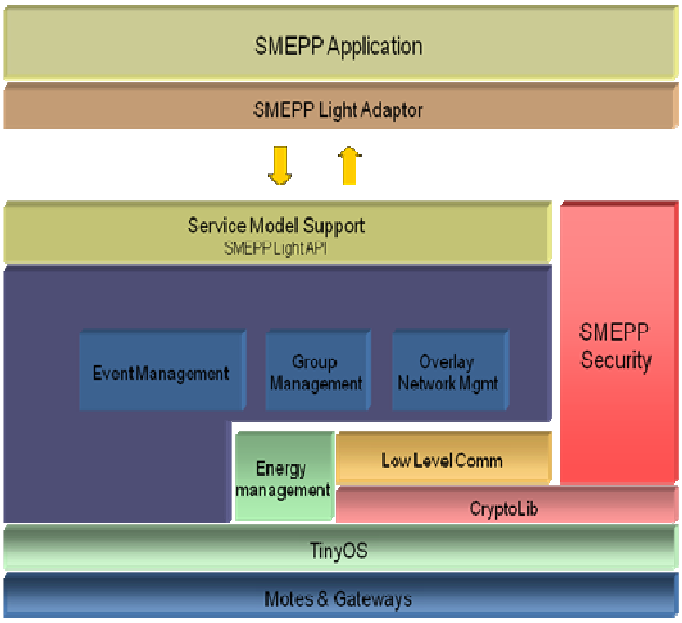


Fig. 1. SMEPP Light Architecture Overview

The SMEPP Common Services are composed by Event Management, Group Management and Overlay Network. The Group Management component manages the peers of each group and the topology of the group, and uses the group management and peer initialization primitives.

The Event Management component maps the event management primitives and it is in charge of subscription and event management. It provides commands to subscribe or unsubscribe to a certain event, generate events and set the reception mode for a certain event.

The Overlay/Network Management component implements the communication between peers. It relies on the Runtime Framework, which is based on TinyOS network/MAC component (CC24240ActiveMessage) that simply allows to address a single node, or to broadcast the message to the entire network.

The Energy Management component manages the duty cycles of the peer. In particular it manages the on/off periods of the radio interface.

The Low Level Communication offers the basic secure-communication primitives to higher layers, by using the CryptoLib library developed for motes using an efficient AES-128 algorithm suitable for wireless sensors [13, 14, 15]. In addition, it manages the keys for all the security issues related to the network and the group layers, and provides commands to encrypt/decrypt messages.

Due to the special communications system in sensors, such as radio protocol used typically, that is 802.15.4/Zigbee, the middleware has to provide some mechanism in order to overcome this problem. In order to provide communication between SMEPP and SMEPP Light (PCs or laptops and computers) it is necessary to use an adaptor application that is the called SMEPP Light Adaptor.

Security is a cross cutting aspect on all layers and is shown on the side so that its relation to other functionalities can be clearly shown in the architecture. The Cryptographic Services component provides the foundations for the higher-level security components by offering implementations of cryptographic primitives; in SMEPP Light by means of CryptoLib library, developed specially for sensors in nesC. Some devices might feature explicit support for security processing. Such device-specific support is captured in the Infrastructure Security component. Secure Topology Management includes the authentication and authorization of new peers joining the SMEPP or SMEPP Light network as well as to the protection of routing data. Finally, group security provides similar services for groups: secure joining of groups and protected communication within groups.

In [17] is shown a complete study of WSN challenges and a classification of the different solutions that exists. It is interesting to compare in some way the present architecture of SMEPP Light with other ones that are similar, and study its suitability for the case study that we are proposing. As said before, SMEPP (and SMEPP Light) belongs to Component-based methodology that proposes software construction by plugging software components [18]. The most remarkable case as a competitor in WSN for SMEPP Light is RUNES. The main differences between them are that RUNNES runs Contiki as Runtime Component Framework, and SMEPP Light runs TinyOS, and the other important difference is that SMEPP provides a secure middleware even for embedded devices, which is a characteristic less important in RUNNES.

3 SMEPP Component Model and Runtime Framework

The SMEPP applications cover several device types ranging from standard PCs to lower capabilities devices such as mobile phones or motes. This way, the election of the component model used in SMEPP is a crucial decision. Component models such as .NET [5] or JavaBeans [8] could be a good election because they are intensively used in standard workstations. However, they are not suitable for embedded systems, since they do not explicitly address memory, real-time or cost constraints. The project RUNES [1] presents an architecture for networked embedded systems based on components. However, the different underlying SMEPP paradigms (P2P and Service Orientation) and the status of the implementation make difficult the reusing of the implementation components and tools.

Middleware configurability and customization constitute an active research area. These topics are addressed from two approaches [6]:

- a dynamic approach, in which the middleware is capable of adapting to the dynamics of the system by reconfiguring itself during runtime, and

- a static approach, that focus on highly customizable middleware architectures capable of fulfilling the requirements of different distributed applications.

SMEPP follows the second approach, providing configuration, adaptation and analysis tools. One of the main contributions of our approach is that of providing analysis tools. The presentation of the tools is out of the scope of this paper.

3.1 The SMCOM Component Model

The component model used for the implementation of SMEPP components is called SMCOM (SMEPP Component Model). This component model derives from a previous work, UM-RTCOM [4] a component model specifically suitable for real-time and embedded systems. Some of its main features have prompted SMEPP to use it for the development of SMEPP components.

One of the main features of SMCOM is the using of the synchronization primitives to carry out the communication between components. This communication is performed through method interfaces and events:

- The `wait` primitive is used to wait for new invocations on services or the raising of consumed events.
- The `call` primitive is used to invoke services of SMEPP components (interconnected with the caller).
- The `raise` primitive is used to create events in an asynchronous way.

In the case of motes, SMEPP Light is implemented on top of TinyOS and nesC [10] which provide a runtime component framework. In this case, we use this component framework to implement SMCOM in such a way that it is possible to communicate components of SMEPP and SMEPP Light.

3.2 Runtime Framework

The Runtime Component Framework (RCF) is the basis for the execution of the different SMEPP and SMEPP Light components. It has to provide an Application Programming Interface (API) based on SMCOM that will be used by the different components of the SMEPP architecture. Basically, the RCF has to provide mechanisms for: connection to the RCF, component interconnection, methods invocation, events publishing, subscribing and creation.

These mechanisms are the only way that components of SMEPP can interact between them. In this sense, the runtime component framework works as a type of virtual machine or scheduler between all the layers of the SMEPP architecture where invocations of components are received, scheduled and executed.

4 Case Study: An Environmental Monitoring Application

In this section we describe a real application used to test the middleware in the field of sensor networks for nuclear power plants monitoring environmental conditions and sending it over the network to PDAs or PCs.

4.1 Application Scenario

The application is focused on the environmental monitoring and remote control of workers in industrial plants. Monitoring the effects of industrial plants on the environment is a key issue in different application domains and very especially in the field of the nuclear energy, where the security aspects of SMEPP could be validated. Moreover, in nuclear industry, exposition to ionizing radiation is a risk present in daily operation and maintenance activities for workers present in the plant. The departments of radiological protection of the nuclear facilities (e.g. power plants, industry, and healthcare) have the responsibility of controlling exposition of workers to radiation. Previous works has been carried out in the field of nuclear power plants, as it is RadMote framework [16] that has attracted the attention of the security departments of power plants. This work implements the study case that was presented in [16] but without using any middleware so, it is easy to see that it is very complex to do it as a commercial product.

The system is composed of one or more wireless sensor networks that will be deployed on some industrial environment. Particularly, the proposed system is focused on nuclear power plants. Sensors will be deployed outside and inside the plant and could measure different environmental conditions.

In the application, radiation sensors will be connected with a small device with compute capabilities (mobile devices, such as motes, PDAs or similar) that workers will wear. On the other hand, static radiation sensors (also known as sensor network's monitoring areas) and environmental monitors (measuring temperature, air quality, etc.) will be also used. It is important to provide a detailed control of some zones sensed by sensors; this will be controlled using wireless high precision cameras. These static sensors and all the personal devices will form a large ad-hoc network.

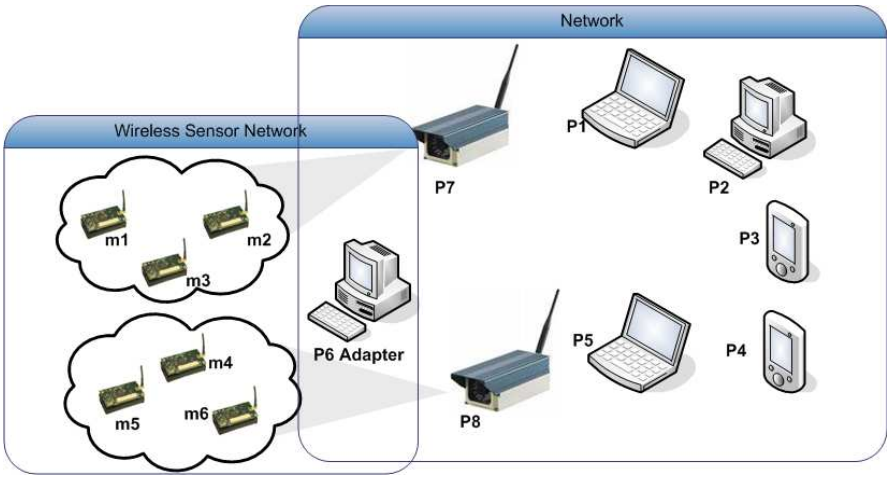


Fig. 2. Environmental Monitoring Application

In Figure 2, a deployment scheme of the proposed system is sketched, where a set of sensors are situated in fixed position into the plant, forming groups. Each sensor is running SMEPP Light, and a special peer will run SMEPP and a special adaptor, in order to enable communication between the two different SMEPP configurations, for sensors and for the rest of devices. Different critical zones are controlled by high performance cameras, that in case of alarm for high radiation values will broadcast images of the risk area to the specific peers. The SMEPP peers will form groups in order to share the environmental information and the alarms of the system.

In the following table we show the tasks for each device:

Table 1. Roles of Participants.

Devices	Description
m1, …,m6	Monitor radiologic measurements raising an event when the measure exceeds a predefined threshold.
P6 Adapter	Enables communication between SMEPPLight and SMEPP. Providing monitor service for the WSN.
P1	Responsible of supervising through video the sensor network when an event is raised.
P2, …,P5	Monitors events on the sensor network.
P7,P8	Provides video streaming service to monitor a critical situation

This application requires working with SMEPP and SMEPP Light, so the adapter takes a key role in this communication. The main efforts for programmers will be the implementation of the services, because the peer implementations using SMEPP API simplifies the code by abstracting the programmer of problems such as communication or security issues. Figure 4 shows the flow diagram of the *camera* peer and *supervisor* peers required for this application where each of the boxes of the diagrams represents an invocation of a primitive provided by SMEPP.

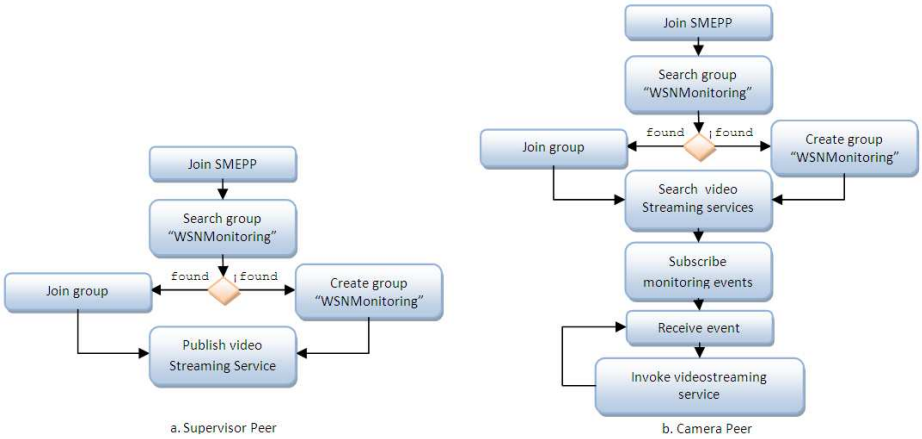


Fig. 3. Camera and Supervisor Peers

In the rest of the section we are going to briefly describe some of the main points to develop a SMEPP-based application in the described scenario.

4.2 Creating Peers and the SMEPP API

The first steps related to the creation of a SMEPP application are usually the "creation" of the peer and the joining or creation of groups. Later, the peer will publish some service or will invoke services of other peers. In order to do these tasks, the SMEPP developer can use several primitives provided by an API. Of course, this API will be different depending on the target language used (e.g. Java or nesC). The following table shows some of the primitives provided by the SMEPP API [7].

Table 2. SMEPP API primitives.

NewPeer	Programs call the newPeer primitive to become peers
CreateGroup	SMEPP groups are logical associations of peers in which services can be published, and messages can be sent (either as service invocations/responses, or as raised events). Peers call the createGroup primitive to start new peer groups.
JoinGroup	Peers use the joinGroup primitive to enter a group (viz., to become members of a group)
GetServices	Returns the list of services which (optionally) matches a given pattern.
Invoke	The invoke primitive serves to call an (one-way or request-response) operation (identified by operationName) of a running service identified by id, which can be either groupServiceId, or peerServiceId, or sessionId.
Publish	Peers offer a service inside a groupId group through publish primitive.
Event	Event raises (viz., notifies to the middleware) an eventName event.
Subscribe	Entities (viz., peers or running services) use the subscribe primitive to

register themselves as event listeners

First of all, the peer must be created, this is done though the newPeer primitive as the following code exposes (the developer provides security credentials):

```
PeerManager peer = PeerManager.newPeer(myCredentials);
```

Once the peer is created, it is part of SMEPP (if the execution success) and it is able to invoke any other provided primitive. After this step it can be provider or consumer of services. In this scenario we have a video-streaming service and consumers of that service pushing video data obtained. To do this, the peer has to be included in a SMEPP group by creating a group or joining an existing group. The following code shows the process of creating a SMEPP group:

```
//group creation
GroupDescription groupDescr = new GroupDescription(
                    "WSNMonitoring",...);

GroupId gid = peer.createGroup(groupDescr);
```

When the peer is joined to one group, it can publish services as the following section explains.

4.3 SMEPP Services and Video-streaming Service

SMEPP Services [11] can be: *state-less* or *state-full* services. The state less services do not keep track of their interactions with clients. Clients can invoke the operations of such services one or more times and in any order.

On the other hand, the state-full services keep track of their interactions with clients. We divide state-full services into *session-less* and *session-full* services. State-full session-less services are services that have only one *virtual communication channel*, which is shared by all clients, and which is active when the service is published.

In our application, we need *session-full* services to provide video-streaming services. In order to transmit video from the wireless cameras (p7 and p8) to the *supervisor* peer, SMEPP developer can carry out the following tasks:

- *Supervisor* peer offers a session-full service which repeatedly receives one-way messages (exposing an operation) carrying video data.
- Cameras start a session of the above service and flushes data by repeatedly invoking that operation. A timeout on each call allows the application to send data without necessarily waiting for an acknowledgement.

The percentage of calls which terminate without raising a timeout exception automatically identifies the amount of data which has surely been delivered. The throughput can be varied independently from the timeout, by concurrently sending data through different threads (each of them calling *Invoke* primitive on different bits of data).

In order to provide a service (e.g. video streaming service), the peer must provide an xml document describing the service, called “Contract” [12]. This contract will be

used for *matching* (which will be in turn used within *discovery*) and also for *verification* and *analysis*. A contract must contain all the information that any client of the service may need to *discover*, to *instantiate*, and to *interact* with the service. The following code shows the publishing of a SMEPP service using a contract for the video-streaming service:

```
serviceId = peer.publish(gid,
                    ContractLoader.loadXMLFromFile(
                    "VideoContract.xml"),

                    new SMEPPServiceGrounding(VideoService.class),
                           myPeer, new HashMap());
```

The service must also provide its behavior. Services are programmed in Java or nesC (depending on the platform) and they must listen to the operations described in their contracts.

```
public void run() {
  ReceivedMessage message;
  Serializable[] data;
  while (true) {
    message = svc.receiveMessage("VideoServiceOperation",
                                        new Class[]{Byte[].class});
      // Store video information ...
  }
```

In the above code we can see the reception of SMEPP messages. These messages have been sent by another peer joined to the same group of the service publisher.

4.4 SMEPP Light sample application and adaptor

We have developed an application for WSN using SMEPP Light. This application has been done using nesC as the programming language and SMEPP Light as a nesC-component that makes easy the programmer's life. The use of SMEPP Light is really similar to SMEPP code that is shown previously, but the main difference is that services are very restricted here because of WSN restrictions, and services in SMEPP Light are very basic ones are implemented using nesC commands and avoiding XML in motes.

In the application, the motes create two groups, and depending on their positions each sensor will join to a particular group.

```
event void Boot.booted() {
     pID = call SensorSmepp.smepp_newPeer(netKey, netMAC, per);
     //Other actions needed when booting the mote
     call SensorSmepp.smepp_getgroups(group);
);

event void SensorSmepp.getGroups_result(uint16_t gID[]) {
     uint8_t i;

     for (i = 0; i < MAX_GROUPS_RECEIVED; i++) {
            if (gID[i] == group.ID) {
                    founded = TRUE;
```

```
            }
        }
        if (found) {
            call TimerJoin.startOneShot(10);
        }
        else {
            call TimerCreateGroup.startOneShot(10);
        }

    }
```

In the code above is shown the code that is executed when the mote turns on, and first of all it creates itself as a peer inside SMEPP Light, and invokes the command asking for the getGroups primitive. This command will raise the getGroups_result as soon as getGroups primitives finishes. Following it will check if the group that the peer wants to join exists, if so it joins and otherwise it creates it.

The footprint is about 3008 bytes in RAM in the mote including the application and the middleware and it is not so big taking into account the set of functionalities that SMEPP Light offers.

SMEPP peers have to subscribe to WSN events in order to receive environmental information so they can make the correct actions depending on the value sensed. To do this, we need to use the so called SMEPP Light adaptor that is a Java-based application that faces with WSN connection; it is component-based as well, so depending on the type of base station (stargate device, MIB520 base station) the adaptor will provide the appropriate component in order to manage the connection with sensors. The adaptor basically manages the subscriptions between SMEPP and SMEPP Light, converting 802.11 b/g packets into 802.15.4/Zigbee packets for the motes, and receiving the events from WSN and sending it to SMEPP peers.

5 Conclusions

SMEPP is a middleware specifically designed for the development of EP2P systems. In this paper we have presented a case study for monitoring environmental conditions in power plants and remote control of workers inside the plant. In order to help the reader to understand how to develop this sample application we explained briefly the component model and the Runtime Component Framework that allows its execution.

The focus of the SMEPP proposal was on the integration of security services and network quality. From the security point of view, security was taken into an approach from the very beginning of the project and it is now integrated into the current version of the middleware. Quality of service has also being taken into account in the design. The specification can be achieved in terms of service contracts and the component model has been designed to be able to analyze and monitor the real-time behaviour of all the components, including those related to basic communication support.

Acknowledgments. This work is fully supported by EU funded project FP6 IST-5-033563

References

1. Costa, P. et al: The RUNES Middleware: A Reconfigurable Component-based Approach to Networked Embedded Systems. 16th Annual IEEE Internacional Symposium on Personal Indoor and Mobile Radio Communications IMRC'05), Berlin, Germany. September 2005.
2. Tarvainen, P.:.An Approach to Evaluate the Adaptability of Software Architectures. Proceedings of the 5th Workshop on System Testing and Validation, (ICSSEA 2007), Paris, France, December 2007.
3. Vairo C.; Albano M.; Chessa S.: A Secure Middleware for Wireless Sensor Networks. Middleware for Mobile Embedded Peer-to-Peer Systems 1st International Workshop. July 2008.
4. Díaz, M.; Garrido, D.; Llopis, L.; Rus, F.; Troya, J.M.: UM-RTCOM: An Analyzable Component Model for Real Time Distributed Systems. J. Syst. Software (2007), doi:10.1016/j.jss.2007.07.010
5. Microsoft .Net Web site. http://msdn.microsoft.com/netframework/
6. RTZen. UCI-DOC Group. http://doc.ece.uci.edu/rtzen/
7. SMEPP Web site. http://www.smepp.org
8. Sun Microsystems. Enterprise JavaBeans Specification 2.1. 2005.
9. TinyOS Web site. http://www.tinyos.net
10. Gay D.; Levis, P.; Von Behren R.; Welsh, M.; Brewer, E.; Culler, D.: A Holistic Approach to Networked Embedded Systems. Proceedings of Programming Language Design and Implementation, 2003.
11. Brogi, A.; Popescu, R.; Gutiérrez, F.; López P., Pimentel, E.: A Service-Oriented Model for Embedded Peer-to-Peer Systems. In proceedings of the 6th International Workshop on the Foundations of Coordination Languages and Software Architectures, Lisbon, Portugal, September 8, 2007.
12. Brogi, A.; Popescu, R.: Workflow Semantics of Peer and Service Behaviour. Proceedings 2nd IEEE International Symposium on Theoretical Aspects of Software Engineering, June 2008, Nanjing, China.
13. Roman R.; Fernandez-Gago, M.C.; Lopez, J.:. Featuring Trust and Reputation Management Systems for Constrained Hardware Devices. Proceedings of the 1st International Conference on Autonomic Computing and Communication Systems (Autonomics 2007), Rome (Italy), October 2007.
14. Baek, J.; Foo, E.; Tan, H.; Zhou, J.: Securing Wireless Sensor Networks - Threats and Countermeasures. Chapter 3 of "Security and Privacy in Wireless and Mobile Computing", ISBN 978-1905886-906, Troubador Publishing, 2008.
15. López, J. Zhou, J.: Wireless Sensor Network Security. Cryptology & Information Security Series, IOS Press, April 2008.
16. Barbarán, J.; Díaz, M.; Esteve, I.; Rubio, B.; RadMote: A mobile framework for radiation monitoring in nuclear power plants. In Proceedings of the XXI International CESSE. May 2007.
17. Rubio. B; Díaz, M.; Troya. J.M.; Programming approaches and challenges for wireless sensor network. In Second International Conference on Systems and Networks Communications. August 2007.
18. G. HeineMan and W. Council. Component-Based Software Engineering: Putting the pieces together. Addison-Wesley: Reading, MA, 2001.

User behaviour in a WLAN campus: a real case study

Enrica Zola, Francisco Barcelo-Arroyo, María López-Ramírez,

Universitat Politècnica de Catalunya, c/ Jordi Girona 1-3, Mòdul C3
08034 Barcelona, Spain
{enrica, barcelo, maria.lopez}@entel.upc.edu

Abstract. European universities taking part in the Education Roaming project (eduroam) provide wireless connectivity through different institutions all over Europe, thus encouraging mobility of students and researchers. The Technical University of Catalonia (UPC) takes part in this project. In order to analyze users' activity inside the wireless network at UPC, syslog information has been captured at the access points in the main library during a period of a week. Even if it is not a long trace period, interesting results can be extracted about user behaviour. Despite the widespread of lightweight devices which facilitate their use while walking inside the campus, we still observe low mobility. Moreover, regardless of the overcoverage provided by the infrastructure, users still experience many connectivity problems. These results may be useful for network developers in order to improve the quality of the wireless service and for software developers in order to create location-based applications.

Keywords: IEEE802.11, handover, measurement, residence time, WLAN.

1 Introduction

By the end of the previous decade, wireless communications have been spreading in many fields of our every-day life, becoming commonplace in many environments and making people used to move around while staying connected. By 2000, users have been increasingly interested in take advantage of the flexibility of such technology and a boom in its implementation to local area networks have been seen till nowadays. Universities have been pioneers in developing infrastructures in order to provide connectivity all over the campuses. An example is the Wireless Andrew at the Carnegie Mellon University campus [1], an enterprise-wide broadband wireless network developed in 1993. Nowadays, it is common for a university to provide wireless connection to its employers and students. To encourage the mobility of researchers and European students, the Education Roaming project (eduroam) [2] has been developed in 2003 [3], providing wireless connectivity through different institutions all over Europe; in this way, users from corporations taking part in the project can access the Internet using the wireless networks of other institutions participating in the same program.

The growing popularity of WLAN encourages researchers to investigate this new scenario. A lot of work has been published providing deep insight on user behaviour and traffic characteristics in a real WLAN, such as university campus [4, 5, 6, 7, 8], corporations [9], conference hall [10], or other industrial environments [11]. In one of the earliest studies [6], Tang and Baker analyzed a 12-weeks trace of 74 users at the Stanford CS Department. Besides a deep insight on traffic characterization is given, they also present few results on handover performance. They find out that the maximum number of handover (HO) at one access point (AP) within 5-minute or 15-minute period varies between 2 to 10 depending on the AP; the AP at the library, for instance, registers a maximum of 5 HO within 5 minutes, and 8 HO within 15 minutes.

Henderson, Kotz and Abyzov provide a complete analysis of a 17 weeks trace obtained during the fall term 2003 at Dartmouth College [5]. Among mobility results, they find out that, although roaming increases from their previous study performed in 2001 [4], users are still not mobile and tend to stay at one home location (i.e. AP to which a user associates during more than half the time he stays online) for most of the time.

A similar study was performed at a corporate WLAN composed of three main buildings from July 20th to August 17th 2002 [9]. They found out that a great amount of users do not move much out of their home location. Since they consider as home location any AP inside a building, mobility means movements among buildings. Moreover, since they use SNMP in order to collect data, mobility results can be affected by the poll period (i.e. 5 minutes), since events among this period are not registered. This is a different approach from the one presented in this paper, so we expect different results.

An interesting study on handover characteristics from a real environment is presented in [7] from data collected in 1997 at the Wireless Andrew network at Carnegie Mellon. This study is based on a set of 9105 sample of residence time, of which 54% is less than 3 seconds and 93% are less than 5 minutes. However, mean dwell time is 50 minutes. Similar results will be presented in this paper, but we will give a different approach in order to understand why the median and the average dwell time are so different. Moreover, from the analysis of the sign-on inter-arrival time (i.e. new users) it is not clear if daily cycles have been taken into account in their work.

In this paper we present results from association information collected during one week at the UPC central library. Unlike other works, we deal with mobility results from one population, that who accesses the library. The aim is to better characterize cell residence time (i.e. time between handovers) and to identify different behaviours among users. Since syslog is used, we do not loose events as other works do [9].

The remainder of this paper is organized as follows. Details about how data has been collected can be found in Section 2. A deep insight on user behaviour is given in Section 3. Results on cell residence time are presented in Section 4. Section 5 concludes the paper.

2 Data Collection

The study has been taken at the Technical University of Catalonia (Universitat Politècnica de Catalunya, UPC), which is composed of many campuses spread out in the city of Barcelona (Spain) and its surroundings. To encourage the mobility of researchers and European students, UPC is taking part in the eduroam project [2], providing wireless connectivity to both its own users and outer eduroam partners. In this way, users from institutions taking part in the project have access to the Internet using the wireless networks of the other participating institutions. Being part of eduroam allows users to access wireless network at a visited institution connected to eduroam by simply using the same credentials (i.e. username and password) the users would use if they were at their home institution. The only requirements are to have a properly configured Wi-Fi compatible device (e.g. computer, PDA, wireless cell phone, etc.) that is compliant with IEEE 802.11b or 802.11g, and to have been given credentials from the home institution.

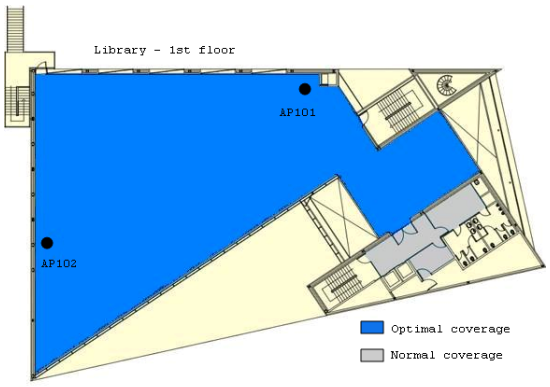


Fig. 1. Coverage at the first floor of the library

In this study, only data proceeding from the central library at Campus Nord in Barcelona have been considered. It is a four floors building with eight AP giving access to library users (i.e. two AP at each floor). The infrastructure provides good coverage all over the building: as an example, Fig. 1 shows the coverage at the first floor and the location of the two APs (i.e. AP101 and AP102). People access the library from the ground floor, where the loan service is placed: apart from books, UPC students and researchers can also borrow laptops for a given period of time. The first and second floors have room for the library collection, which is divided according to specific subjects; students use to spend their time there to work at their subjects while consulting the books and the Internet. The third floor provides specific documentation for PhD students and researchers.

Users were not informed that the study was performed. The only sensitive information that we gathered were the MAC and IP addresses of network cards, as well as the names assigned to AP. To ensure user privacy, information has been anonymized.

2.1 Syslog

Data has been collected with syslog, which is a standard for forwarding log messages in an IP network. Syslog is a client/server protocol where the client sends a small textual message to the syslog server through UDP and/or TCP connections. The AP of the wireless network at UPC are configured to send syslog messages to a central server whenever clients authenticated, associated, roamed, disassociated or deauthenticated. Each message contains the AP name, the MAC address of the card, the time stamp at which the AP received the message (i.e. 1 second's precision), and the type of message. Once logs have been anonymized, association and disassociation messages can be analyzed. We remind here that, after authentication, a WLAN user (i.e. MN, mobile node) chooses the best AP among a list of nearby AP and associates with it; only then, the MN can exchange data. When the MN no longer needs to use the network, it disassociates with its current AP. Disassociation can be due to the MN moving into another cell (i.e. MN performs a handover to another AP), to authentication problems (i.e. Previous Authentication no longer valid as the cause field in the disassociation message) or to the MN leaving the network.

2.2 Settings

The aim of this study is to analyze the cell residence time in a real environment: for this purpose, only association and disassociation frames have been considered. A user can only associate to an AP to which it has authenticated before, so we can ignore authentication messages and only use association/disassociation frames. For each MN, the cell residence time (i.e. the time MN remains associated to the same AP) can be computed as:

$$Cell\ residence\ time = TS_{dis} - TS_{ass}\ , \qquad (1)$$

where TS_{ass} is the timestamp of the MN's association to a given AP (i.e. when the AP receives association request from the MN) and TS_{dis} is the disassociation of the same MN from that AP (i.e. when the AP receives the disassociation frame from the MN).
If two or more association requests from a given MN are received at the same AP, the last one is considered since it is possible that AP's association responses to previous requests got lost and, then, the MN has been trying again to connect to that AP.
In this paper we focus on data collected during a week, from Monday June the 2nd to Friday June the 6th, 2008. Since that was a busy period for the library (e.g. final exams period was going to start on June the 9th), the traffic generated is high enough to obtain good statistical results.

3 User Behaviour

A total number of 1085 different MAC addresses have been observed in the whole library during the period. Since laptops can be borrowed at the library, the actual

number of different users accessing the WLAN can be higher, but in this study we won't consider it. For simplicity, from now on we will refer to users (i.e. MN) instead of MAC addresses.

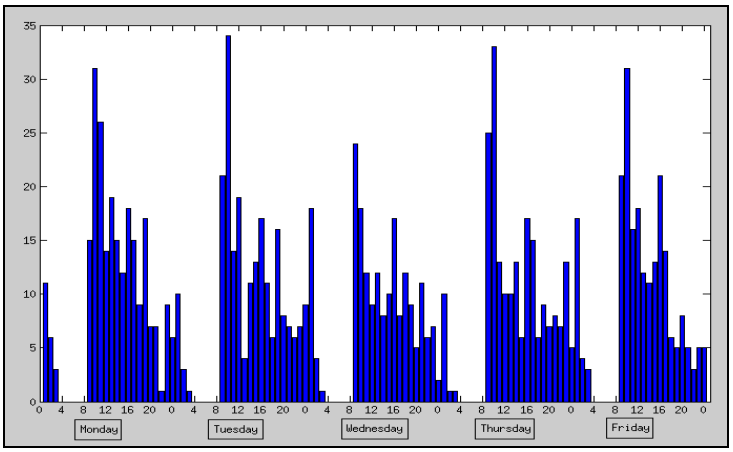


Fig. 2. Number of first association per day and hour.

The number of first association per day to any AP in the library (i.e. the first time a new user enters the library and connects to the WLAN per day) has been depicted in Fig. 2: it displays the distribution of first association per day over one hour periods. The library stayed open from 8.30 a.m. to 2.30 a.m.; therefore, nobody appears between 3 a.m. and 8 a.m. We observe a peak in first arrival when the library opens and a decreasing but nearly constant access of new users to the library during the day. New users appear after 8 p.m. until midnight, reflecting activity in the library also overnight: since the after-dinner pattern is different through days, we use traces from 10 a.m. to 5 p.m. only (i.e. working-day pattern) in order to deal with homogeneous data. Monday is the day at which the highest amount of distinct users appears, while Wednesday the lowest.

The number of distinct users who associate to a given AP from Monday to Friday is displayed in Table 1: since the same user can associate to one AP in the library more than once during the day, here it is considered only once (i.e. the first time he associates to that AP per day). The second column displays the value over the whole period (i.e. Monday to Friday from 0 to 24h), while in the third column only the working-day pattern is considered. The last row represents the number of distinct users who associate to any AP in the library: since a user who associates to more than one AP during his journey is counted only once, the sum of the previous rows does not match the value displayed here. We can observe that the proportion of users appearing during the working-day with respect to those appearing during the whole period is always around 50%, except for the APs on the second floor (i.e. 201 and 202) for which the proportion is lower (i.e. around 33%). The APs in the first and second floor (i.e. 101, 102, 201, and 202) deal with over the 60% of all first associations: since students can seat and study in these floors, it is normal that the

great amount of users connected to WLAN first appears there. AP102 and AP202 have been selected as the APs at which we perform the analysis of the cell residence time, since they are the most loaded during the working-day pattern.

Table 1. Number of user per AP from Monday to Friday

Number of users on: / AP:	0-24h	10-17h
001	316	160
002	89	57
101	300	151
102	431	230
201	411	136
202	476	164
301	211	108
302	187	111
Whole library	1085	473

Table 2. Percentage of users accessing the library one or more days

Number of days	Number of users	%
1	355	57.44
2	136	22.01
3	62	10.03
4	32	5.18
5	33	5.34

To further characterize user population, Table 2 provides the number of users that appears in the library only once in the week or more frequently (i.e. more than 1 day). It is interesting to note that more than half the population accesses the library only one day through the whole trace, which means that we deal with non-frequent users. As the number of days increases, the number of users connecting more days decreases: only 5.34% goes to the library every day. We remind here that we may be considering as frequent users laptops borrowed from the library by different users each day. Since a user who appears more than one day is counted once, the sum of the values in the second column (i.e. 618) does not match the total number of distinct users in the library (i.e. 1085 in Table 1).

The number of users connected to any AP in the library per day and hour is displayed in Fig. 3. Monday is the most populated day, while Wednesday the least: figures shown in Fig. 2 are kept. Every day, except Tuesday, the number of connected users increases from 11 to 13 hours; every day this figure decreases for lunchtime (i.e. 14-15). The number of associated users increases towards the late afternoon every day except Monday. The high number of non-frequent users does not influence the working-day pattern of new arrivals shown in Fig. 2, as also stated in [9]. On the contrary, users' behaviour in terms of permanence in the library varies throughout the week and has an impact on results presented in Fig. 3.

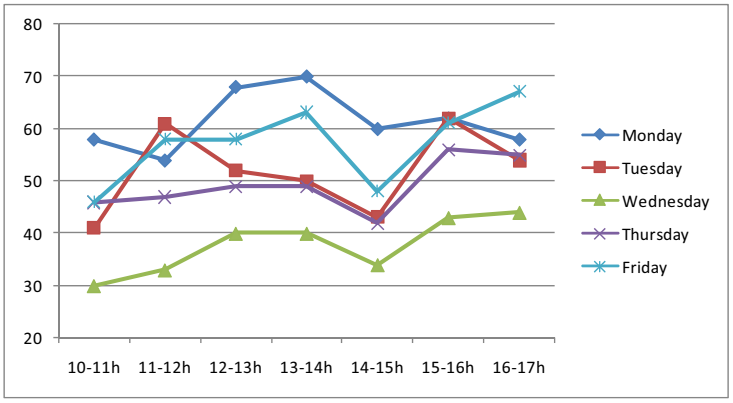


Fig. 3. Number of users connected to any AP in the library per day and hour.

4 Cell Residence Time

As explained before, in order to deal with homogeneous data we analyze traces from each AP during the working day period (i.e. Monday to Friday from 10 a.m. to 5 p.m.). The association and disassociation frames at MAC layer were analyzed for each AP, and AP 102 and AP 202 were selected as the most representative ones (i.e. high and nearly constant load during the working day). From each set of data, we obtain a number of 1036 and 929 cell residence times, respectively.

Table 3 shows statistics for the overall cell residence time: for AP102 the mean is 416 seconds (i.e. 7 minutes) and the standard deviation is 1246 seconds (i.e. 21 minutes), with a high coefficient of variation of nearly 3. For AP202 the average is lower (i.e. 295 seconds, less than 5 minutes), while the coefficient of variation (i.e. CV) is higher (i.e. 3.38). Beside the mean value is high, the 50th percentile is very low (around 1 minute for each AP), reflecting a high percentage of very short cell residence times in the sample.

Table 3. Cell residence time statistics for AP102 and AP202

	AP102			**AP202**		
	Overall	**No HO**	**HO**	**Overall**	**No HO**	**HO**
Mean	415.92	1479.80	272.6	294.67	631.50	258.99
Median	86	218	76	66	136	61
Max Value	14339	14339	10389	15558	15558	11976
Stdv	1246.10	2667.70	798.18	994.74	1736.20	874.64
CV	2.99	1.80	2.93	3.38	2.75	3.38
Sample	1036	123 (11.87%)	913	929	89 (9.58%)	840

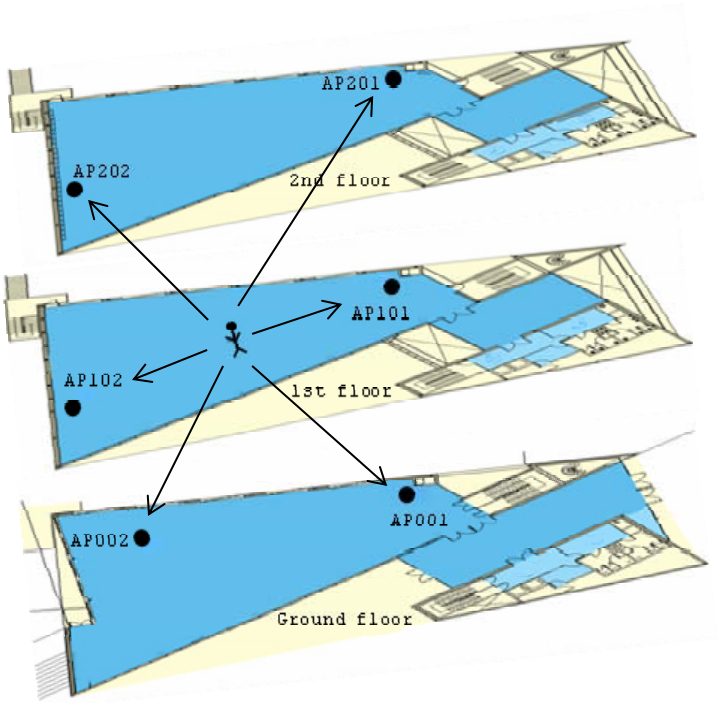


Fig. 4. Ping-pong user: an example.

Fig. 4 illustrates a very usual situation: the client is located at the first floor of the library and he is currently associated to AP102. Due to his position, it is possible that the signal received from other APs (i.e. AP202, AP201, AP101, AP001 and AP002) is better than the one from his AP. It is common that the user will ping-pong among such APs, despite he does not move. This is a common situation in a wireless network with a high concentration of APs (i.e. strong overcoverage) like the one analyzed. We often saw users "ping-pong", associating and reassociating with several APs many times in succession: it is clear that roaming does not imply physical motion, as already observed in [5].

Since it was impossible to determine which situation reflects a real movement and which a ping-pong (i.e. logs do not provide users' locations), we track users who associate to one AP per day (i.e. no handover is performed for such users) and study their cell residence times separately (we will refer to them as No HO users from now on). Results are shown in Table 3: as expected, we obtain higher mean values (i.e. around 24 and 10 minutes, respectively) and more stable results (i.e. CV are lower) for both APs. The 50^{th} percentile is still very low (i.e. 3.5 and 2 minutes, respectively), reflecting that still there are many short cell residence times even if users does not ping-pong among neighbouring APs. Results for users who connect to more than one AP per day (i.e. HO users) are also presented: the mean is around 4.5

minutes for both APs and CV are still high, as for the overall sample. It is interesting to observe that the 90% of the overall sample is represented by HO users, showing that the network has to handle a great amount of signalling due to HO requests.
From Table 3 we can also estimate the average number of HO per AP and hour: for the overall case, it varies between 9 and 12 depending on the AP, while for HO population it is stable between 13 and 14.

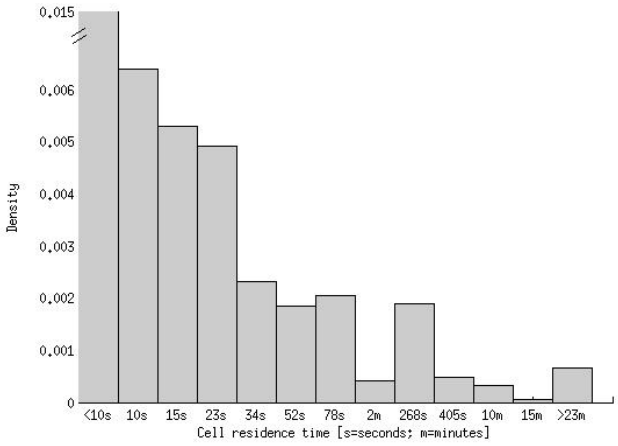


Fig. 5. Distribution of the cell residence time for no HO users for AP102.

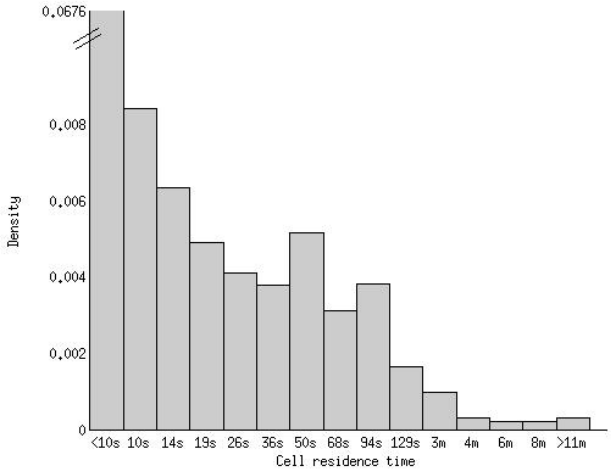


Fig. 6. Distribution of the cell residence time for HO users for AP102.

Fig. 5 and 6 display the distribution of the cell residence time for AP102 and for no HO and HO population, respectively: log-scale is used in order to better show the high concentration of very low values (i.e. less than 2 or 3 minutes) while still

representing the tail (maximum values are provided in Table 3). Results are very different from those presented in [7], where authors reported that the 54% of the cell residence times were lower than 3 seconds, but our results prove that users still suffer from connectivity problems not due to bad coverage or mobility. Further investigation on the causes provoking a user continuously and rapidly disconnecting from the network is needed.

5 Conclusions

Association and disassociation information of WLAN users accessing the library at the main campus of UPC has been collected during a week (i.e. from Monday June the 2nd to Friday June the 6th, 2008) using syslog and then analyzed in order to extract useful information about user behaviour. Even though users distribute among all the AP inside the library, access points at the first and second floors deal with the 60% of all first associations, reflecting that people generally start their connection there. We also found out that about 60% of the users connect to the WLAN only one day during the trace; despite we deal with infrequent users, the working-day pattern is maintained through the trace.
From the analysis of the cell residence time, it is clear that users do not move much: the 12% of the entire population is observed at just one AP per day (i.e. no HO users). No other information about user location and mobility can be extracted by the traces, but we can provide results about the average connection times to a given AP. Cell residence times are extremely variable: while there is a high concentration of connections lower than 1 minute, some users stay connected for up to 4 hours. We then analyzed no HO and HO users for separate, in order to depict different trends. While on average a HO user reside inside a cell for 4 minutes, this value for no HO users can vary from 10 to 25 minutes, depending on the AP. Since not only HO users but also no HO users, who are supposed to be static, experience very short cell residence time, it is clear that some sort of connectivity problem, not due to bad coverage or mobility, causes users to frequently disconnect from the network. Deeper investigation is needed in order to determine the cause of these short connections and, if possible, help network developers to eliminate them and provide a better service. The aim is to deal with connection times that reflect the user behaviour only: from that, we would be able to extract more interesting and useful results that may help software developers in order to create location-based applications.

Acknowledgments

Authors would like to thank UPCnet for providing WLAN traces; a special thank to Sergi Sales Llop, Josep-Lluís Cortés and Margarita Garrido Lorenzo for their help and continuous support during the study.

This work was supported by the Spanish Government and ERDF through CICYT project TEC2006-09466/TCM.

References

1. Hills, A.: Wireless Andrew [Mobile Computing for University Campus]. IEEE Spectrum vol. 6, pp. 49-53. DOI= 10.1109/6.769269 (1999)
2. Eduroam project at UPC, https://upcnet.upc.edu/serveis/servidors-i-xarxes/gestio-de-xarxes/xarxes-sense-fils-upc-eduroam/xsf-upc-eduroam-upc-wireless
3. Wierenga, K., Florio, L.: Eduroam: past, present and future. Computational Methods in Science and Technology 11 (2), pp. 169-173 (2005)
4. Kotz, D., Henderson, T., Abyzov, I: Analysis of a Campus-wide Wireless Network. Wireless Networks 11, pp. 115-133 (2005)
5. Henderson, T., Kotz, D., Abyzov, I.: The Changing Usage of a Mature Campus-wide Wireless Network. Computer Networks 52, pp. 2690-2712 (2008)
6. Tang, D., Baker, M.: Analysis of a Local-Area Wireless Network. In: 6th Annual International Conference on Mobile Computing and Networking, pp. 1–10, ACM Press, Boston (2000)
7. Thajchayapong, S., Peha, J.M.: Mobility Patterns in Microcellular Wireless Networks. IEEE Transactions on Mobile Computing, vol. 5 no. 1, pp. 52-63. DOI= 10.1109/TMC.2006.13 (2006)
8. Hutchins, R., Zegura, E.W.: Measurements from a Campus Wireless Network. In: IEEE International Conference on Communications, ICC 2002, pp. 3161-3167, vol. 5. DOI= 10.1109/ICC.2002.997419 (2002)
9. Balazinska, M., Castro, P.: Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network. In: 1st International Conference on Mobile Systems, Applications and Services, pp. 303-316 (2003)
10. Balachandran, A., Voelker, G.M., Bahl, P., Rangan, P.V.: Characterizing User Behavior and Network Performance in a Public Wireless LAN. In: 2002 ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems, pp. 195–205, ACM Press (2002)
11. Willig, A., Kubisch, M., Hoene, C., Wolisz, A.: Measurements of a Wireless Link in an Industrial Environment Using an IEEE 802.11-Compliant Physical Layer. IEEE Transactions of Industrial Electronics, vol. 49 no. 6, pp. 1265-1282. DOI= 10.1109/TIE.2002.804974 (2004)

Part II

Short papers and extended abstracts

Admission Control for Supporting Active Communication Sessions in Mobile WiMAX Networks

Natalia Vassileva[1], Yevgeni Koucheryavy[2], Francisco Barcelo-Arroyo[1]

[1]Department of Telematics Engineering, Technical University of Catalonia (UPC)
C./Jordi Girona 1-3, mod. C3, 08034 Barcelona, Spain
[2]Department of Communications Engineering, Tampere University of Technology
P.O. Box 553, FIN-33101, Tampere, Finland
natalia@entel.upc.edu, yk@cs.tut.fi, barcelo@entel.upc.edu

Extended Abstract

Most of the call Admission Control (AC) algorithms for supporting on-going communication sessions are designed for conventional cellular telephone networks. A relevant question is whether these algorithms are applicable to wireless networks that feature broadband access. The focus in this research is on admission control implementation in mobile networks that comply with the IEEE 802.16 family of standards [1] (commonly referred to as Worldwide Interoperability for Microwave Access, WiMAX) as it is the de facto next-generation broadband wireless access technology today.

The principal feature that distinguishes Wireless Wide Area Networks (WWANs) from the rest of the networks is mobility support. Common for both traditional telephone and next-generation WWANs is their cellular architecture. Mobility of terminals and cellular architecture of networks lead to common scenario where Mobile Stations (MSs) visit several Base Stations (BSs) during their communication sessions. The problem of guaranteeing continuity of the on-going calls while on the move is multifaceted as it encompasses mechanisms for handover decision-making, algorithms for handing over sessions from source to target BSs and allocation of radio resources. AC algorithms for supporting active sessions address the challenge of providing available resources throughout the whole service area to prevent handover failure events due to lack of available resources in the target BS.

The problem of providing free resources to MSs is widely studied in conventional telephone cellular networks (see e.g., [2] and [3]). A relevant question is whether AC algorithms designed for traditional cellular networks can be readily implemented in the next generation networks. The main conclusion is that a problem induced by network's architecture and terminals' mobility is heavily dependent on implemented technology as explained below.

Capacity is fundamental metric in admission control. System capacity and required service capacity are used to check the basic condition for resource availability. Conventional cellular networks have hard capacity and are voice oriented. These characteristics define the notion of channel – a fixed number of resources (a time slot

or frequency) is assigned to each call. Calls are distinguished based on the originating BS (new and handover calls).

Service capacity provision and resource allocation have different definition in IEEE 802.16 framework. WiMAX technology provides rich service environment with different QoS requirements, therefore service in addition to call-origination differentiation should be considered.

In WiMAX BS controls physical-layer metrics, as modulation and coding, to adapt to time-varying wireless channels, the outcome of which is time-varying radio resource allocation. Note also that total number of resources for a session request not only depends on type of service but does so on MS's current physical layer parameters too. BS scheduler is executed in each frame due to changes in the required resources (after application layer change in data rate or after physical layer change in modulation and coding). Moreover, opportunistic scheduling is present to achieve best use of resources. All of these factors have immediate capacity and therefore AC consequences. The variable nature of the required resources imposes challenges on handover prioritization and/or capacity reservation for future handover arrivals because required resources can change over time as indicated before. In addition, since WiMAX is QoS committed, admission control must satisfy negotiated service-level agreements. The latter implies a tight interaction with BS scheduler.

The outlined succinct discussion concentrates on IEEE 802.16e-compliant networks. However, the main result – AC technology dependence – holds for LTE networks due to commonalities in the physical and MAC layer definitions.

Work in progress is devoted to addressing the listed research issues. Afterwards, the admission control algorithm proposed in [4] will be implemented. The algorithm is based on statistical estimates of metrics available at the BS, which obviates the need for complex measurements, prediction techniques or interchange of control AC messages between BSs. The simplicity and efficiency of the algorithm in terms of execution and performance results respectively make it suitable for practical implementation. In WiMAX defined environment, most of the teletraffic metrics will have different statistical nature, which requires adaptation of the algorithm to the technology and metrics that characterize supported services. OPNET Modeller simulation tool is considered for AC performance evaluation. This requires implementation of the algorithm in the WiMAX module of the simulator.

References

1. IEEE Std 802.16e-2005 and IEEE Std 802.16-2004: Air Interface for Fixed and Mobile Broadband Wireless Access Systems (2005)
2. Hong, D., Rappaport, S.S.: Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized hand-off procedures. IEEE Trans. Vehicular Technology, vol. VT-35, pp. 77--92 (1986)
3. Ramjee, R., Nagarajan, R., Towsley, D.: On optimal call admission control in cellular networks. In: IEEE INFOCOM, pp. 45--50 (1996)
4. Vassileva, N., Barcelo-Arroyo, F.: A New CAC Policy Based on Traffic Characterization in Cellular Networks. In WWIC 2008. LNCS, vol. 5031, pp. 1-12, Springer, Heidelberg (2008)

User-Driven Reputation of Mobile Network Providers

Jean-Marc Seigneur[1], Xavier Titi[1] and Lenny Ridel[2],

[1] Advanced Systems Group, University of Geneva, Switzerland
{Jean-Marc.Seigneur, Xavier.Titi}@unige.ch
[2] Traffix Systems, Israël
{Lenny.Ridel}@unige.ch

Abstract. Nowadays, mobile users can switch between different available networks, for example, nearby WiFi networks or their standard mobile operator network. Soon it will be extended to other operators. However, unless telecommunication operators can directly benefit from allowing a user to switch to another operator, operators have an incentive to keep their network quality of service confidential to avoid that their users decide to switch to another network. In contrast, in a user-centric way, the users should be allowed to share their observations regarding the networks that they have used. In this paper, we present our work in progress towards attack-resistant sharing of quality of service information and network provider reputation among mobile users.

Keywords: Quality of service, reputation, telecommunication network provider, trust.

1 Introduction

On the 10th of September 2008, the European Commission launched its Future Internet Research and Experimentation (FIRE) initiative. We envision the Future Internet as being able to infer the user experience quality of the network services that it provides and to take into account these user-centric observations at time of selection of these network services. As a first step towards this vision, we have been investigating appropriate mechanisms for mobile network selection based on Quality of Experience (QoE). We stress that it is important to make the difference between QoE and Quality of Service (QoS). The ITU-T in its E800 recommendation [1] defines QoS as follows: "the collective effect of service performances, which determines the degree of satisfaction of service users". Thus, QoS has mainly focused on objective technical evidence such as session throughput measurement. QoE goes beyond purely technical evidence and includes asking the opinions of the users about their degree of satisfaction after using the service. Assuming that it is infeasible to have a unique QoE information service trusted by everybody and collecting all QoE information generated in the world, QoE information will be distributed and provided by different entities. We work on an EU-funded R&D project, called PERIMETER [2], where the users share their QoE information in a decentralized peer-to-peer fashion. In this case, traditional telecommunication operators, who are powerful

entities, may invest a lot of resources to influence their level of QoE to keep or attract more users. Telecommunication operators may even try to attack such a user-centric system to protect their market. In Section 2, we present a decentralized model where users will be able to share their QoE without being censored or abused.

2 Towards Reputation-based Network Selection

We consider the level of QoE of a telecommunication or network provider as its QoE reputation. That QoE reputation can be influenced because it is not formally proven and composed of distributed information that may not be complete or from unauthenticated sources. Romans considered reputation as "a vulgar opinion where there is no truth" ("reputatio est vulgaris opinio ubi non est veritas") [3]. Nowadays, there are still many potential attacks on computational trust and reputation metrics [4]. In this section, we first depict the attack model expected in our shared user-centric-generated QoE scenario. Then, we present our work in progress towards an attack-resistant computational reputation model for this scenario.

2.1 Security Aspects of Reputation-based Network Selection

First, an identity layer is needed to be able to build reputation on top of identities. However, the work towards the materialisation of true user-centric mobility requires a paradigm shift from contract based mobile service delivery, that limits the ability of the user to choose the best provider for the needed service to a dynamic, contract-less service delivery based on privacy preserving identity management and pay-pal like billing services. The PERIMETER project [3] pushes these boundaries through the design and development of user-centric privacy and anonymisation mechanisms, that will allow end users to enjoy privacy protection and if required, logically separate their identity and their activities on the network from the billing process, while retaining their ability to autonomously select the best connection and best service from the available choices in each area. These mechanisms will be designed to be independent from the underlying networking technology, so that fast, inter-technology handovers will still be possible.

However, once it is possible to identify the interacting parties, reputation must be computed in these parties and reputation is less objective than identification. QoE may be more subjective than objective technical evidence. QoE may vary between users for the same network or telecommunication provider because users have different tastes and preferences. Thus, network QoE reputation may vary between users for the same network. Of course, in a democratic tolerant world, having a different opinion cannot be considered as an attack. However, one may be tempted to cheat to influence QoE reputation. Worse, it may be a coalition of entities who collude to drive that reputation to the level they wish. This time, this is an attack.

The different types of attackers that we consider are:

- Network Provider: The goal of having Always Best Connected (ABC) may be different between telecommunication operators and end-users. Unless telecommunication operators can directly benefit from allowing a user to switch to another operator, operators have an incentive to bind the user to their networks or service provisioning. In contrast, for end-users ABC may mean saving money by switching to the lowest cost operator.
- End-user: They may attack for different reasons, from playing to making money for example as being paid to take part into a coalition of attackers.
- Coalitions of attackers composed of:
    - End-users only,
    - Network providers only,
    - A mix of end-users and network providers.

In addition, different types of attacks can be carried out at the reputation level:

- Technical attacks:
    - Propagation of false QoE evidence: by many pseudonyms created and controlled by a same entity through:
        - normal pseudonym creation,
        - spoofed pseudonyms,
        - compromised legitimate pseudonyms.
    - Destruction or denial of reputation evidence.
    - Use of the good reputation of:
        - a compromised entity,
        - by a spoofed entity.
- Social engineering attacks: Attackers may propagate false QoE evidence via not directly controlled entities from external information posted on forums to rewarding real influencing end-users or misleading end-users on the network provider that they have been using. For example, the attacker could choose a WiFi SSID with a friendly name mentioning a well-known provider for a bad network...
- Whitewashing attacks: When the entity reaches a high-enough level of reputation either via normal actions or propagation of false evidence, the entity behaves very badly without being prosecuted afterwards. The entity may be able to rejoin under another pseudonym without being detected.
- Privacy attacks: If all would be known about an entity, the reputation would correspond to the truth but the entity would have lost privacy. Mechanisms are needed to protect the privacy of the end-users.

We have been investigating computational trust and reputation management to increase the resistance of our system against the previously listed attacks.

2.2 A Trust Engine for Reputation-based Network Selection

Trust engines, based on computational models of the human notion of trust, have been proposed to make security decisions on behalf of their owner or help their owner to choose with a selection of the most trustworthy targets including their trust

information [5]. These trust engines allow the entities to compute levels of trust based on different sources of trust evidence, that is, knowledge about the interacting entities:

- local direct observations of interaction outcomes, that is in our case, the QoE level given by the user;
- recommendations, that is, sharing these personal QoE level observations with other users;
- and even reputation when it is an aggregation of QoE levels from an unknown number of unauthenticated recommenders.

Reputation is more difficult to analyze because generally it is not exactly known who the recommenders are and the chance to count many times the same outcomes of interactions is higher. Reputation may be biased by faked evidence or other controversial influencing means as mentioned in the previous subsection. Reputation evidence is the riskiest type of evidence to process. We define reputation as follows: reputation is the subjective aggregated value, as perceived by the requester, of the assessments by other people, who are not exactly identified, of some quality, character, characteristic or ability of a specific entity with whom the requester has never interacted with previously. We assume that reputation management only means to be able to perceive and monitor the reputation of an entity without trying to maliciously influence reputation that is considered as attacks. We propose to use a trust engine in our case to manage trust and reputation of the different entities, from end-users pseudonyms to networks to network providers… QoE evidence of potential networks will be weighted with trust values through our trust engine before being returned to the users for their final selection or an automated selection. Since we use a peer-to-peer system for information storage, any peer, including any mobile terminal, will have its own local trust engine that may vary in terms of memory space for evidence and computation power. All trust engines will communicate between them and support peers may host evidence for peers and compute trust values on behalf of peers in case the computations are too intensive.

We define the trust value as follows: a trust value is a non-enforceable estimate of the entity's future behavior in a given context based on past evidence. The reputation value will simply be a trust value when no direct observation has been made previously. A trust metric consists of the different computations and communications which are carried out by the trustor (and his/her network) to compute a trust value in the trustee. Three main trust contexts occur in our scenario:

- Network QoE trust context;
- User Recommending Network QoE trust context, the trust in the recommendation made by a user;
- Friend trust context, manually enforced by the users who will be allowed to specify who their friends are.

The main components of our trust engine are depicted in Figure 1. The identities of the involved entities are the first context elements that we need. To be able to assess the current risk involved in the selection and use of the future network, we need other context information such as the type of the application to be used, for example, a game or a m-banking session…

We assume that context information is given by a Context Recognition component. Another component is the Trust Value Computation component that can dynamically compute the trust value, that is, the trustworthiness of the requesting entity based on pieces of evidence (for example, direct observations, recommendations or reputation) and manually set friend trust relationships.

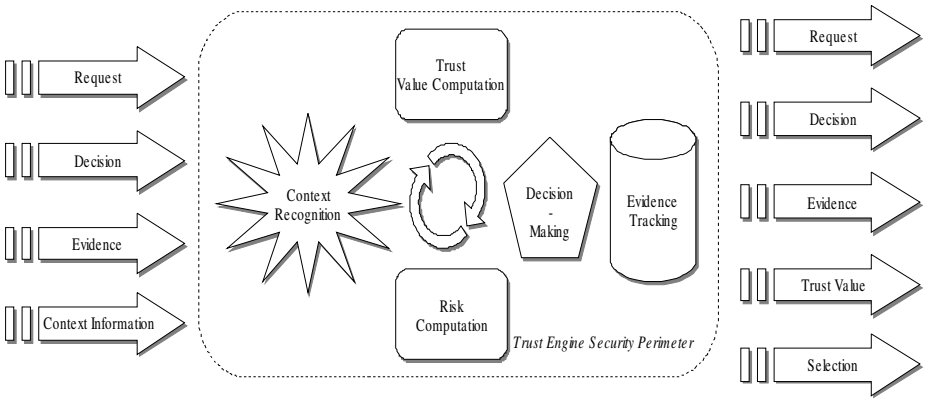


Fig. 1. High-level View of the Trust Engine

The risk component will dynamically evaluate the risk involved in the interaction based on the recognized context. Thus, risk evidence may also be needed in our scenario. In this case, the trust engine may be able to answer to a Risk Request with a Decision related to the risk level. In the background, another component, called Evidence Tracking, is in charge of gathering and tracking evidence: recommendations, comparisons between expected outcomes of the chosen actions and real outcomes… In our case, the expected outcome is a QoE matching the user expectation or user agreed QoE. The feedback from the user will create the real outcome. This evidence is used to update risk and trust information. Thus, trust and risk follow a managed life-cycle in our system.

3 Conclusion

Mobile users can already switch between their standard telecommunication provider and nearby available WiFi networks. In the near future, it will be possible for them to move from fixed subscription-based unique operator to on demand-based operator selection. To facilitate the choice of the best network available, we envision to share network QoE information among the users. To avoid false information propagation, we propose to use computational trust engines and reputation management that are resistant to a number of attacks that we have listed in this paper. Our future work is to implement and fine-tune the attack-resistance of our system.

Acknowledgments. This work is sponsored by the European Union, which funds the FP7-ICT-2007-2-224024 PERIMETER project.

References

1. ITU-T E800, http://www.itu.int/md/T05-SG02-080506-TD-WP2-0118/en.

2. PERIMETER, http://www.ict-perimeter.eu/

3. Bouvier, M.: Maxims of Law. Law Dictionary (1856).

4. Hoffman, K., et al.: A Survey of Attack and Defense Techniques for Reputation Systems. CSD TR #07-013, Purdue University (2007).

5. Seigneur, J.-M.: Trust, Security and Privacy in Global Computing. PhD Thesis, Trinity College Dublin (2005).

The Threat of Mobile Worms

Marc Fouquet[1], Elnaz Eghbali Afshar[2], and Georg Carle[1]

[1] Technical University of Munich
[2] University of Tübingen

Abstract. Mobile devices, such as cellular phones, are becoming more and more powerful. The latest devices assume permanent data connectivity to the Internet and provide the user with a rich set of 3^{rd} party applications. These developments also increase the risk of malware on mobile phones. In this work we first investigate one way of worm propagation on mobile devices. Then we explore the possible harm a mobile worm can do, including a discussion of regional denial of service attacks.

1 Introduction

Mobile phones are becoming more and more intelligent. With fast processors, the ability to install and run a huge amount of client applications and permanent Internet connectivity, today's smartphones have become powerful platforms for work and gaming.

Many mobile phones have some form of short-range communications besides the actual cell phone functionality. This is usually Bluetooth, but WLAN is also not uncommon. For the future, Near Field Communication is planned, mainly to support mobile payment applications.

There have been a number of worms for mobile devices, spreading by Bluetooth like Cabir or MMS like Beselo.A, however no serious outbreak has been observed until today. There is hope that the problem of mobile worms will not become as bad as PC worms, since mobile phones are a relatively closed platform. However phones based on Symbian, Android or Windows Mobile allow the installation of software which has not been tested by the operator. Further it is not unlikely that worms can spread via security holes, as mobile phones are patched rarely.

Cell phones are an interesting target for worms, possible goals of an attacker could be payment systems or spying on the user's location, calls and stored private data.

In this work we provide two main contributions. A study of worm propagation via shot-range communication is presented in Section 3. In Section 4 we will describe, what possible harm a mobile worm can do and what defences look like.

2 Related Work

There have been a number of publications on epidemic spreading of viruses for biological diseases but also for computer worms. In the latter case, the authors

either use simulations [1] or mathematical models of epidemic spreading [2]. Yan et. al. [3] observed that the mobility model plays an important role for worm propagation. Su et. al. [4] did experiments and simulations, showing that large-scale infections with Bluetooth worms are in fact possible.

Today, worms on PCs form highly organized botnets [5]. These networks of drones are used by criminals to send spam, to host phishing sites and to launch denial-of-service (DoS) attacks. This kind of organization has not yet been observed with mobile worms.

Security research regarding cellular networks mostly focuses on encryption and authentication issues, but also DoS attack scenarios have been investigated in the past [6].

3 Worm Propagation

In this section we describe our own worm propagation simulations using the PS-Model framework [7] for OMNeT++. PS-Model provides a radio propagation model with attenuation by walls, and a "strategy-based" mobility model, which allows users to select destinations by mood, i.e. a restaurant when they are hungry. Further, the users prefer to walk in groups rather than independently.

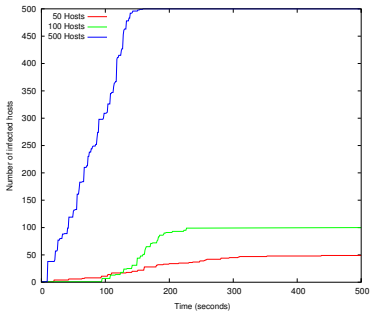


Fig. 1. Influence of the Population Size.

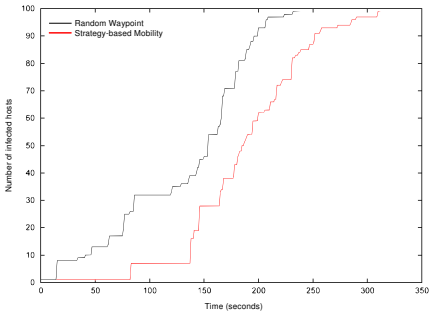


Fig. 2. Group Mobility vs. Random Waypoint.

Figure 1 shows a simple setting, different numbers of users are moving on an open 500 m×500 m playfield according to a random waypoint model. The users are carrying simulated Bluetooth class 1 devices with 100 mW transmission power, which results in a communication range of approximately 100 meters. We assume that the reception of one broadcast (i.e. Bluetooth Inquiry) message is sufficient to infect a new device. Such a message is sent by each infected device in intervals of 10 seconds. Initially only a single device is infected.

One can see that with a low density of only 50 users, it takes more than 400 seconds to infect the whole population. With a higher number of users, the time for spreading the worm decreases. With 500 users, the whole population is

already infected after 150 seconds. A higher density of users means that more potential victims are in range of infected devices and also reduces the need to rely on user movement to bring the worm to new areas.

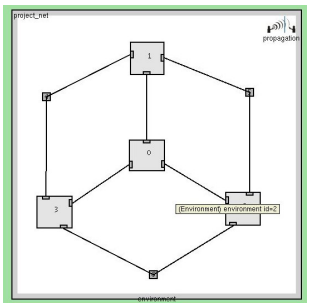


Fig. 3. Map for the Group Mobility Experiment

For Figure 2 we changed the setup to make it more realistic. The map for this experiment is shown in Figure 3. Four rectangular areas (houses) were placed, the walls of these houses attenuate the radio signal. Users can decide to move around inside each of the houses, or they can walk out on roads to move to a different house. The movement model was changed to "strategy-based" mobility as described above.

The result shows that worm propagation is slowed down under these more realistic conditions. All 100 hosts are infected after 325 seconds, while this only takes 240 seconds with the simple random-waypoint scenario. The main reasons for this are radio attenuation and the group mobility, which causes the users to form isolated clusters and therefore reduces the opportunities of spreading the worm.

On the other hand it should be noted that, even though the map has the same size as before, it is not fully occupied by the users as they stay in the houses or on the roads between them. Therefore the general user density can be considered higher in this simulation run.

4 Threat Analysis

Until now we only have considered worms that spread from one device to the next via short-range communications. However there are numerous other possibilities to infect mobile devices: SMS/MMS, drive-by infection on websites or even shipping malware as a trojan in some seemingly useful software. In this section we will explore possible dangers by mobile malware, therefore we will not restrict ourselves to a single method of infection.

4.1 Direct Threats for the User

As already mentioned in the introduction, there are a number of reasons for an attacker to write malware for cell phones:

- Users increasingly use their cell phones for email and processing other documents. Malware can have the goal of harvesting passwords and other user-specific information.
- Mobile phones are used for payment services. Especially banks are sending one-time passwords for PC-based online-banking to cellular phones. Tampering with payment systems could bring a direct financial benefit for the author of the malware.

– If desired, mobile phones allow new ways of spying on people. Not only is it possible to get audio and video from the phone, also tracking the location of the user is possible.
– Mobile malware could call expensive service numbers or use other paid services.

Further a mobile phone could also be made unusable by malware, i.e. by changing connection settings (which would cut the phone off from the network) or by automatically launching an application on startup which immediately crashes the phone. Such a destructive behaviour was common to viruses in the past, however it is rarely observed today as it limits spreading and does not give the attacker a benefit.

Defence against such malware is possible by controling the applications on the device. This would not necessarily mean a complete walled-garden philosophy, the network operator could just force the users to run a virus-scanner.

4.2 DoS Attacks on the Infrastructure

A number of denial-of-service scenarios against the network operator's core network and Voice-over-IP infrastructure in general have already been discussed in the past [6]. In this section we focus on possible flooding DoS attacks against individual cells of the radio access network.

Of course an attacker could perform any kind of jamming attack if he had direct access to the radio interface. In this section we will make the more realistic assumption that any user-installed software on a mobile phone is bound to the implemented MAC protocols for the cellular as well as the short-range interfaces.

Generally cellular networks are subject to admission control and apply QoS mechanisms, so they are more robust against DoS attacks than an unmanaged Ethernet or WLAN. A single device is unable to flood a cell, as its share of the medium would be limited by air-interface scheduling.

When multiple infected devices are present in the same cell, an attack is possible. Again air interface scheduling of the operator strongly determines the effectiveness of such attacks. The authors of [8] investigated the capacity of cells in different operator networks and their reaction to high-load conditions. Their results suggest that 20-80 voice calls are required to deplete the resources of an UMTS cell. They also found out that there is an operator-specific guaranteed minimum bandwidth for each data call and that not all operators prioritise voice higher than data. Highly-loaded cells tend to show strange behaviour, including the possibility that all active connections are dropped. Some network operators also offer video calls which are given a relatively high priority while using 3-4 times more bandwidth than voice calls. These results suggest that it is generally possible to DoS an UMTS cell with less than 30 terminals when choosing a combined attack pattern of voice, data and possibly video.

Such an attack requires coordination between the participating devices which could be achieved in one of the following ways:

– **Decentralized Coordination**: When multiple infected devices are close together, they form a (possibly meshed) network, which has the main purpose of counting the number of nearby infected devices. As soon as this number crosses a threshold which allows the devices to DoS the local cell of the mobile network, the devices launch their attack. This method has the advantage that no communication via the cellular network is necessary for the coordination of the attack. Disadvantages are the lack of control by the malware author and that the fact of enough devices being in the cell may not be detected as they are not close enough to form a common ad-hoc network.
– **Central Coordination**: The malware contacts some central coordination point on the Internet (possibly via some proxy to disguise the communication) and transmits its location information. Location data is available in many mobile phones today; even without GPS, the location of a mobile phone can be estimated by GSM or UMTS. The central botmaster can see the distribution of mobile phones and chose to attack cells with enough infected devices.
– **Hybrid Coordination**: To reduce the amount of detectable traffic via the cellular networks, bots can coordinate themselves via an ad-hoc network, but still contact a central point of control when a threshold regarding the network-size has been crossed. This would still allow central control with only minimal exposure.

We see several options to detect and mitigate this threat:

– Looking at the communication to the Internet-based botmaster, if existent. However this is difficult, as already today bots can disguise their traffic, i.e. as normal HTTP communications.
– Detecting short-range coordination traffic, i.e. by placing probes at crowded places. This has already been suggested in [4] to detect worm propagation.
– Scheduling attack traffic to use no or only very limited bandwidth during an attack. This requires the possibility to distinguish the attack traffic from legitimate traffic, which will again be difficult. In any case voice calls should be given highest priority.
– If the attack started on all devices simultaneously, the operator could use this fact to identify the attacking devices. However this can be disguised by the attacker by using individual start times for each device.
– If the attack is coordinated locally using an ad-hoc network, the operator could detect the participating hosts by looking at the locations of the nodes in the cell. The attacking nodes are expected to be relatively close together. This might allow blocking the attackers, however it may also cause some legitimate sessions to be dropped.
– Detecting devices that issue large numbers of localisation requests. However there might also be legitimate applications that behave similarly.

Users might also become suspicious when noticing fast battery depletion of devices with active bots, as close-range communications, GPS and also cellular localisation consume a lot of energy.

5 Conclusion

In this paper we presented our own investigations regarding mobile worm propagation. We also discussed threats by mobile worms, including a possible DoS attack on cellular networks.

One open question is, whether we really have to expect such attacks. Most of today's worms are used by criminals to earn money. It is unlikely that money can be made from extortion of network operators. However this attack can also be seen as targeting a geographical region rather than a specific operator, so for example public events could be potential victims. Even if this attack is hypothetical now, the possibility should be considered when designing cellular networks.

6 Acknowledgements

The authors would like to thank Andreas Klenk for fruitful discussions and Andreas Monger and Marcel Kronfeld for their help with PS-Model.

References

1. Vogt, T.: Simulating and optimising worm propagation algorithms. (2003)
2. Bulygin, Y.: Epidemics of mobile worms. Performance, Computing, and Communications Conference, 2002. 21st IEEE International (2007) 475–478
3. Yan, G., Flores, H.D., Cuellar, L., Hengartner, N., Eidenbenz, S., Vu, V.: Bluetooth worm propagation: Mobility pattern matters! In: ASIACCS '07: Proceedings of the 2nd ACM symposium on Information, computer and communications security, New York, NY, USA, ACM (2007) 32–44
4. Su, J., Chan, K.K., Miklas, A.G., Po, K., Akhavan, A., Saroiu, S., Lara, E.D., Goel, A.: A preliminary investigation of worm infections in a bluetooth environment. In: Proceedings of the ACM Workshop on Rapid Malcode (WORM), Alexandria, VA, USA (2006)
5. Holz, T., Steiner, M., Dahl, F., Biersack, E.W., Freiling, F.: Measurements and mitigation of peer-to-peer-based botnets: a case study on storm wor. In: LEET'08: 1st USENIX Workshop on Large-Scale Exploits and Emergent Threats, April 15, 2008, San Francisco, USA. (Apr 2008)
6. Zhao, B., Chi, C., Gao, W., Thu, S., Cao, G.: A chain reaction dos attack on 3G networks: Analysis and defenses. In: IEEE INFOCOM 2009. (2009)
7. Petrak, L., Landsiedel, O., Wehrle, K.: Towards realistic strategy-based mobility models for ad hoc communication. In: Proceedings of the 2005 Conference on Software for Communication Systems and Computer Networks. (2005)
8. Tan, W.L., Lam, F., Lau, W.C.: An empirical study on the capacity and performance of 3G networks. IEEE Transactions on Mobile Computing **7**(6) (2008) 737–750

A Scalable Context-aware Solution for Inter-vehicle Communication

Yasar Ansar-Ul-Haque[1], Preuveneers Davy[2] and Berbers Yolande[3],

[1,2,3] Deparment of Computer Science, Katholieke Universiteit Leuven,
Celestijnenlaan 200A, 3001 Leuven, Belgium
{ansarulhaque.yasar, davy.preuveneers, yolande.berbers}@cs.kuleuven.be

Abstract. In a large scale network inter-vehicle communication leads towards a common concept known as 'telematics' which refers to the idea of vehicles equipped with smart computing devices with communication capabilities over certain networks. In this paper we present our relevance backpropagation algorithm for efficient context-aware inter-vehicle communication as prevailing P2P communication protocols and routing algorithms do not well serve the purpose. Our preliminary results show that relevance backpropagation decreases the communication overhead in a large scale vehicular network.

Keywords: Context-awareness, Scalability, Telematic, Inter-vehicle, Communication, Networks, Ubiquitous.

1 Introduction

In the ubiquitous computing paradigm, devices and applications are able to interact with one another and often have an awareness of the context of their users to create a smart environment that is proactive, dynamic and supportive. These applications are characterized by their ability to adapt their behavior to the ever changing environment in which they operate. Such dynamic settings with integrated information and communication technologies are found in intelligent transportation and traffic management systems, for example, which can also be referred to as Telematics. These applications employ wireless and sensor network technologies to create new opportunities for exchange of information within and between vehicles.

Intelligent large scale vehicular networks point towards many application areas such as emergency message transmission, collision avoidance, congestion monitoring and intelligent parking space location. The goal is to acquire relevant context information from other vehicles and manipulate this information to perform context-sensitive tasks. There is, however, a critical aspect in the development of such intelligent applications i.e. getting the *right information at the right time and place*. Current P2P communication protocols like Gossip, Pastry and Chord are not suitable for scalable context-aware information dissemination as the relevancy of information cannot be determined at intermediate nodes during interaction between several nodes and also no routing algorithm takes relevance of context into account. This is the focus of our research work.

2 Contributions and Outcome

In order to provide a solution to the problems mentioned earlier in section 1, we developed a relevance backpropagation algorithm [1] to enable scalable and efficient context-aware communication between vehicles so that only relevant information and services could be delivered to the interested nodes at right time and place. We also incorporated various versatile application requirements in our algorithm to ensure scalable context-aware communication which are as follows:

Spatial coverage: The application must not only take into account the geographic coverage area, but also efficiently route information to other vehicles within the network at a different location, velocity and direction where relevant e.g. in case of a car accident at a highway.

Timeliness of information: It is crucial that only up-to-date context information reaches its destination as it can lose relevance after a certain period of time e.g. free parking spot information.

Routing efficiency and efficacy: This parameter deals with the dynamic nature of the vehicular network and measures effectiveness inter-vehicle communication protocol aspects.

These aspects are integrated into our *Relevance backpropagation* algorithm which relies on feedback of neighboring nodes to reduce the number of peers to forward the information to. The information is initially forwarded to the adjacent nodes unless maximum number of hops is reached. Each forwarding node reduces the hop counter, adds its identification and marks the message relevancy tag if the information is relevant for its purpose. The feedback technique is based on context information like position, velocity, direction, time-to-live, interest etc that decides whether the data that was received is relevant or not and also help determine the information relevancy on the intermediate nodes. The feedback to the delivering node is initiated if the context information *is relevant, irrelevant*, *unused* or *duplicate information* is received reducing the information dissemination only to the interested nodes. A vehicular network is highly dynamic in nature and application dependent. As the context information can be provided by the application itself the routing of the information is adapted accordingly and perhaps different for various applications. So the network re-calibrates itself if a new node sends an *arrival beacon* or *an old node no longer transmits the feedback information*. In this mechanism the goal is to efficiently filter and route the relevant information as close to the source as possible in a dynamic network.

Our preliminary results show that relevance backpropagation decreases the communication overhead between the nodes in a large scale vehicular network [1].

References

1. Yasar, A., Preuveneers, D., Berbers, Y.: Adaptive context mediation in dynamic and large scale vehicular networks using relevance backpropagation, in the proceedings of Mobility conference, Taiwan (2008).

Wireless Mesh Networks for Interconnection of Remote Sites to Fixed Broadband Networks*

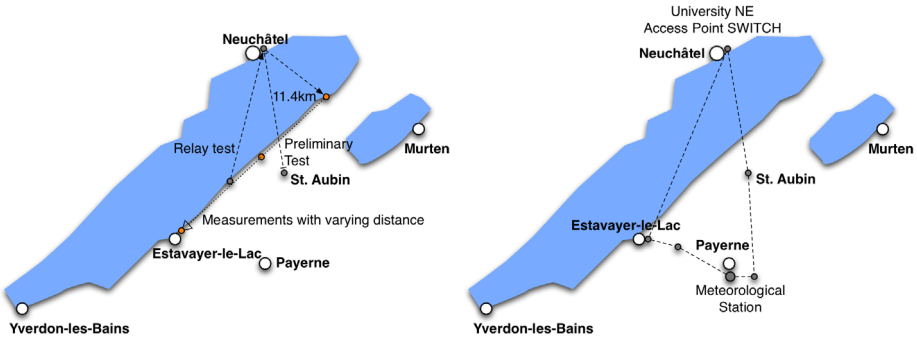
Thomas Staub[1], Marc Brogle[1], Kurt Baumann[2] and Torsten Braun[1]

[1] Institute of Computer Science and Applied Mathematics, University of Bern - Switzerland
staub@iam.unibe.ch
[2] SWITCH - Switzerland
kurt.baumann@switch.ch

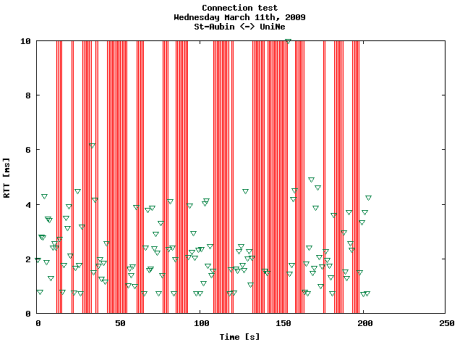
Abstract. The paper describes a technology transfer project that intends to evaluate the usefulness and feasibility of wireless mesh networks (WMNs) in meteorological monitoring applications. We try to identify application and usage scenarios for WMNs. We investigate whether and how the used hardware and software components are appropriate for the intended application scenarios and whether the application requirements such as bandwidth, delay, reliability, recovery times etc. can be met. Potential weaknesses and bottlenecks will also be identified. We performed a preliminary latency test over 11.4 km.

We investigate whether wireless mesh networks (WMNs) are appropriate for connecting sensor networks or other devices deployed in remote areas, where no fixed network access is available, to a fixed broadband network. To support a variety of application scenarios, the WMN must meet reliability requirements and bandwidth in the 10 Mbps range over distances of several 10 km, e.g., by using directional radio transmission. We intend to develop, deploy, and evaluate a WMN based on IEEE 802.11a/h (5 GHz). WMNs would allow SWITCH, the provider of the Swiss national research and education network, to extend the geographic coverage of their fiber network and to offer broadband services to further locations that are not close to the fiber network.
Experiments with real-world deployments have proven the usability of directional antennas for wireless radio networks to connect nodes at long distances (e.g., Heraklion MESH, WiLDNet CATER, and Quail Ridge Reserve WMN). Actual measurement results of far-distance 5 GHz (802.11a/h) links applying directional antennas are rare. The advantage of 5 GHz links is expected in lower interference with existing networks.
We intend to run experiments with different network scenarios. We need to evaluate the attainable signal quality, bandwidth, and delay for a single long distance hop (Fig. 1(a)). Afterwards, relaying the transmission by intermediate mesh nodes has to be evaluated. Moreover, we will measure the QoS characteristics including reliability using redundant paths via multiple intermediate hops (Fig. 1(b)). Finally, performance tests in case of link breaks (recovery times) will be performed. Our goal is to build a reliable and redundant connection between the meteorological testing station of MeteoSwiss at Payerne and the access point to the SWITCH backbone in Neuchâtel over multiple mesh nodes deployed in the region of lake Neuchâtel.

* This work was supported by the Swiss Commission for Technology and Innovation (grant number 9795.1 PFES-ES) and our industry partners (SwissMeteo, SWITCH, and PCEngines).



(a) Preliminary, distance and relay tests (b) Meteorological station ↔ SWITCH



(c) Neuchâtel (d) St. Aubin (e) Measurement results

Fig. 1. Test scenarios and results of preliminary measurements Neuchâtel ↔ St. Aubin

Preliminary measurements have been performed between a mesh node mounted on the roof of the University of Neuchâtel and another temporary node on the other lake side (distance of 11.4km). The nodes and the measurement results are shown in Fig. 1(c) - 1(e). Our first measurements performed using ICMP echo packets show good round trip times (0.8 - 5ms). Periodically, there are some time periods with packet loss (39% of the totally sent packets with vertical lines). Strong winds during the measurements caused periodic movement of the antenna top (5m above ground). By tensioning the antenna to ground with ropes, the movement can be eliminated in future measurements. Moreover, no adaptations have been made to 802.11a/h parameters in the first measurements. Fine-tuning the parameters for the long-distance links may further reduce the packet loss. Two other problems occurred. First, the correct alignment of the antennas without tools is very difficult. Binoculars, a clinometer and an amplitude compass will be used in future measurements. Second, the dedicated 100m Ethernet cable (mesh node on the roof of University of Neuchâtel ↔ border router of SWITCH in basement) and the lightning protector resulted in packet loss due to high attenuation and collisions. The packet loss has been eliminated by reducing the cable distance to the next active network component through an additional VLAN on the building network instead of the dedicated physical connection.

Cooperation Incentives between Wireless Mesh Network Operators

Xenofon Fafoutis and Vasilios A. Siris

Institute of Computer Science (ICS)
Foundation for Research and Technology - Hellas (FORTH)
P.O. Box 1385, GR 711 10 Heraklion, Crete, Greece
{fontas,vsiris}@ics.forth.gr

1 Motivation and Problem

In wireless mesh networks, low transmission rate links create bottlenecks that degrade the end-to-end throughput. To make matters worse, low rate links degrade the performance of neighboring links that operate on the same channel, due to the long duration of low rate transmissions and the per-packet fairness of the 802.11 MAC protocol [1].

One approach for mitigating the performance degradation due to low rate links is to consider routing protocols with metrics such as ETT [2], WCETT [2] and CATT [3], which take the transmission rate into account. In wireless LANs, another approach is to use relay nodes in order improve performance by having multiple high rate transmissions instead of a single low rate transmission [4][5]. The above works either focus on a single network or assume full cooperation between the nodes. Our work, on the other hand, considers the case of wireless mesh networks which belong to different operators, which act in their own self-interest. For such a scenario, we investigate the incentives for cooperation due solely to performance improvements that cooperation can yield for all mesh networks.

2 Performance Incentives for Cooperation

Consider two mesh networks, A and B. An overlapping part of these networks consists of a sequence of four nodes, A_1, B_2, A_2 and B_1. Network A has traffic originating from A_1 and destined to A_2, while network B has traffic originating from B_1 and destined to B_2. Without cooperation, there are two flows from A_1 to A_2 and from B_1 to B_2 respectively, Fig. 1(a). When the two mesh networks cooperate, Fig. 1(b), nodes B_2 and A_2 forward the traffic of the mesh networks A and B, respectively. As a result, the traffic from A_1 to A_2 flows through B_1, and the traffic from B_1 to B_2 flows through A_1.

As a first step, we assume that all mesh nodes operates at the same channel. This scenario arises in dense networks, when the orthogonal channels are limited, e.g. wireless networks operating at the 2.4GHz band. The model we present,

(a) No Cooperation. (b) Cooperation.

Fig. 1. Overlapping Wireless Mesh Networks.

however, can be extended to the case of multiple channels and mesh nodes with multiple radio interfaces.

Assuming that the MAC layer protocol provides long-term fairness in medium access, as can be assumed for 802.11 DCF, the long-term end-to-end throughput of each mesh network is

$$X^a = \frac{1}{\frac{1}{R_{A_1A_2}} + \frac{1}{R_{B_1B_2}}} \quad \text{and} \quad X^b = \frac{1}{\frac{1}{R_{A_1B_2}} + \frac{2}{R_{B_2A_2}} + \frac{1}{R_{B_1A_2}}}\,, \tag{1}$$

where X^a and X^b refer to the no cooperation and the cooperation case respectively and R_{ij} is the rate of the links between the nodes i and j. When the mesh networks cooperate, the same amount of data needs two hops to reach the destination. The ratio of X^b over X^a indicates when cooperation is beneficial for both mesh networks, and gives the corresponding gain. The above expressions can be extended to account for the protocol overheads.

As an example, consider that the wireless links use the 802.11b protocol. Assuming the scenario where the transmission rate of the links A_1A_2 and B_1B_2 is 1Mbps and the transmission rate of the links A_1B_2, B_2A_2 and A_2B_1 is 11Mbps, the model estimates that the cooperation yields 5.5 times increased end-to-end throughput for each mesh network.

References

1. M. Heusse, F. Rousseau, G. Berger-Sabbatel and A. Duda. Performance anomaly of 802.11b. In IEEE INFOCOM, 2003.
2. R. Draves, J. Padhye, and B. Zill. Routing in Multi-Radio, Multi-Hop Wireless Mesh Networks. In MobiCom, ACM Press, pp. 114-128, 2004.
3. M. Genetzakis and V. A. Siris. A Contention-Aware Routing Metric for Multi-Rate Multi-Radio Mesh Networks. In SECON, 2008.
4. L. M. Feeney, B. Cetin, D. Hollos, M. Kubisch, S. Mengesha and H. Karl. Multi-rate relaying for performance improvement in IEEE 802.11 WLANs. In WWIC, 2007.
5. V. Bahl, R. Chandra, P. P. C. Lee, V. Misra, J. Padhye, D. Rubenstein and Y. Yu. Opportunistic Use of Client Repeaters to Improve Performance of WLANs. In ACM CoNEXT, 2008.

SWITCH PWLAN
Proposal of a Multi Provider Enabled Infrastructure

Kurt Baumann

SWITCH, Werdstrasse 21, CH-8021 Zurich, Switzerland
kurt.baumann@switch.ch

Abstract. In a reciprocal approach of SWITCH Public Wireless LAN commercial users gain validated access to the Internet through a Multi Provider Portal (MPP) on universities' hot-spots. In return students and staff of the higher educational institutions have free Internet access by VPN at a large number of public hotspots (railway stations, airports etc.) between their home and universities. The introduction of Eduroam allows a layer two authentication using the EAP-authentication framework. Thus the synergy of the Eduroam project, a seamless roaming based on EAP-SIM implementation at the university campus simplifies the Internet access for commercial users and enlarges the footprint of the Wireless Internet Service Providers (WISPS) as the footprint of the Academic Association of SWITCH PWLAN. The approach of a multi provider/user capable architecture enables on the one hand a centralized multi provider handling and on the other hand a seamless roaming for students, staff, researchers and commercial users.

Keywords: Wireless Access, WLAN, PWLAN, EAP-SIM, Seamless Roaming

1 Introduction

The purpose of SWITCH PWLAN [1] is to represent the standard of the future WLAN-technologies and also to strengthen acceptance within the Mobile Working Group [2] of SWITCH in Switzerland. The vision behind all activities of SWITCH PWLAN is sharing the educational-wireless infrastructure with the WISPS (Wireless Internet Service Provider) infrastructure. Referring to the Bologna activities it is vital to investigate in new Wireless LAN-technologies to become more and more flexible for students, staff, and researchers at commercial platforms and also for commercial customers at the universities. Thus a pertinent question aroused on which wireless network access a future WLAN-infrastructure should be built up using the EAP-authentication framework [3]. The answer for covering heterogeneous WLAN infrastructures of PWLAN-universities was a trial to demonstrate the feasibility of combining a RADIUS [4] infrastructure based on the IEEE802.1x [5] framework for port based access control. Using technical expertise of Eduroam [6] a first EAP-SIM test bed was implemented at the end of 2007.

This paper is divided into five parts. Section 2 shows the basis of SWITCH PWLAN and explains briefly the reciprocal approach. Section 3 investigates the extension by EAP-SIM at the educational institutions in cooperation with the WISPS and in section 4 a revised EAP-SIM [7] implementation will be discussed, the vision of a multi-provider-capable infrastructure. Finally in section 5, PWLAN-traffic statistics will be analysed and a status of SWITCH PWLAN will be given.

2 Basic PWLAN

The basis for SWITCH Public Wireless LAN is a VPN-solution, distributed to the most Swiss universities that allow an "inter-institutional access" to local resources like personal data-bases, libraries, mails, e-learning-tools etc. An additional component, a Multi Provider Portal (MPP) [8] enables the reciprocal approach; that means on the one hand the unrestricted free of charge access on the Internet on commercial hot-spots for students, researchers and staff by VPN and on the other hand a validated access by the MPP on the campus for commercial users.

3. PWLAN extended by EAP-SIM

EAP stands for Extending Authentication Protocol and is primarily developed as a Point-to-Point Protocol (PPP); today we are speaking about a generic Authentication Protocol according to RFC3748. EAP-SIM is a mechanism for mutual authentication and session key agreement using the Global System for Mobile Communications (GSM) Subscriber Identity Module (SIM).

The introduction of EAP with IEEE802.1x in SWITCH PWLAN required significant changes to the current roaming architecture at the universities.

Belong the unencrypted docking network of basic PWLAN an additional VLAN, WPA2-encripted is implemented (Fig. 1). The authorization system, universities' Access Points (APs) have to be configured for handling EAP(-SIM) authentication requests and the authentication system, a RADIUS-server, e.g. Radiator, FreeRadius has to be building in the roaming architecture. The routing method of the EAP(-SIM) authentication is realm-based; SWITCH-user's EAP(-SIM) request, for instance will be peroxided (authenticated) according the realm **@switch.ch** by the RADIUS-server. Now how is the implementation of a seamless roaming for providers at the universities?

A. Architecture

The EAP-SIM Radius requests are originated in the AP management IP-interface, and forwarded to the AAA/RADIUS-server, the authentication system at provider's data center using a Radius protocol message. The fact that the providers' IP range is private a source, destination NAT (Network Access Translation) of the data center's router is mandatory.

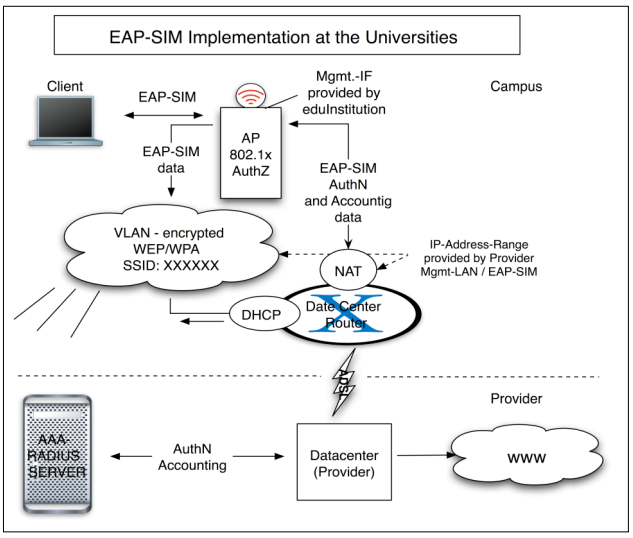


Fig. 1. EAP-SIM Extension Architecture at each University

EAP-SIM is used for seamless and fast, broadband connections in hotspots environment. A hotspot operator also provides access to the GSM network through standard WLAN consisting of 802.1x-supplicants with a SIM-card; Modem or PCMCIA-card. During the authentication process, the AAA/RADIUS server requires information on the SIM card to access the GSM authentication centre by a GSM/MAP/SS7 gateway. If the SIM-card and EAP-SIM software are able to validate the GSM triplets, the AAA/RADIUS server requests that the APs grants network access to the users. The AP connects the user's device to the WLAN and forwards accounting-information to the AAA/RADIUS-server to indicate that the connection has been completed. The current EAP-SIM extension is highly provider based and guarantees appropriate commercial users seamless roaming on the campus.

4. Multi-Provider Enabled Infrastructure

With respect to the current implementation of EAP-SIM, scalability problems will occur, because if there are more then one provider a lot of different SSIDs have to be defined and broadcasted by the educational institutions. Furthermore, an additional EAP-SIM infrastructure provider based will be needed on each campus. Thus a multi-provider/user-enabled-architecture seams to be a valuable approach. That means a common SSID for all providers, a port based access control, EAP/IEEE802.1x and a hierarchical RADIUS-server infrastructure with 802.1x-supplicants, supported by the operational systems. Now let us take a look to the proposal of a Multi Provider enabled Infrastructure:

Architecture:

Common SSID: This approach requires a common SSID, separate VLAN, broadcasted at each participant's locations.

Confederation Top-Level-Domain (TLD)-RADIUS-Server, at SWITCH: SWITCH represents the Academic Association (educational institutions) and maintains the Top-Level-Domain-RADIUS-server of SWITCH PWLAN, which authorizes each user at the location to local policies and authenticates its own users by realm. Furthermore, end-users from other educational institutions, providers will be routed to the appropriate institutions/provider RADIUS-server.

Federation Institutional-RADIUS-(Proxy)-server: These RADIUS-server have the same functionality on confederation level as the TLD-RADIUS server.

EAP-requests: EAPOW-start-message from the supplicant starts the authentication protocol and indicates the APs that the client wants to authenticate using EAP. In response, the APs send an EAP-Identity request message to the supplicant. At this point the client has not yet been assigned an IP-address, and the APs block all messages from the client except for those necessary for authentication; EAP, EAP-SIM protocol messages, challenges. The authenticator-system determines whether the authentication is a success or a failure. After a successful authentication, authorization (layer 2) the user will get an IP-Address (layer 3).

Network Topology: The implementation will be based on IEEE802.1x, wired or wireless. Different approaches could be possible:

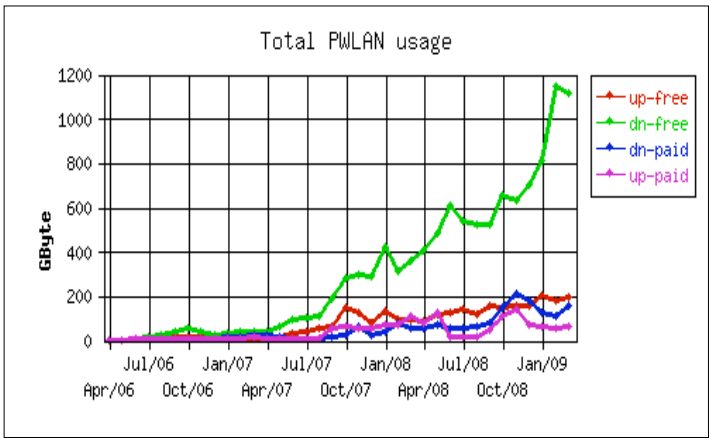
- *In focus of stacking VPN on IEEE802.1x:* A VLAN will be defined at each location. Authenticated at their Home Organization, authorized to the local policy users get a locations' IP-address without routing on the Internet. From this service restricted VLAN only VPN is feasible. Such an approach finally delegates security to the educational institutions and prevents that unknown users getting local IPs on the Internet.
- *In focus of a special IP-address range provided by SWITCH for SWITCH PWLAN:* An IP-address range divided into different subnet handled over the same SSID has to be defined at the educational institutions and providers. This means, for each organization a sub-VLAN has to be defined. Thus after successful authentication/authorization users get an appropriate IP-address and is visible with it on the Internet.
- *In focus of extending the campus to over all participants*: In this case an EoMPLS (Ethernet over MPLS) architecture respective a layer 3 VPN is a logical approach to build up a virtual campus including the educational institutions and all providers.
- *In the focus of introducing a common policy*: This means that a closed community has to be defined, e.g. a confederation including a service organization with a policy-authority, SWITCH and its federation members, the educational institutions and providers [6].

5. Facts / Conclusion

At the beginning of SWITCH PWLAN three strategic objectives were in front of the project:

1. "SWITCH PWLAN extends footprint for the Academic Association of SWITCH and for the WISPS".

Today SWITCH PWLAN is steadily growing; more then thirteen educational institutions have joined SWITCH PWLAN, the Academic Association. Thus a potential of more than 150'000 users, students, staff and researchers were providing a date volume of 1.114 TB traffic (dn-free) on the commercial platforms in March 2009 (Table 1). Between November 2008 and February 2009 the traffic volume provided by educational staff increased to totally 483GB. Such an extreme progress compared to previous moths could be explained by introducing new 3G-mobile devices (e.g. IPhone). Furthermore, new business models of the WISPS supported the development of SWITCH PWLAN.



up-free :traffic uploaded by students, staff and researchers at the WISPS
dn-free: traffic downloaded by students, staff and researchers at the WISPS
dn-paid: traffic downloaded by WISPS' customers at universities
up-paid: traffic uploaded by WISPS' customers at universities

Fig. 2. PWLAN usage volume based

Table 1. Traffic measurement – 6th months period

Month/Year	up-free MBytes	dn-free MBytes	dn-paid MBytes	up-paid MBytes
March 2009	187415 *+10211**	1114050 *-26138**	150183 *+45259**	56510 *+4626**
Feb 2009	177204 *-18146**	1140188 *+332842**	104924 *-12955**	51884 *-3154**
Jan 2009	195350 *+45711**	807346 *+112566**	117879 *-59198**	55038 *-13691**
Dec 2008	149639 *-4470**	694780 *+64171**	177077 *-31700**	68729 *-66114**
Nov 2008	154109 *+11723**	630609 *-21955**	208777 *+64742**	134843 *-31101**
Oct 2008	142386 *-7237**	652564 *+135831**	144035 *+72068**	103742 *+61426**
Sept 2008	149623	516733	71967	42316

* to previous month

As commercial platforms, four WISPS are participating to SWITCH PWLAN. Thus commercial users were providing at the educational institutions a data volume of totally 150GB (dn-paid) in March 2009. One reason for the fluctuations in the last months is suited in the fact of minor notoriety of SWITCH PWLAN into commercial environment.

Until the end of 2009 four new educational institutions have joined SWITCH PWLAN. Actually no further provider is taken in SWITCH PWLAN.

2. "SWITCH PWLAN corresponds technologically to the most current standards; IEEE802.11b/g, IEEE802.1x and EAP/EAP-SIM".

SWITCH PWLAN was built up on classical WLAN technology, IEEE802.11b/g . The technical extension of the infrastructure by EAP-SIM/IEEE802.1x (Fig.1) enables seamless roaming on the campus and is actually an additional infrastructure, because only one WISP grants to their users access on the Internet. Thus one vision should be to offer a PWLAN service only based on EAP/IEEE802.1x for many different providers and educational institutions. For this reason discussions about moving SWITCHconnet [9] on IEEE802.1x are in progress.

3. "SWITCH PWLAN enables a further enlargement of the user population by a "Multi Provider Enabled Infrastructure".

Based on the development of SWITCH PWLAN, a multi-provider enabled infrastructure will be the conclusion of a former reciprocal approach for handling different providers and educational institutions. A layer two authentication through an 802.1x supplicant simplify the access on the Internet at the educational institutions and providers' sites. Furthermore, security is delegated to the appropriate participants and a proper accounting on confederation/federation level allows reconstructing interconnection traffic. The challenge for building up such a

multi-provider enabled infrastructure is to virtalize, geographically as technically heterogenous network environment between the Academic Association and the WISPS.

References

1. (2008) SWITCH PWLAN website. [Online]. Available: http://www.switch.ch/pwlan
2. (2008) Mobile Working Group website. [Online]. Available: http://www.switch.ch/mobile/mobilwg
3. Extensible Authentication Protocol (EAP), IETF RFC3748, June 2004
4. Remote Authentication Dial in User Service (RADIUS), IETF RFC2865, June 2000.
5. Port-Based Network Access Control, IEEE Std 802.1X-2004, 2004
6. (2008) Eduroam website. [Online]. Available: http://www.eduroam.org
7. Extensible Authentication Protocol Method for Global System for Mobile Communication (GSM) Subscriber Identity Module (EAP-SIM), IETF RFC4186, January 2006
8. (2008) Multi Provider Portal (MPP) website. [Online]. Available: http://www.ins.hsr.ch/Produkte.2201.0.html.
9. (2009) SWITCHconnect website. [Online]. Available: http://www.switch.ch/connect

Middleware for decentralized multimedia multiparty applications in the IP Multimedia Subsystem*

Alberto Hernández, Enrique Vázquez,
Pedro Capelastegui, Manuel Álvarez-Campana

ETS Ing. Telecomunicación, Universidad Politécnica de Madrid
Av. Complutense s/n, 28040, Madrid, Spain
{albertoh, enrique, capelastegui, mac}@dit.upm.es

Abstract. Multimedia multiparty applications on the Internet are commonly supported by specific servers containing all the required functionalities in a centralized manner. The IP Multimedia Subsystem (IMS) is the current paradigm of the Next Generation Network (NGN) model. It provides a set of basic service enablers and functions beyond IP connectivity, such as multimedia session control, presence and group management. This paper proposes a middleware aimed to ease the development of server-less multiparty applications over the IMS, providing a high level of abstraction. This middleware manages the communication sessions according to the selected topology (full mesh, cascaded, client-server) transparently for the applications, making the most of IMS service enablers.

1 Introduction

The IP Multimedia Subsystem (IMS) was originally conceived by the Third Generation Partnership Project (3GPP) as the solution for providing IP-based multimedia services on Universal Mobile Telecommunication System (UMTS). The IMS defines a flexible network architecture based on Session Initiation Protocol (SIP) [1] servers and other elements. The IMS provides a set of functionalities that complement the basic media transport capabilities offered by IP networks. Over recent years, the IMS has been adopted by other standardization bodies such us ETSI/TISPAN, 3GPP2, CableLabs and the WiMAX Forum, becoming the paradigm for the provision of multimedia services in fixed and mobile networks.

Multimedia multiparty applications on Internet usually rely on dedicated servers which provide all the required service capabilities. In the case of P2P applications, the lack of user management functions in the underlying network is paid in terms of added complexity in the peers. The IMS model can contribute to overcome the mentioned problems, avoiding the need for specific server-side support and simplifying the management of P2P topologies.

* This work has been partially supported by the Spanish Ministry of Science and Innovation, within the national Researching Formation Program (BES-2006-12803) and the project CASERTEL-NGN (TSI2005-07306-C02-01)

Decentralized topologies help reducing the deployment cost of multiparty applications, though they difficult the development. For instance, in a centralized online game with several users, each player usually establishes a single session with the game server. In a decentralized full-mesh scenario, players establish a session with every other player.

Java is a standard platform for developing IMS applications, both at the server and at the client side. Application servers usually include *SipServlet* containers compliant with Java Specification Requests JSR 116 and/or JSR 289. Java clients can be implemented following the recently defined JSR 281 and JSR 325. Nevertheless, these specifications are based on one-to-one sessions and do not provide decentralized multiparty features.

This paper proposes an optimised Java-based middleware specifically oriented to the provision of highly interactive multimedia multiparty applications (HIMMA) over IMS. The proposed middleware includes the definition of a new Java Application Programming Interface (API) which provides a higher abstraction level than previously standardized APIs for IMS client applications, namely JSR 180 and JSR 281/JSR 325. This middleware enables decentralization from the developers' point of view, automatically managing the sessions among peers required to form a full mesh of nodes, a cascaded hierarchy of nodes and supernodes, or a traditional client-server topology.

The rest of the paper is organized as follows. Section 2 cites related work. Section 3 provides a brief description of the proposed HIMMA middleware and section 4 details the interactions with the IMS nodes that operate underneath the middleware. Finally, section 5 summarizes the contributions of the paper.

2 Related work

We have not found any previous work directly addressing the definition of a middleware for the provision of decentralized IMS multiparty applications.

The most relevant work related to generic development of IMS client software is the IMS Client Platform (ICP) [2], included within the Service Development Studio (SDS) of Ericsson. The ICP Java client interface was proposed as a Java standard in 2005, under the JSR 281 specification led by Ericsson and BenQ. At its final stage of standardization, JSR 281 was separated into two specifications:

- *JSR 281 (IMS Services API)* [3], approved as a standard in July 2008, includes basic IMS features such as registration and establishment of audio and video sessions, as well as a generic framework for accessing to IMS services.
- *JSR 325 (IMS Communication Enablers)* [4], currently in progress (early draft), provides an interface for accessing to specific IMS service enablers, namely presence, group management and instant messaging. This part was removed from the original version of JSR 281.

ICP and related JSRs simplify the development of IMS applications on the client side. However, they are based on one-to-one sessions only, and do not include the concept of multiparty activities. Therefore, they are not adequate for applications that require establishing and managing multiple sessions with several parties.

3 HIMMA middleware

The proposed HIMMA middleware is conceived as an additional layer on top of the ICP implementation (similar to JSR 281/JSR 325). It provides an API for developing IMS applications following an activity oriented approach instead of the session or dialog oriented approach of current Java APIs, as illustrated in Fig. 1.

The HIMMA middleware manages multiparty activities by organizing participants into groups through a friendly Java interface. The key feature of our middleware is that it completely hides the establishment of sessions among participants. The developer selects the appropriate topology for the application, for example client/server if there is specific server-side support, or full-mesh or cascaded for decentralized approaches. As a result, each instance of the middleware automatically manages all the required interactions with the IMS network elements and the rest of participants. In the case of cascaded topologies, the middleware also indicates whether the peer is acting as a normal node or a supernode, so the application behaves accordingly. For instance, in a multiplayer game using a cascaded topology, normal nodes send the actions of each player to their supernodes. Then, supernodes update the game state and send the new state back to their nodes.

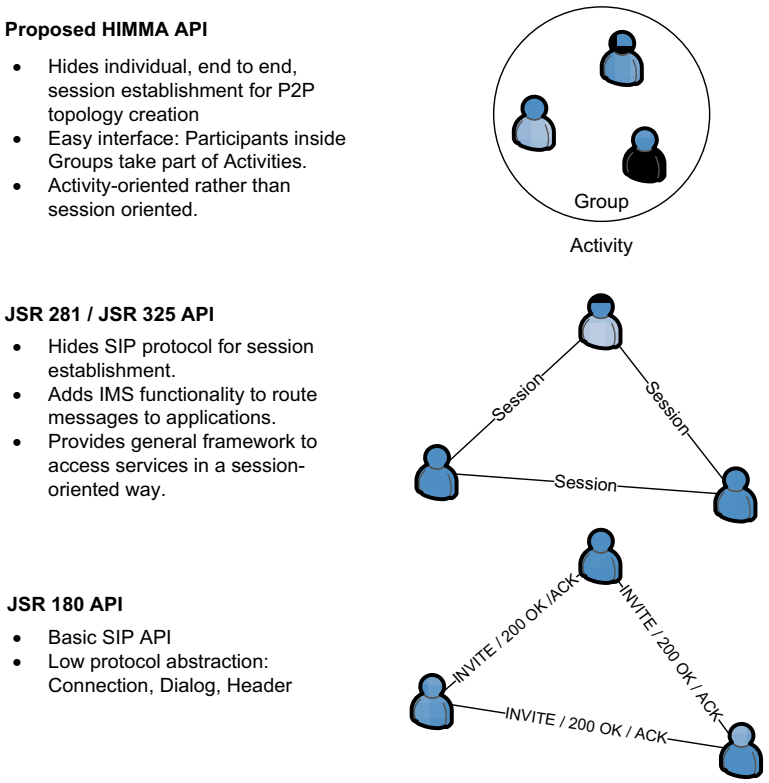


Fig. 1. Abstraction levels in the HIMMA, JSR 281 and JSR 180 client APIs

The HIMMA middleware is structured into two layers as illustrated in Fig. 2. The *Top layer* provides the highest level of abstraction, modelling multiparty applications in terms of activities. The concept of *ActivitySession* differs from the standard SIP session concept. In our middleware, interactions among participants are modelled as a single activity, regardless of the number of underlying SIP sessions actually involved. A *Group* is a specific set of *Participants* interacting in an instance of the activity. Finally, the *Engine* class provides search functions for users and activities, as well as other management features (initialization, management of blocked users, etc.).

On the other hand, the *Bottom layer* provides a lower level of abstraction, including the knowledge about the network architecture or topology in use, as well as the involved individual SIP sessions. This layer is not available through the public API, because it is aimed to extend the logic of the middleware with new or enhanced distributed mechanisms. Currently, the bottom layer supports the three aforementioned types of topologies (client-server, full mesh and cascaded).

4 Interaction with IMS functional elements

Another key feature of the HIMMA middleware is the management of the interaction with the IMS functional elements. This capability facilitates the development of multiparty applications without the need for specific server-side support. More specifically, the middleware interacts with the IMS Call/Session Control Function (CSCF) nodes (nodes acting as SIP signalling intermediaries in IMS), in order to establish the necessary sessions and create the desired topology among participants. It also interacts with the XML Document Management Server (XDMS) [5], which provides essential presence and group management functions. These functions enable the multiparty activity management and community features of the middleware.

Applications usually rely on the presence service to notify the changes in the status of the users included in a list of contacts, for example if a certain person becomes available, away or busy, among other states. However, the presence service is rarely used as an enabler to develop advanced applications. This is an original feature of our proposal. The HIMMA middleware supports the rich presence extensions defined in RFC 4480 [6], allowing users to indicate their willingness to participate in each activity, as well as other useful information like their device capabilities. In this way, the middleware enables community services like finding participants willing to get involved in a particular activity (i.e., an online game). Rich presence status is stored at the XDMS node of IMS, so further server-side support is not needed.

Managing the active participation of users in a distributed multiparty activity also involves intensive use of IMS service enablers. The 3GPP has specified a framework for audiovisual conferencing in IMS [7]. Following a centralized approach, the framework includes the definition of two functional elements: the Media Resource Function Controller (MRFC) and the Media Resource Function Processor (MRFP). The MCRF is responsible for handling the SIP signalling (acts as a conference focus) while the MCRP provides media mixer capabilities. Such elements are usually quite expensive. In our solution, the HIMMA middleware relies on IMS group management

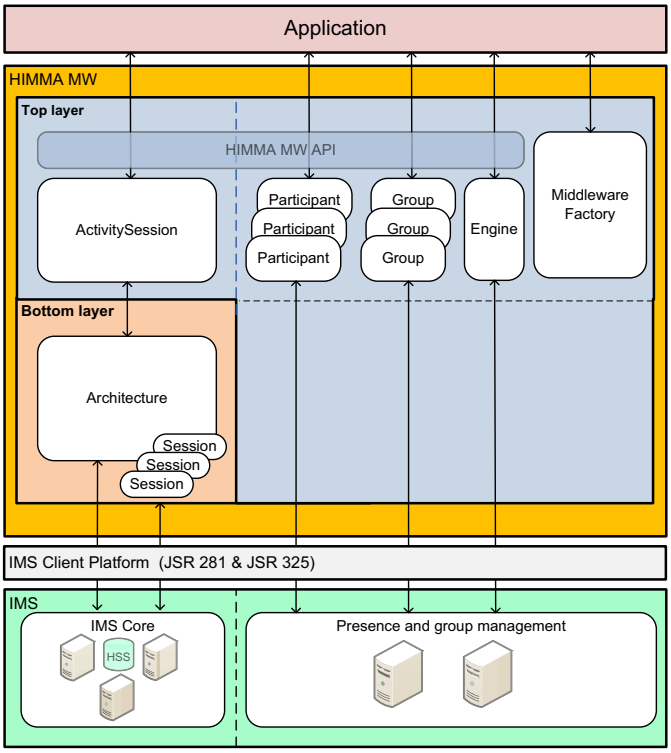


Fig 2. Software architecture of the HIMMA middleware

functions to store the participants of an activity, avoiding the need for conferencing servers. This idea is consistent with the abstraction level previously shown in Fig. 1.

To get involved into a given activity, participants must subscribe to the corresponding Group previously defined in the XDMS. The subscription provides them the capability of being notified every time a participant joins or leaves the activity. A new participant wanting to join an activity in progress must firstly discover at least one participant already engaged to it. This can be done by checking the presence-enabled contact list or by means of a XDMS query. The subsequent participants are known thanks to the discovery mechanism included in the middleware.

Fig. 3 shows an example of full-mesh topology. In this type of topologies, the HIMMA is in charge of establishing an initial SIP session with one of the participants in order to obtain the activity group identifier. This identifier is then used for querying the XDMS in order to get the SIP URIs of the rest of participants and establish a session with all of them.

In a client-server or cascaded topology, the middleware instance that receives the initial invitation may send a SIP REFER message to redirect the new participant to the server or to a supernode. Alternatively, it may accept the session request if that instance is acting as a supernode and is able to accept a new child node.

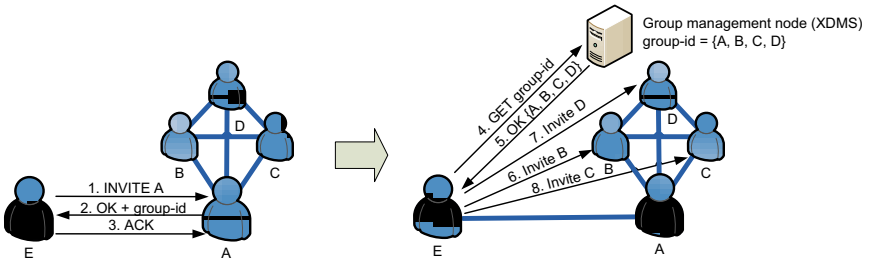


Fig 3. Discovery of participants in a full-mesh topology

5 Conclusions

This paper proposes an original middleware for highly interactive multimedia multiparty applications (HIMMA) over the IP Multimedia Subsystem (IMS). The middleware exploits IMS capabilities in order to ease the development of server-less multiparty applications. The main novelty of our approach is the intensive use of IMS session establishment mechanisms, presence and group management functions. This solution differs from conventional conference models, which usually rely on centralized architectures based on a conference focus.

We have developed a first prototype of the middleware [8] on top of the IMS Client Platform (ICP), which implements a pre-JSR 281 API. As noted in section 2, group management, presence and messaging functions have been moved to JSR 325, which is currently in draft state. Therefore, we plan to publish an updated middleware conforming to both Java standards once JSR 325 is approved.

Further work includes extending the functionality of the middleware to provide advanced functions like pausing and resuming activities, as well as multimodal and multidevice support.

References

1. Rosenberg, J. et al.: SIP: Session Initiation Protocol, RFC 3261, (2002)
2. Kessler P.: Ericsson IMS Client Platform. Ericsson Review 2, 50--59 (2007)
3. Java Specification Request JSR-281, IMS Services API, version 1.0, Ericsson AB (2008). Available from: http://jcp.org/en/jsr/detail?id=281
4. Java Specification Request JSR-325, IMS Communication Enablers (ICE), Early Draft, Ericsson AB (2009). Available from: http://jcp.org/en/jsr/detail?id=325
5. Open Mobile Alliance: XML Document Management Architecture, version 2. 0, (2007)
6. Schulzrinne H. et al.: RPID: Rich Presence Extensions to the Presence Information Data Format (PIDF) RFC 4480, (2006)
7. 3GPP: Conferencing using the IP Multimedia (IM) Core Network (CN) subsystem, Stage 3, TS 24.147 (2007)
8. Marí Diego, A., Hernández, A.: Development of a decentralized network architecture to provide highly interactive multimedia multiparty applications in the IP Multimedia Subsystem. Internal Report (2009), unpublished

Part III

Special Sessions on NoE and V2X Communications

NoE Session Invited Talk: Ubiquitous Content Networking - The CONTENT NoE Approach

Fernando Boavida

University of Coimbra, Portugal
CONTENT NoE WP4 Leader

Abstract

In a world full of content providers, where users can easily have Access to all kinds of content, are there still any challenges in ubiquitous content networking? This talk provides some answers to this question and explains the CONTENT NoE Framework and approach to content networking. The objectives are to provide an overview of issues surrounding content networking and an outlook on the further development of this crucial area.

NoE Session Invited Talk: Emanics - Managing the future Internet

Aiko Pras

CTIT, University of Twente, The Netherlands
EMANICS Research Coordinator

Abstract

Emanics is a FP6 Network of Excellence (NoE) focussing on management of future Internet infrastructures and services. Like any NoE, Emanics has three main objectives: 1) integrating research groups within Europe, 2) disseminating knowledge to the rest of the world and 3) performing research. Emanics is very successful on all three objectives, and may be a good example for future NoEs. Education of (PhD) students is very important, and Emanics members play a key position in organizing our research worldwide. Research within Emanics focuses on scalability, accounting and autonomic management. In this presentation we will focus on some of the Emanics highlights, and provide some ideas for future NoEs.

V2X Communications Session Invited Talk: Vehicle-to-Business: Use Cases, Requirements, Architecture and Technology

Thomas Michael Bohnert

SAP Research, CEC Zürich, Switzerland

Abstract

Recent years did prove it. There is no viable business model beneath Vehicle-to-Vehicle communications. Yet along with the advent of powerful broadband wireless technologies, the whole business of telematic services is enjoying a reincarnation, fundamentally enabled by Vehicle-to-Infrastructure communications. But capable communications alone do not make a business model neither fulfill current and future business model's multi-faceted requirements. In this talk we'll present insights to what we see to be the future of telematics, or how we call it, the concept of Vehicle-to-Business (V2B). Starting with an overview of V2V, V2I, and telematic service offerings, their enablers, and inhibitors, we'll move on with presenting novel ideas along with an overview in what regards inherent requirements. Based on this, we'll present an framework that allows to deliver telematic services to the whole automotive eco-system.

V2X Communications Session Invited Talk: Connect & Drive: On the use of Cooperative Adaptive Cruise Control to increase Traffic Stability and Efficiency

Geert Heijenk

University of Twente, The Netherlands

Abstract

At moderate to high traffic densities, road traffic exhibits an unstable behavior, resulting in decrease traffic efficiency. One of the phenomena that can be observed at these densities is the propagation of shock waves against the flow of traffic, which seriously reduce the speed of vehicles. The Connect & Drive project aims to mitigate these shock waves by designing a cooperative adaptive cruise control system, based on the use of vehicle to vehicle communications. Besides increasing traffic efficiency, such a system could also increase traffic safety, and reduce emission of vehicles. The project considers aspects of traffic dynamics, control systems, human machine interaction, and communication systems in a multidisciplinary approach. Besides a model-based approach, to investigate e.g. scalability aspects, the project will deliver a prototype system that will be tested with cars driving in a highway scenario. In this talk, I will highlight some of the problems tackled in the Connect & Drive project, and outline ideas for solution. The focus of the presentation will be on vehicle-to-vehicle communication system aspects.

Author Index