International Journal of Multimedia and Ubiquitous Engineering
Vol. 2, No. 1, January, 2007

Authentication Protocol for Heterogeneous Networks

Maricel O. Balitanas and Rosslin Robles

Western Visayas College of Science and Technology
jhe_c1756@yahoo.com, rosslin_john@yahoo.com

Abstract

Heterogeneous Networking poses many challenges in several areas. The expansion of link and MAC layers has been accompanied by contracting in the network layer as a core networks use IP packets to facilitate key service including telephony, data and multimedia. In this paper we discuss the Heterogeneous Networks, presented a reference model and scenario and applied an authentication mechanism to this context

Keywords: *Heterogeneous Network, Authentication, WLAN*

1. Introduction

Heterogenous Networking and the devices themselves as hetnet devices. Heterogeneous networking poses many challenges in several areas. At the lowest levels, many new access technologies including 3G, WiMax and UltraWideBand (UWB) will be supported on hetnet devices. This expansion of the link and MAC layers has been accompanied by a contraction in the network layer as core networks use IP packets to facilitate key services including telephony, data, and multimedia.

One of the major capabilities of heterogeneous networking is that of handover. This is necessary because the networks that are currently available and/or their points of attachment may be changing as a mobile node changes its location. Another key issue that needs to be addressed because of heterogenous networking is that of Quality of Service (QoS) in peripheral networks. This is because different wireless networks have varying QoS so vertical handover not only affects the point of attachment but also the QoS of the link as seen by other entities higher up the protocol stack. This in turn affects the ability of the network and transport services to deliver effective performance since these systemsmust respond to changes in QoS in the available channels. At the higher layers how these QoS changes are dealt with and what can be done by the system to minimize their effects on applications must also be examined. Alternatively, future applications will be able, with the help of the system, to structure themselves to make use of QoS changes in wireless networks. In addition, new kinds of application environments which can facilitate new features such as personalisation, personal area networks and location-based services will be built. In order to address the issues above, the authors believe that there needs to be a framework that encapsulates the key challenges of heterogeneous networking for mobile devices. Like a map clearly helps one to plan a journey, a framework is needed to help us move forward in this unexplored area. The approach taken here is similar to the OSI model in which layers are used to specify functionality, allowing a modular approach to the extension of systems and the interchange of their components such that different implementations can be easily explored. The authors are implementing a prototype testbed to explore the proposed architecture.

In then recent years the word heterogeneity has been wide applied and from diverse perspectives in the publications and works on information and communication technologies topics. Researches efforts have been dealt towards maintain the heterogeneous networking transparent and hidden under the powerful all-IP concept. By means of that, technologies of different nature, such as satellite communications, backbone internet, wired and wireless LANs, and cellular networks could cooperate together replying with a wide spectrum of solutions and added-value services to the business model demands. Consequently, this fact has an impact on the development growth of multi-mode terminals capable to inter-work to a variety of access networks. Over that landscape, our work pays special attention to the recent initiative carried out by the 3G Partnership Project in order to specify the 3G cellular system to wireless LAN inter-working [1], and more specifically to the security aspects [2]. The considerations to deploy this case of heterogeneous networking could be summarized in:

a) Since WLAN terminals (STA) are being massively integrated in normal-life and in a diversity of environments, they are far for being considered exclusively for sophisticated users. Decrease of the devices size and power consumption in contrast to the increase of the computation capacity, improvement of security protocol performance [3, 4] and multi-mode radio links capabilities [5].

b) Combining both technologies [6], coverage and data rates suffer an important growth comparing to the capacity of the next future cellular access networks. As target, the WLAN is seen as a complement to the 3G systems to deliver enhanced value to the end-user. Obviously, roaming solutions, vertical and horizontal handoffs among others require special considerations. The known security gaps, mentioned above in this document, could be the weakness aspect of this solution, therefore is object of several studies and part of this work review them.

c) Business model are pending to be well-defined in order to include in this scenario the Wireless Internet Service Provider, WISP functionalities. New revenue from access and services could be disputed between the three parties: WLAN operator, WISP and cellular operator.

d) A global standardization process [1] supports this initiative integrating the interests of vendors, operators, manufacturers and service providers. This in-progress standard proposal could be considered as a step in the roadmap to the 4G systems. Firstly, our work proposes a novel model reference and an example of applicability for secure electronic payment in heterogeneous networking environment. The detailed description of such scenario is presented in the next section. Afterwards, section 3 is devoted to the impact of the new participant entities on the Trust Model and consequently to trust relations. In order to guarantee the security levels in section 4 a set of authentication requirements are derived. In section 5 a new authentication protocols approach is introduced in order to complain such requirements and includes network architecture proposed in our study.

2. Architecture

2.1 Conceptual Framework

As outlined in the previous section, the OSI model does not act as a good conceptual framework for today's networks. A key point is that this does not invalidate a layered approach, but rather indicates that it must be updated. Such an approach must specify the hierarchy of functionalities, but not prescribe detailed interfaces between them. The ordering of the hierarchy is perhaps the most important consideration; this is evidenced by the difficulties that have been experienced in placing, for example, vertical handover functionality above the transport layer. We propose an architecture that, like the OSI model, has seven layers, but that uses a novel hierarchy of functionalities (Figure 1). In our model, vertical handover, with input in turn from policy management, is placed below the network transport layer. Similarly, quality of service is given its own specific layer to separate it from both the application and the network transport modules. This contrasts with current approaches where QoS appears to be more of an "add-on" rather than an integral part of the network stack. It is also important to realize that the layering paradigm does not restrict implementers to rigid modularization. A framework is a useful concept, rather than a detailed design specification. Hence, whilst our model details the necessary ordering of functionality, it does not discount the possibility of cross layer approaches. Indeed, as detailed in Section 4, policy management and vertical handover functionality might well be integrated into a single component. Also, higher layers may well need context not only from the layers directly beneath them in the hierarchy, but also those further down. This can be seen to be the case for the QoS layer, where hints from the network abstraction layer would be of great utility. Hence, we emphasize the need for clear conceptual separation, whilst leaving open the possibilities for vertical integration and trans-layer interfacing in terms of implementation. In the following sections we proceed to describe each of these separate conceptual layers in turn, using a bottom-up approach.

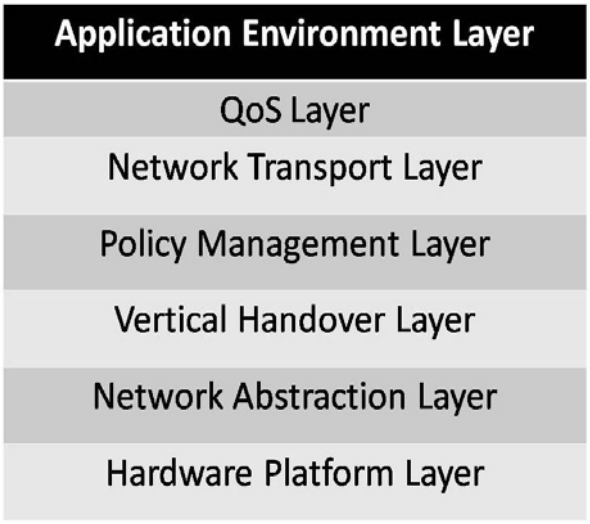

Figure 1. Architecture Model

2.2 Hardware Platform Layer

This layer is used to define the hardware components and technologies required to support wireless networks. This layer defines several characteristics, for example, the electromagnetic spectrum required for a given technology, the different modulation techniques that may be used, as well as the MAC algorithms for acquiring and reserving channels. It is recognized that individual systems may be totally incompatible with each other. Hence, the layer is composed of vertical sub-layers, with each sub layer representing a particular network, e.g. 3G, WLAN, WiMax, etc.

2.3 Network Abstraction Layer

The Network Abstraction Layer is used to support different networking technologies using a common interface. Different wireless device drivers will eventually be written to map onto this layer. The network abstraction layer has to do with controlling and maintaining the network on the mobile node. Recently, the IEEE convened the 802.211 working group to examine the possibility of standardizing the interface to different wireless MACs. This work is relevant to the network abstraction layer of our architecture.

2.4 Vertical Handover Layer

Vertical handover is clearly a key component of heterogeneous networking [23]. There are two distinct approaches to vertical handover. The first is the network-controlled approach in which the network decides when and how the handover will occur. This means that there are mechanisms in the network that maintain all relevant information on the mobile host, including its location relative to different networks, their signal strengths at the mobile node's location, and its direction and speed. However, we do not believe this approach is scalable. As new wireless networks are added, the information about all the networks to which the host is currently attached becomes difficult to maintain. Such an approach also relies on operators sharing detailed information about their networks, a concept that is unlikely in the current cellular telecommunications environment. The second approach is termed client-based handover. With this approach, handover is controlled by the mobile device.

There are clear advantages to using this approach. Firstly the mobile node will keep up to date information about its network interfaces and therefore is in a superior position to decide when handover should occur. Secondly, the mobile node will also be aware of the state of its TCP connections and other higher level issues, and therefore these factors can be taken into account when making the decision to conduct a vertical handover. Finally, by removing handover functions from the core network, it should then be more cost effective to build increasingly robust core systems.

2.5 Policy Management Layer

A policy management system is needed to evaluate all the circumstances when handover should occur, taking into account various factors such as changes in coverage and signal strength, the state of the network, and the state of any transport connections associated with the mobile host. There are two types of policy management which are being explored: reactive and proactive. A reactive policy depends entirely on notification from the network abstraction layer about the presence or absence of networks as the hetnet device changes its location. Almost all the policy management systems so far reported in the literature (e.g. POLIMAND [25], including the most recent, PROTON[24], have been reactive. Proactive

policies attempt to acquire and use information about the likely coverage and signal strength before the mobile actually reaches a given location. A key parameter sought from proactive policies is the Time Before Vertical Handover or TBVH. Knowing the TBVH allows the higher layers of the protocol stack to make maximum use of channels that may soon be unavailable. Proactive policies can be divided into two types: A proactive-knowledge-based policy mechanism makes explicit use of knowledge of the mobile node's location and knowledge about the types of coverage and signal strengths of each network that the hetnet device is likely to encounter at that location. These systems require enhanced location detection and monitoring routines as well as a mechanism for ensuring data on what networks are available at a given location is accurate and easily accessible. Such data must be available in a format that allow it to be processed by hosts with limited resources, and transmitted over potentially low bandwidth networks without significant impact. A few systems [19][20] have utilized data collected from previous journeys. Another approach is a proactive-modeling approach in which a mathematical model is used to determine the TBVH based on simple geometric calculations. Though less accurate than the proactive knowledge approach, this approach is flexible and can be used in simulations as well as in real networks. There is also a growing need to combine proactive and reactive policies, with a view to developing an architecture where it is possible to choose which policy to use in a given situation. Hence, when there is accurate information then a proactive policy may be used. However, when there is no coverage data or the data is unreliable, then the system can fall back to a reactive policy mechanism.

2.6 Network Transport Layer

This layer concerns functions that would normally be assigned to the network and transport layers of the OSI model. Hence this layer examines addressing, routing and transport issues in peripheral networks. The current opinion is that all networks whether in the core or on the periphery should be using TCP/IP. This thinking has been reinforced by End-to-End arguments which have been used throughout the architectural discussions when the Internet was designed [22]. The current evolution of the Internet questions some of these End-to-End arguments. As indicated previously, today the Internet is evolving into a very fast core network with mobile networks on the periphery. This means that characteristics of the core and the periphery are diverging in terms of latency, throughput and error profiles. In the light of this, the assumption that TCP/IP should be used in peripheral networks for heterogeneous networking needs to be carefully re-examined. Firstly we should question whether IP should be used in such networks. An assumption of the current IP infrastructure is that every machine should have a globally unique IP address to use on the network. This has, however, been directly challenged by the success of Network Address Translation (NAT) techniques. In NAT, a private IP address space is employed in the peripheral network while only a few global IP addresses are used to actually communicate with other machines on the Internet. The NAT software then provides the translation between the global IP connection and the local machine with its private IP address. Because all datagram must traverse the server performing the NAT, it provides a point in the network where incoming packets can be analyzed and filtered as necessary. In addition, it increases security by not making local machines visible on the Internet, thus reducing the potential for targeted security flaw exploits, and DoS attacks against specific machines. The success of NAT – which is considered to be far from ideal by network purists, including the authors – questions the assumption that all machines should be assigned a globally unique IP address. NAT makes the case that IP addresses should be confined to moving data within the core network. In the

peripheral networks some other form of local addressing may be used with the translation between the networks taking place at the local gateway. Such an approach, if deployed, will question the (often challenged) assumption that there are insufficient IPv4 addresses. Though the authors support the deployment of IPv6, the key requirement for its deployment – to provide an infinitely large global address space – needs to be re-examined in the light of new realities.

2.6.1 TCP – FoundWanting : It is clear that TCP is unsuitable for wireless networks [23][24] This is primarily due to the fact that TCP interprets packet loss as exclusively due to congestion and reacts by substantially decreasing its send rate, and then employing its slow start mechanism. While such a conclusion may be valid for wired systems such as the core network, peripheral wireless systems continuously lose packets due to channel fading, interference, vertical handovers, and other related effects. Most of these transients are temporary and unrelated to network congestion. There have been several attempts to modify TCP in the light of these effects, such as those described in [18][19]. Recently, there has been a move towards not modifying the TCP protocol engine but making it more responsive to temporary network outages [19]. While this is useful, a clear, generally applicable, and elegant solution has not been found.

2.7 The Case for Network Plurality and Application Conformity

The idea that a different networking infrastructure runs in peripheral networks brings with it many challenges. Most importantly, the ability to translate to different naming and addressing schemes as packets are transmitted through different networks. Some of issues are addressed in Plutarch [17] by the introduction of contexts and interstitial functions. The key requirement is the development of a framework where different networks interwork to the benefit of the applications running on those different networks. However, in order to ensure that every application developed using TCP/IP will not be required to be rewritten, the issue of application conformity must be addressed. An interesting solution is to pursue the idea of TCP/IP not only being a real protocol suite in the core network but also a protocol interface in peripheral networks. This means that on end user devices in these networks, the TCP protocol engine forms an overlay above whatever networking protocol is actually being used on the network. Hence, applications can assume TCP behaviour whilst the actual protocol in use need not be TCP. We believe that this approach should allow network plurality to emerge whilst maintaining application conformance. Though application conformancemay be necessary in the short and medium terms, in the long term we believe that the we need to move to a situation where an application's transport requirements are not specified by the selection of a specific protocol, but by the definition of a QoS vector which specifies the transport requirements of the application, as suggested in [20].

2.8 Quality of Service Layer

Because the hetnet device will be using multiple networks, it will inevitably experience different qualities of service [21].Therefore, a QoS layer is required to manage this variability. To cover this appropriately we now introduce the concepts of downward and upward QoS.

2.8.1 Downward QoS : The intention is to construct a series of mechanisms to handle the different qualities of service that are encountered in vertical handovers in heterogeneous

networks. In order achieve this a mechanism is needed that bundles connections over the different available channels, which themselves may be varying. We term this Downward QoS, which will be required to support legacy applications.

2.8.2 Upward QoS It is hoped that future applications will be able to react to QoS changes as the hetnet device changes its location. These changes will be provided via the QoS layer. This is termed Upward QoS. In this case, an event-based QoS signalling mechanism can be used to inform applications of QoS changes. Applications, when they begin executing, would be able specify routines that should be called in response to event notifications by the QoS layer. This is similar to the X Windows System [24] in which clients of the X-Server can specify what routines should be called on being notified about certain events concerning windows they have created on the screen. In the networking community, the TRIGTRAN [22] and PILC [23] projects illustrate this paradigm, by ower layers providing hints to higher layers. 3.8.3 QoS-AwareMiddleware A hugely useful component in the support of Upward QoS is the development of a QoS-aware middleware platform for Distributed Environments. The concept is to take a well known environment such as CORBA and add support for Upward QoS capability natively to the architecture. This would allow distributed applications to work seamlessly over heterogeneousmobile networks.

2.9 The Application Environments Layer

A Toolkit Approach Some application environments attempt to encapsulate several architectural layers [23]. These systems are therefore so large that compatibility with other systems is not really considered important. This is an issue if we desire true connectivity between all devices and applications.

An alternative approach is to adopt a Toolkit philosophy in which the principal goal is not to build a particular application environment but to specify the components of a toolkit which would aid different groups in building application environments. Hence there would be a degree of compatibility which would also encapsulate the functionality of the lower layers of the architecture.

3. The OSI Model

The most well-known framework in data communications is the OSI model [25] developed by the CCITT. This is regarded as a reference model, and not an implementation plan, but it clearly delineates the functions of the layers to provide a framework for exchanging data between networked applications. The authors believe that the OSI model cannot be considered the dominant model for heterogeneous networking. It is not wrong; it is simply inadequate for heterogeneous networking for several reasons. Firstly in the OSI model, the first three layers, the physical, link and network layers are concerned with the movement of packets between networks. The higher layers of the OSI model – the transport, session, presentation and application layers – are designed to deal with end-to-end issues between application processes. Such a horizontal separation between networking and End-to-End Issues is no longer sustainable in heterogeneous networks. There are several reasons for this. In the OSI model, the network is essentially used to simply forward packets to their relevant destination. However, in heterogeneous networking, there are new functions that the network must also support.

One of these is vertical handover. The point here is that vertical handover requires frequent and intimate communication between the mobile host and the network which cannot be simply incorporated into the OSI model. To go further, vertical handover also involves the reconfiguration of certain parameters in the network, such as allowing for the reservation of resources to ensure quality of service in the receiving network. Such a reality is difficult to model in the OSI context. The other major observation is that the OSI model works well when the characteristics of the networks at the edge are not overly dissimilar from the core network. The early Internet had Ethernet or Token ring systems which were basically wired networks capable of several megabits per second between endpoints. However, looking at network trends since that time, the core network and end systems have taken different evolutionary paths. The core network is being made faster with the use of technologies such as MPLS and single-mode fibre optics, while the peripheral systems are rapidly becoming dominated by the emergence of wireless technologies that have very different characteristics in terms of latency, bandwidth, and availability and error distribution properties. In the light of these observations, the authors believe a new framework is required to better reflect how tomorrow heterogeneous networks should be constructs. This new model is described in the following sections.

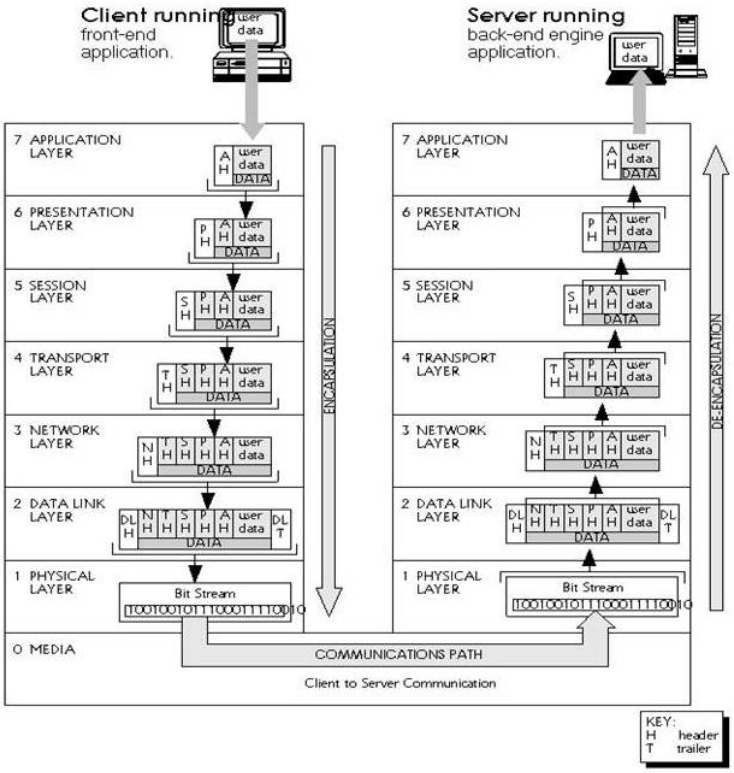

Figure 2. OSI Model

At one time, most vendors agreed to support OSI in one form or another, but OSI was too loosely defined and proprietary standards were too entrenched. Except for the OSI-compliant X.400 and X.500 e-mail and directory standards, which are widely used, what was

once thought to become the universal communications standard now serves as the teaching model for all other protocols.

Control is passed from one layer to the next, starting at the application layer in one station, and proceeding to the bottom layer, over the channel to the next station and back up the hierarchy.

Physical layer defines the cable or physical medium itself, e.g., thinnet, thicknet, unshielded twisted pairs (UTP). All media are functionally equivalent. The main difference is in convenience and cost of installation and maintenance. Converters from one media to another operate at this level.

Data Link layer defines the format of data on the network. A network data frame, aka packet, includes checksum, source and destination address, and data. The largest packet that can be sent through a data link layer defines the Maximum Transmission Unit (MTU). The data link layer handles the physical and logical connections to the packet's destination, using a network interface. A host connected to an Ethernet would have an Ethernet interface to handle connections to the outside world, and a loopback interface to send packets to it.

Ethernet addresses a host using a unique, 48-bit address called its Ethernet address or Media Access Control (MAC) address. MAC addresses are usually represented as six colon-separated pairs of hex digits, e.g., 8:0:20:11:ac:85. This number is unique and is associated with a particular Ethernet device. Hosts with multiple network interfaces should use the same MAC address on each. The data link layer's protocol-specific header specifies the MAC address of the packet's source and destination. When a packet is sent to all hosts (broadcast), a special MAC address (ff:ff:ff:ff:ff:ff) is used

NFS uses Internetwork Protocol (IP) as its network layer interface. IP is responsible for routing, directing datagrams from one network to another. The network layer may have to break large datagrams, larger than MTU, into smaller packets and host receiving the packet will have to reassemble the fragmented datagram. The Internetwork Protocol identifies each host with a 32-bit IP address. IP addresses are written as four dot-separated decimal numbers between 0 and 255, e.g., 129.79.16.40. The leading 1-3 bytes of the IP identify the network and the remaining bytes identifies the host on that network. The network portion of the IP is assigned by InterNIC Registration Services, under the contract to the National Science Foundation, and the host portion of the IP is assigned by the local network administrators. For large sites, the first two bytes represents the network portion of the IP, and the third and fourth bytes identify the subnet and host respectively. Even though IP packets are addressed using IP addresses, hardware addresses must be used to actually transport data from one host to another. The Address Resolution Protocol (ARP) is used to map the IP address to it hardware address.

Transport layer subdivides user-buffer into network-buffer sized datagrams and enforces desired transmission control. Two transport protocols, Transmission Control Protocol (TCP) and User Datagram Protocol (UDP), sits at the transport layer. Reliability and speed are the primary difference between these two protocols. TCP establishes connections between two hosts on the network through 'sockets' which are determined by the IP address and port number. TCP keeps track of the packet delivery order and the packets that must be resent. Maintaining this information for each connection makes TCP a stateful protocol. UDP on the other hand provides a low overhead transmission service, but with less error checking. NFS is

built on top of UDP because of its speed and statelessness. Statelessness simplifies the crash recovery.

The session protocol defines the format of the data sent over the connections. The NFS uses the Remote Procedure Call (RPC) for its session protocol. RPC may be built on either TCP or UDP. Login sessions uses TCP whereas NFS and broadcast use UDP.

External Data Representation (XDR) sits at the presentation level. It converts local representation of data to its canonical form and vice versa. The canonical uses a standard byte ordering and structure packing convention, independent of the host provides network services to the end-users. Mail, ftp, telnet, DNS, NIS, NFS are examples of network applications.

4. Reference Model

Our work takes as start point the reference model defined in [1] and more concretely the scenario1 3 with non-roaming features. Afterwards it redefines an enhanced model adding a new entity with authentication capabilities in user-side, here named Supplicant Equipment (SE). In our reference model in Fig. 1, SE exclusively interacts with the system through the WLAN User Equipment (WLAN-UE) by means of the Reference Point At, RAt. The User Equipment might exchange messages with the SE over the RAt in order to initiate the session and gain access to the rest of the system, establish data traffic transmission/reception and finalize the session. Details of these low-level messages are out the scope of this work. The 3GPP AAA server retrieves authentication information and subscriber profile (including subscriber's authorization information) from the HLR/HSS. Afterwards, it communicates authorization information to the WLAN and Packet Data Gateway, PDG. It should inform to PDG about the authorized W-APN, necessary keying material for tunnel establishment and user data traffics.

The WLAN routing functionalities address packets through PDG and vice versa, providing WLAN UE with 3G Packet Switched based services. External IP network services are accessed via PDG by means of a authorisation process, service selection (e.g. Wireless-Access Point Name selection, W-APN access point for a concrete service) and subscription checking. PDG functionalities includes IP address resolution and allocation.

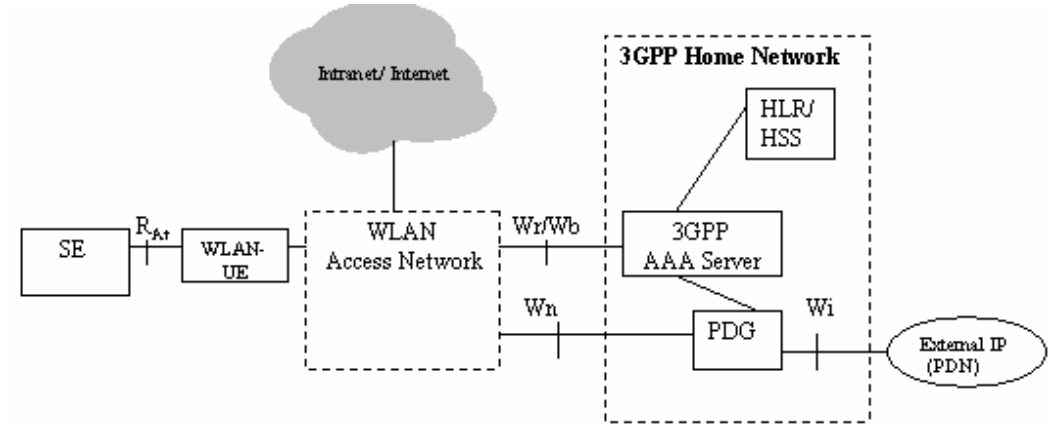

Figure 3. Reference Model [26]

Some of the envisaged functions for the WLAN-UE are

i) associating to an IWLAN,

ii) WLAN access authentication based on EAP methods,

iii) Building an appropriate NAI, iv)obtain a local IP address. And specifically for the scenario 3 v) building an appropriate W-APN, vi) request the resolution of a W-APN to a PDG address and establish a secure tunnel to the PDG, vii) obtain a remote IP address in order to access services provided in the operators PS domain. From our perspective, and such as is mentioned the set of WLAN-UE functions must include the provision to SE of relay access to the rest of the system.

5. Scenario

A smart card featured as a credit/debit card (assuming the presence of a Cardholder) plays the role of Supplicant Equipment, SE. The referred WLAN-UE could be materialized by a Wireless Point of Sale, WPOS, featured as a mobile internet device and provisioned with a SIM/USIM registered with the HLR at the 3G Home PLMN. In our context, the stated 3GPPP AAA server (e.g. RADIUS Server) should be under the Payment Operator control, since it is the responsible entity for authenticating the WPOS in its domain. Agreements between payment operator and cellular operator should be envisaged. This aspect tightly concerns to the conception of the business models. Although the WPOS is located at merchant facilities, the security of the payment procedure is payment operator's responsibility; therefore Access Point (AP) and a potential WLAN AAA Proxy should remain transparent in the end-to-end Authentication process. However, the need of a payment card in this realistic scenario introduces complexity. Thus, for security reasons we propose that the WPOS should be remain transparent in the complete payment process, playing the role as tunneled entity avoiding risks of potential attacks from a manipulated point of sale terminal. This novel perspective is clearly different of the current envisaged process for smart cards payments [7]. The performance for electronic payment through a WPOS could be derived in the following summarized manner:

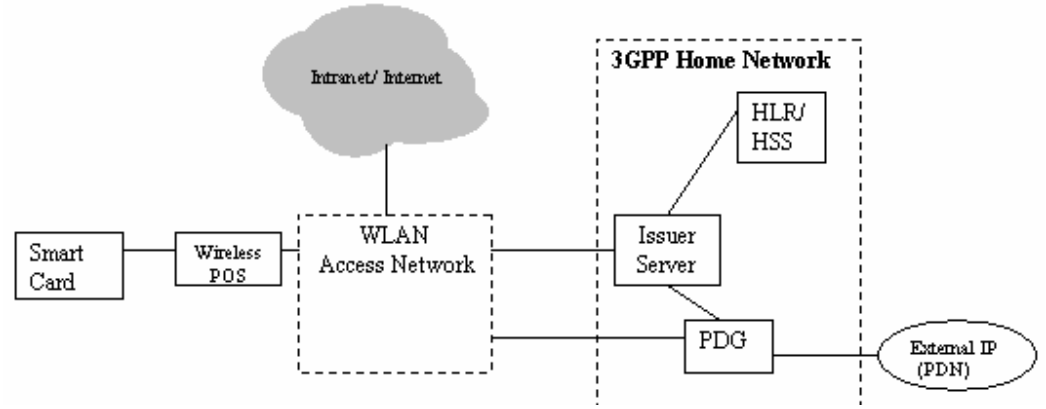

Figure 4. Scenario [26]

Once the Cardholder inserts the debit/credit smart card in the card reader, the Cardholder Client System integrated in the WPOS initialises the chip card and the smart card application, then it communicates with them in order to obtain a cryptogram from the smart card for validation. In the course of communicating with the chip card several exchanges occur. The chip card may request that the Cardholder enter a PIN. If so then the Cardholder Client System notifies the Cardholder using the display and activates or connects, if necessary, the PIN pad or key entry device to permit and capture PIN entry. The application on the chip card returns one or several cryptograms for validation. While sending messages (commands) and receiving responses from the card in the previous steps, the Cardholder Client System was gathering the information necessary to compose the Authentication Response which it now sends to the Issuer Server (issuer bank domain). Finally, the Issuer Server validates cryptogram sent from Cardholder Client System software. In our opinion, at the glance of this performance, the inclusion of improved authentication procedures should be required. Our work proposes partially redefine the functionalities of these entities in this address, in order to secure the payment in our scenario. Therefore, once the WPOS device is authenticated by payment operator then it is secure to accept a payment card. As mentioned, the smart card exploit the Supplicant functionality in a generic authentication context. Thus, to perform the payment the smart card by means of the credentials inside referred to the cardholder should be authenticated directly by the issuer bank server and vice versa. In our proposed scenario, the 3GPP AAA server is represented by such Issuer Server. The envisaged functions for the smart card are:

i) store and execute the application,
ii) communicate with Cardholder Client System software (WPOS),
iii) validate cardholder PIN and,
iv) generate and return cryptogram upon request.

On the other side, the functions provided by Cardholder Client System should be:

i) communicate with issuer server, with smart card and with Cardholder for Card insertion/removal and PIN entry,
ii) exchange of authentication messages both smart card and issuer server, and iii) relay of the cryptogram from smart card to the issuer server.

The Issuer Server (authentication server) functions could be summarized:

i) securely stores keys needed for cryptogram validation,
ii) collect the necessary data to perform smart card authentication processing and initiate this procedure,
iii) validate cryptogram sent from Cardholder Client System software at Hardware Security Module (HSM).

6. Authentication Requirements

In order to make effective the reviewed Trust Model over the proposed scenario, a set of correct authentication mechanisms are required, as described before, enforcing secure payment considering the smart card (SE) and Issuer Server (AS) the extremes of an end-to-end authentication procedure, minimizing the potential attacks from hosts in the transaction

route. Based on this goal, our work detected a lack of authentication requirements to be suited to this scenario, in order to obtain major security guaranties.

First Requirement: end-to-end security over this heterogeneous architecture with multiple nature hosts, capable to be origin or bridge of attacks, attempting against the security guarantee of the whole system. As is claimed, smart card and Authentication Server and should be the end-entities in this authentication scheme.

Second Requirement: mutual authentication once mechanisms for session key establishment have been provided. An attack in this scenario could be seen from both smart card perspective and the Authentication Server one. For instance, the server could be supplanted in order to obtain critical information from the smart card, therefore the server should be correctly authenticated by the smart card, representing to the cardholder. On the other side, the smart card could be used in a fraudulent manner by a malicious user. The repercussion of that in bank card is more than relevant. The mechanisms for session keys establishment should guarantee the freshness of them.

Third Requirement: Protected and flexible cipher-suite negotiation. This could be considered as a general purpose requirement since is defined to cover the different protocol implementations/versions and the end-to-end participant capabilities. But more specifically in our work, this requirement is introduced in order to adjust the cipher-suite complexity or security degree depending on the payment usage context (e.g. wired/wireless, micropayment, mobility context, etc.)

Fourth Requirement: scalable key management capabilities. Obviously, the approach proposed is an direct and strong impact on the server capability in order to attend requests coming from an important number of the on-line smart cards.

7. Authentication Protocol Approach

The conception of this inter-working system is addressed towards an all-IP environment, therefore and as is stated in [1] the mechanisms in order to guarantee the security must deployed when feasible along the whole protocol stack taking into account both signalling and data traffic planes. These mechanisms must be irrespective of others implemented in upper or lower layers. Our research is focused to the layer 2 authentication, but not preclude additional secure mechanisms adopted at IP level (where mobile-IP could be considered, from the WLAN-UE perspective) or transport level, remaining out the scope of this work. Motivation on this focus is briefly stated:

- security strengthening, irrespective of adopted solutions in upper/lower layers.

- to guarantee a strong authentication in absence of IP connectivity. Current limitations in smart cards show them as enough powerful devices capable to perform a strong authentication at lower layers but insufficient to perform protocols at IP and upper layers.

- payment transaction session (messages exhange) is relatively single from the payment card point of view. Likely solutions at IP level are not necessary (other financial application such as SW downloading, customisations, etc. should be desirable in the next future [10]).

- an implementation of an authentication infrastructure [11] with efficient key distribution mechanisms [12] would be easily well-suited to our model. Wireless LAN roaming aspects will be issues to be included in further work.

- important initiatives based on EAP methods have been covered by IETF, IEEE and interested parties.

Far from a detailed analysis of the potential specific protocols to be implemented, our work aims to feature a new layer 2 authentication protocols approach on the basis of the reference model in figure 1 and as result a novel protocol stack architecture is proposed and depicted in figure 4. One of the approaches of this work is to exploit the smart card capabilities as embedded device in terms of lower layer protocol implementations and secure storage, highlighting the differences with the conventional magnetic band card to perform electronic payments. A better performance including IP-like techniques in the next generation smart cards is foreseeable. As mentioned before, our work foresees an autonomous authentication smart card procedure before the cardholder enters the PIN. Therefore, could be considered as an embedded device authentication based on the credentials stored in the card. The user authentication is completed with something that he/she knows. It is not in the scope of our work to define a specific EAP-type for the architecture in fig. 4, but establish the authentication relationship between the participant entities. However, for our new approach, the recent standardization initiative [13] for EAP support in smart card should be considered. Its goal is defining universal ISO 7816interface, supporting most of EAP authentication protocols. Components such as a logical network interface that directly processes EAP messages and allows EAP profiles (types) definition, an operating system interface to provide identity management, storage of cryptographic material, and management interface are envisaged. As result of this effort, in [14] an specific type, EAP-SC defines an EAP method and multiplexing model for the support of smart card based authentication methods. EAP-SC provides for encapsulation of other EAP methods such as EAPTLS among others.

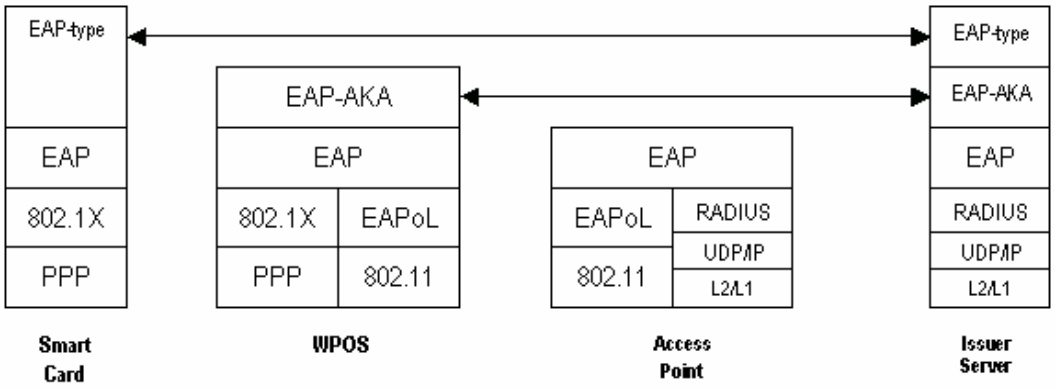

Figure 5. Protocol Architecture [26]

From the point of view of the authentication method, the wireless POS behaves as standard WLAN User Equipment. In particular, the protocols EAP-SIM [14] and EAP-AKA [15] should be considered. Although these protocols have been included in the 3GPP

specifications, weakness are identified in [15, 16] and improvements should be investigated. In [17,18] some signalling UMTS/GPRS messages are embeeded within EAP messages facilitating fast handover. In [9] the authors propose to combine the well-suited to this heterogeneous environment performance of EAPAKA with strong authentication of TLS protocol.

8. Conclusion

A novel Reference model and scenario was presented in this paper and applied an authentication mechanism that was presented to be efficient. The impact in the trust relations has been asses and a set of authentication requirements has been provided in order to guarantee the security process.

References

[1] 3GPP TS 23.234 v6.2.0: 3GPP system to Wireless Local Area Network (WLAN) Interworking; System Description, September 2004
[2] 3GPP TS 33.234 v6.3.0; 3GPP system to Wireless Local Area Network (WLAN) Interworking Security System, December 2004
[3] Argyroudis, P.G., Verma, R., Tewari, H., O'Mahony, D.: Performance Analysis of Cryptographic Protocols on Handheld Devices. NCA 2004: 169-174
[4] Gupta, V., and Gupta, S.: Experiments in wireless internet security. Proc. IEEE Wireless Communications and Networking Conference, WCNC, March 2002, Vol. 1, 859-863
[5]IST Project Brain. Broadband Radio Access for IP based Networks (IST-1999-10050), 2001
[6] IST Project Wine. Wireless Internet Networks. (IST-1999-10028), 2001
[7] 3D –Secure Functional Specification, Chip Card Specification v1.0, Visa Corp., August 2001
[8] IEEE Standard for Information technology, 802.11i-2004 Amendment to IEEE Std 802.11i/D7.0: Wireless Medium Access Control (MAC) and physical layer (PHY) specifications: Specification for Enhanced Security, 2004
[9] Kamborakis, G., Rouskas, A., Hormentzas, G., Gritzalis, S.: Advanced SSL/TLS-based authentication for secure WLAN-3G interworking, IEEE Proc. Communications, Vol. 152, No. 5 October 2004
[10] Lagosanto, L.: Java Card evolution toward service-oriented architectures Fifth e-Smart Conference, Sophia Antipolis, France, September 2004
[11] Aust, S., Fikouras, N., Protel, D., Gorg, C., and Pampu, C. (2003). Policy based mobile ip handoff decision (polimand). Technical Report draft-iponair-dna-polimand- 00.txt, Work in Progress, IETF.
[12] Balakrishnan, H., Padmanabhan, V. N., Seshan, S., and Katz, R. (1997). A comparison of mechanisms for improving TCP performance over wireless links. IEEE/ACM Trans. On Networking, 5(6):756–769.
[13] Chakravorty, R., Cartwright, J., and Pratt, I. (2002). Practical experience with TCP over GPRS. In Proc. IEEE GLOBECOMM.
[14] Chakravorty, R., Vidales, P., Subramanian, K., Pratt, I., and Crowcroft, J. (2004). Performance issues with vertical handovers – experiences from GPRS cellular and WLAN hot-spots integration. In Proc. IEEE PerCom.
[15] Chandra, D., Harris, R., and Shenoy, N. (2003). TCP performance for future IP-based wireless networks. In Proc. IASTED Wireless and Optical Communications Conference.
[16] Cottingham, D. N. and Vidales, P. (2005). Is latency the real enemy in next generation networks? In Proc. ICST CoNWiN.
[17] Crowcroft, J., Hand, S., Mortier, R., Roscoe, T., and Warfield, A. (2003). Plutarch: an argument for network pluralism. SIGCOMM Comput. Commun. Rev., 33(4):258–266.
[18] Dawkins, S. (2003). End-to-end, implicit 'link-up' notification. Technical Report draft-dawkins-trigtranlinkup01 (work in progress), IETF.
[20] Karn, P. (2003). Advice for internet subnetwork designers. Technical Report draft-ietf-pilc-link-design-15 work in progress), IETF.
[21] Laasonen, K., Raento, M., and Toivonen, H. (2004). Adaptive on-device location recognition. LNCS 3001, pages 288–304.
[22] Mapp, G. and Hodges, S. J. (1997). QoS-based transport. In Proc. IFIP Workshop on Quality of Service.
[23] McNair, J. and Zhu, F. (2004). Vertical Handoffs in Fourth- Generation Multinetwork Environments. IEEE Wireless Communications, 11(5).
[24] Meyer, M. (1999). TCP performance over GPRS. In Proc. IEEE WCNC, pages 1242–1252.

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

[26] J. Torres, A Izquierdo, A. Ribagorda, and A. Alcaide, “Secure Electronic Payments in Heterogeneous Networking”, Computational Science and Its Applications-ICCSA 2005.