Illustrating a Publish-Subscribe Internet Architecture

Nikolaos Fotiou, George C. Polyzos
Mobile Multimedia Laboratory
Athens University of Economics and Business
Athens, 113 62 Greece
+30 210 8203 693
{fotiou, polyzos}@aueb.gr

Dirk Trossen
BT Research
Adastral Park, Martlesham
Ipswich, Suffolk, IP5 3RE UK
+44 7918 711 695
dirk.trossen@bt.com

Abstract—**The PSIRP (Publish-Subscribe Internet Routing Paradigm) project is an EU funded project aiming at developing and evaluating a clean slate architecture for the future Internet. PSIRP's ambition is to provide a new form of internetworking which will offer the desired functionality, flexibility and performance, but will also support availability, security, and mobility, as well as opportunities for innovative applications and new market opportunities. This paper illustrates PSIRP's high level architecture, revealing its principles, core components and basic operations through example usage scenarios. While the focus of this paper is specifically on the operations within the PSIRP architecture, the revelation of the workings through our use cases can also be considered being useful to similar work on publish-subscribe architectures.**

Index Terms—**Future Internet, clean slate, networking usage scenarios.**

1. Introduction

The current Internet architecture has remained relatively unchanged since its inception. The Internet was initially designed in a way that resembles to a telephone network, where uniquely addressed endpoints trust each other and exchange data through an internetworking infrastructure. However, nowadays this design does not cope with current networking trends, neither with applications' needs. Viruses and worms have led to a state where two arbitrary nodes do not trust each other anymore. End-to-end communication does not seem to be the prevailing paradigm as data requests are likely to be served by an intermediate element–such as a content delivery network, or a proxy server–and popular applications–e.g., p2p file sharing applications– focus on the information itself rather on endpoints' location. Furthermore, an imbalance of powers exists in the current Internet, where the network is designed to forward data from senders to receivers, whether the receivers want to receive the data or not, leading to problems such as denial of service attacks and spam email. Various mechanisms such as network address translation (NAT) and spam filters have been deployed in order to restore this imbalance; however they do not solve the problem completely. Moreover challenges related to security, mobility, scalability, quality of service, and economics have increased the need for a clean slate approach to a new Internet architecture [1]. A candidate communication paradigm for the future Internet's architecture is publish/subscribe (pub/sub).

Pub/sub is an information-centric paradigm that shifts the power away from the data sender, i.e., data consumers express their interest on specific pieces of information, which are forwarded to them by the network when they become available. As a result information is propagated only to those nodes, which really want to access it. In general, pub/sub architectures consist of three core components [2]: publishers, subscribers and routing nodes–also called brokers.

Publishers are information providers that feed information elements into the pub/sub network by virtue of publications. Subscribers are consumers that explicitly express their interest in a specific published element by issuing subscription messages. These messages contain the criteria that a publication should fulfill in order to be forwarded to a subscriber. Brokers are elements that forward data between subscribers and publishers. A node where the matching of a publisher's content with a subscriber's interest takes place is refered to as the rendezvous point (RP). These elements initiate routing, forwarding, and distribution decisions, eventually leading to the delivery of the content from publishers to subscribers. Publication and subscription operations are decoupled in time and space as they do not have to be synchronized, they do not block each other and publishers do not have to be aware of the subscribers–and vice versa. Moreover multicast is the preferred delivery method.

The Publish/Subscribe Internet Routing Paradigm (PSIRP) project [3] is an FP7 EU funded research effort aiming at creating a clean-slate architecture for the future Internet based on the pub/sub communications paradigm, taking nothing–not even IP–for granted. In PSIRP, performance and efficiency will be achieved with the use of innovative multicasting and caching techniques, and security will be a native component of the architecture. The PSIRP project envisions a robust and scalable architecture where mobility will be the norm and data morphing will allow users to access information anywhere through any medium. The PSIRP suggested architecture is information oriented and it abides by the Trust-to-Trust (T2T) principle [4], i.e., its functions take place at trustworthy points.

The remainder of this paper is organized as follows. Section 2 gives an overview of the PSIRP architecture presenting its core elements and its basic operations. In Section 3 PSIRP usage scenarios are given, illustrating at some level of detail the supported functionality and operations. Finally, Section 4 presents applications considerations for the PSIRP architecture and Section 5 presents our current conclusions, as well as future work.

2. The PSIRP Architecture

Information is the core element in the PSIRP architecture; in PSIRP information is "everything" and "everything" is information [5]. Information is organized in a hierarchical way, ranging from small "meaningless" data chunks, to large documents, or video files. Information becomes available through publications. Each publication is identified by a unique identifier, the rendezvous identifier (RId). Publications–and, therefore, information–are organized in networks, called scopes. Scopes can be physical networks, e.g., a university "campus" network, or logical networks, e.g., a social network. Scopes are also hierarchically organized and each scope is identified by an identifier; the scope identifier (SId). SIds and RIds are flat, location independent labels similar to the labels that are used in ROFL [6]. Scope identifiers can be private or public in correspondence to the (current) Internet's private and public IP addresses. RIds have to be unique within their scope. Scopes, apart from information locating mechanisms can also be used as access control mechanisms. Scope policies define who can access the publications that are published in a certain scope, e.g., only the professors and students of a university can subscribe to publications published to the university's scope. In a similar way policies may define who can issue publications within a scope, e.g., only professors can publish information to the university's scope. Scopes are managed by special elements called rendezvous nodes.

Whenever publishers wish to issue a new publication they have to use two identifiers: RId and SId. A publication's RId can be derived by an application specific function, e.g., a hash function over the data to be published. A publication's SId should denote to which extent the publisher wishes the publication to be available. Prior to publishing an information element, publishers have to locate the rendezvous node that is responsible for managing the desired scope. This rendezvous node will be the rendezvous point for the publication's SId. What is actually published to the rendezvous point is the publication's metadata, which contains information specific to the actual publication; this can be for instance the author of the publication, its size and perhaps a small description of it. It can always be the case that a publication has to be fragmented in smaller chunks. Each chunk is considered as a separated publication, having its own publication identifier. Chunk's RIds can be algorithmically related to the original publication's Rid; this way a potential subscriber can easily find the RIds of all chunks that compose a bigger publication. The metadata of a publication may contain information about its chunks as well.

In order for a subscriber to access a publication she must be aware of its RId and SId. She expresses her interest about a specific publication by issuing a subscription message towards the publication's rendezvous point (RP), identified by the SId. The RP is responsible for matching publications with subscriptions and for initializing the forwarding of a publication from the publisher towards the subscriber. In order for a publication to be transferred to a subscriber, a path toward her has to be created. The topology formation function of the PSIRP architecture is responsible for creating the delivery path, in a way similar to the way MPLS creates forwarding paths. The forwarding is implemented through assigning a stack of path specific identifiers to the publication, named forwarding identifier (FId). These FIds are used by intermediary forwarding nodes to forward the publication to the desired destination. Publications may be cached along the path. In case that more than one subscriber subscribe for a specific RId, a multicast tree is created in order to deliver the publication. Publications may have different versions, and whenever a new version of a publication is created, all subscribers are being informed.

Publish/subscribe is used at all levels of operation of PSIRP, ranging from a node's internal functions to publication transfer throughout the (PSIRP) Internet.

3. PSIRP Usage Scenarios

While we outlined the general PSIRP architecture in the previous section, we now illustrate the workings of the architecture in some typical PSIRP usage scenarios. With this, we attempt to further clarify the relations and operations within PSIRP but also shed some light onto the development of typical PSIRP applications.

A. Scenario Setup

A user, **USERA**, works in university **UNIA**, in the department **DEPTA**. USERA has prepared a presentation, and he wants to make it available to his colleagues in his department. USERA has three colleagues; **USERB**, **USERC** and **USERD**. USERB and USERC want to access USERA's presentation through UNIA local network and USERD wants to access it from his home network, **NETA**. Fig. 1 depicts this scenario setup.

B. Publication

The UNIA network consists of four rendezvous nodes **RN001, RN002, RN003 and RN004**. DEPTA has its own scope with SId **00A1**. Scope 00A1 is managed by RN003 and RN004, i.e. RN003 and RN004 have subscribed to all publications on scope 00A1, and their subscription message has been broadcasted to UNIA network. As a result every potential publication to scope 00A1 will be forwarded to either RN003 or RN004. Scope's 00A1 identifier is private, so it can not be directly accessed from a network outside UNIA network.

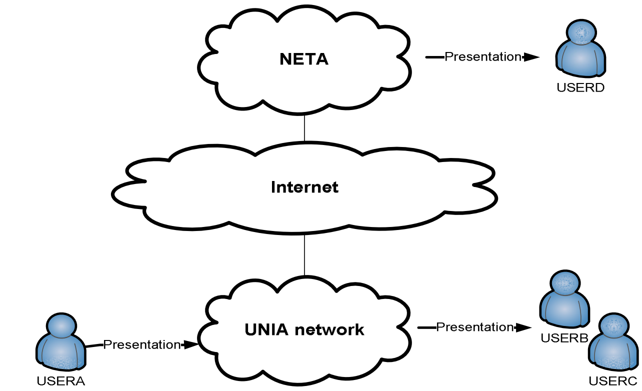

Fig. 1. PSIRP's usage scenario setup

Moreover scope 00A1 implements the following access policy: Only members of DEPTA are allowed to publish information and to subscribe for publications in that specific scope. USERA wants his presentation to be accessed only by members of DEPTA, so he decides to publish it in scope 00A1. A publication is created and an application specific function generates an Rid identification number for this publication, which is **AA12**. USERA's application running on **SERVER01** issues a publish message with SId 00A1 and RId AA12, which is forwarded to RN003, as this rendezvous node has subscribed to all publications in scope 00A1 (through subscribing to the SId itself). The publish message contains, along with the identifiers, USERA's presentation metadata. The publication operation ends with RN003 updating its internal RIds database by adding AA12 and becoming the rendezvous point for this RId. The publication operation is depicted in Fig. 2.

C. *Subscription from the Internal Network*

USERB learns about USERA's publication through some from of discovery, e.g., an internal listing of presentation or some new form of search engine. In order for USERB to subscribe for USERA's publication, he has to authenticate himself to scope 00A1. He achieves that by logging in to SERVER03 with credentials that allow him to subscribe for publications in scope 00A1. USERB's application running on SERVER03 issues a subscribe message towards scope 00A1, denoting that he is interested in the publication with RId AA12. RN003 receives USERB's subscription message and it initiates the creation of a forwarding path between publication's AA12 location, i.e., SERVER01, and SERVER03.

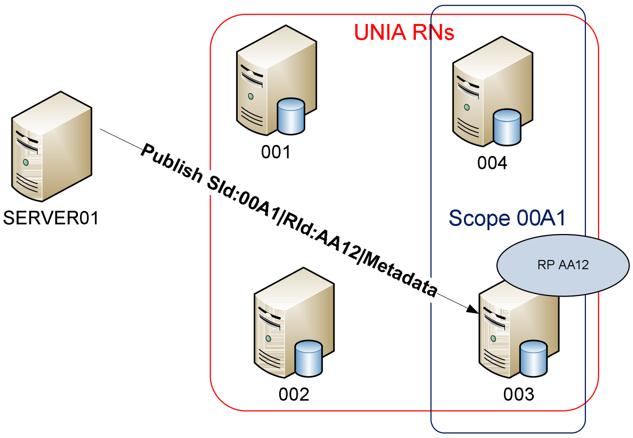

Fig. 2. PSIRP's publication operation

A stack of forwarding identifiers is determined for the links along the forwarding path. Each forwarding node along the path maintains a forwarding table which contains the incoming forwarding identifier, and its correspondent outgoing interface and identifier. RN003 informs SERVER01 that it should start sending publication data using the FId of the first link of the path. The forwarding node that will receive publication data will modify its FId and it will forward it to the appropriate interface. This way the publication will reach USERB. Fig. 3 presents a forwarding example.

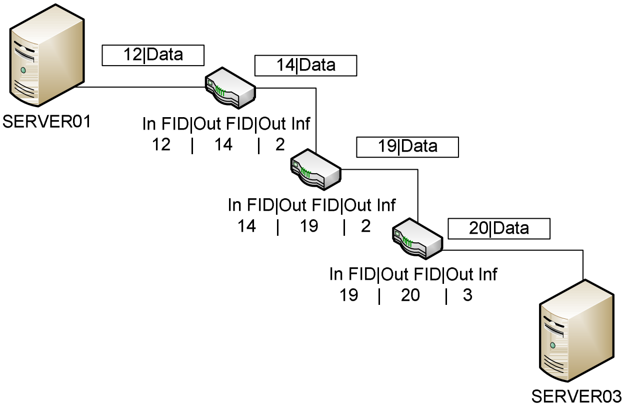

Fig. 3 PSIRP's forwarding.

The box over the lines represents a data packet with its FId. The table below each forwarding node shows its forwarding table. The forwarding table contains the incoming Fid, the outgoing FId and the outgoing interface. SERVER01 sends a packet with FId 12. The first forwarding node checks its forwarding table and it finds out that it has to forward this packet to interface 2 with FId 14. In a similar way the second forwarding node forwards the packet to its interface 2 with FId 19 and the final interface forwards the packet to its third interface by changing the FId to 20

In scope 00A1, a publication with Rid F20E exists, that contains the RIds of every presentation available in this scope. This publication is provided by a presentation announcement service. USERC is interested in every presentation in scope 00A1 so she has subscribed for publication F20E. USERC uses SERVER03 with credentials that allow her to subscribe for publications in scope 00A1. When USERA publishes his presentation the presentation announcement service creates a new version of F20E, which contains USERA's publication RId. USERC receives the new version of F20E and she becomes aware of USERA's publication so she subscribes for it. At this point a procedure similar to the one described for USERB's subscription is followed in order for publication to be forwarded to SERVER03. Fig. 4. depicts PSIRP's subscription operation.

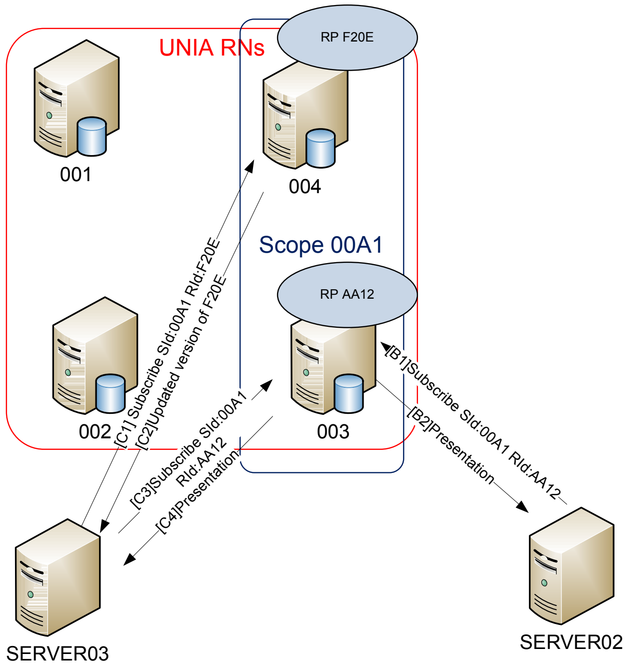
UNIA RNs
RP F20E
001
004
Scope 00A1
RP AA12
002
003
[C1] Subscribe SId:00A1 RId:F20E
[C2]Updated version of F20E
[C3]Subscribe SId:00A1 RId:AA12
[C4]Presentation
[B1]Subscribe SId:00A1 RId:AA12
[B2]Presentation
SERVER03
SERVER02

Fig. 4 PSIRP's subscription operation. The numbers inside brackets denote the sequence of each message, e.g., [C1] is the first message that USERC sends.

D. *Subscription from an External Network*

USERD wants to access USERA's presentation form his home network NETA. However scope 00A1 is private and it can not be directly accessed by an external network. UNIA has a scope with SId 32BE which is managed by RN001. This SId is public and it can be globally accessed. Moreover, 32BE is the parent of 00A1 in the scope hierarchy. USERD has to generate a subscription message that initially has to be forwarded to scope 32BE and then to scope 00A1. In order to achieve this he uses the notion of algorithmic identifiers, more precisely he encodes 00A1 using an algorithm with 32BE being its root. All rendezvous nodes are able to determine the root used in order to generate the algorithmic identifier but only RN003 –which is the RN that manages 32BE– can decode the algorithmic identifier and find out which the destination scope is. Moreover, USERD has to authenticate himself against scope 00A1 so to be able to subscribe for publications in that scope. A subscription message is created by USERD which contains USERA's publication RId, the algorithmic encoded SId, as well as USERD's account information encrypted with scope's 00A1 public key.

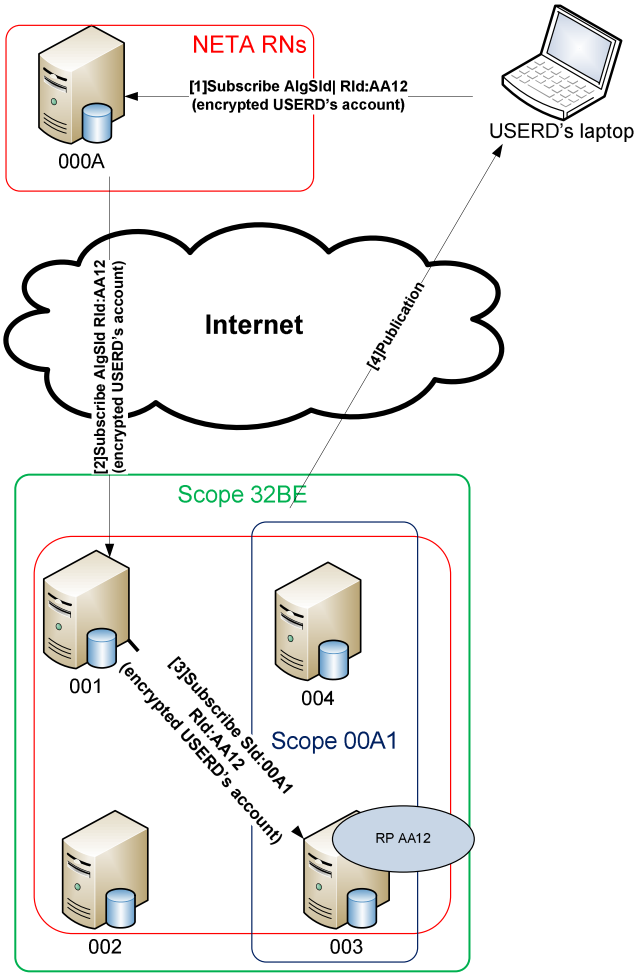
NETA RNs
[1]Subscribe AlgSId| RId:AA12
(encrypted USERD's account)
000A
USERD's laptop
Internet
[2]Subscribe AlgSId RId:AA12
(encrypted USERD's account)
[4]Publication
Scope 32BE
001
004
Scope 00A1
[3]Subscribe SId:00A1
RId:AA12
(encrypted USERD's account)
RP AA12
002
003

Fig. 5. PSIRP's subscription operation from an external network. The numbers inside brackets denote the sequence of each message, e.g.,[1] is the first message that USERD sends.

The subscription message is forwarded to NETA's default RN which is **000A**. 000A is able to determine that the root of the algorithmic identifier is 32BE so it forwards the subscription message to RN001. RN001 is able to decode the algorithmic identifier and finds out that the destination scope is 00A1 so it forwards the subscription message towards this scope. RN003 which is the rendezvous point for this specific RId finally receives the subscription message. However, instead of creating a forwarding path towards USERD it initially checks for the validity of USERD's credentials, send encrypted along with the subscription message. If USERD credentials are valid, a forwarding path from SERVER03 to USERD's computer is created and the desired publication is forwarded using this path. The overall operation is depicted in Fig.5.

4. An Application Developer's View

PSIRP's use cases shed also some light on applications perspectives on this architecture. It seems clear that the identifiers used in PSIRP are not very application oriented and, although their structure is very useful for the overall system effectiveness, it may pose a burden on application developers, as e.g., they should assure that the –publication–identifiers their applications generate are unique within a scope. However, even at these early steps of PSIRP, an in-kernel module that generates publications identifiers is provided. Yet, what is not still provided by the current prototype is algorithmic identifiers. Algorithmic identifiers are going to be a crucial part of PSIRP's architecture as they will denote relationships between various pieces of information. Application developers, although they will be relieved from creating applications that will generate algorithmically related identifiers, they should consider and get familiar with them, since applications should handle them.

Scopes in PSIRP are used for describing the dissemination of information, and not the structure underneath, therefore they are a more general notion than their first level approximation in the current Internet, which are networks. Scopes are managed by lower level mechanisms and their operation is completely transparent to applications. Even more transparent will be the operation of the forwarding function; forwarding identifiers will be completely hidden from applications.

Another fundamental concept that may trouble developers is the use of trust in every transaction. Trust is a core principle of the PSIRP architecture as all functions should take place at trusted points (in the internetwork). PSIRP will provide to applications all the means for determining if another node in the network behaves in a trustful way, however applications are also expected to implement their own–application specific–trust evaluation algorithms since at the application layer there are various trust expectations, which can not be predicted by the lower layers of the architecture.

5. Conclusions and Future Work

PSIRP is an ongoing research effort. During this phase of the development, part of PSIRP's functionality has been implemented (on top of the FreeBSD operating system) and it is being tested (at this point mostly in the local area). Work is being done in order to implement all PSIRP functionality defined in the FreeBSD kernel as well as by adding a transport layer to the current implementation. Further research is taking place in order to enhance security features in the PSIRP architecture utilizing protocols such as PLA [7] and TESLA [8].

Moreover various internetworking routing protocols are being tested, inspired by hierarchical DHTs–such as Canon [9]. Further work explores PSIRP's operation in mobile environments by evaluating overlay multicast protocols that behave in a way similar to PSIRP's [10]. A procedure for attachment to publish/subscribe networks is also being investigated [11]. Future work includes defining an inter-domain topology formation protocol, exploring trust related issues in PSIRP networks, implementing a fast forwarding mechanism, porting PSIRP functionality to other operating systems and incorporating existing applications to the PSIRP environment.

In this paper we presented only a high level overview of PSIRP. With it, it becomes clear that PSIRP completely changes the way a future Internet could operate. Its information centric nature, derived from using a publish/subscribe communication paradigm, completely differentiates it from the current Internet architecture, offering the potential for developing more intriguing applications that will enable end users to harvest the full potential of such future Internet. Moreover, its built-in security features, the usage of flat location independent identifiers and its multicast nature, support the claims for a secure, scalable network which will facilitate mobility and will provide high availability of resources.

The use cases presented in this document offered a clearer view of the PSIRP architecture by analyzing step by step the actions needed in order to complete simple operations, by shading light on the usage of various identifiers, and by presenting in a more comprehensive way the notion of scopes. We believe that the contributions of this work and the use case presentation in particular are two-fold: first, the PSIRP structure and operation was presented in a simple and understandable way and, second, a point of reference in provided for PSIRP developers, in order to further evolve the PSIRP architecture.

Acknowledgments

The work reported in this paper was supported by the FP7 ICT project PSIRP, under contract ICT-2007-216173.

References

[1] A. Feldmann, "Internet Clean-Slate Design: What and Why?" *Computer Communication Review*, ACM SIGCOMM, vol. 37, no. 3, pp. 59-64, 2007.

[2] P.T. Eugster, P. Felber, R. Guerraoui, and A.M. Kermarrec, "The Many Faces of Publish/Subscribe," *ACM Computing Surveys*, vol. 23, no. 2, pp 114-131, 2003.

[3] PSIRP homepage, available at: http://wiki.psirp.org/ (accessed on 15 March 2009).

[4] D. Clark and M. Blumenthal, "Rethinking the Design of the Internet: End to End Arguments vs. the Brave New World," *ACM Transactions on Internet Technology*, vol. 1, no. 1, pp. 170-109, 2001.

[5] "Deliverable 2.2: Conceptual Architecture of PSIRP Including Subcomponent Descriptions," available at http://www.psirp.org/ (accessed on 15 March 2009).

[6] M. Caesar, T. Condie, J. Kannan, K. Lakshminarayanan, I. Stoica, and S. Shenker, "ROFL: Routing on Flat Labels," *Computer Communication Review*, ACM SIGCOMM, vol. 36, no. 4, pp. 363-374, 2006.

[7] C. Candolin, "Securing Mlitary Decision Making in a Network-Centric Environment," doctoral dissertation, Helsinki University of Technology, 2005.

[8] A. Perrig, R. Canetti, J. D. Tygar and D. Song, "The TESLA Broadcast Authentication Protocol," *CryptoBytes*, vol. 5, no. 2, pp. 2-13, 2002.

[9] P. Ganesan, K. Gummadi, and H. Garcia-Molina, “Canon in G Major: Designing DHTs with Hierarchical Structure,” Proc. 24th International Conf. on Distributed Computing Systems (ICDCS 2004), Tokyo, Japan, March 2004.

[10] K. Katsaros, N. Fotiou, G.C. Polyzos, and G. Xylomenos, “Overlay Multicast Assisted Mobility for Future Publish Subscribe Networks,” Proc. ICT Mobile Summit, June 2009 (in press).

[11] J. Kjällman, “Attachment to a Native Publish/Subscribe Network,” ICC Workshop on the Network of the Future, June 2009 (in press).