Architectures for Communication in Personal Networks

R. V. Prasad, Martin Jacobsson, Sonia Heemstra de Groot[‡], Anthony Lo, Ignas Niemegeers

Wireless and Mobile Communications,
Faculty of EEMCS, Delft University of Technology,
Delft, The Netherlands
{vprasad, m.jacobsson, a.lo, i.niemegeers} @ewi.tudelft.nl

‡Twente Institute for Wireless and Mobile Communication
Institutenweg 30, 7521 PK, Enschede, The Netherlands
Sonia.Heemstra.de.Groot@ti-wmc.nl

Abstract— **Personal Networks (PN) [8] is a new concept related to pervasive computing with a strong user-focus view. The key to a successful PN realization is a general network architecture that is capable of bridging different current and future technologies and offers a homogeneous and clear view to the end-user. In this paper, we focus on forming a PN by connecting remote personal devices using infrastructure-based IP networks, including 3G networks and WLAN hotspots. One way is to upgrade the current access networks with new functionality to support PNs. Since many devices in PNs are mobile and battery powered, this may help them to achieve a faster service and to save energy. However, to deploy such functionality is not easy and may hamper the adoption of PNs altogether. Therefore, in this paper we study three possible inter-cluster communication architectures that can use current IP networks. To discern the above proposal we also give a detailed picture of PN network architecture supported by infrastructure. We believe that this detailed discussion will help the success of PNs.**

Keywords- Personal Networks, Gateway nodes, Edge Routers, PN Agent

I. Introduction

Personal Networks (PN) [8] is a new concept related to pervasive computing with a strong user-focused view. PN extends a person's Personal Area Network (PAN) that surrounds him with devices and services farther away. This extension will physically be made via infrastructure-based networks, vehicle area networks, a home network or mobile ad hoc networks (MANET). A person's PN is configured to support his/her applications and takes into account the context, location and communication possibilities. A PN must adapt to changes in the surroundings, be self-configurable and support many different types of networks and devices. Figure 1 shows a future PN scenario. The key to a successful PN realization is a general network architecture that can bridge different technologies and offer a homogeneous and clear view to the end-user. Since a PN should address all communication needs of a person, it must include not only the person's wearable and wireless devices but also devices in the home, in the car, in the office, etc. It will undoubtedly be the network layer that should integrate all these devices and networks into one PN and at the same time co-operate with existing networks such as infrastructure networks and other fixed networks.

Some of the novel application scenarios for PNs include: (a) *A health-monitoring application* – where the person would keep updating his monitoring application in his house by connecting from his PAN. (b) *Business environment extended from the office to the car* – where the devices in car, office and home would all converge to provide necessary support to the person. (c)*Walking through a smart building* – where the building responds to the person by enabling lighting and performing the access controls, etc. (d) *A telepresence session* – extending the meeting environment by adapting to mobility and available resources. PNs are very much centered on a person and his/her needs. As such they will need be dynamic in composition, configuration and connectivity, depending on the time, place and circumstances, the resources required and the partners one wants to interact with.

Thus PNs must be capable and flexible enough to be able to make use of current and future communication networks. For direct communication between a person's devices, PNs must be able to use ad hoc wireless communication technologies such as Bluetooth [2], ZigBee [26], Wireless Local Area Networks (WLAN) [6] in ad hoc mode, low rate WPANs, IEEE 802.15.4 [7] and other future technologies. To interconnect personal devices in different locations, PNs must be able to use infrastructure-based networks such as UMTS and GPRS networks, WLAN hotspots, wired and wireless broadband access networks and other evolved future versions of these access networks.

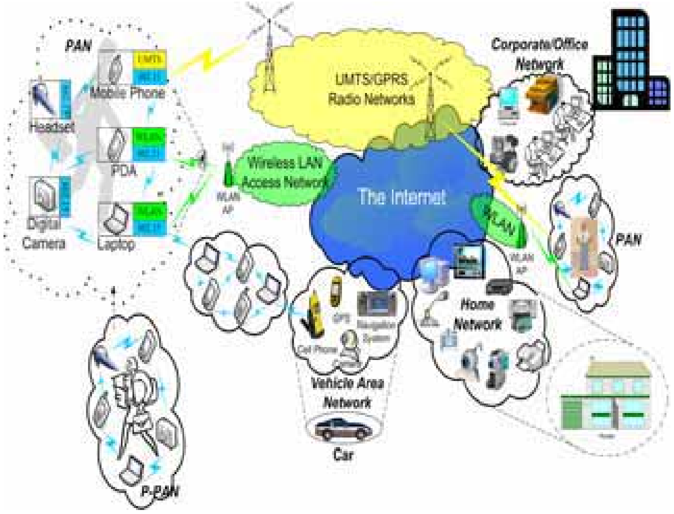

Figure 1. An example of a Personal Network

[1] Some of the ideas in this paper were presented as a poster in WMASH 05 [30].

[2] While a part of this work is sponsored by the IST MAGNETBeyond and Freeband PNP2008 projects, some aspects presented here do not necessarily reflect the views of MAGNETBeyond and PNP2008

0-7803-9792-4/06/$20.00 ©2006 IEEE

Especially IEEE 802.11b and IEEE 802.11g [6] wireless standards have been widely-adopted and used for wireless connectivity. WLAN Hotspots based on IEEE 802.11b or 802.11g are often located in places such as airports, train stations, libraries, marinas, conventions centers and hotels. Devices with newer operating systems can bootstrap to connect to these networks. These networks usually have a DHCP server and can offer Internet connectivity to a user by assigning IP numbers on the fly. On many occasions users are directed through Network Address Translation (NAT) [1, 5, 22] setup to access the Internet infrastructure. PNs should still be able to use this and other setups that are frequently available. This will help PNs to deploy faster and make use of communication networks even more.

Another important criterion for the success of PNs is the *trust* that people can have in the system. Systems like PNs are extra vulnerable because of their mobile and wireless nature. In the world of mobile communication, IT security meets traditional security and this opens up a completely new world of problems in the security domain. The ad-hoc nature of PNs means that a person's PN will encounter many unknown persons and the owner should be able to remain anonymous and properly protected from those parties he does not know or trust whenever he wishes to. PNs achieve this by requiring the user to personalize one's devices so that the system can distinguish between personal devices and devices belonging to someone else. For more details on this we refer to [11].

In this paper we describe a PN architecture that is closely conforming to the IST-MAGNET [12] view of PNs wherein, up to now the focus is on a solution involving special infrastructure support. Whereas this dependency on the infrastructure allows easier implementations of PNs however it makes the deployment slower. This paper discusses the positioning of the essential functions to support PNs in the *network* or *higher layers*. Therefore, we also look beyond, and propose architectural solutions that are oriented more towards the P2P paradigm in nature and can use the existing Internet infrastructure as is. We explain the approaches and possible combinations. We thus form a basis for evaluation of different approaches for discerning readers, network engineers, and service providers.

The rest of the paper is organized as follows. In Section II and III, we describe some related studies and PN architecture. We explain inter-cluster PN network architectures in Section IV. We briefly explain the prototyping efforts in Section V. We briefly discuss our approaches and conclude with our views in Section VI.

II. Some Related Studies

Most technologies focus on a particular aspect of future wireless communication. Here we list proposed solutions that try to meet more of a person's communication needs. In a proposal from the University of Illinois at Urbana-Champaign and the Mobius project [19], they group devices in close vicinity into so called Mobile Grouped Devices (MOPEDs). Each MOPED is connected to a proxy (some kind of home agent) via an infrastructure connection. MOPED is not suitable for PN because it is still too dependent on the proxy and the infrastructure. Furthermore, MOPED does not address direct ad-hoc communication with other person's MOPEDs and is therefore still too limited to support the PN vision.

On the other hand, Contact Networking [3], which is another proposal from the Mobius project, tackles local communication. It provides lightweight, localized network communication to nodes with heterogeneous air interfaces. Link-layer awareness and automatic interface management enables Contact Networking to establish IP connectivity and other crucial networking services with neighboring nodes. The Mobile VCE project has defined a concept called Personal Distributed Environment (PDE) [14]. PDE has a very similar vision to PN, but has no clear network architecture yet.

IXI Mobile [13] has a commercial product around a concept called Personal Mobile Gateway (PMG). It is basically a mobile phone with a WPAN-technology that has been extended to better manage a person's WPAN. PMG-enabled devices can communicate with each other and can also use the PMG-enabled mobile phone to connect to the infrastructure. However, all services are controlled by the operator and all external communication has to go through the operator's networks and this will not be able to meet a user's all future communication needs. We feel that a user should be able to start using his personal devices and home/office network in an efficient way and also without the need for changes in the infrastructure that provides him/her the outside connectivity. In the next section we explain the abstract views of a typical PN and terminology that we use.

Ambient networks [29] aim to achieve edge to edge network control and try to connect heterogeneous networks to provide a homogeneous view of the network to the users of network resources. It tries to connect different kinds of networks such as inter vehicle networks, body area networks, and sensor networks using IP as the basic network protocol while keeping an eye on the mobility requirements. Ambient networks aspires three important features: (a) Composability – by defining a universal framework for building unified communication support using individual networks; (b) Mobility – provides are integrated mobility concept that can act locally (at the cluster level in ad hoc mode) and globally also involving QoS support; (c) Heterogeneity – tries to bring under one umbrella multiple networks of different operators and technologies that bestows a user with the freedom to use the technologies/services that are best suited. It also aims to have a generic link layer (GLL) to abstract the different network layers. While ambient networking is related to PN, it does not address personal and secure connectivity concepts as defined in PN. In a PN, a more person centric view is taken.

III. PN Architecture

A. *The Three Abstraction Level View*

As shown in Figure 2, the IST MAGNET project [9, 12] has proposed a PN architecture, which is composed of three abstraction levels; the connectivity, the network and the service abstraction levels [18]. The connectivity abstraction level consists of various wired and wireless link layer technologies, organized in radio domains, including infrastructure links. The link layer will allow two nodes implementing the same radio technology to communicate if they are within radio range. To allow any two nodes within a PN to communicate, a network abstraction level is needed.

The network level divides the nodes into *personal* and *foreign nodes*, based on trust relationships. Trust relationship is a way nodes can gauge trustworthiness of other nodes with whom they interact. There can be different levels of trust relationships. Only nodes that are able to establish long term (permanent) trust are personal and can be part of a user's PN. personal nodes that are 'nearby' and have such a long term common trust relation form a '*cluster*'. Clusters can communicate with other clusters via infrastructure. The next section will further develop the architectural concepts of the network abstraction level.

The highest level in this architecture is the service abstraction level, which incorporates two types of services; public and private services. Public services are offered to anyone whereas private services are restricted to the owner or trusted persons by means of access control and authentication.

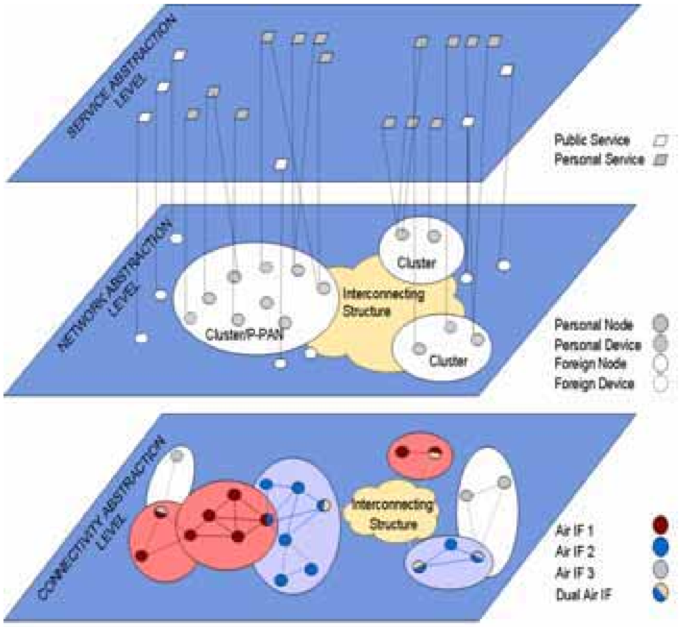

Figure 2. The abstraction level view

B. *Networking Level Architecture*

The network level has to be as independent as possible from the underlying connectivity level so that all current and future wireless communication technologies can be supported. In the Internet, IP was designed to meet this requirement and therefore IP is the underlying protocol for packet transfer also for PNs.

In this architecture [9, 10, 16, 30], the home network of a person will likely a single cluster, the car network another, the PAN of personal nodes around the person (called Private PAN or P-PAN) a third and so on. The link layer technology used to form a cluster will limit the geographical spread and size of a cluster. All clusters work as local networks, therefore need their own independent networking solutions such as self-configuration, self-maintenance, addressing, routing, etc. The formation and maintenance of clusters is a purely local process and does not need any support from infrastructure. Clusters are dynamic in nature. Nodes are switched off and on as well as roam and might suddenly show up in a different cluster. Clusters can split for example, when a person leaves some devices behind and takes some with him similarly, clusters can merge when a person comes back home with his devices. Therefore, the solutions used in all the clusters, including the P-PAN, will be the same so that they can merge and split without extra effort.

An important requirement for cluster formation is the capability to keep foreign nodes out of the clusters and only include personal nodes (see Figure 3). This is done by using a special authentication and authorization mechanism [17, 11]. The cluster formation approach we propose is opportunistic and tries to make the clusters as large as possible. The purpose here is to be able to use intra-cluster mechanisms to provide communication between personal nodes as often as possible, since it is likely to be more efficient than using infrastructure networks.

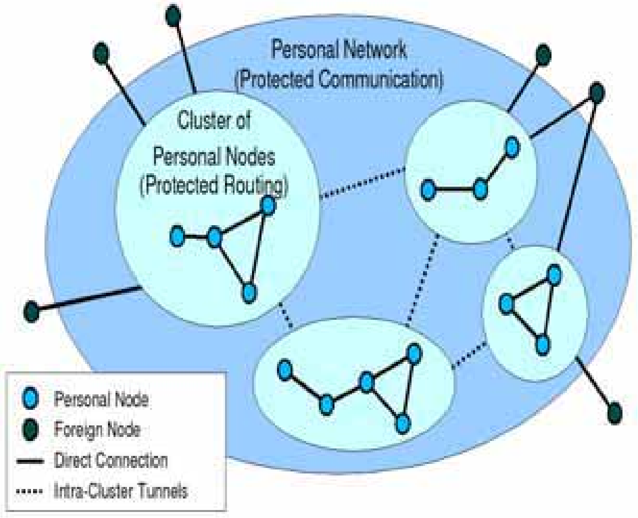
Personal Network (Protected Communication)
Cluster of Personal Nodes (Protected Routing)
Personal Node
Foreign Node
Direct Connection
Intra-Cluster Tunnels

Figure 3. The Network Layer Architecture

Inter-cluster mechanisms involve infrastructure which is not always available and is, in many cases, poorer (in terms of performance, cost, etc.) than direct multi-hop intra-cluster communication. Clusters are defined from the connectivity and trust perspective. Therefore, if there is connectivity, a cluster can be very large with many hops between nodes. However, typically we expect clusters to have a small number of nodes and a limited geographical span, because of the way they will be deployed. To facilitate communication between remote clusters, each cluster will have a special *node* called *gateway* node that can support communication to the other clusters and also to the rest of the outside world. We briefly explain this entity below.

C. *Gateway Nodes*

Gateway (GW) node is a personal node within a cluster that enables connectivity to nodes outside the cluster. GW nodes have links to infrastructure. GW nodes have some special requirements such as address translation, set up and maintenance of tunnels, filtering of incoming traffic, etc. They should preferably be powerful devices since the tasks required by such a node might be quite heavy. The process of finding capable GW nodes with links to foreign nodes or the infrastructure is a joint task of the nodes in a cluster. It is part of the cluster formation process. The selection of GW nodes for a particular data session depends on several aspects not only decided by the cluster but also on the person owning the PN. There can be more than one GW node simultaneously active in a cluster.

After a cluster has discovered which nodes provide connectivity to the infrastructure, it initiates the establishment of tunnels between these GW nodes and the GW nodes of remote clusters. GW nodes are responsible for setup of outgoing or incoming tunnels. A single GW node can also take the sole responsibility in a cluster. We shall not discuss the formation of the secure IP tunnels here. Security related protocols such us IPSec [15, 23] can be used here. More about the security issues are discussed in detail in [11].

When clusters want to communicate with remote clusters through their GW nodes, they first need to locate each other. Further, inter-cluster communication needs to be secure and maintained when clusters merge, split and their nodes roam or are activated/deactivated. This will be accommodated through dynamic tunnel establishment mechanisms. The aim of this is to both facilitate secure inter-cluster communication as well as solve the mobility problem. Each node will have an intra-PN IP address that stays the same as long as the node is part of the PN. Since nodes can roam freely and may shift to another cluster, there is no possibility for hierarchical organization of intra-PN addresses without introducing address changes. If address changes still need to occur within the PN, then the mobility problem is not entirely solved, which is one of the purposes of introducing a flat intra-PN addressing scheme.

IV. SOLUTIONS FOR INTER-CLUSTER COMMUNICATION

There are many solutions to enable inter-cluster communication which is an important task for enabling PNs. We present four different solutions for dynamic tunneling between clusters in the following subsections.

A. PN Agent-Facilitated Inter-Cluster Communication

This is the simplest and most straight forward solution, where the GW nodes of the clusters establish tunnels to a central server, the *PN Agent.* The PN agent is an infrastructure-based management entity that knows the location of all clusters in the PN. The PN agent has a fixed address and can be contacted from any Internet-connected infrastructure network. Each personal node knows the address of the PN agent of the PN it belongs to. This address can, for instance, be distributed during the personalization of the personal nodes. GW nodes, which are always personal nodes, therefore know the address of the PN agent.

In this solution, the GW nodes set up tunnels to the PN agent. Hence, all inter-cluster traffic goes through the PN agent as shown in Figure 4. As soon as a GW node in a cluster detects an infrastructure network with connectivity to the PN agent, it may use it to establish a tunnel to it. As the cluster roams, different access networks may be chosen for the tunnel. As the GW nodes keep at least one tunnel connected to the PN agent at all times, inter-cluster communication will work and the mobility problem is handled. Further, since the tunnels are initiated by the GW nodes, it will work even if the GW nodes are using infrastructure connections with dynamic addressing and NAT, such as from WLAN hotspots.

We should also note that the PN agent can either be centralized under the control of a single provider or operator, or distributed over multiple providers or operators. PN agent functionality can overlap with any other node in a PN and it can be hosted by the user himself if so desired. For instance, the GW node of the home cluster can also act as a PN agent if it has a public IP address.

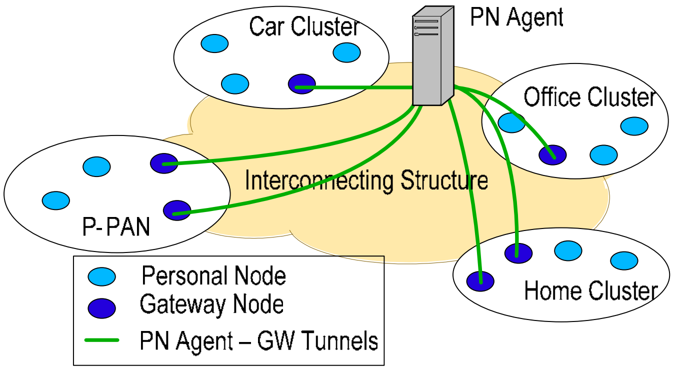

Figure 4. Central server based inter-cluster communication

Furthermore, the PN agent can provide additional functionality. For example, it may assist in other PN-internal mechanisms such as name resolving and service discovery. The PN agent, which is also a proxy agent, can also be used by foreign nodes that wish to communicate with a specific PN. The location of the proxy is the only information foreign nodes need to know in order to communicate with the PN. The proxy will find the appropriate service or node within its PN to connect with. For instance, if someone wishes to establish a voice conversation with the owner of a PN, the proxy will locate a phone-capable node in a cluster near the user and divert the incoming call to that node.

Since this solution is a star topology, the PN agent has to be very powerful – computationally and bandwidth wise as well – which is a major drawback of this solution. Since all inter-cluster traffic has to go through the PN agent, the latency and other QoS characteristics of the inter-cluster communication will be suboptimal. However, this can be a simple and quick solution to realize a PN.

B. Edge Router- Facilitated Inter-Cluster Communication

This section describes how inter-cluster tunnels between the clusters can be established directly and maintained using the concept of *Edge Routers* (ER)[1]. We discuss the ER based

[1] ERs can be seen as functionality and can reside anywhere. The solution considered here assumes it to be implemented at the edge of the providers' network supporting a PN for higher efficiency.

architecture here for the sake of completeness and we refer to [9] for a detailed account. A working prototype has been developed in MAGNET using this architecture. A similar approach can also be found in [28]. Figure 5 shows the PN architecture based on ER with multiple clusters. Each cluster consists of at least one GW node which in turn connects to an ER. The ERs establish tunnels with each other in the Internet and thus enable PN connectivity without sending traffic through the PN agent. Instead, the PN agent is only used to locate the other clusters.

1) *Edge Routers*

Edge routers are endpoints in the interconnecting structure such as the Internet that communicate with GW nodes and support them by offering PN functionality. They are usually managed by a network or service provider. Thus they will probably be owned by the service provider. On behalf of a cluster, an ER can communicate with the PN agent and takes care of the tunnel establishment to the other clusters. ERs can easily make sure that direct tunnels to the ERs of the other clusters are established, even if the number of clusters is large. In this way inter-cluster traffic does not need to go through the PN agent at all. On the other hand, mobility will mean that potentially many tunnels must be updated instead of only one. However, with the support from the ERs, as far as the user is concerned, his PN will be up and running at all times, even if there is no traffic between the clusters.

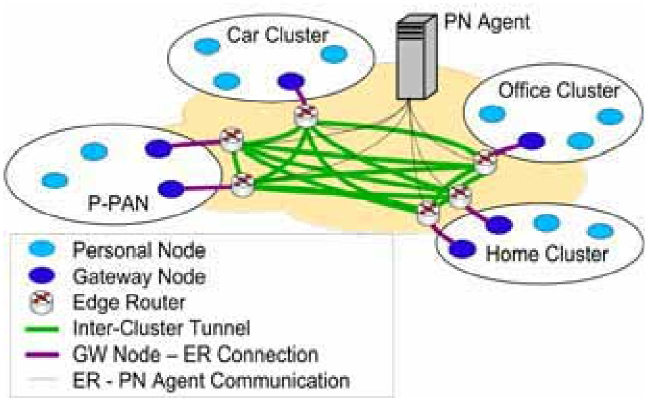

Figure 5. PN connectivity using Edge Routers

ERs being fixed and powerful devices and being in the infrastructure network, can also take other types of loads from GW nodes such as remote service discovery, service repository, etc. However, there are some questions that are raised regarding the concept of having these services at the edge of the infrastructure instead of in the PN itself. Before taking up this dilemma – of moving the functionalities down in the network architecture or up above near the application layer – we shall discuss the reasons for the need for an ER.

If we assume that GW nodes are responsible for the establishment and maintenance of tunnels (for one or multiple PNs) then GW nodes need to build and maintain tunnels all the time. Establishing and maintaining these tunnels consumes resources (processing, battery power, etc.) while GW nodes are usually battery-powered. In addition, a user probably wants to have access to all his/her devices and services at all time with minimum delay and without impact of mobility on ongoing communication. This favors a proactive, always-on, infrastructure supported policy. Further, multiple tunnels (ER non-selective) between two clusters can also improve inter-cluster communication and provide resilience. It is useful for redundancy or fault tolerance. The choice of a specific tunnel can be decided by ERs based on the QoS requirements of the applications and services. Thus the ERs, interacting with the PN agent in the PN framework, may also support QoS management and service access. At the same time, the GW nodes can be lightweight, and almost any PN Node can be able to provide this functionality. The PN setup and maintenance is easy with the help of ERs. Since no overhead is involved in maintaining the connectivity to the network, communication between two nodes in a PN or in fact with nodes in other PNs is simple. Network formation is fast and can easily support session mobility.

ERs also reduce the connection establishment time when a cluster moves. This is enabled by GW nodes connecting to different ERs when they move around. Since the tunnels start and terminate at the ERs, GW nodes need not start a fresh connection establishment process. Only when new ERs are discovered (for instance a new GW node becomes available) or existing ERs become unavailable (for instance due to mobility), this information is propagated and the ERs will dynamically update the tunnels by destroying existing ones and creating new ones.

PN specific applications and services that can operate completely independently of the user, for example a remote backup system between a home cluster and an office cluster, do not require the presence of the user. This implies that PN functionality will also be required even if the user is not present. Consequently, above arguments favor an always-on, ER-based tunnel building and maintenance policy, in which tunnels are established and maintained between ERs as and when a cluster is connected to the interconnecting structure through an ER. In all these cases, an ER having a permanent connection would be useful. Thus ERs provide an overlay network for PNs. This can be harnessed by the nodes of PNs efficiently without bothering about the connectivity with the other nodes. We call this an optimized way of supporting PN connectivity taking into account the nature of PN nodes and their capacities. This is nothing but placing intelligence in the network layer or infrastructure so that higher layers do less complex work regarding network connectivity and use the prowess that can be afforded at the infrastructure layers. An excellent discussion similar to the above can be seen in [28]. The MAGNET architecture proposes to always have tunnels between all the clusters.

The ER concept in infrastructure requires entities that are to be explicitly designed for supporting PNs. They possess the following drawbacks: (1) *ERs need PN functionality:* Public service providers need to build these entities and maintain them. (2) *ERs need to be trusted:* Since, the tunnels are from ERs the security information needs to be shared with them. This may endanger the security of PNs. Furthermore, ERs must trust the PN agents, even when they belong to another operator or the user. (3) *ERs do not reduce the complexity of the architecture:* Its presence does not simplify the architecture it merely shifts the complexity to the network however it reduces connection setup time. (4) *ERs must have public IP addresses:*

It is for the purpose of building direct tunnels amongst themselves. It will enable always-on connectivity. But it would also mean that ERs are to be in the public domain and maintained as such in the public domain. This is a difficult proposition for individual users (without the help of service providers) in the realm of IPv4.

Given the above characteristics it is clear that ERs require a lot of support from the ISPs. It is not evident that ISPs may be willing to offer this support on a large scale and in the short term. Therefore, if we can avoid ERs in the network it will be faster to deploy and easier to operate though *un-optimized* at times. The PN agent does not have this drawback as it can be located anywhere in the infrastructure as it has a fixed public IP address. Moreover a PN agent can support many more PNs if necessary since it is not handling data traffic and thereby increases scalability.

2) *The PN agent in ER-based inter-cluster communication*

In order to offer PN functionality, ERs need to be aware of the location of the other clusters in the PN. This is something that the PN agent can provide the ERs with. However, the ERs, usually, would not know the address of the PN agent before they get a query from a GW node. Thus it is necessary to have a mechanism for the ERs to get this address. However each GW node knows the address of the PN agent of the PN it belongs to and therefore can provide the ER with the address.

This is how the ERs build an overlay network of ERs for a particular PN with the help of PN agents assisting them. In case the PN agent functionality is distributed and under the control of an operator/provider, the operator needs to provide the ERs with the location of the PN agent dynamically (comparable to providing information with DHCP, for instance). In addition, a distributed system requires communication between PN agents across different provider domains, as information needs to be exchanged between the PN agents. Alternatively, a hierarchical system is also possible. The choice of how to implement a secure PN agent concept will mainly depend on the operators' viewpoint on deploying PNs commercially, but in all cases the same functionality needs to be provided.

Clusters that have obtained access to the interconnecting structure announce their presence to a PN agent as shown in Figure 5. More precisely, the ERs send a registration to the PN agent. The registration messages sent by the ERs need to contain at least the following essential information: PN identification, cluster identification and IP addresses of the attachment points of the cluster to the infrastructure (public IP address of the ER). The PN identification is needed since PN agents must be able to check the credentials of the cluster wanting to join a certain PN. This identification must specify which PN the cluster belongs to and which PN it wants to join . It must also prove that it is authorized to join. The cluster identification is needed for tunnel management. Based on this information, the PN agent, who knows the ERs of each cluster, will initiate the process of building the tunnels to the new ER that wants to join the PN. Finally, the IP address of the ER is of course needed, as this will represent the endpoint or starting point of the tunnels. The IP number can be explicitly taken from the ERs or from lower layer connection points. The information contained in the registration messages must be transferred to the PN agents in a secure (information may not be altered by or visible to the outside world) and reliable way (information may not get lost).

When the PN agent receives a registration message, it will verify the identification information and store its information in a secure database. The PN agent will also compare this new information with the already stored information for this PN in its database and will decide if new tunnels have to be created. If so, the PN agent will trigger the establishment of new secure tunnels between the new and existing clusters. The tunnel triggering messages contain all information of the tunnel endpoints.

The end result of the PN establishment through the PN agent and ERs is a PN that consists of multiple remote clusters that are directly interconnected by secure tunnels between ERs, where the central PN agent has the complete view of the PN topology. This PN can be seen as an always-on virtual network sitting above the ER-based overlay network, with its own solutions to naming, addressing, and routing and mobility management. All of these operate on top of existing infrastructure, with IP as the common base.

C. *Gateway Node-Facilitated Inter-cluster Communication*

From the discussion in the previous section it is clear that the concept of having ERs in the providers' network requires a lot of support from the ISPs. It is not evident that ISPs may be willing to offer this support on a large scale and in the short term. Therefore, if we can avoid ERs in the architecture it will be faster to implement and easier to operate.

In this section we explore an alternative PN architecture without ERs. Since the GW node now needs to be available to the other clusters, it will need to be a little more intelligent. Furthermore, good amounts of support need to be sought from PN agent(s) when the ERs are no longer in the picture. In the following subsections, we list the most important differences.

1) *Changes in the GW nodes*

In the previous solution, the ERs are to provide a seamless overlay network on behalf of the PN clusters. Now the GW nodes are to be the inter-cluster tunnel endpoints instead. In this modified solution, the architecture of a PN is as shown in Figure 6. There are two ways a GW node can connect to the interconnecting structures, either using a public IP address or through a NAT [1, 5, 22].

A) GW nodes on public IP numbers

If the GW nodes have publicly routable addresses, they would be able to form the route with the help of the PN agent. The PN agent is updated by each GW node with their location and care of address (CoA). The PN agent therefore knows the address of all GW nodes with infrastructure access. Lightweight IPSec tunnels can be established between the GW nodes to transfer the intra-PN communication over the interconnecting structures. The tunnels are secured using shared secrets distributed during the personalization step of the GW nodes [11]. The connection between two GW nodes can be handled in two different ways:

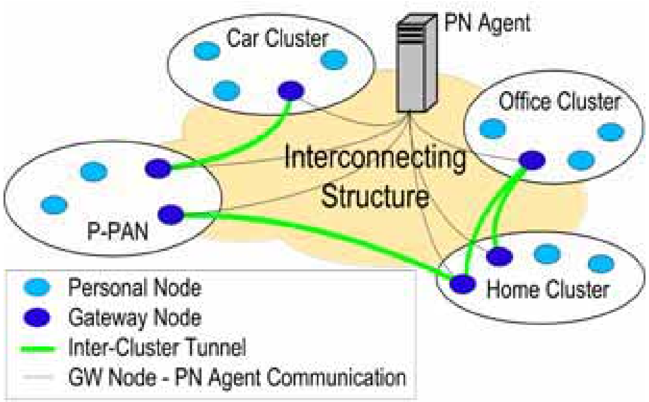

Figure 6. PN formation with GW node (without ERs)

(1). Always on inter-cluster tunnels: Here the GW nodes initiate the tunnels with help of the PN agent and build a quasi permanent connection with all present GW nodes in the PN and keep these tunnels intact as long as possible. When the attachment point of a GW node changes due to mobility or other reasons, the GW node must not only inform the PN agent about the new CoA, but also all other GW nodes in the PN. Upon receiving such an update, the tunnel can be diverted to the new CoA. The process is a bit similar to the bind updates in Mobile IPv6 [31]. If the tunnels are implemented using IPSec in tunnel mode, the tunnel will survive this type of update of one endpoint. The idea here is to maintain the tunnels between all GW nodes for as much as possible so that the delay in building new communication channels between them is minimal.

(2). On-demand inter-cluster tunnels: In this case, every GW node keeps a connection with the PN agent and updates it with its location. Tunnels between GW nodes are only established when needed. This means that a GW node only needs to update other GW nodes to which it has active tunnels with its new CoA. Except for this, it will work exactly the same as the always on inter-cluster tunnels.

The tunnels are created above the IP layer and at IP layer the tunnels do not mean anything as IP packets are not identified as a flow at IP layer. The tunnels are actually formed by authorizing the GW nodes with each other using the shared trust relationship. Therefore, if GW nodes can keep track of the trust relationships with each other, then they can communicate with each other as and when required. The basic idea here is to use lightweight tunnels between the GW nodes and do away with ERs.

In the always on inter-cluster tunneling solution, a normal ad hoc routing protocol can run unmodified over the inter-cluster tunnels as they are always up and look like normal links to the routing protocol. However, in the on-demand inter-cluster tunneling, this is not possible. Each GW node should know which cluster to establish a tunnel to before establishing that tunnel. The easiest solution to this problem is to let the PN agent also know the member nodes of each cluster. A GW node can then consult the PN agent before establishing a new tunnel. The PN agent will be able to tell which cluster and which GW node can be used to establish a tunnel to the cluster of the destination node.

To send a lot of updates to various nodes, including PN agents and GW nodes, will require extra bandwidth from the used access technology. In always-on inter-cluster tunneling, there may be many CoA updates to transmit, while in the on-demand inter-cluster tunneling, also cluster member node updates must be transmitted over the infrastructure interface. However, the trend in wireless infrastructure-based access technologies is clear; with UMTS and WLAN, we are increasing the available bandwidth. The future should be able to offer us even more bandwidth and at a lower price. We therefore believe this problem will grow away.

B) Gateway Nodes behind NAT

When GW nodes are not on publicly routable IP, the PN agent has to take care of establishing the connectivity between them. Popular methods such as STUN [20] and TURN[21] may be used here for NAT traversal. These methods have the least overhead and are transparent to the IP layer. Therefore it is worthwhile to look into this scheme. However, there will be some difficulties if and when the NAT is symmetric/restrictive on either side. Then the PN agent can be used to relay the connection between the GW nodes and the transactions are still going to be secure. If this is happening often, the PN agent may need to be powerful and have a good network connection. Several strategically placed PN agents may be a good option.

There are other solutions for these problems without the security at the network layer, however they are using the concepts of super nodes, for example, overlay P2P networks such as CAN, Pastry, etc. A popular example of such an application is Skype. In this case the PN agents are actively involved. Here, the PN agents will be sharing the connections for many other PNs and in turn security may be at stake. In this situation, GW nodes themselves have to find a way to secure their transactions by moving the security aspects above the IP layer using protocols such as IPSec etc.

2) Functions of the PN agent without ER

The functions of the PN agent in this scenario consist of all the functions explained in Section IV.B.2. The additional functions include handling the trust relationships for the GW nodes. A database for holding the addresses of GW nodes, and thus the clusters of a PN is also needed. A PN agent has to switch the packets in case two GW nodes could not communicate directly with each other. In this scenario the PN agent will have to route the packets through. The PN agent might have to run STUN servers etc., for P2P connectivity when required. This means that they need to be more powerful in order to enable more traffic to go through it.

The PN agent can be a centralized system maintained by the user himself. It can also be maintained by a service provider. PN agent functionality may be distributed. It can be a Reliable Pool of Servers (ReSerPool [24, 25]). Here the choices of PN agent broaden so as to attract the (application) service providers to start thinking in terms of PN concept. In this sense, the PN agent will be handy and can be implemented by anyone without expecting a change in network support entities. However concerning security here, we will have to take care of it at the higher layers.

D. *Complete Peer-to-Peer Inter-Cluster Communication*

In the absence of a PN Agent, the cluster needs to take an even more active role in building the tunnels for communication between them. The GW node would be keeping track of the various Care-of Addresses (CoA) of other clusters and GW nodes. The idea here is to remember the last connected point or CoA of these Clusters. The assumptions are:

1. Some nodes and clusters will almost never move.
2. Not all the GW nodes will change their point of contact or CoA simultaneously.

With these pragmatic assumptions, one can easily see that it is possible to communicate with the other clusters/GW nodes, since at least one of the addresses is the same as the one at the previous instance. This unchanged GW node can act as the starting point for refreshing the point of contact of other GW nodes. The only condition is that all the nodes/clusters will start their operation together from the same location so that they can bootstrap by sharing the current CoAs. When they move (not all), they keep updating their new address to the GW node of the cluster which does not move frequently (like the home cluster in Fig. 7). Each GW node stores the new CoAs of all the other clusters to maximize the possibility of successfully re-connecting later.

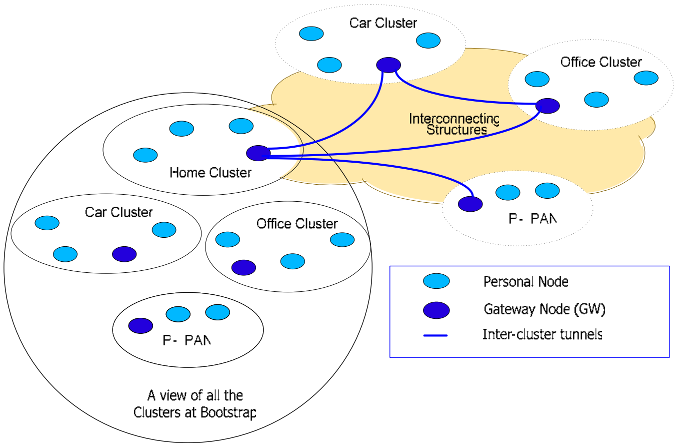

Figure 7. Inter-cluster tunnels without PNs and ERs

In case of the assumptions being not valid any more, the clusters have to be re-bootstrapped. Since all clusters have different CoAs, the only way is to manually give them the CoAs to all the other clusters or to re-bootstrap the PN again as above. The idea here is to distribute the task of the PN agent to all the clusters and thereby updating their CoAs. In this sense if more and more nodes cache the CoA of other nodes then re-bootstrapping can be avoided. This approach helps in distributing the PN functionality and avoids dependency on some server/entity in the infrastructure.

V. Prototyping

A prototype of a PN has been developed that can handle all the necessary intra-PN network level mechanisms. The prototype is built around Linux 2.6 and IPv6 using Ethertap virtual interfaces [27]. The PN network level is realized on each of the node as a virtual interface. Any packet sent to the virtual interface is intra-PN and uses intra-PN IPv6 addresses. Special software handles the packets as they arrive on the virtual interface and adds security and other intra-PN information before sending them to a physical interface for transmission to the next hop. The same software listens for intra-PN packets from other personal nodes on each physical interface. It uses the intra-PN routing table to forward them to the next hop if it is not the final destination. Otherwise, it removes security and other intra-PN headers before passing the packets to the intra-PN applications listening on the virtual interface. In this way, applications can remain unmodified and use both intra-PN communication as well as normal IPv6 communication at the same time.

At this moment, we are currently extending this prototype also to enable inter-cluster communication. A process on each gateway node will listen to both the virtual interfaces and thereby take part in the intra-PN communication and at the same time monitor its infrastructure-based physical interfaces for potential connections to the infrastructure. When such connections are found, this process will coordinate with the PN agent and establish tunnels to the other GW nodes when needed according to the GW node-facilitated inter-cluster communication solution discussed earlier in this paper.

VI. Discussions and Conclusions

MAGNET [9] has up to now mainly focused on the ER-based and with infrastructure support for the sake of improving performance but it assumes infrastructure support. There are some advantages in having the alternate solutions without ERs. Without ERs, a user does not need to have any support from the infrastructure. A PN agent can be run on any public IP address so that it can be customized according to the users requirements. In fact, users should be able to have and maintain their own servers if they wish to. The spin off of this technique is that the service providers need not run special services near to each cluster thereby avoiding deployment investments. No modification is necessary to the infrastructure since the intelligence is in the higher layers. Thus many of the existing Internet 'Presence' service providers can popularize the concept of the PN resulting in a wide reach. Connecting to another PN or foreign nodes would be very easy if the service is offered by well known 'presence' enablers when there is no common radio link. The architecture without ERs can attract application service providers like Yahoo, Skype, etc. It has an outlook similar to that of P2P solutions – in the sense that GW nodes talk to each other directly at PN level – and thus can be universal in nature. It is nothing but having the 'presence' of our personal devices/nodes similar to what we have in the Internet at present.

We can also use a mix of many of the combinations of the solutions presented in Section IV. We believe that multiple techniques can co-exist, for example, we can use the solution in Section IV.D when the PN agent is temporarily unavailable along with the other solutions. Another aspect is to select a solution that is best suited for the situation. When one can not find ER functionality in the infrastructure one may use an alternate solution, for example as in IV.D. The idea is to have more flexibility such that one can use the best possible combination of the approaches.

In this paper, we have described various techniques for enabling inter-cluster communications for PNs. We have provided solutions that do not require the support of network service providers. While we can achieve better services sometimes with infrastructure support, we also think it is necessary to look into the other ways of sustaining the PNs without the support of service providers. We surmise that having many alternate solutions can help speedy deployment and testing of PN concept. We also think that if the popular 'presence' applications can support PN concepts it would have far reaching effects. The next logical step is to scrutinize the security aspects of all the solutions in detail though we do not expect surprises.

Acknowledgment

The authors were partially supported by the IST MAGNET and IST MAGNET Beyond research projects during this work, as well as Freeband PNP2008 research projects. The authors would like to thank all the partners from these projects for their fruitful discussions on the topic of Personal Networks and helping us to crystallize our ideas.

References

[1] Aboba, B. and W. Dixon, "IPsec-Network Address Translation (NAT) Compatibility Requirements", *IETF RFC 3715*, 2004.

[2] Bluetooth SIG, "*Specification of the Bluetooth System – Version 1.1 B*", http://www.bluetooth.com/, 2001.

[3] Casey Carter, Robin Kravets, Jean Tourrilhes, "Contact Networking: A Localized Mobility System", *MobiSys 2003*, San Francisco, USA, May, 2003.

[4] Committee on Research Horizons in Networking, Computer Science and Telecommunications Board, National Research Council, Free Executive Summary, *Looking Over the Fence at Networks: A Neighbor's View of Networking Research*, ISBN: 0-309-07613-7, 2001.

[5] Huttunen, A., Swander, B., Volpe, V., DiBurro, L. and M. Stenberg, "UDP Encapsulation of IPsec ESP Packets", *IETF RFC 3948*, January, 2005.

[6] IEEE P802.11 - The Working Group for WLAN Standards, http://www.ieee802.org/11/.

[7] IEEE 802.15 - Working Group for WPAN, http://www.ieee802.org/15/.

[8] Ignas G. M. M. Niemegeers, Sonia M. Heemstra de Groot, "Research Issues in Ad-Hoc Distributed Personal Networking", *Wireless Personal Communications: An International Journal*, Volume 26, Issue 2-3, Pages 149-167, Kluwer Academic Publishers, August, 2003.

[9] IST-507102 MAGNET/WP2.1/INT/D2.1.2/R/PU/001/1.0, "*Overall secure PN architecture* ", January, 2005.

[10] IST-507102 MAGNET/WP2.4/IMEC/D2.4.1/PU/001/1.0, "*Architectures and Protocols for Ad-Hoc Self-configuration, Interworking, Routing and Mobility*", December, 2004.

[11] IST-507102 MAGNET/WP4.3/UNIS/D4.3.2/PU/1.0, "*Final Architecture of the Network-Level Security Architecture Specification*", March, 2005.

[12] IST MAGNET, *http://www.ist-magnet.org/*.

[13] IXI Mobile, *http://www.ixi.com/*.

[14] John Dunlop, R.C. Atkinson, James M. Irvine, D. Pearce, "A Personal Distributed Environment for Future Mobile Systems", *In IST Mobile and Wireless Communication Summit*, Aveiro, Portugal, June, 2003.

[15] Kent, S., and R. Atkinson, "Security Architecture for the Internet Protocol", *IETF RFC 2401*, November, 1998.

[16] Martin Jacobsson, Jeroen Hoebeke, Sonia Heemstra de Groot, Anthony Lo, Ingrid Moerman, Ignas Niemegeers, "A Network Layer Architecture for Personal Networks", In the Proceedings of the Workshop on "*My Personal Adaptive Global Net: Visions and beyond*", Shanghai, China, November, 2004.

[17] Martin Jacobsson, Ignas Niemegeers, "Privacy and Anonymity in Personal Networks", In *Proc. International Workshop on Pervasive Computing and Communication Security (PerSec'05)*, Kauai Island, Hawaii, USA, March, 2005.

[18] Martin Jacobsson, Jeroen Hoebeke, Sonia Heemstra de Groot, et al., "A Network Architecture for Personal Networks", *14th IST Mobile and Wireless Communications Summit*, Dresden, Germany, June, 2005.

[19] Robin Kravets, Casey Carter, and Luiz Magalhaes, "A Cooperative Approach to User Mobility", *ACM Computer Communications Review*, Volume 31, Pages 57-69, October, 2001.

[20] Rosenberg, J., Weinberger, J., Huitema, C. and R. Mahy, "STUN - Simple Traversal of User Datagram Protocol (UDP) Through Network Address Translators (NATs)", *IETF RFC 3489*, March, 2003.

[21] Rosenberg, J., "Traversal Using Relay NAT (TURN)", *Internet-Draft draft-rosenberg-midcom-turn-06*, October, 2004.

[22] Srisuresh, P. and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)", *IETF RFC 3022*, January, 2001.

[23] Thayer, R., Doraswamy, N. and R. Glenn, "IP Security Document Roadmap", RFC 2411, November, 1998.

[24] M. Tuexen, et al, Architecture for Reliable Server Pooling, Internet draft, *draft-ietf-rserpool-arch-09.txt*, February, 2005.

[25] Tuexen, M., Xie, Q., Stewart, R., Shore, M., Ong, L., Loughney, J. and M. Stillman, "Requirements for Reliable Server Pooling", *IETF RFC 3237*, January, 2002.

[26] ZigBee Alliance, http://www.zigbee.org/

[27] Ethertap, http://vtun.sourceforge.net/tun/

[28] W. Louati and D. Zeghlache, "Network based Virtual Personal Overlay Networks using Programmable Virtual Routers", *IEEE Communications Magazine*, 86-94, July 2005.

[29] N. Niebert, A. Schieder, H. Abramowicz, G. Malmgren, J. Sachs, U. Horn, C. Prehofer, H. Karl, "Ambient Networks - An Architecture for Communication Networks Beyond 3G", In IEEE Wireless Communications, 11(2): 14-22, April, 2004.

[30] R.V. Prasad, M. Jacobsson, S. Heemstra de Groot, A. Lo, I. Niemegeers, "Architectures for intra-personal network communication", In the 3rd ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots (WMASH'05), Cologne, Germany, Sep. 2, 2005.

[31] D. Johnson, C. Perkins, J. Arkko, "Mobility Support in IPv6", *IETF RFC 3775*, http://www.ietf.org/rfc/ rfc3775.txt, June, 2004.