

Advances in Mobility Management for the NG Internet

Piotr Pacyna

Universidad Carlos III de Madrid, Madrid, Spain

AGH University of Science and Technology, Kraków, Poland

ppacyna@it.uc3m.es

ABSTRACT

Mobility of users and their demand for unrestricted communications creates opportunities as well as challenges for telecom industry. Current situation points at solutions based on IP protocol suite, since they are likely to bring together different environments for uniform support of mobility across the global Internet. In the paper we provide motivation for various approaches to IP-based mobility and identify research challenges. We also make an inventory check of framework architectures and interworking models, and characterize some prospective solutions. Our focus is specifically on solutions which have recently emerged and are under study now. We refer to standardisation processes and directions in standardisation bodies. Our considerations focus specifically on advances in IETF and IEEE.

Key words: IP, mobility, multihoming, security, seamless handover, convergence, Internet

I. INTRODUCTION

Since the introduction of personal mobile communications a high degree of uncertainty about the future of new mobile technologies has been accompanying standardisation, development and deployment efforts.

Today, users of mobile communications services have positive experience resulting from the past use of technologies and services - they are more technology-aware, more interested and more demanding. Nevertheless, a huge uncertainty remains about the actual service offer and the resulting network use patterns. This uncertainty has an influence on research objectives and priorities, resulting in various architectural approaches and solutions.

A wide range of radio access technologies with continuously expanding capabilities are becoming available, calling for high interoperability between systems and services. Hence the ongoing effort to harmonize technologies with end-to-end IP protocol, so that mobility can ubiquitously be provided over various radio access technologies.

Multimode terminals allow simultaneous use of different technologies depending on user preferences and network availability. Hence, support for mobility needs to account for flexibility of service deployment, resource management.

The importance of security for control- and user data in a radio environment is well understood. Securing the users and the network infrastructure with acceptable costs in terms of additional processing and latency needs to be supported by default. However, deployment of security has an impact upon design of almost every function supported by

the network, resulting in difficulties to provide security in a consistent way.

II. GENERAL NETWORK ARCHITECTURE

A general network architecture for the next-generation Internet is presented in Fig. 1. Its characteristic feature is a very high level of heterogeneity. The future Internet architecture will accommodate multiple interworking models, specifically in the access part of the network, where fixed broadband access, single-hop radio access, multi-hop radio access, broadcast technology and moving networks will coexist. Each of these models will require some specific supporting functions to exploit their potentials, while some functions will be common. Mobility is considered to be among the important common functions.

III. MOBILITY MANAGEMENT

Today, the networking model of Internet mandates to refer a network node by means of IP address which is used as a name for the node and an indicator of its point of attachment to the network. A change of point of attachment to the network, e.g. resulting from mobility of the node, typically implies the need to update its IP address in order to reflect the new point of attachment. Consequently, the identity of the node changes making it impossible to sustain running sessions which use addresses and port numbers for identification of session endpoints. It is also not possible to receive new session setup requests from correspondent nodes, since they are not aware of the new locator, nor identifier.

Mobility in the Internet architecture refers to the capability of a node to change point of attachment

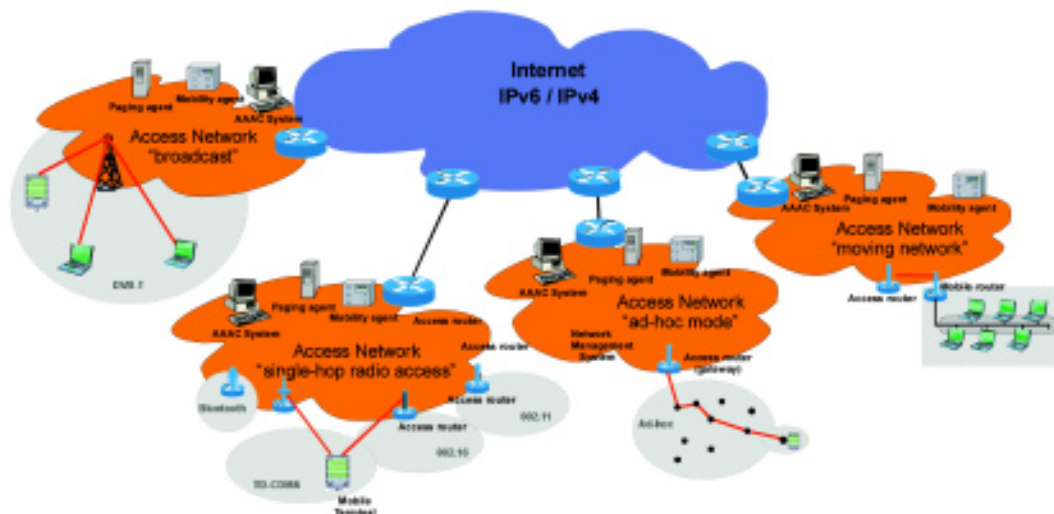


Fig.1 Heterogenous network environment for next-generation internet

Technology domains and administrative domains have the most influence on the interworking solutions and network use schemes. Numerous parameters such as available bandwidth, symmetry or asymmetry of the channels, the level and granularity of control over the resources, reliability, latency, availability of multicast, and availability of cross-layer functions impact design of prospective solutions.

while preserving the communication context for running sessions, and at the same time preserving the identity for subsequent incoming calls. Therefore, mobility is a problem of decoupling identifier from the locator and managing a binding between the two. An identifier, used to uniquely identify an entity, is supposed to be constant and to conform to naming conventions mandated by a name-space. The dynamic locator is a topologi-

cally correct address in the Internet address-space that provides information on current location of the entity or of its legitimate proxy. Different mobility management schemes are basically different methods of managing identifier-to-locator mapping by means of bindings. Binding updates are necessary to reflect essential changes in location of the entity in the network. The 'granularity' level for a change to become significant is typically determined by changes of an IP network prefix, however hierarchical approaches to mobility management or localised mobility schemes can 'scale' that parameter.

Most of the current proposals for mobility management are located at the network layer, in order to make support for mobility transparent for upper layer protocols and applications, like Mobile IP^[1], but some contain proposals positioned between network and transport layers, like Host Identity Protocol^[2], or at the transport or session layer, like extensions to Session Control Transport Protocol^[3] and to Next-Steps In Signalling^[4] protocols.

A dynamic nature of a binding results in considerable threat, that an unauthorised entity can issue a false binding update with the intention to launch a traffic redirection-based DoS attack on the mobile node in the form of flooding attack, disconnection attack or man-in-the-middle attack. Therefore, all mobility management schemes (both at network and data-link layer) must provide sophisticated protection of mobility management.

Protection scheme for bindings depends on availability of security association between the sender of an update information and the recipient. In the absence of such association, a classical method is

based on Return Routability procedure of MIPv6^[1], which allows the recipient of a binding to authenticate the sender by referring sender in two different ways: by means of its locator and with its identifier. It allows the recipient to relate the new locator with identifier for a proof-of-address ownership and thus authenticate the binding update. Other forms involve e.g. self-certifying property of CGA addresses.

3.1 Address derivation

Apart from the well known techniques for address configuration such as DHCP, DHCPv6 or manual, IPv6 allows for dynamic address discovery by means of stateless address autoconfiguration with Neighbour Discovery protocol (ND)^[5]. In ND an IPv6 address is created as a concatenation of a network prefix valid on a link and the interface identifier derived from MAC address. In addition to these well-known methods Cryptographically Generated Addresses (CGA) have been introduced in order to replace the interface identifier with a dynamically generated one, that encodes user identity, and hence provides implicit linkability of the identifier and the locator inside the IPv6 address itself^[6].

The work on cryptographically generated addresses resulted from the observation that IP-based security protocols, such as IPSec, operate on top of IP protocol, so the network layer itself has no means to authenticate the received IP packets. As a result, there is no way to check at the IP layer that a node that claims to be at a given address is really the node at the address. With CGA the receiver can verify

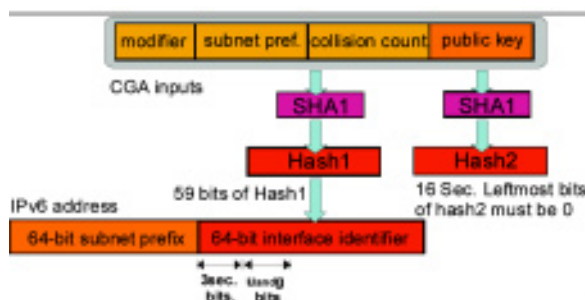
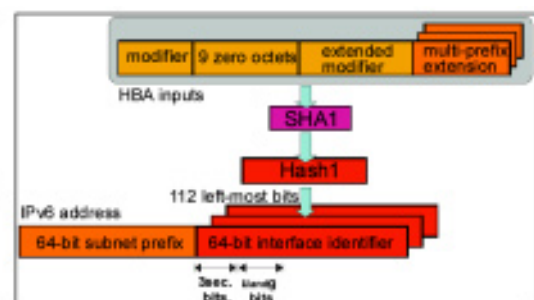


Fig.2 a) Address generation with CGA.



b) Address generation with HBA (simplified)

that IP packet was indeed originated at some source. The CGA solution uses RSA key pair for generating interface-identifier part of the address by hashing the public key of a user with some other parameters (Fig. 2a). Therefore, CGA address has a valuable property to be self-certifying and is used for proof-of-address ownership.

Hash-based addresses (HBA) are a different form, to provide a secure binding between multiple IPv6 addresses used by a multihomed node^[7]. The binding is achieved by generating interface identifier- part of each address as a cryptographic hash of all the network prefixes to be used by the node (Fig. 2b). As a result, the information about the prefixes is digested and included in each address, so the addresses are bound to a prefix set rather than to a public key. The set of prefixes needs to be known in advance. New prefixes cannot be accounted for afterwards, nor old prefixes removed which makes HBAs the right tool for static multihoming scenarios but restricts its use in mobility-enabled multihoming. HBAs have been designed as a CGA extensions for compatibility between the two, and to allow for their simultaneous use e.g. with SEND^[8]. Interface identifiers that are derived by different CGA or HBA are statistically unique. IPv6 Duplicate Address Detection (DAD) check is used to verify that the resulting addresses are outstanding on a certain link^[1].

IV. MOBILITY MANAGEMENT SCHEMES AT NETWORK LAYER

The major reason to locate mobility management schemes at the network layer is that this layer can provide end-to-end support and at the same time a uniform interface to upper layers. Therefore any successful solution operating at the network layer will likely be able to provide transparency to upper layer, specifically the transport layer. It is considered essential for the seamless support of legacy applications.

4.1 Global mobility management

The primary scheme for global mobility management is Mobile IPv6^[1], a protocol which maintains reachability of a mobile node with its home address, independent of whether it is attached to its home link or is away from home. Mobile IPv6 introduces a home agent on the home network to serve as a proxy for the mobile when it is out of home. While away, mobile node registers its topologically correct care-of address with the home agent. The home agent uses proxy Neighbor Discovery to intercept IPv6 traffic addressed to the mobile node on the home link, and tunnels the packets to the mobile node's primary care-of address with IPv6-in-IPv6 encapsulation. With Mobile IPv6, a mobile node can notify corresponding nodes about its current locator for route optimisation purpose, that results in direct traffic routing between correspondent nodes and the mobile node. Another scheme for global mobility, Host Identity Protocol (HIP) is discussed in Section 7. on multihoming.

4.2 Localised mobility management

Localized Mobility Management (LMM) is a new, general approach to manage IP mobility within a restricted, topologically and geographically compact fragment of the network, by exploiting the locality of the movement^[9]. Recent proposals^[10],^[11],^[12] differ conceptually and technically from the old, well known schemes of hierarchical mobility management schemes in that implications of changing the location are concealed both from the global Internet and from the terminal. Here, the mobile node continues to use the previously configured IP address much independent of its current point of attachment, as long as it remains within the localised management area (Fig.3).

The research on LMM is driven by the expectations that multiple mobility-enabling protocols will co-exist in future and that versatile requirements will push the operators to use different solutions that fit best their networking, service and business models. The work is also accompa-

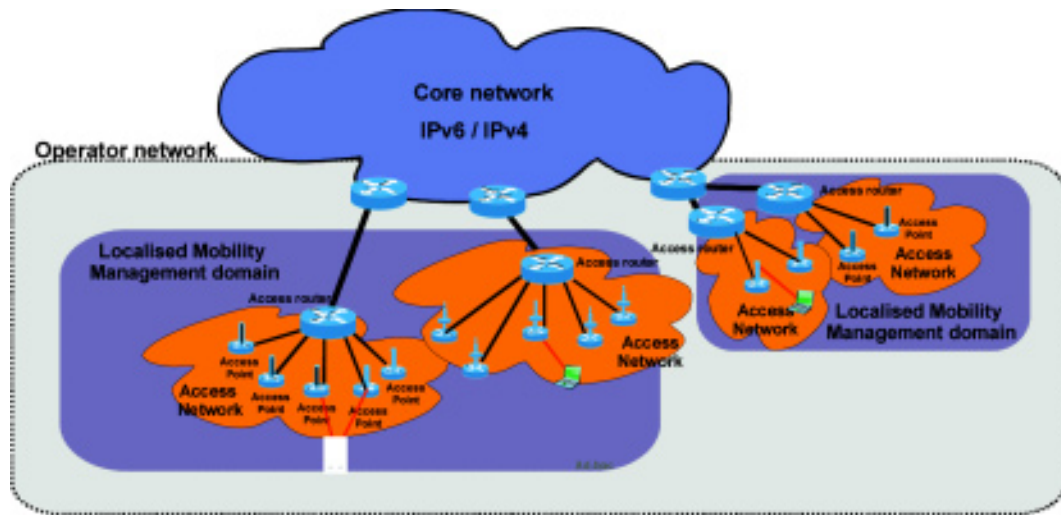


Fig.3 Localised mobility management

nied by change in design paradigms that favour solutions with minimal support from a mobile node^[13] at the cost of increased functionality at the network side. A side effect of minimal host support is that security of communications across the access link is impaired when compared to classical mobility approaches.

V. MOBILITY MANAGEMENT AT DATA-LINK LAYER

The technology often used as a reference for presenting mobility at data-link layer is IEEE 802.11 family of standards. In 802.11 data-link layer handoffs are supported with association, disassociation, distribution, integration and re-association services. Unfortunately, handover delays are prohibitive to support applications such as voice or video, mainly due to time-consuming scanning for candidate network Access Points within proximity of a station. IAPP 802.11f^[14] helps to compensate for this deficiency by letting APs on a common DS to share information about stations in the Extended Service Set (ESS). IAPP allows for secure exchange of station's security context between the current access point (AP) and a new AP during handoff, and enforces a single association for a MT throughout an ESS. Remote

Authentication Dial-in User Service (RADIUS) protocol is used to acquire session keys between APs for secure exchange of data. Proactive caching is introduced to avoid handoff latency caused by IAPP communication. With caching, the serving access point distributes context of a mobile station to neighbouring access points before the station actually MOVES between access points. IAPP relies on a STA making use of reassociation request, hence its applicability depends on the use of these frames.

Mobility management can further be enhanced with CAPWAP protocol^[15] which is based on the assumption that a module located in the AR, which understands 802.11 management frames, can provide control and optimization of the ESS. CAPWAP is a secure protocol to enable AP-to-AR communications for control and provisioning of wireless APs. Supported functions include AR discovery, capability negotiation and AP monitoring (such as signal strength or MT statistics), management and control. The messages can be used as triggers for mobility, providing optimised mobility across a BSS domain without the need for client software. Quality of Service can be supported since the AR can manage the radio links to allow for efficient load balancing. Enhanced support for mobility is achieved by terminating management interface for

the WLAN network in the AR.

IEEE 802.11r^[16] is the work-in-progress for fast BSS transitions aimed to allow connectivity aboard vehicles in motion, with fast and seamless handoffs between APs. Fast Transition is an optional capability in a mobility domain. The domain is defined as a set of all 802.11 APs between which a STA can roam but maintain the same domain MDIE identifier. The primary application currently envisioned for the 802.11r standard is VoIP via mobile phones in WLAN. Fast Transition requires active or passive scanning for candidate APs in the area, re-authentication and re-association with the target one, key derivation by means of 4-way handshake with use of 802.1x and QoS admission control to re-establish QoS. During the past months the standard body Task Group was combining TAP (Transition Acceleration Protocol) and Just-In-Time 2 Phase Association protocols into a single solution.

IEEE 802.11u is the early proposal for a future standard aimed to allow for a common and generic approach to interwork IEEE 802.11 access networks to external networks. The specification is expected to include amendments to the PHY and MAC layers on seamless handoff, session persistence, authentication, IP addressing and data rate changes.

There is also some early work on IEEE 802.11h which is supposed to address radar detection, transmit power control and dynamic channel selection.

A fast growing segment of mobility solutions which are already available in products is based on the concept of dynamic re-configuration of AP to adjust the 'coverage' of a 'virtual LAN' segment so as to make the station believe that it is continuing to operate within the old segment. Some of these solutions transfer station associations between APs and maintain forwarding entries in Ethernet switches to adjust for the current point of attachment of the station within a link-layer mobility domain. The entries allow to forward traffic to and between stations. These solutions resemble link-layer versions of a per-host routes in network layer, but restricted in scope to

a certain topological domain. This idea, after some transformations, is now re-implemented in network-layer mobility in some approaches to localized mobility management.

VI. CROSS-LAYER APPROACHES FOR MOBILITY MANAGEMENT

Substantial handoff latency is inherent in IP environment due to latency of movement detection of a mobile, latency of router discovery on new link, address configuration and validation, registration in a new network and binding update. Additional latency is introduced by security schemes employed for (re)authentication of a new network access with various schemes, based on handshake protocols, that can be roughly classified as pro-active, just-in-time and credit-based.

6.1 Handover latency management

Numerous optimisations have been proposed for improved handoff performance ranging from reactive handoffs with- or without support from network, to proactive handoffs with movement anticipation, and to network assisted handoffs^{[17], [18], [19]}. Furthermore, substantial effort has been made to combine handover preparation with handover execution^[20]. The problems and some interesting solutions are well documented in research papers, such as,^{[21], [22], [23], [24], [28], [48]}. The paramount number of solutions resulted in a debate on feasibility of disruptive and non-disruptive solutions, with argumentation that fall beyond the scope of this paper. Nevertheless, we would like to focus on some standardisation-related research aimed to alleviate the problem of handover latency.

6.2 Link-layer event indications

A principle of layered networking and current TCP/IP protocol stack implementations leads to a problem of a slow movement detection of a

mobile node entering the new IP network^[23]. The problem is under study in various groups. IETF DNA WG^[25] suggests two complementary solutions. For interworking with legacy routers a Complete Prefix List protocol^[26] allows a mobile node to maintain a list of known on-link network prefixes, typically learnt from router advertisement messages. Reception of a link-prefix which is not on the list is considered an indication of a change of network attachment and results with a mobility trigger. The DNA protocol augments this approach with solicited Fast Router Advertisements that provides a unique link identifier, agreed upon by neighbouring routers operating on that link. In addition, a mobile can solicit a regular IPv6 RA message after a L2 handoff indication to verify if the network prefix is still valid on a new link.

A more systematic approach is reflected in IEEE 802.21 standard which defines a Media Independent Handover (MIH) architecture for an optimized mobility management in heterogeneous networks^[27]. The proposed architecture is different from many detection-based schemes with strong decoupling of layers. MIH natively supports various cross-layer indications from link-layer to upper layers. The objective is to enhance the handover process by supplying information from lower layer to upper layers to indicate events such as detection of new carriers and to supplement and control the handover process. This task is accomplished by the MIH Function that resides in the protocol stack of the mobile node as well as in the access network node. Its primary role is to provide indications in the form of triggers to the network layer mobility management entity and to higher layers. MIH Function provides convergence of link-layer state information to the upper layers from multiple heterogeneous access technologies available in the terminal in a consistent and technology-independent way. Link indications support 3GPP and 3GPP2 specifications and IEEE

802 family of standards, including wireless and wired. The MIH enhances mobility management with a wide range of handover supporting functions. Access to the MIH is available throughout a set of

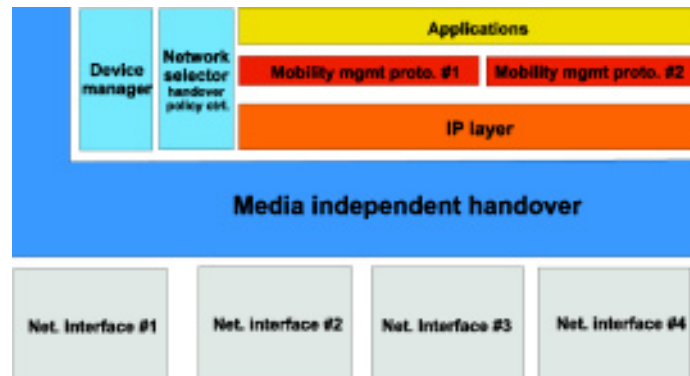


Fig.4 Media Independent Handover architecture (simplified)

media-independent Service Access Points (SAPs) and related primitives as presented in Figure 4.

The MIH Function is supported by three services that allow the management entity to receive information from link-layers available in the mobile node as well as to retrieve information directly from the network or other nodes and to dispatch the handoff accordingly.

Media Independent Event Service (MIES) facilitates handover detection in the network layer mobility entity. To receive this information the entity registers with the MIH Function for selected events. The events include Link Up, Link Down, Link Going Down, Link Parameters Change, triggers for imminent data link layer handoffs and other. When registered, network layer receives this information selectively for every link. Local events are conveyed from link layer to MIH and further to network layer or directly to mobility management entity. Remote events propagate from MIH layer or network layer mobility entity in one node to the corresponding layer in another node. Based on this information, layer 3 mobility management protocol can prepare e. g. an imminent handover.

The Media Independent Command Service (MICS) supports commands from the higher layers to MIH layer and from there down to lower layers or, where appropriate also to a peer node. The command car-

ries decisions of a higher layer or the mobility management entity responsible for the lower layer in order to control handover process. The set of commands to be supported from higher layers to MIH layer is defined in the standard as well as the commands from MIH to lower layers

Media Independent Information Service (MIIS) provides a framework and the mechanisms for MIHF (Media Independent Handover Function) to retrieve network information needed as input to handover decisions. MIIS defines information elements as well as request-response mechanism for information retrieval. The information may be stored in the MIH entity or in an external repository. The mobility management entity can use that information for multiple purposes, including e.g. dynamic view of network topology.

The information service can be used to access static and dynamic information about the network such as neighbor reports, channel activity and security information, or about higher layer services in order to optimize network discovery and selection.

Currently there is a substantial effort to harmonise 802.21 solution with other groups within 802 and other standardisation bodies by means of liaison.

VII. GENERALISED MOBILITY WITH MULTIHOMING

In the context of support for mobility, multihoming allows a node to make use of multiple attachments (interfaces) to the global Internet without the need to insert mobile node-specific entries in the global routing system.

Recent proposals for multihoming allow the node to use multiple network attachments in parallel, or to select dynamically the preferred one based

on: local preference, information exchange with the communication peer, or based on a status of attachments. Multihoming is typically supported without intervention or interaction with the transport and application layer, hence making it transparent to applications as well.

At present, there are multiple approaches to multihoming^[29]. A large group of them is based on an assumption that nodes will explicitly provide support for multihoming, while other solutions assume minimal involvement of mobile nodes.

The first group of solutions introduces additional endpoint identity protocol element (EIP) to be inserted into existing protocol stacks. The EIP would present endpoint identifiers to upper layer protocols (ULP) to identify the local and the remote stacks. By doing this, stable identifiers could be maintained throughout duration of the session. The lower layer protocols (LLP) would be presented with locators rather than identifiers (Fig. 5). The identity protocol element would have to maintain a dynamic mapping between the two, to be updated on triggers from the network environment, from remote endpoints and from local control applications. The peering between identifiers would be preserved throughout duration of a session while the locators would be agreed upon with the peer by signalling exchange to maintain reachability.

Alternative proposals are centered around the concept of modified protocol elements (MPE), where an existing protocol stack element (transport or network layer) is modified to allow for binding multiple locators to a single session and then communicating this locator set to the remote transport entity^[29]. This

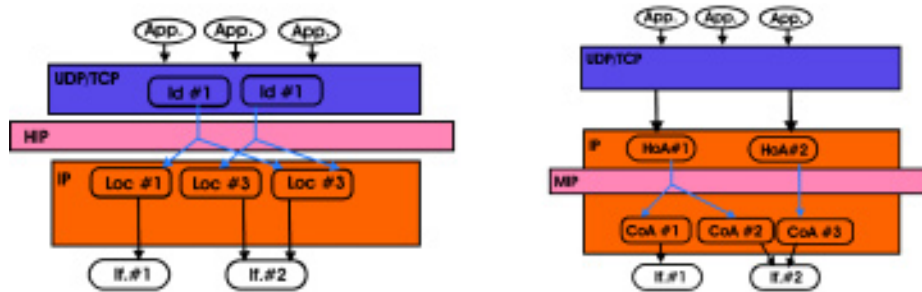


Fig.5 Architectural approaches to multihoming

allows the local MPE to switch the mapping to a different locator for either the local endpoint or the remote endpoint. The switching can be implemented with e.g. dynamic rewriting of the packet header when processed by the protocol element, or with various types of encapsulation. While this approach is feasible at the network layer, a change in the transport layer implies a change to application programming interface (API).

Approaches with minimal host involvement assume that certain type of packet transformations will be performed in the network elements, e.g. in the access routers. The local hosts may write its endpoint identity as the source address in a packet, and the access router will rewrite the source to appropriate locator. Packet header rewriting methods by network elements have large number of associated security concerns.

The common thing about all the multihoming approaches with support for identity is that there is a certain exchange of information that allows communicating peers to be aware of the set of locators of the peer.

One of recent solutions, shim6 protocol^[30], allows for change of a locator below the transport protocols, so that the session can survive loss of connectivity on one of interfaces. The node can use shim6 to setup state with its peer, which can later be used to switch to a different locator pair, to preserve established communications. The protocol does not imply any latency during connection setup because signaling exchange is carried during normal session operation. The protocol provides limited support for mobility, because of lack of support for simultaneous renumbering on all of network interfaces.

The Host Identity Protocol (HIP) is a well known secure mobility and multi-homing protocol^[2]. It introduces a new layer between the transport and network layers of the TCP/IP stack that maps host identifiers to network locations. HIP defines host identifiers for naming the communication endpoints thus leaving the IP addresses a role of locators, and performs authentication and creation of IPSec security associations between the identifiers. The host

identity in HIP is its public key. It is self-certifying in the sense that it can be used to verify digital signatures without access to certificates or a public-key infrastructure. The identity is usually represented by a host identity tag, a 128-bit hash of the host identity.

VIII. SECURITY COMPONENTS

8.1 Bootstrapping network attachment

Network attachment is the initial step taken by a node connecting to a network. Attachment process typically involves negotiation of link layer parameters, authentication, access control, and configuration of network layer information. These steps usually induce large but tolerable latencies, because the process is performed at connection setup only. Authorisation can be carried e.g. with IP-agnostic EAP^[31] or a link-layer agnostic protocol such as PANA^[32].

EAP is an authentication framework with support for various authentication methods. EAP typically runs directly over data link layers such as, e.g. IEEE 802, without requiring IP. It is a common practice to use higher-layer authentication on top of link-layer authentication. PANA complements link-layer authentication with a network-layer scheme to allow for a mutual authentication of a host and a network for network access on multi-access links. It operates between a client's device and a proxy to authentication infrastructure, acting as a front-end, and a carrier for a EAP or mechanism which is used there^[33].

PANA allows clients to interact with the infrastructure AAA in a protocol-neutral way that does not require the client to understand the AAA protocols.

8.2 Protection on the local link

Nodes and routers in IPv6 network use Neighbour Discovery Protocol (NDP)^[5] in order to discover peers on the local link. Nodes transmit solicitations when they are looking for another node to provide

a particular service or receive some data. Neighbour advertisements are transmitted by other nodes to indicate the addresses they own and to provide related information. Advertisements are transmitted on a regular basis so that all nodes are aware of the current network configuration. Advertisements are also transmitted in response to solicitations which allow new nodes to learn details of the network without having to wait for regular advertisement or to allow a node to obtain information which it has not cached. Neighbour discovery process depends on nodes accurately advertising their capabilities and services. Although neighbour discovery is critical for the operation of a local network, it is easy for an insider to disrupt this process by transmitting false advertisements and solicitations resulting with a DoS attack for some stations on the local link. Secure ND (SEND) defines multiple extensions to secure NDP with use of Cryptographically Generated Addresses (CGA) for authentication of a sender of a message, and a chain of certificates linking back to a "certificate anchor" which is used to ensure that a device is authorised to act^[8]. Proof-of-address ownership is supported with CGA addresses and message signatures to link the message contents to the sender address. SEND also defines messages to allow for managing key chains by means of a delegation chain.

8.3 Protection on the access link

From the systems point of view protecting an access link is currently supported by abstracting various radio access network security capabilities in order to provide uniform security interface for higher layers. Abstracting involves use of native security support of the link layer and additional network layer enhancements.

Current approaches also recommend cryptographic separation of access technologies to avoid the domino effect when one of the links is compromised. Such a separation poses a problem for handover management. Therefore, research concentrates on developing EAP keying framework to provide au-

thentication services and key management by generating Master Session Key (MSK) and Transient keying material to a pass-through authenticator entity^[34]. The MSK is supposed to be used for derivation of transient session keys (TSK) with the use of a Secure Association Protocol for protecting individual links. Although the EAP Keying framework provides the tools, the problem is in topologically decoupling of the authenticator from access routers (and from access networks) in order to protect the MSK and thus avoid time-expensive re-generation of the MSK during handover^[35]. Recent proposals aim at bringing the concept of a pass-through authenticator and key holder from 802.11r to handover keying at network layer.

8.4 Authorisation for service access

Recent approaches recommend separation of network and service access control, even though similar authentication methods can be used in both cases. Currently four major schemes are mainly used in various mobility approaches for service access control. Authorisation with cryptographically generated identifiers is based on self-certifying property of identifiers derived from public key of the key pair. Knowledge of the secret (i.e. corresponding private key) gives an implicit authorisation for certain actions, such as updating bindings during mobility. A similar approach is used in authorisation scheme based on token generated by an off-line third-party including certificates or SAML assertions. Authorisation certificate can assert that the owner of a particular public key is entitled to take some actions. Another is the authorisation based on protocol exchange with an online third party, like RADIUS/Diameter. These are typically slow due to multiple round-trip message exchanges^[36]. Authorisation certificates can be used for delegation purpose, where the issuer gives the subject the authorisation to execute certain actions or to transfer that privilege. One use case for such behaviour is e.g., when a mobile node delegates authorisation to a local access

router to act on behalf of a mobile. The disadvantage often is that certificate revocation needs to be implemented for such scenario as well.

8.5 End-to-end protection

Message authentication, confidentiality and integrity at the IP layer is supported with ESP and AH subprotocols of IPSec^[37]. IPSec make use of Security Associations (SA) between communicating peers, which are established and maintained by the IKE protocol and referred to by a Security Parameter Index (SPI). The SPIs are maintained inside hosts in a lookup table typically indexed with IP addresses of peers.

In the baseline IKEv2 protocol^[38], the IKE SAs and the resulting ESP IPsec SAs are created implicitly between the addresses pair that is used during the protocol execution. This means that in each host only one IP address pair is stored for the IKEv2 SA as a result of a single IKEv2 protocol session. Currently, it is not possible to change any of the two addresses because lack of support for such a change in IPsec. In principle, the IKE SA and the corresponding IPsec SAs could be re-established after the IP address change, but it is a computationally prohibitive, and user interaction might be required every time as part of the IKEv2 authentication procedure. MOBIKE^[39] introduces an automatic mechanism to update the IP addresses associated with the IKE SA and the IPsec SAs during an IPsec session, so that a mobile node can continue connection when changing location. MOBIKE provides updates of IPsec tunnel endpoints, but does not support simultaneous change of both endpoints nor initial discovery of endpoint location. MOBIKE supports only one pair of addresses at a time, hence the solution is better suited for mobile VPN where the gateway is fixed.

8.6 Security of multihoming

The ability to switch running sessions between interfaces by the host raises concerns about security implications. The concerns depend on type of identification of session end-points, which can be identifier-

or locator-based. Many of the mobility protocols de-couple the two and use locators to identify endpoints of a communication channel but identifiers for session control and data presentation purpose. A number of solutions have been proposed along these lines, like e.g. shim6^[30]. In these approaches threats are related to attacks against identifier-locator binding. One can view these as an attempt to modify session context originally established when the communication was initiated. In order to protect against attacks on locator-identifier binding it is important for the endpoint to be able to authenticate the communication peer and verify (or presume) its authorisation to control its locators. Such authentication has to be included in every multihoming solution, but sometimes may be restricted to a simple check that the communicating peer continues to be the same one with which the session was originally set up^[40].

This type of check is usually carried after receiving signalling for a change of the preferred locator by the peer or after signalling the change in the set of valid locators. Signalling for a change should also support explicit de-registration of the locator(s) once they're no longer owned by the peer.

The other group are attacks against session management schemes. These typically exploit vulnerabilities of the transport (and upper) layer protocols resulting from the fact some of protocols are not designed to cope with implications of intentional misuse of one-to-many nature of identifier-to-locator bindings. Any of the locators, both on the near- and the far-end can potentially be exploited to launch a redirection-based or MITM attacks leading to loss of integrity, confidentiality or denial of service for the session^[40].

The ease to bootstrap such an attack depends on location of an attacker with respect to session flow and end-points. The attacker can be out-of-path, on-the-path or in the peer end-point.

The last location is particularly vulnerable since a multi-homed station readily discloses its locators to a peer exposing itself to such attacks. Vulnerability to session attacks also depends on

the protection level of the session flow and the relative location of the protection protocol (e.g. IPSec) with respect to multihoming-supporting protocol in the protocol stack.

It is worth adding here that the vulnerabilities discussed above are inherent to multihoming per se. Mobility of a multihomed node just adds some new threats resulting from increased instability of network attachments, latency in detections of network attachments and frequent locator changes.

IX. PRIVACY COMPONENTS

Privacy encompasses, among others, anonymity, pseudonymity and unlinkability. In a mobility context anonymity assures that neither the correspondent in a session nor an attacker are able to find a relation between the entity identifier and its real identity. Pseudonymity allows that even though entity can not be identified, it is still possible to relate two separate pseudonymous acts. Unlinkability implies that successive actions taken by the same entity cannot be correlated. In the context of mobility it means, for example, that successive handover can not be related^[41].

Correlations can typically be found with address-based or control-information-based matching schemes. The former inspect correlations between network and data link layer addresses during activity periods of a mobile node, while the latter focus on tractability of sequence numbers of frames, security parameter index values (SPI) or other indicators which are stable over time or easy to predict.

In Mobile IP scenario, disclosure of HoA implies disclosure of user identity while disclosure of CoA reveals current location^[42]. Locators and identifiers need to be partly disclosed to entitled entities to allow for communications. For route-optimization mode, the locator has to be revealed to Correspondent Node but not to third parties. Disclosure of location alone is usually meaningless, specifically when it is not linkable to identity. Therefore various schemes suggest to use pseudo-identifiers and temporal mobility identifiers^{[43], [44], [45]}. Stable pseudo-

identifiers allow to index security associations (SA) with the communication peers internally in the mobile node, and allow to avoid re-numbering SAs after location change of the peer.

X. SYSTEMS AND ARCHITECTURES

Numerous standardization bodies are working on IP-based approaches with support for mobility in systems beyond 3G. Mobility management schemes incorporate handovers at different layers (session-, network- and data link layer), intra- and inter-technology handovers, within and between administrative domains.

ITU-T is working on Next Generation Network (NGN) to support migration to an IP-based infrastructure, by seeking for viable paths to converge internet protocol networks (IP), public switched telephone network, digital subscriber lines, cable networks, wireless local area network and mobile technologies under global standards. Networking related aspects including frameworks, functional architecture, evolution and convergence into NGN are studied in SG13, SG17 and SG19.

3GPP and **3GPP2** are concentrating on IP Multimedia Subsystem (IMS), a component to provide packet-switched multimedia services. These activities are carried under All-IP Network (AIPN) System Architecture Evolution /Radio Access Network, Long Term Evolution. The working group is also trying to improve throughput with more efficient use of the 3G spectrum between the mobile- and the base station (SUPER3G). The study will be conducted by possibly incorporating WLAN as one physical layer in order to provide alternative to 802.16^[46].

The **ETSI TISPAN** is promoting NGN-FBI solution which builds on a 3GPP IMS for service provisioning and handover between fixed and mobile operators^[47]. The NGN user equipment receives IP connectivity via the transport layer, with support from Network Attachment Subsystem (NASS) and the Resource and Admission Control Subsystem (RACS).

The **IEEE 802** is gaining momentum with media-

independent handover in 802.21^[27].

At the same time handover enhancements are elaborated in IEEE 802.11 Task Group R, while 802.11u is working on inter-technology handover.

IETF has been proposing frameworks and protocols in mip6 and netlmm working groups while enhancements have been evaluated in multiple groups including mobopts, seamoby, mipshop, mobike and dna. Multihoming is studied in multi6 and monami6. Network mobility is studied in nemo working group. Security related to mobility is being covered in multiple groups and bofs, including momipriv, hoakey, ike24multi, just to name a few. New bofs are also set up like, e.g. 16ng.

The WIMAX Forum Network Working Group of is defining a reference architecture for a mobile WiMax supporting flexible mobility schemes with security component^[46]. Their focus is on reference points, protocols, procedures and primitives. The overall goal is to facilitate deployment of services for end-users based on IEEE 802.16 air interface. Scalability, extensibility and interworking with other systems are accounted in the design. Networking functions for radio access to a WiMax system are defined in the Access Service Network (ASN), while baseline IP services are covered in Connectivity Service Network (CSN).

Along with standardization activities, various operators and vendors are also involved in developing harmonized solution for heterogeneous networks, like, e.g. UMA and several other.

XI. SUMMARY

We have addressed several problems related to IP-based mobility in a heterogenous network environment. For better presentation of the problems, we have been referencing work-in-progress documents that set out major research directions.

Observation of the developments clearly shows, that the designers have to re-think in repetitive cycles their design paradigms, revisit assumptions and priorities as well as verify the networking models for the next generation Internet because of the changing expectations and the high complex-

ity of the solutions that have to account properties of underlying infrastructure. It also becomes evident, that the deployment of a flexible and robust infrastructure, resilient to many threats, requires sophisticated use of a set of mobility-enabling and accompanying protocols, many of which are not fully standardised yet.

The proposed solutions, due to the requirement for harmonization across the Internet, need to be endorsed by the industry and accepted by the Internet community. The solutions need to provide a high degree of freedom of use to allow for competitive, but interoperable deployments. Therefore, many prototype deployments still have to undergo a long way, although some interim solutions will probably be deployed soon.

ACKNOWLEDGEMENTS

This work has been done in Daidalos II Integrated Project (<http://www.ist-daidalos.org>).

The author wishes to thank the anonymous reviewers for their support.

REFERENCES

- [1] D. Johnson C. Perkins, J. Arkko, "Mobility Support in IPv6", RFC 3775, 2004.
- [2] R. Moskowitz, P. Nikander, P. Jokela, T. Henderson, "Host Identity Protocol", Work in progress, draft-ietf-hip-base-05, IETF, March 2006.
- [3] T. Dreibholz, J. Pulinthanath, "Applicability of Reliable Server Pooling for SCTP-Based Endpoint Mobility", Work in progress, draft-dreibholz-rserpool-applic-mobility-00.txt, IETF, March 2006.
- [4] S. Lee, S. Jeong, H. Tschofenig, X. Fu, J. Manner, "Applicability Statement of NSIS Protocols in Mobile Environments", work in progress, draft-ietf-nsis-applicability-mobility-signaling-04.txt, March 2006.
- [5] T. Narten, E. Nordmark, W. Simpson, "Neighbor Discovery for IP Version 6 (IPv6)", RFC2461, IETF, 1998.
- [6] T. Aura, "Cryptographically Generated Ad-

dresses (CGA)", IETF RFC 3972, March 2005.

[7] M. Bagnulo, "Hash Based Addresses (HBA)", work in progress, draft-ietf-shim6-hba-01, IETF, October 2005.

[8] J. Arkko, Ed., J. Kempf, B. Zill, P. Nikander, "SEcure Neighbor Discovery (SEND)", RFC 3971, IETF, March 2005.

[9] Kempf, J., Leung, K., Roberts, P., Nishida, K., Giaretta, G., Liebsch, M., and Gwon, Y., "Problem Statement for IP Local Mobility" Work in progress, draft-kempf-netlmm-nohost-ps-01.txt, IETF, January 2006.

[10] I. Akiyoshi, M. Liebsch, "NETLMM Protocol", Work in progress, draft-akiyoshi-netlmm-protocol-00.txt, IETF, October 2005.

[11] G. Giaretta, I. Guardini, E. Demaria, "Network-based localized mobility management (NETLMM) with distributed anchor routers", draft-giaretta-netlmm-protocol-00.txt, IETF, October 2005.

[12] S. Gundavelli, K. Leung, "Localized Mobility Management using Proxy Mobile IPv6", draft-gundavelli-netlmm-mip6-proxy-00.txt", IETF, November 2005.

[13] J. Kempf, K. Leung, P. Roberts, K. Nishida, G. Giaretta, M. Liebsch, "Requirements and Gap Analysis for IP Local Mobility", Work in progress, IETF, January 2006.

[14] "IEEE Trial-Use Recommended Practise for Multi-Vendor Access Point Interoperability via an Inter-Access Point Protocol Across Distribution System Supporting IEEE802.11 Operation", IEEE Std. 802.11F-2003, IEEE, 2003.

[15] P. Calhoun, M. Montemurro, D. Stanley, "CAPWAP Protocol Specification", Work in progress, draft-ietf-capwap-protocol-specification-01, IETF, May 2006.

[16] "Wireless Medium Access Control (MAC) and physical layer (PHY) specifications: Amendment 2: Fast BSS Transition", IEEE P802.11r/D2.0, March 2006.

[17] R. Koodli (ed.), RFC4068, "Fast Handovers for Mobile IPv6", IETF, July 2005.

[18] M. Liebsch, A. Singh, H. Chaskar, D. Funato, E. Shim, "Candidate Access Router Discovery

(CARD)", RFC4066, IETF, July 2005.

[19] T. Melia, R. Aguiar, N. Senica, "Network initiated handover in fast mobile IPv6", Work in progress, draft-melia-mobopts-niho-fmip-01, IETF, July 17, 2005.

[20] R. Aguiar, A. Banchs, C. J. Bernardos, M. Calderon, M. Liebsch, T. Melia, P. Pacyna, S. Sargento, I. Soto, "Scaleable QoS-aware Mobility for Future Mobile Operators", accepted for publication in IEEE Communications Magazine, due June 2006.

[21] Ch. Vogt, "A Comprehensive Delay Analysis for Reactive and Proactive Handoffs with Mobile IPv6 Route Optimization", Tech. Report, TM-2006-1, University of Karlsruhe, ISSN 1613-849X, January 2006.

[22] Lila Dimopoulou, Georgios Leoleis, Iakovos S. Venieris, "Fast Handover Support in a WLAN Environment: Challenges and Perspectives", IEEE Network, vol. 19, no. 3, May 2005 .

[23] S. Narayanan , G. Daley, N. Montavont, "Detecting Network Attachment in IPv6 - Best Current Practices for Hosts", draft-ietf-dna-hosts-00.txt, Work in Progress, April 2005.

[24] G. Lampropoulos, N. Passas, L. Merakos, A. Kaloxylos, "Handover Management Architectures in Integrated WLAN/Cellular Networks", IEEE Communications Surveys, Fourth Quarter 2005, Volume 7., No. 4.

[25] A. Yegin, E. Njedjou, S. Veerepalli, N. Montavont, T. Noel, "Link-layer Event Notifications for Detecting Network Attachments", draft-ietf-dna-link-information-01.txt, February 2005.

[26] J. Choi , E. Nordmark, "DNA with unmodified routers: Prefix list based approach", Work in progress, draft-ietf-dna-cpl-00.txt, IETF, April 2005.

[27] 802.21 draft P802-21-D00-03, October 2005.

[28] I. F. Akyildiz, J. Xie, S. Mohanty, "A Survey of Mobility Management in Next-Generation All-IP-Based Wireless Systems", IEEE Wireless Communications, vol. 11, no. 4, August 2004.

[29] G. Huston, "Architectural Approaches to Multihoming for IPv6", RFC4177, IETF, September 2005.

[30] [shim6] E. Normard, M. Bagnulo, "Level 3 multihoming shim protocol", Work in progress, draft-ietf-shim6-04.txt, March, 2006.

[31] B. Aboba, L. Blunk, J. Vollbrecht, J. Carlson, H. Levkowitz, "Extensible Authentication Protocol (EAP)", RFC3748, IETF, 2004.

[32] D. Forsberg, Y. Ohba, B. Patil, H. Tschofenig, A. Yegin, "Protocol for Carrying Authentication for Network Access (PANA)", draft-ietf-pana-pana-11, Work in Progress, IETF, March 2006.

[33] A. Yegin, Y. Ohba, R. Penno, G. Tsirtsis, C. Wang, "Protocol for Carrying Authentication for Network Access (PANA) Requirements", RFC4058, IETF, May 2005.

[34] B Aboba, D Simon, J. Arkko, P. Eronen, H. Levkowitz, "Extensible Authentication Protocol (EAP) Key Management Framework", Work in Progress, draft-ietf-eap-keying-13.txt, IETF, May 2006.

[35] K. Chowdhury, J. Bournelle, M. Nakhjiri, "AAA based Keying for Wireless Handovers: Problem Statement", draft-nakhjiri-aaa-hokey-ps-01, February 2006.

[36] K. Chowdhury, J. Bournelle, M. Nakhjiri, "Problem Statement for the AMSK", draft-chowdhury-hoakey-amsk-ps-00, Work in Progress, IETF, February 2006.

[37] S. Kent, R. Atkinson, "Security Architecture for the Internet Protocol", RFC2401, IETF, 1998.

[38] C. Kaufman, Ed., "Internet Key Exchange (IKEv2) Protocol", RFC4306, December 2005.

[39] P. Eronen, "IKEv2 Mobility and Multihoming Protocol (MOBIKE)", Work in progress, draft-ietf-mobike-protocol-08.txt, IETF, February 2006.

[40] E. Nordmark, T. Li, "Threats to IPv6 Multihoming Solutions", RFC4218, IETF, October 2005.

[41] [momipriv] W. Haddad, E. Nordmark, F. Dupont, M. Bagnulo, B. Patil, "Privacy for Mobile and Multi-homed Nodes: MoMiPriv Problem Statement", draft-haddad-momipriv-problem-statement-02, October 2005.

[42] R. Koodli, "IP Address Location Privacy and Mobile IPv6: Problem Statement", Work in Progress, draft-ietf-mip6-location-privacy-ps-01.txt, IETF, March 2006.

[43] T. Narten, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6", Work in Progress, draft-ietf-ipv6-privacy-addr-v2-04, IETF, May 2005.

[44] C. Castellucia, F. Dupont, "A simple privacy extension for Mobile IPv6", Work in progress, draft-dupont-mip6-privacyext-03.txt, IETF, January 2006.

[45] W. Haddad, S. Krishnan, F. Dupont, M. Bagnulo, H. Tschofenig, "An Anonymity and Unlinkability Extension for OMIPv6", Work in progress, draft-haddad-privacy-omip6-anonymity-01, March 2006.

[46] "WiMAX End-to-End Network Systems Architecture (Stage 2: Architecture Tenets, Reference Model and Reference Points)", WiMax Forum, April 2006.

[47] "TISPAN Network Attachment Subsystem; Release1", DES/TISPAN-02021-NGN-R1, ETSI ES 282 004, March 2006.

[48] Janise McNair, Fang Zhu, "Vertical handoffs in fourth-generation multinet environments", IEEE Wireless Communications, vol. 11, no. 3, Jun 2004 pp. 8-15.

BIOGRAPHY

Piotr Pacyna received M.Sc. degree in computer sciences in 1995 and Ph.D. in telecommunications in 2005 from the AGH University of Technology in Kraków, Poland, where he has been working as a lecturer. He has been working as a



lecturer in the Department of Telecommunications of the AGH University of Science and Technology. His research focuses on next-generation IP networks, mobility and security. He has spent his sabbatical leaves in Loracom in Nancy, France and in CNET France Telecom, Paris, France. He is now a visiting professor at Universidad Carlos III de Madrid in Spain.

Piotr Pacyna has been active in several EU funded ACTS and IST research projects: BTI (1997-2000), Moby Dick (2001-2003) and IST Integrated Projects Daidalos (2003-2006) and Daidalos II (2006-).

Piotr Pacyna authored several research papers.