International Journal of Computer Theory and Engineering, Vol. 4, No. 5, October 2012

Towards an Aggregated Biomonitoring Traffic Model for Wireless Networks

O. Andrei Dragoi and Nadjia Kara

Abstract—**The use of cellular networks in telemedicine applications, in particular for biomonitoring ambulatory patients, is becoming very popular. A prerequisite for deploying such emerging services is a good understanding of their requirements, in terms of network resources. In the network traffic domain, models are needed to help estimate these needs, and to help in planning services for the mass market. This paper proposes a model for biomonitoring applications based on the experience gained from implementing a functional prototype. The model can be used to estimate the mean bandwidth aggregated over the entire cellular network needed by the biomonitoring service needs, as well as the number of clients the network operator can accept as subscribers.**

Index Terms—**Biomonitoring, traffic model, service performance.**

I. Introduction

The ubiquity of 2.5G and 3G cellular networks makes them ideal for biomonitoring applications. To support mass market deployments, cellular operators need traffic models that can quantify the impact of such services on production networks, helping them with network planning, and service pricing and marketing.

This paper proposes a traffic model to analyze the aggregated impact of the non-streaming traffic of biomonitoring services on a cellular network, and to determine the total number of subscribers who can be supported on the network. Parameters of both a technical and a medical nature are taken into account: the size and frequency of data exchanges (e.g. reading sensors, and sending alarms and images), the features of subscribers' medical profiles, how often and for how long the user's terminal remains connected to the data network and the central monitoring server, etc.

The model starts with the analysis of a prototype system intended for monitoring ambulatory patients with chronic conditions and for the prevention of health problems (among populations like athletes or baby boomers, for example). No attempt is made to characterize per-cell traffic, since that would have required the analysis of a large user population to determine mobility patterns, which was beyond the scope of a prototype system.

In the next section, we give a brief overview of the architecture of the prototype. Section III describes the proposed traffic model, followed by a discussion of sample input parameters for the model in section IV. Section V shows how the model can be used and analyzes the impact of the input parameters on the estimations obtained with the model. Section VI positions this paper among other work in the telemedicine and e-health fields. Section VII concludes the paper and identifies avenues for future work.

Manuscript received May 4, 2011; accepted July 30, 2012.

O. A. Dragoi is with the Nuance Inc, Montreal, Canada (e-mail: oadragoi@gmail.com).

N. Kara is with ETS, University of Quebec, Canada (e-mail: Nadjia.Kara@etsmtl.ca)

II. Prototype Overview

The traffic model described here was developed within the framework of a project investigating the feasibility and potential for acceptance by the medical community of a remote biomonitoring service which uses 2.5G and 3G access technologies and off-the-shelf devices. Such a service can help prevent critical medical conditions and facilitate timely assistance from the main players in the health-care area. At the core of the system is a biomonitoring center (BMC) server, which collects and manages biomedical and environmental data about subscribers to build their individual medical and environmental profiles. The BMC reacts automatically to the criticality of the situation of a subscriber, and takes decisions to execute actions (see Fig. 1).

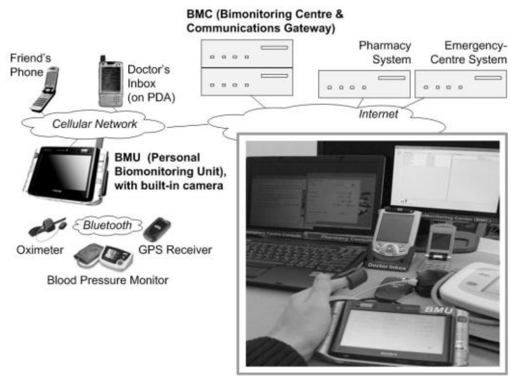

Fig. 1.Biomonitoring prototype overview

Subscribers receive a personal biomonitoring unit (BMU) and a set of sensors. The BMU acts as a filter, preprocessor, and wireless transceiver for all the contextual data that the sensors capture. In the current prototype, the BMU is an ultramobile PC with cellular connectivity.

Subscribers receive a personal biomonitoring unit (BMU) and a set of sensors. The BMU acts as a filter, preprocessor, and wireless transceiver for all the contextual data that the sensors capture.

In the current prototype, the BMU is an ultramobile PC with cellular connectivity. The BMC communicates with the subscriber mainly through notifications, warnings, and alarms displayed by the BMU.

Depending on the condition of the subscriber, the BMC can automatically contact designated friends and family members with cell phone messages generated on the fly, and it can also decide to route pertinent information to the subscriber's family doctor, in ways that doctors already communicate in their day-to-day work: email and phone calls.

For emergency situations, the BMC can channel emergency aid requests to the appropriate emergency center. The requests are annotated with medical and environmental details relevant to the recipient's role and to the subscriber's situation. Further details about the biomonitoring system prototype are given in [1], [2].

III. TRAFFIC MODEL

This section proposes a traffic model for biomonitoring applications, starting with the prototype. The pattern of cellular network usage of a BMU is highly dependent on the biomonitoring scenarios specific to a certain subscriber population, which in turn depends on the medical requirements of that population. As a result, the model is parameterized on medical profiles.

A. Assumptions and Observations

For the purposes of describing the model, a connected BMU is defined as one that is currently connected to the network. An actizve BMU is defined as a BMU that is currently connected to the network and, in addition, is currently receiving an alarm, sending images, or transmitting sensor data.

To facilitate modeling, the following simplifying assumptions are made:

1) Only the traffic on the data channel is part of the model. Any associated traffic carried by the signaling channels, like the mobile pilot or the synch channels, is ignored.
2) The traffic of the biomonitoring application can be characterized by observing the frequency of several well defined exchanges between the BMU and the BMC. Called primitives, these exchanges are:
   a. signing in and signing off with the BMC (grouped together as one primitive) – p_c
   b. receiving an alarm from the BMC – p_a
   c. sending sensor data from the BMU – p_s
   d. sending an image from the BMU – p_i
3) A BMU connects periodically to the BMC, and p_a, p_s, and p_i always end in the same connectivity interval in which they start.
4) The connectivity pattern of the BMU is to connect when there is information to send or receive, and then to remain connected for a fixed period of time.
5) The probability that any of p_a, p_s, or p_i needs to be executed while the BMU is disconnected is null.
6) For a BMU, the combination of primitives executed is specific to the medical profile of the subscriber. However, the average (typical) duration and the number of bits transferred upstream/downstream for a given primitive execution event is assumed not to depend on the medical profile. Devising medical profiles is beyond the scope of this technical document, and the profiles to which this paper refers are included for illustration purposes only.
7) How often a BMU connects is a variable which is independent of the average interval the BMU stays connected.
8) The arrival rates p_a, p_s, and p_i are independent of one another for a connected BMU. The arrival rate of p_c cannot be considered independent.

B. Traffic Model

With these simplifying assumptions, we build a steady-state model for the traffic generated by a BMU. The model is parameterized on network aspects, on application aspects, and on medical aspects, as described below. A network aspect is the total maximum number, Nc, of BMUs that can be simultaneously connected to the cellular network, as observed on the operator's network. This is the total number of data channels available.

The application aspects are captured by the application primitives and their characteristics, which include:

- La, Lc, Li, Lc the average number of bits, exchanged while receiving alarms, sending sensor data, sending images, and signing in and signing off respectively, observed experimentally; upstream, and downstream (identified by the superscripts up or down).
- d_a, d_s, d_i, d_c the mean delay observed experimentally for receiving an alarm, sending sensor data, sending images, and signing in and signing off respectively, measured for a specific wireless network technology. The medical aspects are modeled through medical profiles 1 to M, each with the specific arrival and departure rates for the various application primitives:
- λ_{cj}, μ_{cj} the assumed arrival and departure rates of connected BMUs of medical profile j.
- $\lambda_{aj}, \lambda_{sj}, \lambda_{ij}$ the assumed arrival rates for the event that a connected BMU of *medical* profile j becomes an active BMU executing the primitive p_a, p_s, or p_i respectively..
- $\mu_{aj}, \mu_{sj}, \mu_{ij}$ the assumed departure rates for the event that a connected BMU of medical profile j becomes an active BMU executing *the* primitive p_a, p_s, or p_i respectively.

Other characteristics of a medical profile are:

- Δ_j the assumed average time a BMU of medical profile j remains *connected* to the network,
- N_{cj} the estimated average number of BMUs of medical profile j that, at any given time, are connected to the network,
- α_j assumed proportion of subscribers of medical profile j out of the *total* subscriber population N.

With these, the model can estimate N the maximum number of subscribers the operator can accept, a number which is substantially more than N_c; and B the average aggregated bit rate the service generates over the entire network.

In Fig. 2, the states S_{offj}, S_{idlej}, S_{aj}, S_{sj} and S_{ij} correspond respectively to a BMU that is not connected to the network, to a connected BMU that is not currently executing one of the application primitives that generate network traffic, a BMU currently receiving an alarm (executing p_a); a BMU that is currently sending a sensor-reading update to the BMC (executing p_s) and a BMU currently sending an image to the BMC (executing p_i).

It can be observed that an *active* BMU is always in state S_{aj}, S_{sj}, or S_{ij}, while a *connected* BMU is in state S_{aj} ,S_{sj} ,S_{ij}, or S_{idlej}. It is important to note that, with the assumptions previously stated, it is impossible to move from state S_{offj} to state S_{aj}, S_{sj}, *or* S_{ij}, without first passing through state S_{idlej}.

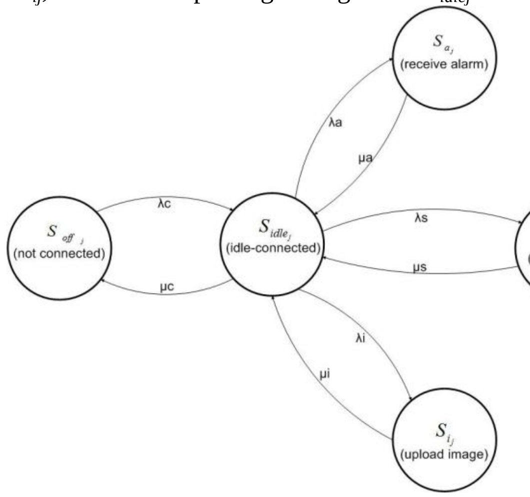

Fig. 2. Steady-state model for the BMU states

Let P_{idlej}, P_{offj}, P_{sj}, P_{ij}, and P_{aj} be the probabilities of being in state S_{idlej}, S_{offj}, S_{sj}, S_{ij}, and S_{aj} respectively. The steady-state probabilities are given in (1) that can be used to estimate the total number of users N, given in (2), an operator can accept as subscribers to the service.

$$\begin{cases} P_{offj} = \left[1 + \lambda_{cj}\Delta_j\left(1 + \sum_{o=a,s,i}\lambda_{oj} d_o\right)\right]^{-1} \\ P_{idlej} = \lambda_{cj}\Delta_j P_{offj} ; P_{ij} = \lambda_{ij} d_i \lambda_{cj}\Delta_j P_{offj} \\ P_{sj} = \lambda_{sj} d_s \lambda_{cj}\Delta_j P_{offj} ; P_{aj} = \lambda_{aj} d_{aj} \lambda_{cj}\Delta_j P_{offj}. \end{cases} \quad (1)$$

$$N = N_c\left[\sum_{k=1,\dots,M}\alpha_k\left(1 - P_{offk}\right)\right]^{-1} \quad (2)$$

Equation 1 can also be used to estimate the mean aggregated bit rates B^{up} and B^{down} given in (3).

$$\begin{cases} B^{up} = \frac{\sum_{j=1,\dots M}\alpha_j\left(1-P_{offj}\right)B_j^{up}}{\sum_{k=1,\dots M}\alpha_k\left(1-P_{offk}\right)}N_c ; B^{down} = \frac{\sum_{j=1,\dots M}\alpha_j\left(1-P_{offj}\right)B_j^{down}}{\sum_{k=1,\dots M}\alpha_k\left(1-P_{offk}\right)}N_c \\ B_j^{up} = P_{onj}\frac{L_c^{up}}{\Delta_j} + \sum_{o=a,s,i}\lambda_{oj}L_o^{up}P_{oj} ; B_j^{down} = P_{onj}\frac{L_c^{down}}{\Delta_j} + \sum_{o=a,s,i}\lambda_{oj}L_o^{down}P_{oj} \\ P_{onj} = \left(1-P_{offj}\right) \end{cases} \quad (3)$$

IV. Some Common Mistakes

Network-related characteristics and application-related characteristics to feed the model have been measured on a proof-of-concept implementation, but outside a specific clinical framework, and they are used only to illustrate how the model can be used for traffic estimations. Different values might be obtained when using an optimized production system, within specific medical constraints.

For each application primitive, the total number of bits exchanged upstream, the total number of bits exchanged downstream, and their duration are recorded. Duration is defined as the period of time over which the BMU generates, expects, receives, and processes packets. This includes the wireless network latency and other delays due to tunnelling. It also includes the processing and brokering time at the BMC. The measurements have been performed on a lightly loaded BMC and exclude the user’s reaction time. Wireshark, traffic capture and trace analysis tool, and a combination of bash scripts have been used to capture the traces. Ten measurements have been taken for each value and the resulting averages were considered.

The measurements were repeated over commonly deployed wireless access technologies: 2.5G/EDGE and 3G WCDMA/HSDPA, CDMA2000/EV-DO. The 2.5G network was a heavily used public network, and the biomonitoring traffic was shared randomly with other traffic. The 3G networks were lightly-used lab networks.

The quality of the traffic estimations depends on how realistic the *chosen* arrival rates are for the application primitives in the context of a specific medical profile. Devising such medically sound values is beyond the scope of this article.

Three medical profiles are considered: diabetic, cardiac, and elderly subscribers. Table I gives the assumed percentage of subscribers for each medical profile. It also summarizes the assumed variation intervals for the arrival rates of each primitive and medical profile. These values are used only as examples.

As a network-specific parameter, the total number of BMUs that can be connected simultaneously to the entire cellular network is fixed to $N_c = 80$, as an expected number of patients for a metropolitan-sized cellular network. This number represents 80% of elderly persons over the age of 65 with at least one chronic disease who have access to a 2G 28 km^2 metropolitan area network [3].

A. Bandwidth Estimation

Equation (2) can be used to estimate an upper bound for the mean aggregated bandwidth upstream and downstream. The worst case in terms of bandwidth used is when the BMU connects before each primitive and disconnects immediately after it. This means that the arrival rate λ_{cj} for each medical profile j is the maximum among the other arrival rates for the profile, and that the average period the BMU remains connected is the smallest value that allows for the execution of both p_c and any of p_a, p_s, or p_i. The maximum taken over all the bandwidth estimations is obtained by varying the arrival rates within their respective variation intervals is obtained by varying the arrival rates within their respective variation intervals (Table I).

Using the measures previously discussed as input (Table I), the model gives an upper bound for the average aggregated bandwidth of approximately 1.2 Kbits/sec in both directions and for all the technologies assessed. Fig. 3 summarizes the upper-bound estimations obtained for UDP, TCP, and SSL, over a 3G network.

Not surprisingly, the estimations obtained for UDP are somewhat lower than the ones for TCP. This is especially true for the downstream traffic.

TABLE I: ARRIVAL RATE VARIATION AND MEDICAL PROFILE RATIOS

Medical profile	Ratio	Alarms/day λ_a	Readings/day λ_s	Images/day λ_i
Diabetic	20%	[1, 3]	[1, 3]	[0, 1]
Cardiac	40%	[1, 6]	[1, 50]	N/A
Elderly	40%	[1, 10]	[1, 50]	[0, 12]

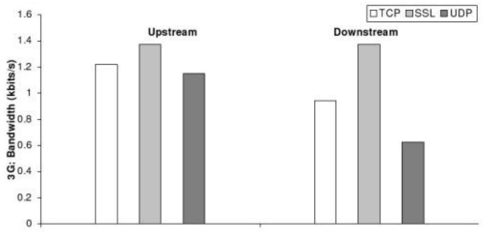

Fig. 3. Estimated average aggregated bandwidth over 3G (TCP/SSL/UDP).

The volume of the non-overhead traffic upstream (sensor readings and images from the BMU) is greater than the non overhead traffic downstream (alarms from the BMC), and this results in a higher overhead downstream with TCP.

For SSL, the bandwidth estimations are higher than for TCP. The difference between upstream and downstream SSL estimations is small, as the SSL overhead dominates.

So far, it has been assumed that the BMU connects just long enough to execute any of the primitives, that is, for a period of *max* $\{d_a, d_s, d_i\} + d_c$.

To assess the impact of the connection interval on the estimations of the upper bound of the average aggregated bandwidth, we compute the upstream and downstream estimations assuming different equal values Δ=[10,30,60,90] in minutes for the average connection intervals for all the medical profiles:

$$\begin{cases} B^{up} = \max\left\{B^{up}\left(\lambda_{aj}, \lambda_{sj}, \lambda_{ij}, \lambda_{cj}, \Delta\right)\right\} \\ B^{down} = \max\left\{B^{down}\left(\lambda_{aj}, \lambda_{sj}, \lambda_{ij}, \lambda_{cj}, \Delta\right)\right\} \\ \lambda_{cj} = \max\left\{\lambda_{aj}, \lambda_{sj}, \lambda_{ij}\right\}, \\ \forall \lambda_{oj} \in \left[\lambda_{oj}^{\min}, \lambda_{oj}^{\max}\right], \forall o \in \{a, s, i\}, \forall j \in \{1,2,3\}. \end{cases} \quad (4)$$

Fig. 4 gives an upper bound for the case using 3G access. The longer the BMU stays connected, the smaller the average aggregated bandwidth required to support the subscribers. A longer connection interval for a BMU potentially results in less bandwidth use for the individual user (as it can reduce the traffic generated by the sign in/sign off primitive), which can be attractive to users with billing plans that charge per data volume.

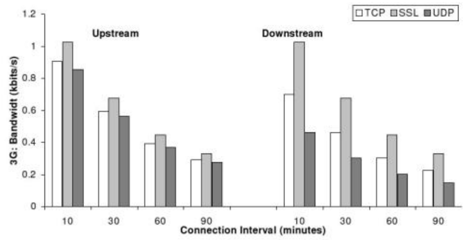

Fig . 4. Estimated average aggregated bandwidth based on average connection interval Δ, over 3G.

B. Estimation of the Total Number of Subscribers

The model can also be used to estimate a bound for the number of clients the operator can accept as biomonitoring service subscribers, by using the conditions given (4), and a minimum over all estimations of the total number of subscribers computed using (2).

Fig. 5 shows the resulting estimations for 3G, using the UDP, TCP, and SSL transports. (Similar trends were observed for 2.5G.). The estimations consistently decrease in magnitude as the connectivity interval increases. This means that longer average connection intervals, while yielding both a smaller bandwidth usage for the individual user and a smaller aggregated bandwidth over the entire network, also result in a smaller bound to how many service subscribers can be accepted. Fig. 5 also shows that, for more connection intervals, the estimated bounds for different protocols are similar. This holds only for connection intervals longer than approximately 10 seconds.

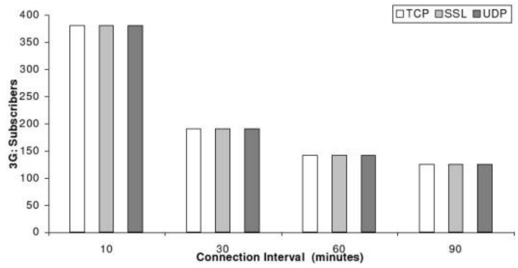

Fig. 5. Number of subscribers as a function of the average connection interval Δ, over 3G.

V. RELATED WORK

Research in the area seeks to establish the suitability of the various wireless technologies for telemedicine, and how to properly dimension and optimize the network use, e.g. the use of satellite links for controlling a telemedicine robot [4], the performance of transferring X-ray images and video in emergency orthopaedics cases over GSM and GPRS [5].

Tuning up the performance can help in evaluating and improving the user experience, and help convince potential users to adopt the application, but this can only be done once the application has been built and deployed [6], [7]. Alternatively, implementation and testing [8] can be limited to the typical data transfers, but this does not take into account the aggregated behaviour of multiple users.

One methodology for devising a traffic model is to mine data traffic traces to determine patterns and properties. In [9], Crovella and Bestavros find the self-similarity of Web traffic, and Paxson and Floyd [10] observe that connection arrivals for user-made connections differ from those for machine-made connections. This methodology assumes that representative packet-level traces are available, either from an already mass deployed application or from a realistic large-scale trial deployment, which is usually not available for medical applications. Another option is to analyze the specifics of the application domain to identify the patterns and properties of the actions the application takes, which are then used to build a traffic model.

In [11], Camorlinga et al. analyze the work flow standardized by medical protocols for working with X-rays, and the characteristics of the required data exchanges, while Clarke et al. [12] identify typical data transfers and analyze the performance of each individually and when they are combined.

Devising traffic models starting with the specifications of the application domain, the approach taken in this paper seems better suited to medical applications. The paper also takes into account the way the prototype was implemented, so the resulting traffic model accounts for the implementation-specific caveats and also the particular features of the application domain.

VI. CONCLUSION

The traffic model described in this paper estimates bounds for the average aggregated bandwidth generated upstream and downstream, and for the average number of users that can be accepted as subscribers. An upper bound of about 1.2 kbps is obtained for the average aggregated bandwidth over the entire network. A trade-off needs to be made between configurations that result in less aggregated bandwidth used (lower costs) and the possibility of accepting more subscribers on the service (higher revenues). The choice depends on the business model of the service operator and on the nature of the target medical profiles to support.

As more becomes known about how remote biomonitoring applications will be built and used, we plan to enrich the traffic model. In particular, a more complex characterization of the medical profile might be needed. With the collaboration of medical specialists, we plan to go beyond using illustrative medical profiles, by finely tuning the model and devising specific profile and parameter values within the bounds of a concrete medical project and a concrete legal framework.

REFERENCES

[1] O. A. Dragoi and N. Kara, "Ubiquitous and personalized contextual biomonitoring," *International Journal of Healthcare Technology and Management (IJHTM) – Special Issue on Pervasive Healthcare*, vol. 10, no. ½, pp. 27-48, 2009.

[2] N. Kara and O. A. Dragoi, "Reasoning with context data in Telehealth applications," *WIMOB 2007 – Special Session Health Care and Ubiquitous Computing*, New York, pp. 69-79, Oct. 2007.

[3] J. L. Gerberding, "Healthy aging preventing disease and improving quality of life among older Americans," *Technical Report*. Department of Health and Human Services Centers for Disease Control and Prevention Coordinating for Health Promotion, 2006.

[4] S. Garawi, F. Courreges, R. Istepanian, H. Zisimopoulos, and P. Gosset, "Performance analysis of a compact robotic tele-echography e-health system over terrestrial and mobile communication links," in *Proceedings of the 5th IEEE International Conference on 3G Mobile Communication Technologies*, 2004, pp. 118–122.

[5] S. Voskarides, C. Pattichis, R. Istepanian, C. Michaelides, and C. Schizas, "Practical evaluation of gprs use in a telemedicine system in cyprus," in *Information Technology Applications in Biomedicine, 2003. 4th International IEEE EMBS Special Topic Conference on*, April 2003, pp. 39–42.

[6] Y. Chu and A. Ganz, "A mobile teletrauma system using 3g networks," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, no. 4, pp. 456–462, December 2004.

[7] C. Lee and C.-Y. Shih, "A streaming management CSCW consultation system for telemedicine," in *3rd International Conference on Information Technology*, June 2005, pp. 303–306.

[8] I. Widya, A. van Halteren, V. Jones, R. Bults, D. Konstantas, P. Vierhout, and J. Peuscher, "Telematic requirements for a mobile and wireless healthcare system derived from enterprise models," in *Telecommunications, 2003. ConTEL 2003. Proceedings of the 7th International Conference on*, vol. 2, pp. 527–534, June 2003.

[9] M. E. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: evidence and possible causes," *IEEE/ACM Trans*. Netw., vol. 5, no. 6, pp. 835–846, 1997.

[10] V. Paxson and S. Floyd, "Wide-area traffic: the failure of poisson modeling," in *SIGCOMM '94: Proceedings of the Conference on Communications Architectures, Protocols and Applications*, New York, NY, 1994, pp. 257–268.

[11] S. Camorlinga and B. Schofield, "Modeling of workflow-engaged networks on radiology transfers across a metro network," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 10, no. 2, pp. 275–281, April 2006.

[12] M. Clarke, A. Fragos, R. Jones, and D. Lioupis, "Optimum delivery of telemedicine over low bandwidth satellite links," in *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, vol. 4, October 2001, pp. 3606–3609.

O. Andrei Dragoi received the PhD degree in software engineering from University of Waterloo, Canada. He worked as a researcher in the International Telecommunication Institute in Montreal, Canada. Currently, he is principal software engineer at Nuance communications in Montreal, Canada. His research interests are in mobile computing and networking, and multimedia communications.

Nadjia Kara received the PhD degree in electrical engineering from Ecole Polytechnique of Montreal. She worked as a researcher and system architect in the industry for more than 10 years. She is an associate professor at ETS, University of Quebec, Canada. She is also adjunct professor at University of Concordia, Canada. Her research interests are in Service and network engineering, and are mainly focused on service and network architectures such as context-aware services, multimedia communications in virtualized and non-virtualized environments, and Quality of Service architecture and management.