

MOBILITY AWARE PATHS: THE IDENTITY CONNECTION

Alfredo Matos
Instituto de Telecomunicações
Universidade de Aveiro
Aveiro, Portugal

Rui L. Aguiar
Instituto de Telecomunicações
Universidade de Aveiro
Aveiro, Portugal

ABSTRACT

This paper discusses the challenges that arise from today's mobility management architectures, how they are restricted to specific identifiers and routes, and how they have disjoint control layers. It describes a novel approach to tackling the shortcomings of current network models, describing the challenges and solutions revolving around Identifiers and Identity based Mobility, associated with the more generic Path concept. It describes how to position communication and mobility, Identity centric, in terms of paths, and in the process solving the challenges raised by the unicast/multicast dichotomy, effectively bridging the gap between peer-to-peer overlays and point-to-point communications.

The paper also discusses the need for cross-layer resolution and routing mechanisms that enable simple and consistent access to an Identity Layer, which is in fact the new control layer, leveraged by policy driven architectures enabling the necessary granularity for future network evolution.

I. INTRODUCTION

Current networks are gradually fulfilling more complex functions. As we stack layer on top of layer, protocol on top of protocol, we build an increasingly complicated network of paths, signals, triggers or controls planes. Every step we take in integrating specific services into today's network adds complexity that components and protocols must leverage in order to cope with these new requirements.

One basic service, which is increasingly being taken for granted at different levels, is mobility. As we require that different layers provide this service we add yet another protocol, with similar features that solves a particular problem at that layer. There are a panoply of examples, going from the link layer to the application layer and above, that tackle and solve each mobility requirement in its own autistic way. We witness IEEE developing a batch of protocols in the 802.11 [1] family, such as 802.11k [8] or 802.11r [9], which are able to provide fast access point transitions and handover target information. This resembles what is being achieved with 802.21 on a heterogeneous level, and is done over IP with IETF protocols such as the Candidate Access Router Discovery (CARD) [15] protocol and Fast Mobile IPv6 [4] – which then resembles SIP [14] redirect functions. Nevertheless, each protocol has its own merit, which should not be discarded, but there is clearly a lack of integration across each protocol, coupled with the fact that each of the mentioned mechanisms relies on its own independent layer for control, and interact with each other through, sometimes ill defined, inter-layer triggering.

Currently we can envision concepts that can provide the solution to these complicated challenges. We can put all the

mobility control together, choosing the right abstractions. This does not necessarily break the layered approach to communications networks, and further enhances it, since all interactions are viewed as centered on the data, on the information itself, and not on the specific communications technology supporting that data exchange.

To provide such a driving force in networked environments, the control plane must have access to a full set of information ranging from user oriented policies, attributes and data to link properties, network conditions, and path information among many others. Network control plane should be to the best extent nearly omniscient, able to factor all variables into determining the best network conditions possible, and establishing the best connection path.

In this paper we present the mobility challenges that arise developing a simplified mobility architecture, in Section II. We continue discussing the current trends on mobility management and the unavoidable links to resolution and identity layers in Section III. Later, in Section IV, we pursue this identify trend and further discuss the design of an identity oriented mobility architecture, along with the paradigm shifts and possible features. The discussion is both at the external level, how mobility becomes transparent to the outside service users, and internal, how this is handled at the technology level. We finished by presenting our conclusions in Section V.

II. MOBILITY CHALLENGES

When early mobility discussions appeared a couple of years ago, the key point was how and when to use Mobile IP [2, 3], either version 4 or version 6. Meanwhile, many improvements have appeared and deployment has matured. The mobility landscape has come a long way since that limited scope of applicability that we have initially considered. We have come to understand that mobility occurs on several levels, requiring different abstractions, and all of them using different identifiers. Each of these levels usually proposes its own control space, enabling specific mechanisms and procedures.

A. Mobility Levels

As mobility occurs on different levels, each proposes its own control space, enabling specific protocol mechanisms to operate without much coupling with the neighboring layers. So we can easily identify technology-mobility, network-mobility, transport mobility and application and service mobility. In this paper, we will focus on the first types.

For next generation networks, technology-level mobility will always exist, and will depend on the specific technology that it will be used. Much more interesting is the structuring of mobility that is now usually accepted. Currently mobility is usually classified in two large sets, global and local (e.g. the

MobiSplit [6] architecture). This is commonly defined as macro and micro mobility. While these two types can be decoupled and independent, linking them together onto the same control layer leads to hierarchical mobility management protocols. The Mobisplit environment is shown in Figure 1, where we clearly see the difference between local and global mobility domains. In this structure, technology-level mobility can be considered a local mobility issue.

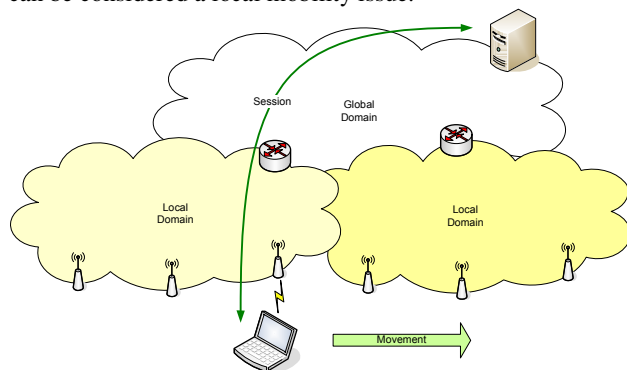


Figure 1: Different mobility levels that occur in complex architectures.

This environment addresses network layer mobility, which in fact means we are moving devices by changing their network attachment points.

If we consider transport layer mobility, the endpoint at the transport level may change, which can be handled by protocols like SCTP [10], even without any change at the network level. This is a situation which will become increasingly common as multihoming becomes widespread.

Another mobility level on top of the previous that must be considered is session mobility, where we consider a session to be part of an application, such as a video stream. This can be accomplished using the SIP [14] protocol.

All of the previous mechanisms present disparities in both purpose and means, each of them considering different methods of performing mobility and differing in mobility subjects. Furthermore, it is very difficult to harness all the different levels of mobile, since they are mostly disconnected from each other and interact through (some) triggers.

While triggers, such as connectivity loss or address renumbering might be realizable for a bottom up approach, the contrary produces no effect, i.e. moving an application level session does not produce any effect in the lower layer, even though there can be obvious and meaningful consequences. Also, this sequential bottom up approach leads to large overhead and delay since each layer most consume the lower layer event, reacting with the appropriate trigger for the above layer.

B. Abstractions

One of the biggest challenges around mobility is that it takes so many different shapes since the objects that are mobile depend on the level or layer we are considering.

In order to harmonize mobility in a truly cross layer architecture, tackling most of the described challenges, it is necessary to abstract the object of mobility, as well as the

functions that provides the control operations, enabling specific protocol level mobility events or triggers.

If we want to achieve the aforementioned uniformity then the real challenge lies in determining the common denominators for mobility.

C. Identifiers

As there is no consistent approach to the mechanisms that enable mobility, the identifiers used in the process also suffer from the same problem. Each layer presents a different namespace (MAC, IP, URI, ...) and consequently a different identifier set with disjoint characteristics and no particular cross layer uniformity. There is no relationship between the identifiers used at different layers and each mobility procedure or protocol has its own meaning and conveys no useful information beyond the namespace where they originate.

Tables, mappings, explicit links and implicit links are among the mechanisms used to link different namespaces, so that the network can cope with identifiers that have only localized meaning. This set of workarounds does not solve the issue of enabling cross layer functions nor paves the ground for more elaborate solutions.

D. Control

While discussing the different mobility levels that can be observed in more complex network architectures, we have implied that there is also a mobility control problem inherent to the multiple disconnected mobility tiers.

Each protocol usually leverages its own control layer, which is governed by different motivators depending on the layer at which mobility is instantiated. The main drive for link layer mobility is Signal to Noise Ratio (SNR) and connectivity availability. In 802.11 we see this in the form of the Received Signal Strength Indication (RSSI). But, as we climb up the protocol stack, we witness different phenomena. At the network layer, the choice is normally based on connectivity alone, while at the application layer it is normally based on user preferences.

But, as networks evolve, we see a constant effort to make all these decisions, at all levels, as informed as possible due to the increasing number of available networks, and user choices. To this end, different protocols, such as 802.11k [8], 802.21 [7] or the Candidate Access Router Discovery (CARD) [15] protocol aim at providing more information than that normally available at the terminal. Furthermore, other platforms are beginning to bring user policies into these decisions, leveraging information such as cost or quality of service (QoS).

So, this is a twofold problem that must be tackled right now. On one hand, a wide range of information must be present for mobility decisions at all levels, to make the best decision possible. But on the other hand we must realize that these are in fact different control spaces that normally exist independently of each other and rely on triggers between them. The challenge that we must face right now is how to make the layers work together, reducing the overhead and the

delay of mobility operations, so that we can obtain the holy grail of mobility, the seamless handover.

If different layers are aware that mobility, regardless of the specific layer, is occurring, then both terminal and network can prepare in advance and parallelize several processes in order to achieve a seamless and secure handover. On the other hand, we aim for these mobility processes to be as much hidden from the final application as possible.

These challenges are best summarized by the need to unify handover decision points, and to distribute the decision results in order to have immediate repercussions on different levels, while keeping this process internal to the communications network.

III. EXISTING TRENDS

There is already much work on these areas, although not with a consistent view. Several protocols, each on its own layer, are attempting a decoupling between layers, so that protocols are not bound to immutable identifiers that cripple protocol operation.

The Host Identity Protocol (HIP) [5] attempts to provide such a separation by introducing a new namespace that yields identifiers that properly map to network identifiers. These identifiers, which are the hash of a public key, in a cryptographic asymmetric key pair, replace locators at the transport layer, leaving IP addresses to its pure function of network locators providing topological routing information. These semantics are capable of solving the proposed challenges, but only if applied vertically across the architecture with the namespace providing identifiers to all the protocol layers.

Based on HIP, the Node ID [17] architecture takes advantage of the features introduced by the indirection level designing a routing scheme that uses the identity layer. It assures that packets traverse the network without the IP address being the prime function for routing, within scoped domains. Also, it uses the concepts of Distributed Hash Table (DHT) based resolution of identifiers, which in fact is basing the routing protocol on identity information, even though it happens within a constrained scope.

Another approach that aims at providing an indirection on existing identifiers and protocol boundaries, but at a higher layer, is the Stream Control Transmission Protocol [10] (SCTP). It provides several advanced transport features and indirection capability that allows address renumbering, along with multi-homing, and therefore supporting mobile environments. But, while SCTP can dynamically change the transport bindings, it requires explicit identification of ports and addresses, diminishing its applicability in general.

The Internet Indirection Infrastructure (I3) [16] provides a set of identifier based triggers. It uses the notion of communication and of peers. It also dissolves the unicast/multicast dichotomy by absorbing the definition that communications are established around endpoints. In fact it proposes that sender registers trigger for a particular information flows, and that receivers register to partake in the announce trigger communication, by advertising that they want to receive packets directed at a particular identifier. This

work has been mostly neglected in its true potential, since it provides tools for a new communication paradigm. But, it has no strong identifier concept, and fails to harness cross layer architectures by devoting the efforts to the trigger based system, and in its essence is still an overlay network which runs atop of the network layer.

On another level, tackling heterogeneity problems that arise from using multi-technology terminals and networks, IEEE 802.21 [7] provides abstract mechanisms for information distribution. It provides the ability of exchanging events, commands and information without actually limiting the applicability scope of such services. Even though 802.21 is a 2.5 Layer protocol i.e. it resides between Link and Network layers, it is the first real effort to model mobility as a set of common operations around objects, even though the original goal is far more humble than what it can actually deliver and what it can solve if the same principles are applied in designing mobility support for other layers.

The Layered Naming architecture [13] is a proposal that trails the path in the right direction. By harnessing existent proposals, it integrates different indirections, which lead to a multi-tier resolution system that relies on DNS for the network layer, but introduces DHT for resolution of flat identifiers which are used as end-point identifier and as session identifiers. The important aspect of this proposal is in fact the introduction of Endpoint Identifiers, based on the host identity protocol, and of session Identifiers, both of which when coupled provide the correct semantics for a truly agile network stack. But, there is unexplored potential of coupling this approach with an identity architecture, which enables managing the introduced complexity and enhancing the communication paradigm, as discussed in the next sections.

On the other hand, Matos et al [11] describe an environment focused on identity driven identifiers that enable a new design of both network and terminals. The usage of specific identifiers derived from an identity namespace enables an architecture built around a common layer. In such an environment, different network protocols and functional boxes can have a consistent view over the user, regardless of the level at which interactions occur. It also provides the initial conditions to build Identity based Mobility solutions, which put identities at the endpoints of sessions instead of standard identifiers. The work in this paper is already a good starting point, but the covered work does not abstract the necessary challenges, functions or objects that provide the full richness of a truly identity oriented design, which needs to be more disruptive and introduce several resolution mechanisms. Sarma et al [12] describe a Virtual Identity Framework that begins to address some of the challenges described earlier. The main focus of this approach is that it derives a set of identifiers that act as references to an omnipresent namespace. The framework itself relies on using the implicit identity pointers to achieve a backwards compliant integration. Also, one of the key points is the integration with an application identity model, which provides attribute resolution. While this is an important issue, this paper lacks a properly design identity oriented and user centric mobility architecture, therefore failing to solve some of the described challenges.

IV. IDENTITY BASED MOBILITY

One way to tackle the proposed challenges is to design an identity based mobility architecture that explores a new path abstraction. This Identity Based Mobility is fueled by four core aspects:

- The definition of a connection path between the communication end-points;
- Path Mobility coupled with identities and identifiers, instead of normal numerical and string addresses;
- A common namespace that aggregates information;
- Meaningful identifiers that reference the identity namespace through resolution.

Following these four cornerstones, detailed bellow, we can effectively devise new network architectures that are user centric and put a common object at the heart of mobility procedures, protocols and identifiers.

A. Paths

The communication path between sender(s) and receiver(s) needs to be properly defined. This is the logical link(s) that connects the two (or more) communication end-points. Mobility management will operate inside the network and the communication stack in such a way that these end-points are kept logically stable, regardless of all the actions needed to be performed at the different networks layers.

This communication path is then the external vision exported by the networking infrastructure to the entities that are using the network. This path is then supplied by a synergistic operation of all the communication layers supporting this path. In particular, several points need to be considered: the namespaces to be used, and the resolution of identifiers; the required mobility management of the paths; the management of the number of end-points, in group communication; and the definition of the way to establish the paths on the network (routing).

B. Namespace and Identifiers and Resolution

Most of the challenges mentioned in the previous sections indicate that one of the main needs of future architectures is a common point of information gathering. It seems obviously necessary that a common namespace must be used to unify information flowing from different layers, especially regarding mobility. We must have a clean namespace that is addressable by all layers and protocols. This namespace needs to contain information, or means to retrieve it, ranging from signal level to the user's favorite color. In fact, all the user and network context and relevant information along with the associated policies for the user, devices and network should be able to coexist in this namespace. Note that it is not required that a single entity keeps all the information because rather than an information holding, the purpose of a namespace is resolution, so that this information can be distributed across the network, eliminating single points of failure and performance problems.

The best candidate that provides the necessary flexibility is an identity-related namespace. Due to its user centricity and information traits, such as being attribute oriented (where attributes are generic bits of information that pertain to the

user or any associated object) an identity namespace is the prime candidate to harbor such generic functionality. Also, identity namespaces and layers, due to the sensitive information held at their level, are usually designed to enhance privacy and security, endowed with strong multi-tier access control.

A pitfall that must be carefully avoided is to assure that the new identity layer inclusion does not break current protocol semantics and mechanism. The design of such a namespace must take into account that it's enhancing existing protocols, creating a large need for legacy support. The alternative being new protocols at all levels, which is the object of clean slate designs – which simplifies the problem.

The best way to establish this blanket characteristic of such an identity namespace is through special constructed identifiers that carry meaning in the identity layer and are easily resolved. When discussing identifiers we are in fact discussing information bytes that bear resolvable meaning. A namespace is only as useful as its associated identifiers.

It does not matter that a namespace harbors a large set of information if it cannot be easily accessed. Therefore it becomes important to provide identifiers that provide an easy method of generation and resolution. As shown in Figure 2, it must be possible, and hopefully easy, to go from a specific layer identifier to the identity namespace, simply by applying a particular function, which can involve resolution mechanisms.

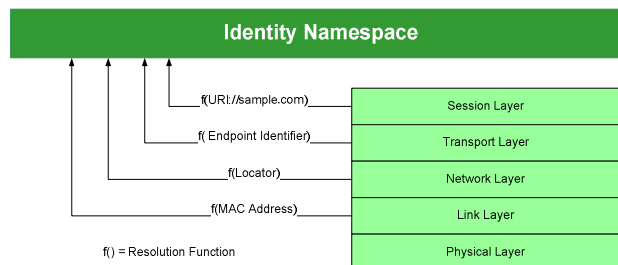


Figure 2: Deriving links to the Identity Namespace.

Identifiers must be easily created because the information associated with them is generally dynamic. Even if the information itself is static, it can be combined into different sets, requiring different access semantics and hence a different identifier. Also, reusing identifiers presents tracking and correlation risks, which mean that for privacy reasons the identifiers themselves should be temporary in nature.

Identifiers which are transient in nature impose the added requirement, on the identity namespace, of providing identifier generation functions. Where the identifiers are generated depends on the provided functions. Either they can be generated by users, or they are built and authorized as part of a functional operation inside a box associated with the namespace. In either case the identity namespace must register or allow registration of identifiers that are to be used by the endpoints.

An implicit assumption is that identifiers with long life span should be supported, such as an email address, divulging the minimum information required, hence protecting the user's

privacy to the best extent possible. Longer identifiers should also be protected by stronger security mechanisms, and should ideally be a first step resolution, implying a double resolution process, such as an email address resolving to a temporary identifier which leads either to the user, e.g. to the user’s mail server, or to the identity namespace for controlled and authenticated attribute resolution.

Even though existing resolution mechanism should be reused, such as DNS, these can only provide initial mappings leading to resolvable flat identifiers that link to the identity namespace. The simplest way to provide such a resolution is to provide a Distributed Hash Table (DHT), similar to those present in NodeID [17] and flat identifier protocols, providing the second step in the resolution process.

Such an example of the previously described steps is shown in Figure 3, where we present a potential resolution process that involves resolving a Uniform Resource Identifier (URI), into a flat identifier (steps 1 and 2) that provides the link to the Identity Namespace. The newly acquired identifier is used to retrieve the location of a specific attribute (steps 3 and 4), which is then resolved into a specific value (5 and 6).

Even though this process is difficult, there are studies that propose completely flat namespaces for routing, and show that it the complexity and delay is feasible.

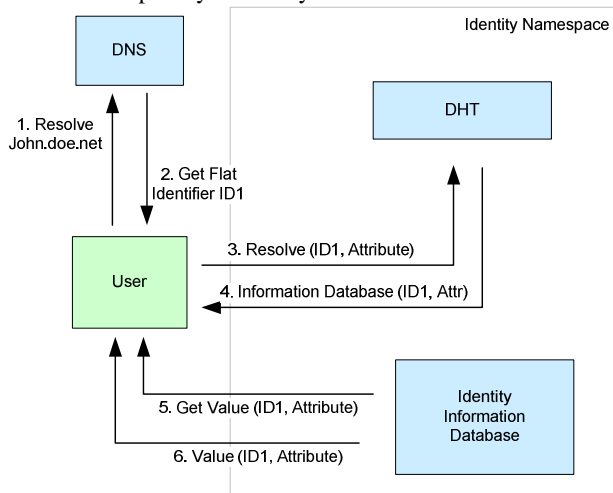


Figure 3: From identifiers to information resolution.

The process described in Figure 3, can be optimized by introducing hierarchy and by varying the identifier semantics. Different identifiers can have different urgency bits or different hierarchy tiers, which could lead to different resolution structures e.g. different tables. It is important to be aware that different identifiers have different properties, since they can be resolved into information that has totally orthogonal purposes.

C. Mobility

To derive the mobility functions of an identity oriented architecture we must draw upon the complex identifier and resolution system mentioned in the previous section.

The most important property to understand about mobility in such an environment is the paradigm which is already applied

today in some protocols to a limited extent, which is the fact that identities are independent of their point of attachment. This is valid at the network, as shown by the Host Identity Protocol, but is also verifiable at upper layers.

Therefore, the paradigm to understand is that every protocol layer should have bindings towards the identity layer, and use the identity namespace as the common drive for both mobility and control.

Relying on ephemeral identifiers such as IP addresses, or location dependent identifiers, such as an address and a communication port, proves to be extremely inefficient when it is necessary to change them. The entire protocol stack should be linked to a durable notion of identity.

While this may seem as the antonym of the previous section, since when we argue that the identifiers used today are highly dependent on network and service conditions along with the usage of temporary identity driven identifiers, this is a misperception. In reality, the benefits of identity, which is proof of ownership, outweigh this issue. If a particular identity changes its endpoint, such as the transport address, if the new endpoint is proven to belong to the same identity, then it can be quickly updated through dynamic mappings, drawing upon the concepts of locator agility as seen in the network layer. The problem then simply becomes that of reachability, solved with the strong resolution mechanisms presented in the previous section.

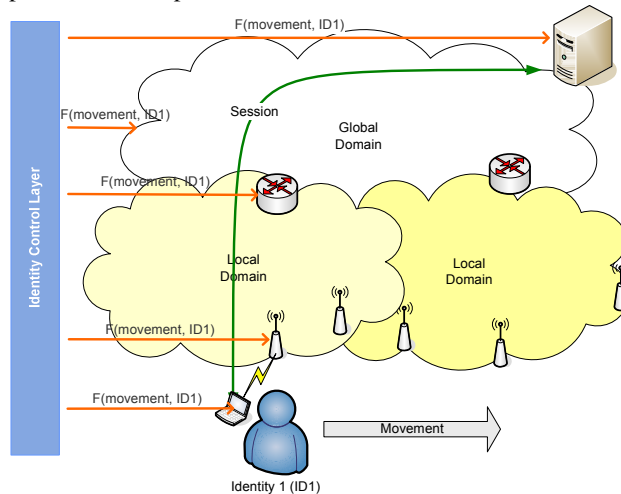


Figure 4: Mobility control by the Identity Layer.

Applying the aforementioned semantics for mobility presents the opportunity of defining common grounds for the different mobility levels. A control layer which is aware of multiple mobility events is now possible, since it is the source for the identifiers that are used. It is also possible to erode the discrepancies that exist across layers, and to provide a unified platform that links mobility triggers. A sample procedure is shown in Figure 4, where we show how an identity layer is able to address several different network functional boxes, with the same semantics ordering or preparing a mobility event relating to a specific identity. It does not matter to which layer the mobility object or protocol belongs, since

they can be notified in the same way. How that notification translates into specific operations is the part that is protocol dependent, and should be analyzed on a case-by-case basis. Mobility events can now be created that cross several layers and even devices and networks, since we are able to describe them as a function of an identity object.

D. Groups

An interesting derivation of the proposed model is that since we are centering communication around identities, we can seamlessly support the notion of group communication, eroding the notion of unicast and multicast communications.

On the delivery path, multicast is a way of transporting packets to several subscribers. But, with an identity model multicast can be treated as any normal communication that has two different identities as endpoints.

As mentioned earlier, through locator agility, multihoming is seamlessly supported within an identity oriented framework, since it is possible to register more than one locator for a particular identity. This property opens the door to the erasure of multicast communication, since a group is also an identity. The fact that an identity points towards more than one user is in fact more of a legal than architectural problem.

We can replace entirely the notion of group by the notion of a multihomed identity, which has many points of attachment.

Since multihoming is already a common target for next generation networks, the presented paradigm successfully breaks the current multicast/unicast dichotomy, leaving only identity communication.

Using specific protocols for multicast distribution becomes only a transport issue that can be tackled at a routing level only, whereas before we had a group management issue.

E. Routing

As discussed in Section IV.C, the proposed paradigm shifts towards mobility centered on identity objects instead of locators and polluted namespaces.

This impacts not only how endpoint mobility is managed, but also how the network organizes around such identifiers. It is perceivable that lower layer identifiers, such as link layer or network layer address, are used in their purest form to provide the functionality they were meant to serve, which is getting a packet or frame from one point to another.

But, by introducing a new layer of coherent cross layer identifiers we enable a simple way of proving policy based packet routing. This can be seen as overlay routing based on identities acting as a point of entry to the overlay management layer. The same concepts used for mobility apply, and allow very flexible mechanisms that can be used to provide features which are complicated to do so in today's networks, such as location privacy. The paradigm used for this is that even though a packet may traverse any number of routes, guided by identity oriented policies, and go through any number of transformation, in the end it can be proved that it was originally sent by a particular identity in a specific context, that of an end to end communication.

V. CONCLUSIONS

This paper presents a set of challenges that need to be met in future networking architectures, in particular from the point of view of multi-level mobility. We further depict a possible set of solutions to these challenges, resorting to a strong identity concept in the network as a binding tool to the definition and management of generic paths, which can provide unicast or multicast communications inherently. The definition of the proper namespace and its associated control mechanisms becomes the largest challenge in these proposed solutions.

REFERENCES

- [1] IEEE Standard 802.11. IEEE standard for information technology - telecommunications and information exchange between systems - local and metropolitan area networks - specific requirements, part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, 1999.
- [2] C. Perkins, "Mobility Support for IPv4." RFC 3220 (Proposed Standard), IETF, January 2002.
- [3] D. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6." RFC 3775 (Proposed Standard), June 2004.
- [4] R. Koodli, "Fast Handovers for Mobile IPv6." RFC 4068 (Proposed Standard), IETF, July, 2004.
- [5] R. Moskowitz, "Host Identity Protocol (HIP) Architecture", RFC 4423, IETF, May, 2006.
- [6] Julien Abeillé, et al, "MobiSplit: a scalable approach to emerging mobility networks", First International Workshop on Mobility in the Evolving Internet Architecture (MobiArch 2006), pp 17-22, Dec 1, 2006.
- [7] IEEE P802.21/D04.00, "Draft IEEE Standard for Local and Metropolitan Area Networks: Media Independent Handover Services", IEEE, March 2007.
- [8] IEEE P802.11k/D13.00, "Draft IEEE Standard for Local and Metropolitan Area Networks: Radio Resource Management", IEEE, March, 2008.
- [9] IEEE P802.11r/D09.00, "Draft IEEE Standard for Local and Metropolitan Area Networks: Fast BSS Transition", IEEE, March, 2008.
- [10] R. Stewart, "Stream Control Transmission Protocol", RFC 4960, IETF, September, 2007.
- [11] A. Matos, S. Sargento, and R. Aguiar, "Embedding identity in mobile environments," in Second ACM/IEEE International Workshop on Mobility in the Evolving Internet Architecture, (Kyoto, Japan), MobiArch2007, October 2007.
- [12] A. Sarma, A. Matos, J. Girão, and R. Aguiar, "Virtual identity framework for telecom infrastructures," in Wireless Personal Communications, (Netherlands), Springer, February 2008. ISSN 0929-6212.
- [13] H. Balakrishnan, et al. "A Layered Naming Architecture for the Internet". *SIGCOMM*, Portland OR, Aug 2004.
- [14] J. Rosenberg, et al, "SIP: Session Initiation Protocol", IETF RFC 3261, June 2002.
- [15] Liebsch, M., Singh, A., Chaskar, H., Funato, D., Shim, E., "Candidate Access Router Discovery (CARD)", IETF RFC 4066, July 2005.
- [16] Ion Stoica, Daniel Adkins, Shelley Zhuang, Scott Shenker, and Sonesh Surana. "Internet indirection infrastructure". In *Proceedings of ACM SIGCOMM*, 2002.
- [17] B Ahlgren, J Arkko, L Eggert, J Rajahalmel, "A Node Identity Internetworking Architecture", INFOCOM 2006, Barcelona, April, 2006.