Bilhanan Silverajan (Editor)

Pervasive Networks and Connectivity
Seminar Series on Special Topics in Networking, Spring 2008
The papers comprising this volume were undertaken by students in the course TLT 2656: Special Topics in Networking. This annual course deals with topical, pertinent issues addressing current advances in networking. In Spring 2008, the theme of this course was Pervasive Networks and Connectivity. The theme was chosen for its timeliness and relevance to current and future directions in network research.

By very definition, pervasive networks are all-encompassing environments. These networks target ambitious goals, including a) possessing enough intelligence to allow not only humans but even devices to actively interact in the network as full citizens, b) providing networking and application support for challenging terrestrial and non-terrestrial environments, and c) incorporating connectivity support for wide-scale user, terminal and vehicular mobility. The consequent extent of research seems endless.

The resulting presentations and reports in the course represented only a fraction of work in this very interesting and popular area. Nevertheless the scope was broad enough that all the presentations catered for good discussion and unique perspectives. The final reports have been structured into five main categories to reflect this scope:

- Part 1 Pervasive Connectivity: Users and Providers
- Part 2 Near Field Communications and Ubiquitous Interaction
- Part 3 Trends in Overlay Networking
- Part 4 Delay Tolerant Networking
- Part 5 Location and Service Discovery

The course attracted students from diverse backgrounds which included Ph.D, M.Sc and International Exchange programmes. Some students were employed as full time research personnel in academia or the industry. Several chose to present work that was in their area of specialty.

We had two invited speakers. I firstly wish to thank Mr Kari Vääranen, Business Development Advisor at Technopolis Ventures Professia for his talk on Wireless Tampere. I would also like to thank Mr Juha Miettinen, Director of the Ubiquitous Computing Cluster Programme at Hermia Technology Centre, for his talk on the Finnish UBI ecosystem and future directions for national Finnish ICT R&D.

Finally I give my deepest thanks to all the course participants who maintained a good level of enthusiasm in this course and had actively presented, discussed their work, diligently performed peer reviews and prepared their final reports based on peer feedback. It was a pleasure for me to serve as the supervisor for this seminar course.

Bilhanan Silverajan
Department of Communications Engineering
Tampere University of Technology
Table of Contents

Part 1
Pervasive Connectivity: Users and Providers

High Speed Network Connectivity for Homes and Metropolitan Areas
Esa Lähteenmäki

2

Wi-Fi and WiMAX-based Mesh Networking
Piotr Mitoraj

8

Community Networks
Karri Huhtanen

16

Vehicular Ad-hoc Networks (VANETs)
Jakub Jakubiak

24

IP Multimedia Subsystem (IMS)
Daniel Brouquet

33

Roaming, Handover and Security Issues with Wireless Operator Networks
Antti Mäkelä

41

User Centricity in Pervasive Networks
Sari Kinnari

47

Part 2
Near Field Communications (NFC) and Ubiquitous Interaction

The Internet of Things
Jani Lahti

58

Pervasive RFID
Erkki Vaaja

65

Security, Trust and Privacy in Pervasive Networking
Israa Frej

72

Smart Phone Interactivity and Pervasive Connectivity
Carlo Iorio

81

Wireless Personal Area Networks: ZigBee
Asif Bhatti

90

Wireless Personal Area Networks: 6LowPAN
Vesa Salo

96

Part 3
Trends in Overlay Networking

Reachability via Overlay Networks and Tunnelling
Tuure Vartiainen

102

Content Networking and IP Mobility
Florent Becattini

108
Peer-to-Peer Communication in Pervasive Networks
Jani Peltotalo

Social Networking and Virtual Communities
Pasa Maharjan

Part 4
Delay Tolerance versus Seamlessness

Intermittent Connectivity and Delay Tolerance
Vincent Battaglia

Connectivity in Challenged Environments
Fausto Grossi

Seamlessness and Vertical Handovers
Antti Leinonen

Part 5
Location and Service Discovery

Localization in short-range wireless networks
Ville Kaseva

Naming and Service Discovery Protocols for Pervasive Networks
Jukka Suhonen

Device and Service Discovery in Home Networks
Petros Belimpasakis
PART 1

PERVASIVE CONNECTIVITY : USERS AND PROVIDERS
HIGH SPEED NETWORK CONNECTIVITY FOR HOMES AND METROPOLITAN AREAS

Esa LÄHTEENMÄKI

Summary

The purpose of this paper is to give the reader an overview of available technologies for high speed connectivity in homes and metropolitan areas. The technologies presented in this paper include Asymmetric Digital Subscriber Line (ADSL), Very high speed Digital Subscriber Line (VDSL), Data Over Cable Service Interface Specifications (DOCSIS), Fiber To The Premises (FTTP), Broadband over Power Lines (BPL), Wireless Fidelity (Wi-Fi) and finally Wireless Interoperability for Microwave Access (WiMAX). Bluetooth is also briefly introduced as its next version will be very interesting.

1. Introduction

There are many different ways to provide high speed network access for the customer. These days the most dominant technology for providing broadband access via wires is ADSL with DOCSIS coming far behind to second place [1]. BPL on the other hand is still a new technology and has not yet gained a lot of popularity. FTTP is coming more attractive due to fiber’s huge speed compared to traditional telephone wiring. Building costs of a fiber connection are huge and therefore Internet Service Providers (ISP) are limiting their efforts to mostly metropolitan areas where they have many customers close to each other.

In the wireless world, Bluetooth provides low power consumption communication for a really short range while WiMAX offers long range communication. Wi-Fi falls in the middle of these two, covering a bigger range than Bluetooth, but yet considerably shorter than WiMAX.

The purpose of this paper is to give the reader an overview of technologies present in this field and also provide some information about the different versions and development of these technologies.

2. Digital Subscriber Line

This chapter will introduce the reader some of the most common technologies that provide broadband access via telephone wiring.

2.1 Asymmetric Digital Subscriber Line

ADSL is the most commonly used technology these days for providing broadband access for customers. ADSL standards are the work of International Telecommunication Union (ITU) [2]. This technology offers broadband access via telephone wires up to 5-6 kilometres. This is the maximum distance between users ADSL-modem and ISP’s Digital Subscriber Line Access Multiplexer (DSLAM), where the connection will work. The transmission speed achieved by this technology depends on the quality of the wiring and the distance of the wire. There are currently three main versions of this technology, ADSL, ADSL2 and ADSL2+, which is the newest version. The first version of ADSL used a frequency band of 1.1 MHz and it was divided into different sections for upstream and downstream. Upstream utilizes 25.875 kHz – 138 kHz band and downstream utilizes 138 kHz – 1104 kHz band. These frequency bands offered a maximum upstream speed of 1 Mb/s and a
maximum downstream speed of 8 Mb/s. Keep in mind that these were the maximum speeds obtainable and were achievable only at very short distances. It is also worth mentioning that ADSL technology does not interrupt or disturb a normal telephone conversation in any way. This coexisting of voice and data is possible because voice uses around 0-4 kHz frequency and data uses much higher frequencies. There are many additions to ADSL standard, which provide small upgrades to various aspects of the basic version. These upgrades can for example offer slightly bigger downstream or upstream speeds and provide a little longer maximum operating range. All of the existing ADSL standards and additions are show in table 1.

<table>
<thead>
<tr>
<th>Standard name</th>
<th>Common name</th>
<th>Downstream rate</th>
<th>Upstream rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANSI T1.413-1998 Issue 2</td>
<td>ADSL</td>
<td>8 Mb/s</td>
<td>1.0 Mb/s</td>
</tr>
<tr>
<td>ITU G.992.1</td>
<td>ADSL (G.DMT)</td>
<td>12 Mb/s</td>
<td>1.3 Mb/s</td>
</tr>
<tr>
<td>ITU G.992.1 Annex A</td>
<td>ADSL over POTS</td>
<td>12 Mb/s</td>
<td>1.3 Mb/s</td>
</tr>
<tr>
<td>ITU G.992.1 Annex B</td>
<td>ADSL over ISDN</td>
<td>12 Mb/s</td>
<td>1.8 Mb/s</td>
</tr>
<tr>
<td>ITU G.992.2</td>
<td>ADSL Lite (G.Lite)</td>
<td>4.0 Mb/s</td>
<td>0.5 Mb/s</td>
</tr>
<tr>
<td>ITU G.992.3/4</td>
<td>ADSL2</td>
<td>12 Mb/s</td>
<td>1.0 Mb/s</td>
</tr>
<tr>
<td>ITU G.992.3/4 Annex J</td>
<td>ADSL2</td>
<td>12 Mb/s</td>
<td>3.5 Mb/s</td>
</tr>
<tr>
<td>ITU G.992.3/4 Annex L</td>
<td>RE-ADSL2</td>
<td>5 Mb/s</td>
<td>0.8 Mb/s</td>
</tr>
<tr>
<td>ITU G.992.5</td>
<td>ADSL2+</td>
<td>24 Mb/s</td>
<td>1.0 Mb/s</td>
</tr>
<tr>
<td>ITU G.992.5 Annex L</td>
<td>RE-ADSL2+</td>
<td>24 Mb/s</td>
<td>1.0 Mb/s</td>
</tr>
<tr>
<td>ITU G.992.5 Annex M</td>
<td>ADSL2+M</td>
<td>24 Mb/s</td>
<td>3.5 Mb/s</td>
</tr>
</tbody>
</table>

Table 1  ADLS standards and additions with corresponding speeds [3]

2.2 Asymmetric Digital Subscriber Line 2+

ADSL2+ is the newest version of ADSL used nowadays. This version doubles the original frequency band of ADSL to a total of 2.2 MHz. The maximum achievable speeds are 1 Mb/s for upstream and 24 Mb/s for downstream. As you can see, the increase in used frequency band only affects downstream and upstream still uses the same narrow frequency band. These top speeds can only be obtained at a maximum of 1.5 kilometers distances.

2.3 Very high speed Digital Subscriber Line

VDSL has been developed based on ADSL standards. The main goal has been to increase transmission capabilities of ADSL. The original version was only for short distances, but there is a newer and better version already standardized, which is called VDSL2.

2.4 Very high speed Digital Subscriber Line 2

VDSL2 utilizes up to 30 MHz of frequency band. This drastic difference compared to ADSL means a huge increase in transmission speed. VDSL2 can achieve up to a 100 Mb/s symmetric speed with distance under 500 meters. Symmetric speed means that upstream and downstream can both get the same top speed of 100 Mb/s at the same time. Unlike the first version of VDSL, VDSL2 offers long range communication. Speed up to 50 Mb/s can still be achieved at 1 kilometer distances and performance is equal to ADSL2+ with distances over 1.6 kilometers.

3. Data Over Cable Service Interface Specifications

LÄHTEENMÄKI, E: High speed network connectivity for homes and metropolitan areas
DOCSIS is a standard developed by CableLabs with contribution from many other companies. DOCSIS has been approved as an international standard by the International Telecommunications Union Telecommunications Standardization Sector (ITU-T). This standard defines how communication is done via cable television system. Cable television system was originally a one-way only, so upgrades were needed to get the upstream direction available in the same infrastructure. In Europe we are using EuroDOCSIS versions, which are a little bit different from the original DOCSIS versions. This is because the television broadcasts in Europe are different from what the Americans use. Also in Japan, they are using their own modified versions of DOCSIS.

There are three main versions of DOCSIS (and EuroDOCSIS) called 1.x, 2.0 and 3.0. The newest version 3.0 introduced channel bonding to further increase the transmission capabilities of this system. In table 2, are all the maximum synchronization speeds with maximum usable speed mentioned in brackets.

<table>
<thead>
<tr>
<th>Version</th>
<th>DOCSIS</th>
<th>EuroDOCSIS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Downstream</td>
<td>Upstream</td>
</tr>
<tr>
<td>1.x</td>
<td>42.88 (38) Mb/s</td>
<td>10.24 (9) Mb/s</td>
</tr>
<tr>
<td>2.0</td>
<td>42.88 (38) Mb/s</td>
<td>30.72 (27) Mb/s</td>
</tr>
<tr>
<td>3.0</td>
<td>+171.52 (+152) Mb/s</td>
<td>+122.88 (+108) Mb/s</td>
</tr>
<tr>
<td>3.0</td>
<td>+343.04 (+304) Mb/s</td>
<td>+122.88 (+108) Mb/s</td>
</tr>
</tbody>
</table>

Table 2  DOCSIS and EuroDOCSIS transmission speeds [4]

FTTP

FTTP means that the ISP provides a fiber connection to the basement of the building you are living in. This is expensive work and therefore fiber connections will first be provided to apartment buildings, where the ISP is likely to receive many customers from. The distance from the basement to the actual customer’s equipment is covered by already existing wiring, either telephone wires or cable television wires. Different technologies will be used depending on the media available, for example ADSL or VDSL over telephone wires.

4.1 Fibers

There are two types of fibers available, singlemode fiber and multimode fiber [5]. Multimode fiber is cheaper, but because it has a wider core, it offers slower transmission speed than singlemode fiber. The wider core provides many different paths for the light signal to travel and this result in a need to use slower transmission speeds. The singlemode fiber has a small core, where there exists only one path for the light beam to travel through. Therefore the travel time of the signal is always constant and we can put more data through the fiber. Due to the capacity of singlemode fiber, it is suitable for backbone networks where speed is very important. Multimode fiber can be used inside a building for example where the speed is not as big issue as in backbone networks.

Modulation technique used is Wavelength Division Multiplexing (WDM). Researchers in Lucent Technologies have found that a single fiber can carry at least 150 Tb/s using Dense WDM (DWDM) [6]. This technology allows sending many colors of light (wavelengths) at the same time, where each color corresponds to a separate information signal. So the limiting factor is not the fiber itself, but the
equipment at both ends of the fiber. These days the equipment available outside laboratories can support at least 2.56 Tb/s for distances up to 1000 km [7].

5. BPL

Power Line Communication (PLC) involves using electric wires in telecommunications. There are three distinct types of electric wires: high voltage transmission lines, medium voltage distribution lines and low voltage building wires. The whole system was originally designed for transmission of AC power. Therefore the power wire circuits have a limited ability to carry higher frequencies. Signal propagation is a problem when considering PLC. Data rates achieved depend on the type of platform being used. Only a few hundred bits can be achieved in transmission lines, but inside buildings even megabits per second is possible. Using this method could eliminate the need to build dedicated network cabling inside a building.

There are some efforts concentrated on BPL these days. Basically the idea is similar to DSL and DOCSIS standards; customer could plug a modem into any outlet in the house and gain high speed Internet access. This system has many big problems, main one being that the power lines are a very noisy environment. Connection speeds obtainable with BPL are starting to be very attractive, up to 400 Mb/s can be achieved these days [8]. There are companies around the world who are already providing BPL access to customers. The biggest deployment country seems to be the United States at the moment.

6. Wi-Fi

The Institute of Electrical and Electronics Engineers standard 802.11 [9] is widely referred as Wi-Fi. The name comes from Wi-Fi Alliance [10]. 802.11 is a set of standards for Wireless Local Area Networks (WLAN). 802.11b was the first widely accepted standard and it is still being used. The next important standard was 802.11g, which is currently also the most used one. The major difference between these two standards is of course the speed; 802.11g offers a much higher speed. There is a new and highly anticipated new standard coming which is called 802.11n. The standardization is not yet complete, but vendors are already releasing equipment with 802.11n pre-draft support. This new version will bring a lot more speed. The speed increase is done by adding a Multiple-Input Multiple-Output (MIMO) system, increasing frequency band, using channel bonding and payload optimization. In table 3 is a comparison of some of the standards in the 802.11 family.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Frequency</th>
<th>Throughput</th>
<th>Max data rate</th>
<th>Range indoors</th>
<th>Range outdoors</th>
</tr>
</thead>
<tbody>
<tr>
<td>802.11a</td>
<td>5 GHz</td>
<td>23 Mb/s</td>
<td>54 Mb/s</td>
<td>~35 m</td>
<td>~120 m</td>
</tr>
<tr>
<td>802.11b</td>
<td>2.4 GHz</td>
<td>4.3 Mb/s</td>
<td>11 Mb/s</td>
<td>~38 m</td>
<td>~140 m</td>
</tr>
<tr>
<td>802.11g</td>
<td>2.4 GHz</td>
<td>19 Mb/s</td>
<td>54 Mb/s</td>
<td>~38 m</td>
<td>~140 m</td>
</tr>
<tr>
<td>802.11n</td>
<td>2.4 &amp; 5 GHz</td>
<td>74 Mb/s</td>
<td>248 Mb/s</td>
<td>~70 m</td>
<td>~250 m</td>
</tr>
</tbody>
</table>

Table 3 Comparison of different 802.11 standards

These standards are used in access points (AP's) that cover a small area, for example a coffee shop. Customers that visit the coffee shop can connect with their wireless equipment to the AP and get connected to the Internet. If we want to provide connectivity to a bigger environment like universities, we will have to use many AP's and link them together using other methods such as WiMAX.
7. WiMAX

The IEEE 802.16 family of standards for Wireless Metropolitan Area Networks (WirelessMAN) has been named as WiMAX by the WiMAX Forum [11]. This forum promotes and certifies interoperability of wireless products. They call WiMAX as a standards-based technology enabling the delivery of last mile wireless broadband access as an alternative to cable and DSL. At this moment, there are two main versions of WiMAX, 802.16d [12] and 802.16e [13]. 802.16d is named Fixed WiMAX and it is an older version that has no handoff between base stations and therefore no mobility. 802.16e is called Mobile WiMAX and it has both fixed and mobile services. This version offers a maximum symmetric speed of 70 Mb/s at very short distances, but a more realistic speed is 10 Mb/s symmetric bit rate to 10 kilometers. 802.16e is interesting because it offers good Quality of Service (QoS). The standard defines five different QoS classes. With these classes, WiMAX can offer a guaranteed bandwidth to Voice over IP (VoIP) and live video streaming as well as a best effort type of channel for web browsing for example.

WiMAX is suitable for connecting Wi-Fi hotspots to each other and providing a last mile access for homes as an alternative to cable or DSL technologies. This can come handy especially in rural areas where the telephone wires can be so long that a normal ADSL connection would not be possible. One possibility for the future is that you have your WiMAX equipment at home that is connected to the nearest operator’s base station. The operator’s base stations can form a mesh network that provides Internet access through it. Your WiMAX equipment can also be connected to a Wi-Fi hotspot inside your house to provide fast wireless access to the Internet anywhere inside and even outside your house.

In table 4 is a comparison of some of the standards in the 802.16 family. The propagation column means that the protocol needs or does not need a Line Of Sight (LOS) to the base station. The cell size can be increased with some versions up to 50 km with directional antennas. The terminal size makes a big difference also when comparing these versions. 802.16e terminal can fit into a PC card, while 802.16a [14] and 802.16d need an external device with external antennas.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Spectrum</th>
<th>Propagation</th>
<th>Speed</th>
<th>Mobility</th>
<th>Cell size</th>
</tr>
</thead>
<tbody>
<tr>
<td>802.16</td>
<td>10-66 GHz</td>
<td>LOS</td>
<td>32-134 Mb/s</td>
<td>No</td>
<td>1-5 km</td>
</tr>
<tr>
<td>802.16a</td>
<td>&lt; 11 GHz</td>
<td>NLOS</td>
<td>1-75 Mb/s</td>
<td>No</td>
<td>5-8 km</td>
</tr>
<tr>
<td>802.16d</td>
<td>&lt; 11 GHz</td>
<td>NLOS</td>
<td>1-75 Mb/s</td>
<td>No</td>
<td>5-8 km</td>
</tr>
<tr>
<td>802.16e</td>
<td>2-6 GHz</td>
<td>NLOS</td>
<td>Up to 15 Mb/s</td>
<td>Yes</td>
<td>1-5 km</td>
</tr>
</tbody>
</table>

8. Bluetooth

Bluetooth is an industrial specification for Wireless Personal Area Networks (WPAN). It is a way to exchange information between devices at short range. These devices can be almost anything; mobile phones, laptops, printers, digital cameras, video game consoles, wireless mouse, wireless keyboard and hands-free headsets. One of the best parts of Bluetooth is its low power consumption. This means that Bluetooth has to operate only on low ranges. There are three classes of Bluetooth with different operational ranges; 1 m, 10 m and 100 m. The current version is 2.1 and it offers 3 Mb/s speed. The next version 3.0 is being developed and it is highly anticipated. Version 3.0 will use ultra-wideband radio technology to achieve up to 480 Mb/s data rates to close proximity and 100 Mb/s to 10 meters. It will operate in 6-9 GHz frequency band rather than the current 2.4 GHz frequency band to prevent interruptions from other networks and devices using this 2.4 GHz bandwidth. All this increase in speed and yet they can still keep a very low power consumption. Bluetooth will certainly be a hot topic in the near future. [15]
9. Conclusion

The trend now days seem to be building fiber connections to apartment buildings in metropolitan areas and using DSL technology to cover the last part towards the customers equipment inside the building. In the near future DOCSIS 3.0 might gain an advantage over traditional ADSL connections due to higher speed. There are not many ISP’s offering BPL yet, but their number is increasing. BPL will not, however, surpass the popularity of DSL or DOCSIS in its current state. In the wireless world, homes can be connected with WiMAX to the Internet. Wi-Fi can then be used inside the homes and near it to access the Internet. Many cities around the world have also started to construct a city wide wireless network using Wi-Fi hotspots to cover the whole city. It remains to be seen if some technology can reach and even surpass the popularity of ADSL and rise as the dominant broadband access technology for homes and metropolitan areas.

10. References

[1] OECD broadband statistics (December 2007), http://www.oecd.org/document/54/0,3343,en_2649_33703_39575670_1_1_1_1,00.html, 14.6.2008
Wi-Fi and WiMax-Based Mesh Networking
Piotr MITORAJ

Summary

Wireless Mesh Networks (WMNs) are regarded as a future means to provide cheap network access in developing areas lacking communications infrastructure. The structure of WMN resembles an Extended Service Set (ESS) in which a Distribution System (DS) is made wireless. In this paper the mesh paradigms and mechanisms used in Wi-Fi and WiMAX are examined. The most important usage scenarios and major research challenges are presented. Finally, an architecture for the mesh testbed deployed at Tampere University of Technology is described along with the preliminary performance measurements.

1. Introduction

Mesh concepts proved to be a perfect way to increase the capacity of the wireless network and to extend its range. IEEE 802.16 standard, known as WiMAX, will have its mesh extension as well. WiMAX mesh standard differs from Wi-Fi and the standardization work is carried out in 802.16j work group. As the Wi-Fi Mesh technology is much mature than its WiMAX counterpart, the paper is focused on IEEE 802.11 solutions.

To keep pace with the latest research trends, Tampere University of Technology (TUT) deployed its own mesh testbed called tutmesh. This paper gives insight into its hardware and software internals.

The remainder of the paper is organized as follows: II section describes Wi-Fi Mesh Networking with focus on 802.11s standard and its concepts. The WiMAX Mesh, namely standardization efforts of 802.16j Task Group are depicted in Section III. Finally, the ideas behind tutmesh are presented in Section IV. Section V concludes the paper.

2. 802.11-based Mesh Networking

In the recent years Wireless Local Area Networks (WLANs) based on IEEE 802.11 standard have become ubiquitous. The Access Points (APs), commonly called "hot spots" can be found not only at universities, airports and railway stations: wireless technology found a way even into our homes as a convenient and reliable means to access the Internet. Basic Service Set (BSS) and Extended Service Set (ESS) are the most popular type of deployment of Wi-Fi networks. Apart from the simplicity, which is a large advantage, this solution has certain drawbacks, especially in terms of scalability. Due to the limited transmitter power and narrow frequency bands, the current deployments have very limited range and capacity. To address those issues, more APs are added to the network. However, WLAN APs are typically wired, which is the biggest wireless paradox. Each AP needs to have a wired connection to the Distribution System (DS). This approach poses constraints on AP placement and increases the deployment cost. Hence the highest cost is in cabling and infrastructure to interconnect the APs. To decrease the deployment cost and allow flexible AP placement, the concept of Wireless Mesh Network (WMN) has arisen. It replaces expensive wired backbone with its wireless counterpart introducing Wireless Distribution System (WDS).

2.1 802.11s

This paradigm is now under standardization process. IEEE 802.11s Task Group (TG) was created in November 2003 and the final standard is going to be approved in August 2009. The aim of the group
is to design routing techniques for the Link Layer and hence propose a WDS that enables transmission in broadcast, multi-cast and unicast modes. The efforts are focused to provide functionality similar to wired DS [1].

2.1.1 Overview

The outcome of the TG ’s’ works will be the amendment to the IEEE 802.11 standard. Although, many issues have to be resolved, no changes to the Physical Layer are considered. As the routing in Mesh Networks will be performed at Link Layer, the focus is on MAC techniques. The Figure 1 presents how the existing ESS will be modified.

IEEE 802.11 standard describes only two types of devices: Station (STA) and Access Point (AP). Therefore, new device classes were introduced and concept of AP is not used. There are 4 device classes:

- Station (STA)
  Connects to the Mesh Network via Mesh Access Point. Functionality is kept similar to IEEE 802.11 STA definition.

- Mesh Point (MP)
  Establishes links with other MPs and creates the Wireless Distribution System - the core of the network.

- Mesh Access Point (MAP)
  Acts like MP with Access Point functionality. Connects STAs to the Mesh Network.

- Mesh Portal (MPP) Point at which Mesh Network is connected to Internet or other external network.

However, the stated previously descriptions denote the functionality of the nodes, it is common that one device may belong to two or more classes, e.g. MPP that acts also as MAP.

2.1.2 Security

Standard IEEE 802.11 Basic Service Set (BSS) provides hierarchical security concept. AP serves as an Authenticator and often implements Authentication Server, unless it is a separate entity in the network. The security mechanisms in Mesh Network does not differ from the STA point of view. However, due to the shared wireless medium (WDS), dynamic topology changes and lack of infrastructure, the network’s

MITORAJ, P: Wi-Fi and WiMAX based Mesh Networking
security is easier to compromise. Therefore, special means must be taken to provide the user secure network access.

802.11s does not provide end-to-end security. It is granted on a Link-basis[citation needed]. The new security mechanism introduced is called Efficient Mesh Security Association (EMSA). EMSA reuses the 802.11i link level authentication model with 802.11X authentication known from single hop wireless LANs. The essential difference in security approaches is in the role of AP. MAP must act both as Supplicant and Authenticator [2].

The security threats in Mesh Networks appear in different protocol layers at the same time. Hence, the cross-layer security framework is highly desired. However, the practical implementation of the system addressing all the security concerns remains an open research issue [3].

2.2 Usage Scenarios

Originally there were 4 usage scenarios supporting the services and deployment areas that would benefit the most from what Mesh Networking provides. At the end of 2004 MITRE Corporation together with military research organizations submitted a "Military Usage Scenario" to 802.11s Task Group. It described the Mesh Networking requirements and scenarios for military usage and was well received. Since November 2004 Military scenarios is included in the official set of scenarios supported by 802.11s [4].

Residential/Consumer Electronics This scenario is intended to provide high performance wireless coverage throughout the building or a single flat. As the high-bandwidth applications, like audio and video streaming, are likely to be used within the home network, the Mesh Networks will help to eliminate the dead spots and low quality areas.

Office The motivation behind office usage scenario lies in the cost reduction and ease of deployment. It applies in the premises where Ethernet cabling is not possible or too expensive. Small and middle sized companies, which are most likely to grow, may benefit from this affordable and easily scalable solution to provide network access.

Campus/Community/Public Access Office This scenario for instance offers wider coverage than the others. The Mesh Networks here are regarded as the lower cost and higher bandwidth alternative to classical Internet access methods e.g. dial up, DSL or cable connections. It enables to implement the location-based services.

Public Safety This scenario aims to provide ubiquitous network coverage for emergency response personnel, such as fire, police and emergency workers at the accident scene. The Mesh Network can be used for video surveillance, voice communication, collection of the data from sensors, tracking and monitoring the condition of emergency workers and coordination tasks.

Military Military usage scenarios may be described as both non-combat and combat. The first resemble previously described use cases. The latter takes into account extreme mode mobility, fully automated network management, and for nodes temporary outside of a coverage area also ability to preserve power.
3. WiMAX Mesh

IEEE 802.16 standard defines the Wireless Access Network widely known as WiMAX. WiMAX network consists of Base Stations (BSs) and Mobile Stations (MSs). The BS provides network access for MS. Currently, there are two topologies supported: point-to-multipoint and mesh. In the first one each MS communicates directly with the BS whereas the latter one allows MS to connect to BS through another MS [5]. Mesh topologies can be divided into client mesh and infrastructure mesh [6].

3.1 IEEE 802.16 Mesh Mode

IEEE 802.16 mesh mode is client mesh. Every MS is capable of forwarding other nodes’ traffic. In mesh mode MSs are also capable of exchanging traffic directly among themselves. Here each node has a 48-bit MAC address however for the sake of efficiency each node is assigned a 16-bit ID after the authorization phase. Whatever tempting the assumptions may be, this mode is less secure, less predictable and harder to manage than the infrastructure mesh described in the next section.

3.2 IEEE 802.16j Mobile Multihop Relay

Slightly different mechanisms for mesh topology are within the scope of the IEEE 802.16j Task Group. The fundamental motivation behind the work of this task group is, that the Relay Stations (RSs) extending the network coverage can be deployed at significantly lower cost than BSs. RS is assumed not to generate any traffic itself and each Subscriber Station (SS) may reach BS through more than one RS, hence load balancing mechanisms are also considered. IEEE 802.16j amendment, the Multihop Relay Specification for 802.16 will be fully compatible with 802.16e-2005 mobile and subscriber stations. This means that the changes are applied only to BSs. Compared to 802.16e, resource allocation algorithm is more complicated as also RSs take part in resource allocation process. The amendment shall support the hop count equal or greater than 2. This means more concern should be put on fairness aspects to ensure that the nodes hops away from BS are still granted a fair share of resources [7].

3.3 Summary

However, most of the research is focused on Wi-Fi mesh networking. Partially, because of availability and low cost of compliant wireless equipment and maturity of IEEE 802.11 standard. Therefore, this technology has bigger potential in rapid and cost-effective deployment. Due to above mentioned reasons and availability of open source drivers for 802.11 wireless network interface cards, Wi-Fi was picked as a platform of choice for the tutmesh deployment presented in Section 4.

4. Tutmesh

This section describes an attempt to build an IEEE 802.11-based testbed for multihop wireless networks. The work has been carried out at the Department of Communications Engineering at the Tampere University of Technology by the author of this paper under the supervision of Yevgeni Koucheryavy.

4.1 Related Work

There are numerous attempts, both commercial and academic, to build hardware and design a testbed for wireless mesh networks. Meraki [8] and LocustWorld [9] are examples of a few companies on the
market that already offer the mesh products. These proprietary solutions, however, are not compatible. This encouraged the creation of the earlier mentioned IEEE 802.11s task group.

While commercial implementations provide the systems that just work, academic approach aims at scientific gain. Therefore testbeds developed by scientific communities are more sophisticated and research-oriented. They are much more flexible in terms of configuration possibilities. Only these were considered as the reference during a design of our testbed. In most of the cases a two-tier network was used. It means that the router-to-router backbone communication is separated from the client-to-router traffic. However, simpler solutions, where each node has only one wireless interface, such as the MIT Rooftop [10] and UCSB MeshNet [11] are popular as well. The latter is the testbed where the mesh router is implemented on small form-factor PCs and on Linksys WRT54G wireless router running OpenWrt. AODV is the routing protocol used for the backbone communication. The second, widely deployed routing protocol is OLSR, like in MCG-Mesh [12] where also a modified DYMO was tested as these two protocols represent the two main routing approaches: proactive and reactive. Here the testbed is assumed to have central configuration. The OS image is loaded via NFS. We believe it is not the best solution as it introduces a single point of failure and requires constant wired connection. The example of the testbed with implemented mobility support [13] is quite similar to our approach, except the hardware used, namely ARM-based platforms. In this small-scale testbed, which contains only 3 nodes, configuration is static. Although the majority of testbeds aims to develop a solution that provide network access for non-real-time applications, the mesh concept was also used at time-critical scenarios such as traffic control [14]. The performance of a mesh network based on IEEE 802.11b standard with MeshAP mesh networking software from Locust World proved to be sufficient also in road conditions [15].

4.2 Deployment

There are two types of nodes in WMN: Mesh Routers (MRs) and Mesh Clients (MCs). Mesh Routers are the nodes with little or no mobility that form the multi-hop wireless backbone and their main objective is to perform the routing function. The MR with such a basic functionality is called Mesh Point (MP). If MR is connected to the Internet or to another access network, e.g. WiMax, it acts as a gateway. It may serve as an access point for Mesh Clients and then we refer to it as Mesh Access Point (MAP). Mesh Clients may be either desktop computers or highly portable hand held devices. For the sake of power consumption the complexity of MC is low as the routing function is not implemented. This is in contrast to ad-hoc networks where all the nodes have similar, high complexity. Testbed topology is comprised of four Mesh Routers and two Mesh Clients. Additionally two MRs are connected to the Internet.

4.3 Mesh Router

The testbed proposal assumes simultaneous usage of separate wireless channels for backbone and user communication. Hence the important requirement for the Mesh Router is to be equipped with two wireless interfaces.

4.3.1 Hardware

The Mesh Router is deployed on small 100 x 160 mm system board. Each board is be equipped with two IEEE 802.11b/g wireless interface cards with 5 dBi nominal gain omni-directional antennas. There is also a slot for 1GB Compact Flash memory card. A System board is alix3c1 from PC Engines. The System board may be described as Single Board Computer (SBC) based on 433 MHz AMD Geode
LX700 CPU. SBC provides 128 MB RAM memory, one Fast Ethernet port and DB9 serial port. The board architecture is denoted in Figure 2. The low power consumption of the board of roughly 4–6 Watts in normal operating mode (without wireless cards) makes it a perfect platform for a wireless router. A configuration connection will be established via serial or Ethernet connection. The Wireless cards are Compex WLM54G 200mW 802.11b/g cards with Atheros AR2413 chip.

4.3.2 Software

The Mesh Routers are implemented on above mentioned alix3c1 boards. The operating system is Voyage Linux 0.4 [16] which is a lightweight Debian-based Linux distribution optimized to work on embedded systems and tested with alix boards. It provides madwifi-ng drivers that are suitable for our Wi-Fi Atheros cards. The size of the Compact Flash card allows us to place two complete OS images. During a remote upgrade of the first system, a router is rebooted to use the second OS. In this case, a CF card accommodates three partitions: a small boot partition and equally-sized sys1 and sys2 partitions for OS images.

4.4 Mesh Client

A mesh Client may be any immobile or mobile station from a desktop computer to a PDA. As there are no special requirements for an MC, two laptops with wireless interface cards are used in this testbed. Both stations are running Ubuntu 6.06. From the client point of view, there are no significant changes. It establishes the connection to a MAP just like to the ordinary Access Point.
4.5 Preliminary Performance Measurements

This experiment is an excerpt from the extensive measurements conducted on the testbed. We evaluate the performance improvement achieved by introduction of second radio on the mesh node in the noisy environment. To add interference, an external traffic was generated on the same channel on which the backbone tier is operating. In a single radio scenario, access tier, backbone tier and interfering traffic were utilizing the same channel. In a dual radio scenario non-overlapping channel was set for an access tier. To obtain the results presented in Table 1, a set of 3 test runs with varying packet size was performed over 3 hop path in single/dual radio environment. Each test run lasted for 10 minutes.

<table>
<thead>
<tr>
<th>Packet Size [B]</th>
<th>Single Radio</th>
<th>Dual Radio</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>500</td>
<td>1470</td>
</tr>
<tr>
<td>500</td>
<td>60</td>
<td>500</td>
</tr>
<tr>
<td>1470</td>
<td>60</td>
<td>500</td>
</tr>
<tr>
<td>60</td>
<td>500</td>
<td>1470</td>
</tr>
<tr>
<td>500</td>
<td>60</td>
<td>500</td>
</tr>
<tr>
<td>1470</td>
<td>60</td>
<td>500</td>
</tr>
</tbody>
</table>

Table 1: Dual Radio vs Single Radio in noisy environment [17]

According to the results obtained, dual radio brings significant gain in delay, jitter and Packet Loss Ratio (PLR). For small packets (single VoIP packets) dual radio does not bring valuable improvement over single radio. With increasing size of the packet (aggregation of VoIP packets from different flows), the positive effect of dual radio becomes stronger. Therefore, introduction of dual radio along with packet aggregation techniques will improve the performance of VoIP in WMNs. More detailed information about packet aggregation in context of VoIP traffic can be found in [17].

5. Conclusion

This paper presents an overview of Wireless Mesh Networking. It provides insights into the introduced new concepts and techniques. Focus is put on the standardization efforts, namely IEEE 802.11s in Wi-Fi and IEEE 802.16j in WiMAX. The paper contains also the testbed architecture intended to accommodate different performance test scenarios. The testbed provides measures to investigate the behavior of numerous networking aspects, like MAC algorithms, routing protocols or Transport Layer mechanisms. A Mesh Router is deployed into an embedded system. This, along with Linux OS guarantees high performance and provides numerous configuration possibilities. Most probably it will require only a software update to meet the requirements of IEEE 802.11s.

6. References


MITORAJ, P: Wi-Fi and WiMAX based Mesh Networking


COMMUNITY NETWORKS

Karri Huhtanen

Summary

The community networks, especially wireless ones, can be built on top of several different technology and business models. As not one of these models has proven the most successful and dominant one, this paper presents briefly the most popular community network models, their advantages and disadvantages. The paper also presents an emerging cooperative community network model and its implementation in more detail to give an opportunity to compare different models and if needed, select the ones for further analysis.

1. Introduction

The community network term is widely used both in wired and wireless networks. While in the wired networks, a community network usually means fiber-to-the-home initiatives or building a community owned network infrastructure and acting this way as a network provider, in the wireless networks there exists more diversity in the business and architecture models. A wireless community network usually means by definition a wireless network either deployed for the community or a network, which is formed by community members in cooperative fashion. In the wireless community networks a clearly dominating business or architecture model has not yet been found, instead there exists several different attempts to find and implement one. This paper presents few common architecture and business models for wireless community networks and analyses briefly their potential advantages and disadvantages. The paper also describes one of the existing models, a cooperative wireless community network, in more detail, and gives an use case example of utilising it in the Wireless Tampere community network [1].

2. Wireless Community Network Models

The wireless community networks are generally founded, and funded, by cities or new startup companies, although there exists initiatives and attempts for personal community networks or internet service provider supported community networks. Some of these community networks are build completely open with initial public money, some based on advertisement based revenue and some as overlay networks over existing Internet subscriptions. Each of them have their advantages and disadvantages which are presented in the following sections.

2.1 The City as a Service Provider Model

In this model, the city provides the full or significant portion of the funding to build wireless coverage around the city. Usually this is done by creating and deploying a new wireless network infrastructure without utilising the existing wireless networks. Wired network connections can be utilised and often there exists an Internet Service Provider (ISP) partner to provide wired network connections and bandwidth for the community network. The costs of deploying, maintaining, extending and developing the network are usually divided by city, potential partners and sponsors. Usually at least the initial deployment is funded typically with public or regional funding support. The panOulu initiative [2] is an example of this kind of network model in Finland. Most of the network deployment is paid by City
of Oulu with the help of regional development and EU support funds as well as sponsors. Oulun Puhelin provides then the Internet service for the panOulu network.

One of the key issues is, what happens to network maintenance and development once it has been deployed. The main ideology behind this wireless community network model is that wireless network should be a public service funded by taxes like for example roads and other public services. The problem is that after initial deployment, which can already cost a lot by itself, the network infrastructure still needs administration, maintenance, upgrades and development, which consumes resources and funds without generating any direct revenue to its founders, ISP excluded. The panOulu project have been facing this from beginning of 2008 according to their blog [3] Because of this kind of cost structure, many cities have adjusted their plans to share some or all the deployment and maintenance costs by outsourcing the community network service to external service providers.

2.2 The Outsourced External Service Provider Model

Cities and communities willing to offer wireless community networks, but not willing to pay themselves all the deployment and maintenance costs, usually try to find a commercial partner to share the risk and costs of the wireless community network. A common way for finding this kind of commercial partner is to hold a competitive bid, where ISPs, network service providers and community service provider compete for the official community builder's status. The winner of the competitive bid usually gets the preferred partner status and right to use the community brand for marketing purposes. Usually only one company acquires this kind of official community network service provider status and the deployment, maintenance and development of the community network is on the responsibility of the winner of the bid. The city or community may provider some marketing support and funds for initial deployment, but it is the company's responsibility to make a business out of the community.

One of the advantages of this model is that a single external service provider provides stability and consistency in network access. There is only one network service provider, so all service areas are similar in functionality and connectivity so the network use experience stays same for the users. One of the disadvantages is that giving a monopoly for single service provider leads to a situation, where the future network and service development as well as pricing depends heavily on the service provider. If the network service provider cannot make enough business out of the network service, the development of the network and services might suffer. At the same time, other companies and community cannot do anything to improve situation, because the network service provider has the exclusive rights to develop the community network and services. One can also question how quickly do the network and services develop when there is no competition for the service provider position.

Examples of this kind of community networks are several attempts to build city wide wireless networks in the United States like for example San Francisco. The city of San Francisco analysed the various wireless community network models and their commercial feasibility [4]. The decision was then made to outsource the network to an ISP called Earthlink, which was to in cooperation with Google to deploy wireless network around the Bay Area. In Autumn 2007 the Earthlink decided to reconsider deploying the wireless network and retreated from the initiative demonstrating thus one disadvantage of having an exclusive service provider [5].

2.3 The Community Network Service Provider Model

A community network service provider is usually not a city or a traditional ISP, but instead a commercial startup (for example Fon [6], Sparknet [7]) or a new operator (for example Wippies [8])
trying build an overlay network on top of the existing Internet infrastructure (see Figure 1 for Wippies example). This way the community service provider (CSP) does not necessarily have to use a lot of money to build the actual wired network infrastructure. The CSP typically only provides a firmware or a network appliance configured and usually locked in certain community. The community members then buy or install free hardware utilising their existing Internet connection and by sharing this bandwidth, the community members gain the right to utilise the community network wherever its available. The CSP gains this way a free transmission network on the cost of the service provider providing the Internet connectivity to the community member. Some of the community network service providers like Fon provide revenue sharing models for the temporary user accounts sold in community member hotspots. There exists also non-commercial community service providers like NoCat and Coova, which offer free software to be installed on Linux-based access points and in return the network access wherever the community has coverage.

![Wippies Network Architecture](image_url)

Fig. 1: Wippies Network Architecture as presented in Wippies TechBlog [9]

One of the problems with free communities and partially with some commercial ones is where the coverage generation is focused. If the coverage is mainly some private hotspots far from the interested user, this kind of community cannot compete with the ones having for example multiple hotspots and access areas in pubs and restaurants around the world in major cities. From the user perspective also the additional services the community service provider offers matter. Wippies for example has decided to offer 4GB email space, blog, 500MB WWW accessible network space among other services to make the Wippies community interesting for potential members. Fon and Sparknet on the other hand offer only basic WiFi connectivity and management and limited services from their partners. The difference to the outsourced external service provider is that the community can have some choice and

Huhtanen, K: Community Networks
opportunities to influence the development and development priority of different services or functionality. Some of the CSPs like Fon and Wippies, have already announced to open their platforms for open source development [10], but not completely, because of the control, security and intellectual property reasons. The CSP still keep the control of community and if they are not inclined to for example fully open their firmware, a community member cannot change that. The viability of the CSP relies heavily on how well does it develop new services and support third party service development.

2.4 The Cooperative Community Network Model

A cooperative community network attempts to combine the advantages of city provided community network model and of the community network model. One of major differences being that in cooperative community network model there exists multiple service and network equipment providers in addition to the community members to develop the services and network solutions together. A separate entity, the coordinator, monitors and controls the community network brand and defines the set of rules and interfaces how the service and network solution providers interact with the community and each other. The network coverage is built and expanded by connecting existing and new networks together with common configuration rules and standard interfaces. The companies and community members can this way utilise their existing hardware and network solutions to join to the community network without the need for vendor or community lock-in.

The basic infrastructure costs in cooperative community network are covered by company membership fees and visitor user account sales. Each member maintains its own network, so the network maintenance costs and resources are shared between the community members reducing the infrastructure running costs for the coordinator or city supporting the model.

While the initial coverage is usually obtained by connecting first large organisation networks, city guest networks and other major networks, the coverage is additionally expanded by network solution providers selling new hotspot and company guest network solutions. The network solution providers get their revenue from selling network equipment, services and expertise, ISPs from selling broadband ready configured connection packages and support companies from helping small companies and individuals to join to the network. The competition between these companies is not limited in any way, which works as an advantage for the community members as the companies are inclined to offer new and interesting solutions and services to expand and keep their customer base.

The important issues with this model are finding the correct balance between benefits and costs to get organisations and individuals to participate as well as identifying those correct incentives to encourage and ensure participation. Having wrong cost-benefit balance or not enough incentives for community members or service/network solution providers hinders the development of the community as the motivation to participate and develop community fades.

3. Wireless Tampere – a Cooperative Wireless Community Network

Wireless Tampere is the first attempt to design and implement a community network based on the Cooperative Wireless Community Network Model. The goals of the project are encouraging constant development and free innovation of services and solutions together with the Creative Tampere program [11]. The project and Wireless Tampere community is coordinated by Technopolis Ventures Professia [12]. The infrastructure components were mainly developed by Arch Red [13] or joint ventures between the coordinator and the company based on the WiFi roaming technology already utilised in the eduroam(tm) WiFi authentication and roaming federation [14].

Huhtanen, K: Community Networks
In addition to the infrastructure developer and coordinator, several service and network solution as well as support service providers participate by offering their products and services to community members. The concept was also approved by several major ISPs (TeliaSonera, Elisa, Tampereen Puhelin), which usually has been one obstacle for community network service providers as Internet subscriptions generally deny sharing network access with third parties.

The initial coverage for getting organisations and individuals interested to participate was build by first connecting major WiFi networks like the city of Tampere network, both university networks and Tampere Polytechnic network together. This coverage was soon extended by individuals connecting their access points to the community network as well as neighbouring smaller cities deciding to participate. Currently (May 2008) the major focus is to get pubs, restaurants and cafes to join to the community with the help of the Wireless Tampere Open Source AP described in detail later in the paper.

3.1 The Architecture of Wireless Tampere

The architecture of Wireless Tampere is presented in Figure 2. To make connecting organisations and individuals easier and faster several new components were developed in addition to the RADIUS based [15] authentication hierarchy already utilised in Funet and eduroam(tm) WiFi roaming [16].

Fig. 2 The Architecture of Wireless Tampere Community Network

The Roaming Root RADIUS server (Figure 2) is the RADIUS server functioning as a RADIUS proxy and connecting the RADIUS server capable organisations together and handling also the roaming traffic between Wireless Tampere members and potential other community networks and roaming partners. The same kind of authentication root server can be found in eduroam(tm) networks as well as community networks like for example Sparknet. The difference in Wireless Tampere is that organisations can utilise their existing RADIUS server and wireless equipment as long as they are configured accordingly to Wireless Tampere rules. In addition to the RADIUS server, the participating organisations are required to provide both a WWW-authenticated Wireless Tampere -network and
WPA-authenticated [17] LANGATON-WPA network or choose either one. This network then becomes a part of the Wireless Tampere coverage area for also other members to use.

As individuals cannot be expected to host their own RADIUS servers and because there was not an interest to build yet another proprietary firmware for Wireless Tampere, a new component called Consumer Auth. Server (Figure 2) was developed. With the Consumer Auth. Server individuals can via WWW interface register their own any WPA-Enterprise capable access points as part of the Wireless Tampere network and gain in return a Wireless Tampere roaming user account in addition to WWW-based user management for their home access point. Like the organisational networks earlier, by registering the access point to the service and configuring LANGATON-WPA as a network name, these individual access points contribute to the coverage of Wireless Tampere. Utilising WPA-Enterprise authentication ensures that the user credentials cannot be captured by home users even if the network equipment they provide is their own.

After initial phase it was realised that companies cannot be expected to have the expertise, time or money to setup RADIUS servers on their own. A Company Auth. Server (Figure 2) was created to offer the same kind of service to companies as the Consumer Auth. Server was. For companies a more feature complete guest network management service was then developed, but with the basic principle remaining the same. The company administrators would utilise the WWW interface to register access points and create organisation users and guest users for their network. The company authentication service is sold to companies via Wireless Tampere network solution providers and this way a new at the same time an outsourced authentication service business was developed to encourage network solution provider participation to the community development and expansion.

The third new community network component was also developed based on the user need. For large companies or organisations with enterprise grade wireless networks it easy to support both WWW authenticated and WPA-Enterprise authenticated network with same access points, but for smaller companies and businesses like cafes, pubs and restaurants, these kind of access points are clearly out of their price range. To get Wireless Tampere adopted in this kind of environments and enlarge this way the coverage, a reasonably priced alternative, an open source based access point firmware, hardware combination was developed.

3.2 The Wireless Tampere Open Source Access Point

The Wireless Tampere Open Source Access Point was developed like many other community network access points based on embedded Linux platform. The software was integrated utilising OpenWRT Linux platform [18] as a software platform and the first hardware implementation was build on top of Meraki hardware. This was recently changed to Accton hardware based Linux platform as Meraki changed their hardware licensing rules to prohibit using custom firmware. Building the firmware on open source platform made it possible to only change the hardware as long as the software platform supported it.

The Wireless Tampere Open Source AP required also other supporting elements to be added to the Wireless Tampere architecture (See Figure 2 and Figure 3).
A common problem with community network access points is ensuring the connectivity between the access point and the community network service providers servers. In Wireless Tampere this was solved by integrating OpenVPN [19] open source VPN software to both provide the connectivity over any connectivity disrupting obstacles and to secure the management traffic between Hotspot Auth. Server (Figure 3) and Wireless Tampere Open Source AP.

The Hotspot Auth. Server provides the centralised WWW authentication page for community access points removing this way the need for certificates for each access point and making using the user credentials a bit safer when security of the credentials was considered. This way the access point was able to provide both the Wireless Tampere WWW authenticated network for visitor access and the WPA-Enterprise authenticated network for regular community member access under 100 euros / access point. Having WWW authenticated network was an important thing for small businesses like cafes, pubs and restaurants since most of their customers were only visiting and could not be expected to configure WPA settings to their terminals only for the brief visit.

Based on the Wireless Tampere openness principle, the access point firmware is to be released as open source and already uses open and standard interfaces like RADIUS and Wifidog [20] Captive Portal Interface for communication. By utilising these same interfaces any access point or access controller capable of running OpenVPN and Wifidog client could be connected to the Hotspot Authentication service. An open source firmware, on the other hand, enables open and free network service and equipment development already on an existing free platform.

4. Conclusions

As the dominant design for wireless community networks has not emerged there exists several different models on which to build a new community network. Each of the models have its advantages and disadvantages, the question is more like which fits the needs of the founding participants. A closed, own community may suit some private business interests while a cooperative community network model may be more useful to encourage regional service and innovation development. If the cooperative community network model seems feasible, the Wireless Tampere concept has already developed some of the needed components and would be based on its openness an option, on which to
build new services and components on. There are several options, the builder of a community network only has to choose which way to go and which model to utilise.

5. References

VEHICULAR AD-HOC NETWORKS (VANETs)

Jakub JAKUBIAK

Summary

Recent advances in wireless communications technology and car industry made it possible to consider wireless Ad-hoc (or mixed Ad-hoc/infrastructure) networks, providing connectivity among vehicles on the road. Such network, often referred to as VANET (Vehicular Ad-hoc Network) is seen to be one of the most valuable concepts for improving efficiency and safety of the future transportations, as well as boosting operators’ revenue. This report covers current research challenges and future applications for such networks, discusses possible deployment scenarios and provides some insights to the state-of-the-art of the current standardization of interfaces and protocols.

1. Introduction

Vehicular connectivity can be fairly considered a future killer application, adding extra value to the car industry and operator's services. Taking into account the constant growth of automotive market and the increasing demand for the car safety, also driven by regulatory (governmental) domain, the potential of car-to-car connectivity is immense. Such system should be suitable for a wide spectrum of applications, including safety-related, traffic and fleet control, and entertainment. First, issues concerning architecture, security, routing, performance or QoS need to be investigated. Standardization of interfaces and protocols should be carefully planned to ensure interoperability, as vehicles coming from different vendors must communicate seamlessly. Having different competing systems would result in decreased market penetration and poor overall system efficiency, thus only one common system can be deployed. And finally, wise deployment strategy has to be proposed, as most application would become functional only after certain market penetration is reached.

The first milestone of standardization process was the allocation of 75 MHz of DSRC (Dedicated Short Range Communications) spectrum to accommodate Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communication for safety-related applications by US Federal Communications Commission (1999). Commercial applications are also allowed to operate in this spectrum, as long as they do not interfere with its primary purpose.

The rest of this article is organized as follows: the next section briefly mentions potential applications of car-to-car networking. Then the current research challenges and their possible solutions are discussed. Section 4 introduces market penetration and deployment strategy issues. Further the current standardization process is presented, with main focus on European, USA and Japanese activities. The concluding remarks are presented in the last section.

2. VANET applications

Integrating a network interface, GPS receiver, different sensors and on-board computer gives an opportunity to build a powerful system, capable of gathering, processing and distributing information. Numerous applications can be deployed in a network established with such equipped vehicles and proper infrastructure. Generally, from the connectivity point of view they could be divided into four main groups: car-to-car traffic, car-to-infrastructure, car-to-home and routing based applications. These applications are either safety-related or comfort-related (commercial).
2.1 Safety-related applications

Safety-related applications may be grouped in three main classes: assistance (navigation, cooperative collision avoidance, and lane-changing), information (speed limit or work zone info) and warning (post crash, obstacle or road condition warnings). They usually demand direct communication due to their delay-critical nature. One such application would be emergency notifications, e.g. emergency braking alarms. In case of an accident (the airbag trigger event) or sudden hard breaking, a notification is sent to the following cars. That information could also be propagated by cars driving in the opposite direction and, thereby, conveyed to the vehicles that might run into the accident.

Another, more advanced example is cooperative driver assistance system, which exploits the exchange of sensor data or other status information among cars. The basic idea is to broaden the range of perception of the driver beyond his field of vision and further on to assist the driver with assistance applications. Transmitting this data to cars following on the same road, drivers get information about hazards, obstacles or traffic flow ahead, hence driving is more efficient and safer. Some applications of this kind are operating only when certain penetration of VANET enabled cars is reached.

2.2 Comfort (commercial) applications

General aim of these applications is improving passenger’s comfort and traffic efficiency. That includes nearest POI (Points Of Interest) localization, current traffic or weather information and interactive communication. All kinds of applications might be applied here, e.g. online games or instant messaging. Another application is reception of data from commercial vehicles and roadside infrastructure about their businesses (‘wireless advertising’). Enterprises (shopping malls, fast foods, gas stations, hotels) can set up stationary gateways to transmit marketing data to potential customers passing by. Furthermore, these services could be integrated with electronic payments.

The important feature of comfort/commercial applications is that they should not interfere with safety applications. In this context traffic prioritizing and use of separate physical channels is a viable solution.

3. Research Challenges in VANETs

When deploying of a vehicular networking system, several issues have to be resolved, often from distant fields of expertise, ranging from applications development up to economical issues. VANET could be considered as an instantiation of MANET network (a Mobile Ad hoc NETwork); however their behavior is fundamentally different. These unique characteristics of these networks are as follows:

- rapid topology changes and fragmentation, resulting in small effective network diameter
- virtually no power constrains
- variable, highly dynamic scale and network density
- driver might adjust his behavior reacting to the data received from the network, inflicting a topology change

Here we briefly mention some of the core research challenges that need to be addressed.

3.1 Wireless Access technology

There are several wireless access standards that could be used as a base for VANET connectivity. In general the aim is to provide a set of air interface protocols and parameters for high-speed vehicular communication using one or more of several available media. Some of the core technologies include:
A. Cellular technology (2/2.5/3G)

The main advantages of 2/2.5G technology are coverage and security, and 3G, slowly but steadily taking over, provides improved capacity and bandwidth. Several telematic and fleet management projects already use cellular technology (e.g. SMS reports), however the relatively high cost, together with limited bandwidth and latency make it impossible to use as a main communication means.

B. IEEE 802.11p based technology

IEEE is working on a variation of 802.11 standard that would be applied to support communication between vehicles and the roadside, or, alternatively, among vehicles themselves, operating at speeds up to 200 km/h, handling communication ranges as high as 1,000 meters. PHY and MAC layers are based on IEEE 802.11a, shifted to the 5.9 GHz band (5.850-5.925 GHz within US). The technology is promoted by the car industry both in Europe (Car2Car CC) and US (VSCC, VII). Estimated deployment cost is foreseen to be relatively low due to large production volumes.

C. Combined wireless access

One of the most significant efforts in combining those wireless access technologies is done by ISO TC 204 WG16, called CALM M5 (Continuous Air Interface for Long and Medium range). It builds on the top of IEEE 802.11p, incorporating a set of additional interface protocols. Currently supported standards include: Cellular Systems: GSM/HSCSD/GPRS (2/2.5G) and UMTS (3G), Infrared Communication and wireless systems in 60 GHz band. Using all those interfaces in a single, uniform system would result in increased flexibility and redundancy, thus improving applications' performance. Apart from interoperability issues, CALM is also engaged in the standardization of the protocols, network layer and the management services.

3.2 Spectrum issues

The intended usage period for V2V communication system is estimated for at least 20 years and within this time the spectrum availability has to be guaranteed. In the USA the FCC has already allocated 75 MHz of spectrum at 5.9 GHz (from 5.850 to 5.925 GHz) band. As agreed by VSC and VII Consortiums, the best technology available for the communications systems using this spectrum would be a derivative of IEEE 802.11. Thus the before mentioned IEEE 802.11p WG and ISO TC204. Unfortunately a continuous spectrum of 75 MHz in DSRC band is not available in Europe. Hence the Car2Car CC has proposed a derivative of the US approach. The proposal allocates the 5.9 GHz band range (5.875 – 5.925 GHz) for primary use of safety critical applications, and 2 x 10 MHz at 5.855 – 5.875 GHz for non safety applications. Those bands are used in the US as well, and their allocation in Europe would allow for world-wide harmonization. Earlier the 5.9 GHz band was allocated for military radar systems and fixed satellite services; however recently the CEPT/ECC Short Range Device Maintenance Group (SRD/MG) has recommended placing the 10 MHz control channel in 5.885 - 5.895 GHz, to align with the US approach, and the second 10 MHz channel in the upper part of the ISM band (5.865 - 5.875 GHz) to take also into account radiolocation services below 5.85 GHz. The decision about bandwidth assignment for safety applications was taken by CEPT/ECC in March 2008.

3.3 Broadcasting and message dissemination

The foreseen applications will require a vast amount of information to be broadcasted, thus several broadcasting techniques are taken into account. That includes some narrow bandwidth solutions like
FM radio (used for RDS/TMC), but also wider bandwidth digital services such as DAB, DVB, DVB-H, S-DMB, T-DMB and DDB. Satellite broadcasting also emerges as a possible solution, as it already includes real-time traffic information services [15].

Broadcasting appears to be an attractive solution due to its low cost and large potential volumes of data. Location-aware broadcasting would limit the broadcast range only to the site of interest, thus reducing overhead (avoiding the broadcast storm problem). Clustering is another approach to optimize the message dissemination process: neighbor nodes form clusters, manageable units that limit the broadcasting range. E.g. in [1] a clustering method called Local Peer Groups (LPGs) is proposed, where nodes can either form static or dynamic clusters.

### 3.4 Routing issues

MANET routing protocols have brought a lot of attention during the last years, however in case of vehicular networking certain network characteristics make these protocols unsuitable. This is due to main MANET routing assumption: intermediate nodes can be found between source and destination and end-to-end connection can always be established. But frequent network partitioning in VANETs requires a different approach, e.g. the 'carry and forward' idea [2], where, if no direct route exists, a packed is carried by a node until it could be forwarded to a node being closer to the destination. The 'carry and forward' concept can be combined with one of the 3 main routing algorithm categories suitable for VANETs: opportunistic forwarding [3], trajectory based forwarding [4] and geographic forwarding [5]. Also a hybrid solution, mixing 2 or 3 different approaches, could be developed.

In opportunistic forwarding, a message is stored and forwarded whenever given the opportunity. This algorithm works efficient in broadcasting mode, but fails when the target is a single node. Some analysis of message dissemination on top of opportunistic forwarding can be found in [6].

Geographic forwarding and trajectory forwarding work similarly in the context of VANETs, as vehicular traffic follows the road layout. In the first case (e.g. GFG/GPSR, [5]) packets are forwarded towards the destination based on node geographical location. That approach offers good scalability, but is problematic with dead-ends and voids even if a path towards a destination exists (perimeter routing partially solves the problem).

In trajectory routing the road infrastructure serves as an overlay directed graph, with intersections seen as graph nodes and roads as graph edges. Messages move following predefined trajectories and distance is defined as a graph distance (unlike in the case of geographic forwarding, where a simple Cartesian distance is used). Trajectory routing could be seen as the most natural message forwarding algorithm for VANET networks.

### 3.5 Power management

Power management in VANET is not concerned about energy efficiency, but rather about the transmission power - when too high, the ongoing transmission could disrupt another transmission at a distant node due to interferences. Thus the denser the network is, the lesser TX power should be used. This issue is also important from the routing point of view: how to adjust the transmission power to maximize the overall throughput, minimizing interferences? Several algorithms could be employed here, e.g. in [7] the power is adjusted to keep the number of neighbors within the max and min thresholds. On the other hand [8] concentrates on improving the 1-hop broadcast coverage by TX power adjustments; however in this study nodes are static and all of them use the same TX power (it was adjusted per all nodes).
3.6 Security and Privacy

Security is an issue that needs to be carefully assessed and addressed in the design of the vehicular communication system. Several threats potentially exist, including fake messages causing disruption of traffic or even danger, compromising drivers' private information, etc. The issues to be addressed include trust (vehicles are able to trust the messages they receive), resiliency (resiliency for interference, easy maintenance) and efficiency, e.g. real-time message authentication.

Privacy is also a major issue that will need to be addressed. Anonymity must be preserved - the communications should not make the vehicle tracking or identification possible for non-trusted parties. The lack of taking into account the privacy concerns at the early design stage could result in multiple law suits after the network is deployed.

If, as it is in the networking world, each node (vehicle) would carry a unique, permanent MAC address, then it could be possible to trace such a car and its driver. For that reason IEEE 802.11p introduces dynamically assigned MAC addresses, along with a mechanism for duplicate MAC address discovery.

3.7 VANET modeling and simulation

Road traffic has certain properties that can not be easily modeled in a straight-forward way, using the classical MANET approach. Vehicles do not move randomly but rather follow the road infrastructure; road signs, traffic lights and other cars influence nodes' behavior. Nodes move at high relative speed, network density changes very dynamically, depending on location, recent events (e.g. accidents) or time of day. Thus, one could either build a sophisticated road traffic mobility model on top of some popular network simulator (NS-2, OPNET, GloMoSim), or use mobility traces from another source. This could be either measurement-based road traffic traces or traces obtained from a third party vehicular traffic simulator (e.g. CORSIM, VISSIM).

An interesting attempt has been made in [9], where authors managed to interlink NS-2 simulator with VISSIM, a vehicular traffic simulator, together with application simulator based on Matlab/Simulink environment. This approach gives one more opportunity - a chance to observe how the VANET functionality affects the drivers' behavior, hence influencing the network parameters. Yet this solution is not optimal from efficiency point of view: instead of 3 different environments (running on different operating systems), having a single, uniform simulation environment with network, vehicular traffic and driver behavior models would improve computational efficiency and decrease complexity.

4. Economic issues

The Car-to-Car Communication technology bears direct network effects, i.e. to be able to profit from it, first a certain market penetration is required. Therefore cooperation among car manufacturers and even other parties such as government agencies is a must. It appears that when perceiving the C2CC technology as a platform for a variety of applications, different user groups, of which some are looking for specific solution to their problems today, can be satisfied [10].

There are two mechanisms that lead to a successful market introduction for consumer technologies: either there is a visible added value of the technology for the customer or a regulative order that does not leave alternatives, requires its use.
The problem with a regulative introduction is that, to be issued, the effectiveness of the technology has to be proven first. But in case of car-to-car communications, a certain market penetration is required before any effects or improvements can be shown. Hence, it cannot be expected that a regulative order is issued on the basis of promised safety and traffic flow improvements before the penetration is reached. The car-to-car communications market is thus unlikely to be driven by such a force. And for the added value there raises another problem: when a consumer can only take advantage of a technology once a certain market penetration is reached, no one will invest in this technology before this is the case, which again means that this penetration might never occur. It was estimated that in order to make the network usable, at least penetration of 10% is needed. Provided that 50% of all newly produced cars are C2CC enabled, reaching that 10% should take about three years. In comparison, company cars are resold in average after about 2.5 years. The respective owner would thus resell a car with a technology he never had the chance of profiting from. And 10% equipment rate has been mentioned here just as a lower bound [10].

The strategic idea is to introduce C2CC with help of Car-to-Infrastructure Communication (C2IC) applications, which cover also the areas of comfort and infotainment. The communication with infrastructure has the advantage that the fixed nodes can be installed independently from the market penetration rate. That way all the users could get some functionality right from the start. The system should be divided into basic C2CC functions, which is integrated in all vehicles, and optional C2IC applications. C2CC applications can be sold only once the required penetration rates are reached. Despite all the problems, there is a potential to successful introduction of the C2CC technology into the market. It requires though that all the interested players are coordinated, so that they can base their concept on the VANET platform and take fair part in the market acquisition. The respective harmonization should be brought about as soon as possible.

5. Current projects and standardization

The following section shortly explains main characteristics of the standardization process and research projects initiatives, focusing on current developments in Europe, US and Japan. It is foreseen that these solutions will eventually converge, leading to a common, worldwide VANET platform.

5.1 Europe

There were several projects held in Europe, joining partners from the industry, governmental agencies and academia. Topics covered within these activities include (among others) hazard warnings triggered by hazard flashers, elaborated within Inter-Vehicle Hazard Warning project (IVHW); cooperative driving which was addressed in CarTALK 2000, PROMOTE-Chauffeur and INVENT VLA projects, and driver information and warning issues addressed by PReVENT WILLWARN, SAFESPOT and FleetNet (extended in NoW and Car2Car CC). Currently the three top-priority challenges in Europe are: frequency allocation, protocol definition and infrastructure deployment. As for now it has been agreed to use a frequency spectrum for vehicular safety applications similar to the US (fraction of the DSRC band). Also adaptation of US/international protocols should be applied wherever possible.

Network on Wheels [11] is an ongoing (2004 - 2008) German research project funded by the German government. Its main partners include academia, automotive and IT industries. NoW aims at designing widely defined communication system, including various functionalities from earlier EU projects (based mostly on the FleetNet project). Standardization is to be done on European level in cooperation with the Car2Car CC. Radio system is based on IEEE 802.11 adapted to European market.
The Car2Car Communication Consortium [12] is a non-profit organization initiated by European vehicle manufacturers in 2004, pushing for further increase of road traffic safety. Its mission is to create an open European industry standard for Car2Car communication systems based on wireless LAN components and guarantee European-wide inter-vehicle operability. That includes proposing of realistic deployment strategies and business models to speed-up the market penetration.

5.2 USA - VSCC and IEEE 802.11p WG

Since 2002 the Vehicle Safety Communication Consortium (VSC) has been working on the development of DSRC standards, protocols and applications, applying inter-vehicle and road-to-vehicle communications. System proposals for the near future include traffic violation warning, curve speed warning and emergency electronic brake lights. Middle future proposals include pre-crash sensing (prepares for unavoidable collisions), lane change and cooperative forward collision warnings, left-turn and stop sign assistance.

VSC Consortium decided to adapt an existing IEEE 802.11 WLAN standard. Namely it incorporated 802.11a PHY layer, using orthogonal frequency division multiplex (OFDM) as well as MAC layer, CSMA/CA, with some adjustments. It operates in 5.9 GHz DSRC band, consisting of seven 10 MHz channels, one of which is assigned for C2CC [13].

IEEE 802.11p, also referred to as Wireless Access for the Vehicular Environment (WAVE) defines enhancements to IEEE 802.11 required to support Intelligent Transportation Systems (ITS) applications. This includes data exchange between high-speed vehicles and between vehicles and the roadside infrastructure in the 5.9 GHz band (5.85-5.925 GHz). It will be used as the groundwork for DSRC (Dedicated Short Range Communications), a US Department of Transportation project, looking at vehicle-based communication networks, particularly for applications such as toll collection, vehicle safety services, and commerce transactions via cars.

The 802.11p Task Group is still active. Per the official IEEE 802.11 Work Plan predictions the formal 802.11p standard is scheduled to be published in March 2008. The WAVE technology is often referred to as WLAN or CALM M5 in the subsequent documents, although the CALM acronym (Continuous Air Interface for Long and Medium range) originates from ISO TC204 WG 16 [14, 15].

5.3 Japan - AHS and ASV

The situation in Japan regarding the development and deployment of Co-operative Systems is more complex. Mainly because Japan is more technologically advanced than US or EU countries, thus there is already a very large installed base of in-vehicle systems with navigation (mostly VICS, real-time traffic and travel information service) and Electronic Toll Collection (ETC) on-board units [16]. In the Japanese approach the public sector builds the basic services for information and safety, which makes buying the on-board unit more attractive. The announced plan for the 2nd Stage of ITS Deployment (2007) is based on the concept of Universal On-Board Unit, on which all the services are based.

There are two main initiatives regarding vehicle co-operative systems: one is vehicle-based (ASV) and the other infrastructure-based (AHS). These are driven by the Ministry of Land, Infrastructure and Transportation (MLIT). The initiative aiming at combining both of these is called Smartway. The word smartway symbolizes a road or highway which enables a wide range of information to be exchanged among all its users, creating a platform for ITS services. The Smartway project consists of several
communicating systems, and a variety of sensors, like short and mid-range radars, cameras, ultrasound, infrared and others.

ASV stands for an Advanced Safety Vehicle and builds on C2CC. Its current generation is ASV-3. AHS stands for Advanced Highway Systems, and is promoted by Advanced Cruise-Assist Highway System Research Association (ASHRA). AHS is based on similar technologies to those of ASV, but unlike the ASV, it is concerned only with C2IC. For the time being the level of co-operation and integration between ASV and AHS programs is not known yet. Neither is it clear if their components will be promoted individually, or under the Smartway concept. However, it has already been decided to assign the 5.9 GHz DSRC band for vehicular communication in Japan.

6. Conclusions

This report lists some of the open research challenges that need to be addressed in order to complete the standardization process, adapting the system to VANET requirements. Also economical, legal and institutional issues still remain unresolved. Only after they are dealt with, the deployment phase can be initiated. The system could become fully functional within a few years, after certain market penetration is reached and sufficient roadside infrastructure built.

The technology promoted by the car industry both in the US (VSC Consortium, VII Initiative) and Europe (Car2Car Communication Consortium), will inevitably lead to one universal standard, based on WLAN/CALM M5/IEEE 802.11p, a vehicular version of the IEEE 802.11 WLAN technology. Although the IEEE 802.11 standards family was initially developed for the use of laptops and PDAs in hot-spots, it is relatively easily converted for vehicular use, moving to a different, licensed frequency band, common for all participating countries. Another benefit of this technology is that it would be cheap, due to large volumes, thus simplifying deployment and accelerating the market penetration.

7. References

IP MULTIMEDIA SUBSYSTEM

Daniel BROUQUET

Summary

IP Multimedia Subsystem (IMS) is a set of specifications that describes the Next Generation Networking (NGN) architecture for implementing IP (Internet Protocol) based telephony and multimedia services. IMS defines a complete architecture and framework that enables the convergence of voice, video, data and mobile network technology over an IP based infrastructure. It fills the gap between the two most successful communication paradigms, cellular and Internet technology.

1. Introduction

One of the most discussed challenges that worldwide carriers are facing today is the telecommunication networks convergence. Third generation mobile networks and broadband access to the Internet are spreading together with new challenges and opportunities. The IMS is the architectural model that is guiding the transition to the next generation of IP based telecommunication services.

The vision of IMS is to provide cellular access to all the services that the Internet provides. [1] IMS is defined by the 3rd Generation Partnership Project (3GPP) as a new subsystem, a new mobile network infrastructure that enables the convergence of data, speech, and mobile network technology over an IP based infrastructure. IMS was designed to fill the gap between the existing traditional telecommunications technology and Internet technology, which increased bandwidth alone cannot do. This will allow operators to offer new services that shareholders and end users are expecting. IMS enables these user-to-user communication services via a number of key mechanisms such as session negotiation and management, quality of service (QoS) and mobility management. [2]

2. Overview of IMS

2.1 Standardization bodies

IMS was initially defined by 3GPP, which is a collaboration agreement among a number of telecommunications standards bodies, as part of their standardization work for supporting GSM networks and radio technology evolution. IMS was first introduced in 3GPP Release 5, where "Session Initiated Protocol" (SIP), defined by the Internet Engineering Task Force (IETF), was chosen as the main protocol for IMS. It has been further developed in Releases 6 and 7 of 3GPP to include additional features like presence and group management, interworking with WLAN (Wireless Local Area Network) and CS (Circuit Switched) based systems, and Fixed Broadband access.

Another standards body, 3rd Generation Partnership Project 2 (3GPP2), also standardized its own IMS. 3GPP2 was born to evolve North American and Asian Cellular Radio-telecommunication Intersystem Operations into a third-generation system.

In addition to the 3GPP and 3GPP2, Open Mobile Alliance (OMA) plays an important role on specifying and developing IMS service standardization. The services defined by OMA are built on top of IMS infrastructure, such as Instant Messaging (IM), Presence service, and Group Management Service. [3]
2.2 Architecture of key functional nodes

The IMS architecture is a collection of standardized functions linked by standardized interfaces. This means that implementers are free to merge two functions into a single node or to split a function into two nodes. The figure 1 shows the IMS architecture.

The nodes included in the IMS are as follows:

- One or more databases, called HSSs (Home Subscriber Servers) and SLFs (Subscriber Location Functions). Those databases contain all user-related informations.

- One or more SIP servers, known as CSCFs (Call/Session Control Functions). The CSCFs, which are SIP servers, process SIP signaling in the IMS. There are three types of CSFCs:
  - The P-CSCF (Proxy-Call Session Control Functions), which is the first point in the IMS for the User Equipment. The major role of P-CSCF is to route information to the correct S-CSCF (Serving-Call Session Control Functions). This node can also host the PDF component (Policy Decision Function) which is responsible for policy decisions.
  - The S-CSCF, this is the central signaling node, which registers the users and provides service to them
  - The I-SCSF (Interrogating-Call Session Control Functions) is able to determine the S-CSCF from which users should register

BROUQUET, D: IP MULTIMEDIA SUBSYSTEM
one or more Ass (Application Servers), an AS is a SIP entity that hosts and executes services, it interfaces S-CSCF using SIP

one or more MRFs (Media Resource Functions), which provides a source of media, the ability to play announcements, mix media streams, transcode between different codecs, obtain statistics, and do any sort of media analysis

One or more BGCFs (Breakout Gateway Control Functions), only used in sessions that are initiated by an IMS terminal and addressed to a user in a circuit switched network. Its main functionality is to select an appropriate PSTN/CS (Public Switched Telephone Network) gateway or an appropriate network where interworking with the CS domain is to occur

One or more PSTN/CS gateways, which provides an interface toward a circuit switched network, allowing IMS terminal to receive and make calls from and to the PSTN/CS network. The gateway is decomposed into three different functional nodes, the MGW (Media Gateway) which interfaces the media plane of the CS network, the SGW (Signaling Gateway) which interfaces the signaling plane of the CS network, and the MGCF (Media Gateway Control Function) which is the central node of the gateway, it controls both MGW and SGW. [4]

3. Signaling in IMS

3.1 SIP

IMS uses the SIP protocol for multimedia session negotiation and session management. IMS is a mobile SIP network able to provide routing, network location, and addressing functionalities.

In contrast to the CS and PS (Packet Switched) Domains, the IMS domain enables any type of media session to be established (e.g. voice, video, text, etc.). IMS also allows the service creator the ability to combine services from CS and PS domains in the same session, and for sessions to be dynamically modified ‘on the fly’ (e.g. adding a video component to an existing voice session). This capability opens up a number of new and innovative user-to-user and multi-user services such as enhanced voice services, video telephony, chat, push to talk (PoC) and multimedia conferencing, all of which are based on the concept of a multimedia session. The figure 2 shows a simple SIP transaction between two SIP terminals. [3]

![Figure 2: Basic INVITE-ACK SIP transaction](image-url)
There are several major advantages to building a new SIP-based feature or service:

- **Simplicity** — SIP is based on a straightforward request/response interaction model, making it a simple and comprehensible protocol for developers to implement. The messages are also text-based, which makes them easy to parse, create, read, understand and debug.
- **Extensibility** — SIP can set up sessions for any media type, such as voice, video and application sharing.
- **Flexibility** — SIP allows developers to interact with individual protocol messages without breaking interaction boundaries. This allows developers to concentrate on creative application development within the context of IP communications.
- **Familiarity** — SIP borrows heavily from HTTP and other Internet standards from the IETF, which allows SIP applications to be developed using web-like technologies. SIP development looks and feels a lot like Web development.

IMS has leveraged the SIP signaling to allow delivery of multiple applications to users across any access network. The Call Session Control Function is a central component to signaling and control within the IMS network. [5]

### 3.2 Session Description Protocol

The main goal of SIP is to deliver a session description to a user at their current location. Once the user has been located and the initial session description delivered, SIP can deliver new session descriptions to modify the characteristics of the ongoing sessions and to terminate the session whenever the user wants. [3]

The most common format for describing multimedia sessions is the Session Description Protocol (SDP), SDP is a textual format to describe multimedia sessions. Figure 3 shows an example of an SDP session description that Bob sent to Alice.

```plaintext
v=0
o=bob 2890844526 2890844526 IN IP4 ws1.domain2.com
s=let's work at the library today
c=IN IP4 192.0.100.2
t=0 0
m=audio 20000 RTP/AVP 0
a=curr:qos local none
a=curr:qos remote none
a=des:qos mandatory local sendrecv
a=des:qos mandatory remote sendrecv
```

Figure 3: Example of an SDP session description

This SDP session description contains the version of the protocol (v=0) the owner of the session (o=bob), the subject of the session (s=let's work at the library today), the connection informations (bob uses IPv4 and his address is 192.0.100.2), the time of the session (t=0 0 means that the session is supposed to take place at the moment this message is received), and media level informations (line m=audio for a stream audio, and the others lines a= for this stream informations, in this case QoS prerequisites).
3.3 Registration

IMS provides mobility management. A user is identified by a SIP URI, we can distinguish the public URI and the current URI. Let us take a look at the example of the Figure 4.

Alice owns a public SIP URI (sip:Alice.Smith@domain.com) and two possibilities for her current SIP URI (at the university sip:alice@pc12.university.edu and at the company sip:asmith@ws1234.company.com). With her public URI Alice can be reached whatever her current URI. The registrar stores her current URI and forward incoming request to Alice wherever she is at her current URI. [3]

4. IPv4 and IPv6 in IMS

When IMS was initially designed, 3GPP has proposed Internet protocol version 6 (IPv6) as the IMS IP version. 3GPP thought that by the first IMS implementation would go to operation it would be the common IP version. And because SIP and its associated protocols were well known examples of protocols that suffered problems when traversing NATs (Network Address Translation), which is required for a large scale deployment of IPv4. [3]

But the industry was not ready to migrate to IPv6 when the first IMS product appeared, and even now the IPv4 is the common IP version. The 3GPP was forced to look for viable alternatives and they decided to allow early deployments of IPv4 for IMS architecture.

To avoid a dual stack implementation of IP versions on each node of the architecture, 3GPP introduced two new functional nodes to the IMS architecture, the IMS Application Layer Gateway (IMS-ALG) and the Transition gateway (TrGW).

- IMS-ALG processes control signaling
  - This is an application specific functional entity that allows an IPv6 node to communicate with an IPv4 node and vice versa. For IMS, an IMS ALG provides the necessary application functions for SIP/SDP protocols in order to communicate between IPv6 and IPv4 SIP applications.
The TrGW processes media plane traffic

- TrGW is a NAT-PT/NAPT-PT and uses a pool of globally unique IPv4 addresses for assignment to IPv6 nodes on a dynamic basis as sessions are initiated across the IP version boundaries. NAT-PT binds addresses in IPv6 network with addresses in IPv4 network and vice versa to provide transparent routing between the two IP domains without requiring any changes to end points. NAPT-PT provides additional translation of transport identifier (TCP and UDP port numbers). [6]

5. IP Connectivity Access Network and PDP Context

Before getting access to the IMS services the user has to get access to an IP CAN (Connectivity Access Network). An IP CAN can be a UMTS/GPRS, ADSL or WLAN network. IP CAN provides access to the IMS network. [3]

In the case of GPRS, the IMS terminal first gets the connectivity to the GPRS network due to so-called GPRS Attach Procedures and then establishes a PDP (Packet Data Protocol) context for IMS related signaling, registration, and other procedures for IMS sessions. The figure 5 shows the signaling transactions before getting access to the IMS services in the case of GPRS.

A PDP context means an IP configuration and bearer QoS attributes. In the case of IMS a first PDP context is send for the signaling, it is the primary PDP context. The IMS terminal can establish additional PDP contexts, referred to as secondary PDP contexts, related to resource reservation flows to send and receive data with QoS characteristics. When a user wants to start a new service in his current session he can send another PDP context, for instance if the user is currently in an audio session and wants to start a video telephony, a chat, a push to talk, etc.

6. Core IMS within NGN

Next Generation Network is a packet-based network able to provide telecommunication services and able to make use of multiple broadband, QoS-enabled transport technologies and in which service-related functions are independent from underlying transport-related technologies.
3GPP have focused on making sure that their radio access networks were ready to accept IMS services, and NGN provides easily access to IMS services from fixed broadband access. We find a core IMS in the NGN architecture, this core IMS is the equivalent of the IMS architecture designed by 3GPP. But some changes can be found. [7]

The figure 6 shows the Core IMS architecture in NGN.

**Figure 6: Core IMS in NGN**

The Core IMS in NGN contains:

- Two new functional nodes for the transport layer which complete the function of the PDF (Policy Decision Function) in the P-CSCF: NASS (Network Attachment SubSystem) and RACS (Resource and Admission Control Subsystem)
  - **NASS** is responsible for supplying the terminal with an IP configuration, providing authentication at the IP Layer, authorization of network access and access network configuration based on users’ profiles, and the location manager at the IP layer.
  - **RACS** is responsible for providing resource management and admission control (among others, gate control functionality, policy enforcement, and admission control based on users’ profiles)
Two new functional nodes to facilitate the interworking with other networks:

- **IBCF** (Interconnect Border Control Function): This new functional entity introduced by NGN acts as a separation between two networks. IBCF provides more formalize boundary between a Home Network and a Visited Network and it may provide functionality similar to the IMS-ALG in case of IP version interworking. It is also responsible to insert an IWF in the path if needed.

- **IWF** (Inter-Working Function): IWF provides signaling protocol inter-working between the SIP-based IMS network and other service provider networks using H.323 or different SIP profiles.

### 7. Conclusion

In this report, a description of IMS architecture and some of its features are given. An IP–based services architecture would be a key element for future cellular networks, and SIP will be a standard protocol to cater to all 3GPP multimedia applications. It ensures that future multimedia services will be truly IP–based. All the key components for IP convergence are commercially available today, and IMS infrastructure equipment is also available.

### 8. References


http://www.iec.org/newsletter/may07_2/analyst_corner.pdf , 25.05.2008


Summary

The operating environment of network and mobile operators is changing. This paper examines the evolution of the current operator model due to pressure from access network technologies diversifying. As a case example, 802.11 Wireless LANs are examined in more detail.

1. Introduction

Currently, network operators have been in tightly conond roles: They are usually either traditional fixed-line ISPs, mobile operators providing mobile phone access, or backbone providers. However, as network access technologies become more heterogeneous and terminals support multiple access technologies, these traditional borders are coming down. On the terminal side, problems arise from network selection and roaming/handover decisions.

In this report, the viewpoints of the operators and their roles are first considered in Chapter 2. End-user viewpoints are considered in Chapter 3. Chapter 4 includes a case-example of the problems rising in the context of IEEE 802.11 based wireless LANs. Finally, a short conclusion is done in Chapter 5.

2. Operator viewpoint

Traditional operator business has been closely based on “Walled garden”-approach; Services such as voice or data have their distinct billing methods and pricing models. Increased pressures from the Internet’s “flat-service” model where all services are founded on IP-based technologies are causing a split from this model. Ambient networks project [1] has supported the view that operators will be split to three basic roles, as shown in Figure 1.

As of now, there are service operators and network operators. However, the relationships are different; a group of service operators is bound to a single network operator as their “home network”. There is no true multi-access.

In the future, the roles will become more clearly orthogonal; Service providers only manage user identities, multi-access operators will provide access technologies to various service providers and network operators will simply act as bit-pipes to the multi-access operators.

When deploying and implementing a network, the operators have certain preferences that help in maintainability, cost structure and scalability. These requirements are common for all technologies and include

- Centrally managed access deployment
- Low cost, “lightweight” access points/antennas
- Security moved away from access points
- No restrictions for host-to-host communication when both are within the managed access network
- AAA (Authentication/Authorization/Accounting) routing
- Network selection problem
- Minimal handover signaling, especially avoiding re-authentication
- Minimal bootstrapping signaling

In the traditional GSM/3GPP world these targets have been, for most part, reached, or they have been worked around (e.g. using out-of-band signaling). A case example of problems when applying these preferences to wireless LANs is given in section 4.

3. End-user viewpoint

When end-user’s multi-access terminal has numerous available networks to choose from, network selection problem[2] becomes prevalent. Furthermore, the end-user may have multiple identities or roles affecting the decision.

In traditional cellular networks such as GSM or UMTS[3], where there usually is only one provider, wireless point of attachment choice is generally easy; it is based almost solely on signal strength. In roaming cases, there is a long-standing, clearly defined method for choosing the correct roaming partner [4] without requiring any end-user intervention. Furthermore, the only vertical handovers are 2G <-> 3G handovers, which are also defined by 3GPP.

MÄKELÄ, A.: Roaming, Handover and Security Issues with Wireless Operator Networks
When multi-access terminal supports network technologies that have not been designed to work together from the start, problems tend to arise. Furthermore, there may be an abundance of networks to choose from which escalates the issue further. Decision points affecting network choice include, but are not limited to, the following:

- Target network’s traffic characteristics
  - Latency
  - Bandwidth
  - Free capacity/degree of congestion
- Coverage
- Pricing model and cost
  - Roaming partner of home network
  - Data-transfer-based
  - Time-based
- User’s chosen identity
  - Personal, corporate, anonymous identity
- Services that the user intends to access

The user may, for example, prefer to pay for better quality of service if he is intending to transfer large files; On the other hand, simply checking for e-mails can be conducted via congested, cheaper network. Roaming agreements may cause pricing schemes offered by networks to change depending on chosen identity as well.

The roaming agreements may also cause conflict in other ways; Local operator might see more profit in offering network access to the end-user directly instead of via the roaming partner.

While the decision-tree is much more complex, the end-user experience should still be relatively automatic; User sets his preferences and the terminal automatically chooses the best network.

4. Examples of technical issues related to IEEE 802.11 wireless LANs

IEEE 802.11 wireless LAN standard has evolved significantly since its inception. The original standard[3] was planned for 1-2 megabit transfer rates and even included support for additional transports besides radio interface; PHY specification includes, amongst other things, infrared transmission, which as fallen into disuse.

In subsequent evolution, the standard has been amended with increases to data rate (802.11b, a, g and n) and security (802.11i). At the time of this writing, all of these amendments except 802.11n have been included in the published IEEE 802.11-2007 protocol.

While 802.11 is very interesting technology from operator point of view, the requirements given in section 2 still apply. 802.11 as a basic standard does not take these into account; The solutions are still partially in development.

4.1 Centrally managed access deployment

The original standard does not in any way support the concept of centralized deployment. All access points are considered as stand-alone units; only the extended service set (ESSID) attribute can be considered spanning multiple access points. From deployment point of view, this architecture results in a considerable management overhead, as each access point requires full IP reachability, access controls,
AAA configuration and fault management. The access points also must have costly hardware to support all the intelligence within.

As a solution, “Wireless LAN switches” or “controllers” have become available. Most prominent vendors are Cisco Systems and Aruba. These are based on a lightweight access points and a centralized controller. The access points communicate with the controller using a special protocol that effectively tunnels all the end-user traffic throughout the network and includes a separate control channel for management. Cisco has published their proprietary LWAPP protocol as an Internet-Draft[6] as part of the IETF’s CAPWAP working group. CAPWAP WG intends to standardize the protocol between controllers and access points. At the time of this writing, the protocol is still heavily work-in-progress; No RFC’s have been published.

4.2 Signaling and security

The original 802.11 attachment protocol is relatively simple, requiring, at best, only two roundtrips total, consisting of authentication and association requests and respective responses. However, the original built-in security mechanism, WEP, which utilized this simple signaling has been proven broken[7] and is now deprecated by IEEE. Furthermore, WEP was based on static, pre-shared keys. Adding functionality to per-user authentication and encryption has caused the amount of signaling traffic to increase heavily.

The current security model for Wireless LANs was originally introduced in 802.11i, now part of the 802.11-2007 protocol. The model is based on 802.1X authentication; The authentication procedure also distributes encryption keys to be used between clients and access points. Cipher suites include TKIP and AES.

The 802.1X is essentially a transport for Extensible Authentication Protocol, or EAP. As the name implies, EAP is extensible, resulting in numerous authentication methods; These can be based on traditional usernames and passwords, certificates or even SIM cards. Each EAP method defines a distinct signaling pattern. After EAP authentication is complete, the encryption keys will be delivered using EAPOL-KEY frames.

With these considerations in mind, establishing a secure, authenticated connection to 802.11 wireless LANs requires at least the following roundtrips:

- (Probing), 1 roundtrip
- Open authentication + association, 2 roundtrips
- EAP signaling (2-20 roundtrips, depending on method)
- EAPOL-KEY frames for personal key (2 roundtrips)
- EAPOL-KEY frames for group key (1 roundtrip)
- IP layer signaling: DHCP and/or IPv6 autoconfiguration (1-4 roundtrips)

The foremost problem with the signaling is not initial attachment; Primary problem is the requirement to complete the full signaling after each and every handover.

A currently existing solution was included in the original 802.11i amendment; a mechanism known as Opportunistic Key Caching. When utilizing OKC, the client still has to perform the full signaling against any previously unvisited access points; However, if performing a handover to an already visited access point, bulk of the signaling may be skipped. In practice, OKC has been shown to work
only in very limited scenarios, such as clients constantly moving back and forth within same premises. OKC does not assist a client which visits each access point only once.

There is lots of research going on to improve the handovers; primary driver as of now is 802.11r amendment, known as “Fast BSS Transitions”. Another approach, based on HIP is presented in [8].

4.3 Network selection problem

Choosing the network which to attach currently presents a huge problem in 802.11. The only distinct identifier for a network is SSID. Apart from the SSID, only characteristics available to an end-user scanning for networks are the security level in use and signal strength. There is no indication on roaming agreements, prices, or even if a particular network provides Internet access.

The RFC 5113[2] provides a comprehensive overview of the network selection problem, and also includes comments applicable to 802.11.

The network services information can essentially be distributed in one of three phases: Before association, before authentication or post-authentication. As seen in section 4.2, authentication process causes heavy signaling load, thus post-authentication is not a valid option. Pre-association, e.g. embedding network information in 802.11 beacons is not feasible due to 802.11 beacons always being sent at base rate, usually 1Mbps. Thus, the beacons, which are sent 10 times/second, cannot hold much information. The remaining choice, post-association, thus becomes the desired option.

In IEEE, amendment 802.11u is a work-in-progress standard that essentially defines[9] a Generic Advertisement Service (GAS) mechanism, which can be used to carry network service information. The definitions include information elements on pricing model, network location, access, and services.

Another issue stemming from network selection is multiple roles for a single user. The RFC 5113 includes, as an example, a user which has an account from an ISP as a private user, and a corporate account from workplace. Both the corporation and the ISP have varying roaming agreements. Thus the user needs to make a decision on what is the optimal network to roam into.

4.4 AAA routing

The end-user must always be authenticated at the home AAA provider. However, compared to the current GRX (GPRS Roaming eXchange) roaming architecture, where the number of providers is longer limited, the huge number of Wireless LAN providers quickly causes scaling to become an issue. As of now, the AAA architecture is based on RADIUS, which supports only static routing. If a new provider joins the roaming network, all partnering operators have to update their RADIUS routing rules.

The DIAMETER protocol[10], intended to become successor of RADIUS, fully supports dynamic routing, thus solving the basic issues related to scalability. However, certain issues still remain; The base protocol basically assumes full-mesh connectivity. In practice, networks usually have numerous policies in place detailing which specific AAA messages from which specific realms are actually forwarded. Furthermore, since there is no guarantee of messages applying to a single session always following same route, accounting by intermediate operators becomes difficult.

One solution is presented in an Internet-Draft[11] submitted for the dime working group in IETF. The draft addresses these issues by defining two new Attribute-Value Pairs, “Explicit-Path-AVP” and
“Explicit-Path-Record-AVP”. These two AVPs essentially implement loose source routing to DIAMETER; When initial request from access point (typically Authentication request) is sent, the traversed route is recorded. On subsequent messages, the initial route can be reused. However, discussions on the dime mailing list resulted in the draft being withdrawn. The issue will be addressed at a later date.

5. Conclusions

Operator role will change as the network technologies will become more heterogeneous. The amount of choices available to the end-user will increase rapidly; however, to alleviate the complexity of these issues, architecture to transparently provide best service has to be implemented. Operators must strive to find a profitable business model in a world where network access is ubiquitous; possible choices include competing with better Quality of Service, pricing, ease of access or comprehensive identity management.

6. References

USER-CENTRICITY IN PERVASIVE NETWORKS

Sari KINNARI

Summary

In this article, we discuss user-centric aspects, designing usability and security in pervasive networks. The importance of the user-centric approach in designing pervasive networks should not be underestimated. Since data security mechanisms are efficient only if they are used correctly, we need to design the systems having the user in mind: as long as any security measures and/or other configuration activities are required from the user, we need to take into account the user and the context of use already from the beginning of the design process.

1. Introduction

Why do we need to talk about user-centricity in pervasive networks? Aren’t the context aware environments “intelligently monitoring the objects of a real world (persons, things, places) and interacting with them in a pro-active, autonomous, responsible and user-authorized way”? [1]. Answer is yes and no. Although living in pervasive networks is already reality to some extend, e.g. households have their own home networks, people share personally created content in the Internet, etc., the end-user is too often forgotten. There are numerous examples of bad design solutions that force the users to either lose their nerves or to contact a skillful computer geek to get help. Even nowadays, configuring a 3G Nokia handset to send emails may require the overweening 78 steps from the user [2].

The research area of Human-Computer Interaction (HCI) strives for understanding the intersection between humans and computers. It has traditionally concentrated on discovering usability problems that complicate the use of a product or a service. In pervasive networks user’s privacy, trust and security issues are at least as important as traditional usability. However, security and usability are often considered opposite goals that can not be designed nor achieved betweenwhiles, although the need to do that is increasing rapidly. In the future visions, networks and technologies will be highly intelligent observing e.g. the context of use while the mobile user will be able to concentrate on his/hers tasks and forget the complex cryptographic security configuration of the devices [3]. While waiting for those days to come some basic characteristics of human beings and their behavior should not be forgotten. Cognitive psychology has long tradition in research of those issues. As long as, for example, security measures and/or other configuration activities are required from the user, we need to take the user into account when designing the pervasive networks and devices used in them.

European Commission has defined its future vision to contain user generated content creation, management and consumption systems, where the content follows the user, regardless of the physical location or the device used for consumption [4]. Although in the future, many smart environments will act on behalf of their users, the user must be given the control over his/hers privacy and security management.

This report is an overview of user-centricity in the broad area of the pervasive networks. It is organized as follows. First, we briefly present the user-centric approach in designing interactive systems. We then continue by discussing the main challenges in pervasive networks from the user’s point of view. After that, some views of the development of future home networks are given. Finally, some concluding remarks are made.

KINNARI, S: User-centricity in Pervasive Networks
2. User-centered design in brief

The core of the interaction design is said to be the following: put the user first, keep the user in the center and remember the user at the end [5]. In this section we discuss the methods of designing products for the user.

2.1 Standards for design

The user-centered design processes and goals have been tried to standardize. In one part of ISO 9241 standard [6], usability is defined as the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments. Having to go through the previously mentioned 78 steps is neither efficient nor satisfying from the user’s point of view.

ISO 13407 standard [7] specifies the human-centered design processes for interactive systems. It emphasizes active involvement of users and clear specification of user and task requirements. The design solutions are tested and iterated several times along the design process. Figure 1 presents the approach of user-centered design by VTT, Technical Research Centre of Finland [8]. The design process starts by defining the user requirements and the initial context of use by which the design requirements will be produced. Usage scenarios, paper prototypes and mock-ups are being produced and evaluated. The final prototype is typically evaluated in laboratory tests and field trials to produce a final, usable product.

![Fig. 1 VTT’s Human-centered design approach](image)

Although the ISO standard has been criticized for not taking modern software engineering aspects into account, it still is valid for designing products and services because it is based on understanding the context of use and user requirements [8].

2.2 Multidisciplinary methods

HCI is a highly multidisciplinary science and the research methods are versatile. For example, cognitive psychology has studied user’s mental models already decades and HCI field has continued
that work. The user’s mental models determine how s/he acts when using the product, system or service. Models describe how the user perceives the problem space and the functionalities of the system, how s/he reasons and what are the goals of the user [9]. By studying mental models we can understand what makes the user interface easy to learn and use. If we design products against these models obeying only the functions and logic of technologies, problems will certainly occur.

User scenarios are typically short stories of the user acting in certain context and using the system or device. Personas are rich pictures of imaginary persons who represent the core user group [5]. However, creating personas is challenging nowadays since the users of different media are very heterogeneous, for example within the context of user generated content creation or new social interaction technologies.

HCI and user-centered design utilize versatile research methods, many of them originally developed in social sciences. When designing new forms of interaction in the modern societies, field studies are valuable source of information, as shown for example in [10]. It is essential to identify users’ needs starting from their everyday life. Thus, ethnographic style methods are natural for this research. In [11] they presented an example of how the results of ethnographic studies of home routines can be applied in practical technology design to support information management. Venkatesh et al. [12] have addressed specific needs of families in the context of their domestic activities and routines, and designed an information infrastructure that uses Internet technologies for home management and external networking, as well as meeting the emotional needs of the family.

Traditional ethnography requires large amount of resources, especially time and money, to be spent on fieldwork. However, there are methods, such as Quick and Dirty Ethnography that take traditional anthropological fieldwork techniques and apply them to contemporary applied research problems, as in [13]. Its goal is to seek relevant information about the users as quickly as possible and it also accepts the impossibility of gathering a complete and detailed understanding of the setting. Contextual inquiry introduced by Beyer and Holtzblatt [14] is a field study method for understanding customers, their motivations and strategies. It helps a cross functional design team to come to agreement on what the users need and how to design a system for them. This is crucial for designing new products for households.

To fully understand current practices and motivations for recording and sharing experiences through mobile media, Jacucci et al. [15] performed an ethnographic Investigation at a large-scale event, the FIA World Rally Championship in Finland in August 2004. They wanted to study how technology can support social agency operating through phenomenal and functional consciousness, and how it can change the cultural relation between experience and expression. Jacucci et al. argued that “multimedia is no more seen only as a mean to record and later re-live experiences but as a way to actively exercise agency and construct experiences with others.” Two participant groups used mobile phones while following the rally and the content created was analyzed afterwards by [15].

Each group was shadowed by a researcher using video camera. Different categories of temporality in which situation the creating and sharing of experiences was most relevant were found. Also, multiple levels of mobility, and notices of sociability, group identity and active spectatorship were found. The participants used mediated multimedia in story-telling, competition, joking, communicating presence and portraying others. These results produced guidelines for designing interaction with the devices themselves and also as a general design tactic technology could be made more prominent in the new mobile services instead of hiding or moving it to the background as usually suggested in pervasive environments. [15]
3. Design challenges

What are the biggest challenges in designing for users in pervasive networks? Abowd and Mynatt [16] list some issues: firstly, designing a continuously present computer interface. So far, these opportunistically acting computational systems have been presented as agents. Secondly, we need to create interfaces that can operate at different levels of the user’s periphery of attention with mechanism for the user. Thirdly, in pervasive networks, human activity is coordinated across two spaces: the virtual, digital world and the physical world. It is necessary to understand how people conceptualize their endeavors in these spaces. Finally, HCI methods lack the support for designing for informal, peripheral, and opportunistic behavior. This support has to be created as an on-going effort, and the methods should also contain feasibility evaluation of the systems driven from a user-centric perspective and the social implications of the pervasive networks. Admittedly, one of the biggest is also the heterogeneity of the users.

Figure 2 presents the basic components of a context-aware system interacting with a user by [17]. Although it originally emphasizes the research of context aware homes, it also illustrates the main building blocks of pervasive networks from the user-centered point of view: Technological elements, sensors, wireless networks and user interface collect context information and enable interaction. Middleware provides decoupling from the actual implementation of the sensors and networks, processes it and manages context information and privacy. Applications gather knowledge to infer what user expects and then deliver the expected services. To assure an enjoyable user experience, the context-aware system should meet the expectations of the user, with minimal frustration. The privacy of the user must be taken into account from the beginning of design process.

![Figure 2 Basic components of a context-aware system by [17]](image)

3.1 Security

Security in the pervasive computing environments is certainly essential. However, it has been argued that hackers pay more attention to the human link in the security chain than security designers do [18]. Different detailed scenarios of smart environments have been written and they seem to have one major property in common: they focus on the interaction of the end users with the system [19]. Very often an average user is unskilled, but s/he is required to understand security and privacy models of pervasive
networks. Additionally, most users just want to finish their task, not caring about technology or security solutions on the background of the system, as shown is Figure 3.

Data security mechanisms are efficient only if they are used correctly. A reasonable level of system security cannot be achieved without taking into account the users and their interaction with the system [19].

Höhn [19] has classified the security-related usability problems in four classes depending on the range of damage:

1. Usability problems not endangering the system security (e.g. lack of ergonomics)
2. Usability problems that compromise the system security despite the user’s adequate security competence
3. Usability problems arising directly from the user’s insufficient security knowledge
4. Security problems not arising from user interaction

Clarke [21] has argued that many of the social consequences of poor interaction design could be solved if the key guidelines were followed, first of them being that the target of all design and development activity should be the majority of users, not the extreme end of technologically able spectrum. Also, by default only the most basic functionality should be enabled at startup. Thirdly, new ICT products and services should not be deployed without assessing their social impact on society.

Improving the observability of the system has been suggested as one way of increasing security from the user’s point of view. We should remember that “if the user can not understand the functionality designed into the system, it doesn’t exist” [22]. Asking user to decide for the security actions is a double-edged sword; on one hand giving the control to the user is desirable, on the other hand, building something that relies on user education to be effective can be a recipe for disaster. As security researcher Amir Herzberg puts it: “Defend, don’t ask” [23]. Höhn [19] has presented one solution for visualizing the security model to the end users (their scenario took place in smart supermarket environment). Users were given a mock-up interface where individual devices were represented by symbols in the map of the covered area and mobile devices were represented by their actual location. The application also allowed a trial and error approach of clueless people; the users were able to simulate certain actions and observe their outcome before actually performing them.
3.2 Privacy

Privacy is usually considered one of the biggest threats with ubicomp and pervasive networks. As noted earlier in Section 2.2, HCI research typically utilizes theories and frameworks in a multidisciplinary way. For example, Lehikoinen et al. [24] have used Irvin Altman’s existing framework of privacy in social sciences [25] in studying privacy in ubicomp arena. Altman emphasizes the role of the environment in the process of privacy regulation. The framework has been widely referred in various contexts and according to [24] it was also applicable in ubicomp environments. The salient idea in Altman’s framework is that privacy is “a two-way process, in which the self and the others interact with each other”. A person (the self) desires an ideal level of privacy with others at particular time and social setting [24].

Lehikoinen et al. analyzed their typical user scenarios based on Altman’s framework. They presented the examples of control mechanisms in different forms of interaction and different communication context with some extensions to Altman’s original framework (Table 1).

<table>
<thead>
<tr>
<th>Other humans</th>
<th>Other’s devices</th>
<th>Interactive environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self</td>
<td>Personal spacing</td>
<td>Personal spacing (proximity).</td>
</tr>
<tr>
<td></td>
<td>Territorial responses</td>
<td>Verbal and non-verbal, behavior</td>
</tr>
<tr>
<td></td>
<td>Verbal and non verbal behavior</td>
<td>(audio, gesture UI’s)</td>
</tr>
<tr>
<td>Personal device</td>
<td>Personal spacing</td>
<td>Filtering,</td>
</tr>
<tr>
<td></td>
<td>(Hiding the device, adjusting volume, brightness of display, etc.),</td>
<td>Context awareness,</td>
</tr>
<tr>
<td>Te rritorial responses</td>
<td>Management of access rights,</td>
<td>Management of “visibility”,</td>
</tr>
</tbody>
</table>

Table 1 Privacy control mechanisms in ubicomp interaction by [24]

Not surprisingly, Lehikoinen et al. found that ability to control interaction diminishes in mediated communication in ubicomp environments, since many ubicomp services follow predefined polices with limited means to regulate the interaction. However, when proposing new privacy regulation mechanisms, this kind of user centric aspects have to be considered thoroughly.

3.3 User experience

On this decade the HCI research has been turning to the direction of user experience design. According to the concept, using technology should be a truly satisfying and uplifting experience. Jordan [26] has proposed the consumer needs hierarchy to define the goals for design: on the first level, there is functionality which contains the necessary qualities to perform a function. Above it, there is usability which defines how easily a product can be used on a satisfactory way. Third level contains the large concept of user experience; context of use, pleasure, meeting the expectations of the user, creating a relationship with the product and building the social identity of the user, etc.

Various definitions of user experience has been written, e.g by [27] and [28]. According to Spohrer and Stein [29] the key challenge in pervasive networks is minimizing transcriptions and redundant data entry of information into multiple systems. Frustration prevents pleasurable user experience. In the future, the user experience will increase its importance, since technologies have evolved and matured and the competition for the users requires the proper design of user experience.
4. Models of evolving home networks

Home networks begin to be reality in many households. Distinctive for them is that the end users are expected to become their own systems administrators. The support and security systems of the possible employer don’t cover private mobile devices and home networks. This section discusses the models proposed by Shehan and Edwards [30]. They have stated that HCI community is required in shaping the usability of current and future networking technologies and have defined six models for how networking in the home environment may evolve.

The Fresh Start Model proposes that networking protocols should be rethought for ensuring better usability. The authors do emphasize that naturally this kind of change does not happen drastically. The Bandage Model relies on improving the existing technology by clever UI design and extensions to interfaces and tools. Not only householders themselves, but also e.g. service providers and consumer electronics producers, might benefit from the results. Nevertheless, the researchers warn that deeper usability of a home network requires more than visualization tools and network analysis tools, since they only show what is functioning. If part of the home network is broken, it is very difficult to visualize it with any of the tools.

The Gateway Model proposes that a single gateway device be placed between the home network and the Internet. Thus, the home network could run different, “user-friendly” protocols not dependent on the outside world. The idea of new protocols breaks the compatibility with existing networking devices, which some users might find annoying. The Outsource Model describes a trend similar to maintaining domestic appliances and cars: let a professional do the work. However, paid outsourcing of home network may encounter problems. Technicians would need a thorough documentation of the network and configurations. Also, Shehan and Edwards remind us about deeply private information that home networks often store. Would we want a technician to see everything?

The Utility Model refers to the ease of use of our everyday utilities such as electricity and the telephone. However, there are such fundamental differences between those and home networking technologies that the Utility Model may not be realistic. The Ubiquity Model proposes that the network is provided by either a commercial or community-sponsored entity. It is based on the idea that householders would not have to concern themselves with the details of the network itself, having to only manage the connectivity to the network. Nevertheless, Shehan and Edwards underline that the users would need new mechanisms to “add back the walls” of their networks if they use municipal-scale WiFi, 3G data services or WiMAX-based systems.

Shehan and Edwards suggest that the HCI community explore and create visualization systems and tools that are better aligned with the needs of householders. Presently, many network analysis tools give far too detailed technical information to an average householder. In addition, new interaction techniques could be layered on top of existing configuration and management protocols. Shehan and Edwards conclude that in order to improve the usability and accessibility of networking for all members of society, the HCI and networking communities need to reflect deeply on these possible models and to engage in future networking issues.

There are examples of success stories in developing services and products by having the user in mind. Elmore et al. [31] have developed a software guide for setting up a secure home network with Linksys routers while guaranteeing a successful and pleasant user experience. They wanted to create a solution that does not require a high degree of technical knowledge. The central theme in the solution was to imagine a networking expert advising a user on how to set up, configure and secure a home network; what would this person tell the user to do? They wanted to avoid the “educate the user about home
networking” mentality. The EasyLink Advisor was a case study which was started by investigating users’ needs and desires related to home networking. The research was triangulated by ethnographic style observations, field interviews, content analysis and heuristic analysis of existing tools. The analysis revealed that most of the problems can be addressed during initial setup and configuration. Also, most customers do not employ wireless security and are not aware of the risks.

5. Conclusion

In this report we presented the some of the main challenges of user-centricity in pervasive networks. Some solutions were also presented as examples of user-centric approach. Although the challenges span across technical, social and pragmatic domains, we believe that by raising the awareness of user needs and possible problems, we can improve the design of technologies and products in pervasive networks.

6. References


PART 2

NEAR FIELD COMMUNICATIONS (NFC) AND UBIQUITOUS INTERACTION
THE INTERNET OF THINGS

Jani LAHTI

Summary

The idea of the Internet of Things is based on physical objects having a digital identity, so that in minimum the status or whereabouts of objects can be tracked and processed as a part of more complex data systems. This paper is a summary of the underlying ideas and enabling technologies on very general level. The scope of the Internet of Things is widened towards active objects which are more complicated than passive objects having just the identity.

1. Introduction

The Internet of Things consists of everyday physical objects which have been given, at least, an electronic identity. Enumerating objects can be quite a challenge since it has been estimated that in average every human being is surrounded by 1000 – 5000 physical objects. If all the objects would have an identity, the comprehensive Internet of Things might consist of about 50 000 – 100 000 billion objects. [1]

There are several numbering schemes to tackle this challenge. The old and widely used Internet Protocol version 4 (IPv4) numbering can not be used since it only allows around 4 billion addresses using 32-bit addressing. The new IPv6 standard, for one, increases this limit to 128 bits and to about $3.403 \times 10^{38}$ addresses or IDs without need to resort to private addressing. [2]

The Internet of Things is not one particular technology, but a collection of enabling technologies and ideas. One of the enabling aspects is miniaturization of electronics which also lower costs. More computation power can be harnessed into smaller and more effective units. Wireless data transfer technologies enable easy communication for objects. Advances in sensor technology allow more precise, smaller and cheaper sensors which consume minimal power. Positioning technologies provide objects with information of their geographical or relative location. [3]

When making the Internet of Things possible to become widely deployed, one important factor is solid standardization of, for example, the wireless data transfer technologies used. Also the usage and format of data and protocols need to follow some sort of agreements to enable interoperability between different objects and back-end systems. [3]

Having a lot of enumerated smart objects generating huge loads of data is of no value by itself. There need to be ingenious applications which leverage the possibilities of this data. On the other hand, when the new mash ups of technologies become widely into use, they will attract hackers and criminals, too, and may ultimately lead even to life-threatening situations if security precautions are not strong enough. On more human side, some individuals may feel threatened by the emerging new applications of technologies which allow tracking of people’s everyday life more thoroughly than ever before.

The next chapter of this paper covers some aspects of the Things which form the Internet of Things. Chapter three discusses some suitable networking technologies and chapter four covers sensor technologies. Chapter five brings up some challenges the Internet of Things concept is facing, and chapter six concludes the paper.
2. Thing

The Internet of Things consists of physical objects with some sort of electronic identity or active features. Identity can be given basically to any object, from toothbrushes to tires. When the objects are connected to each other or to other systems, they make up a pervasive, ubiquitous network which can span large areas and object types.

When building the Internet of Things, many objects would have more added features than just an electronic identity. Objects could have an option to react changes in their surroundings using sensors, own processing power and autonomous data transfer capabilities. Adding enabling technologies into objects creates possibilities not only to communicate between people and objects, but between objects themselves. This creates truly pervasive networks which offer more possibilities than for example standard mobile access networks alone (Figure 1). [3]

![Diagram](image.png)

Fig.1 Connected: any time, any place, any thing [3]

Active objects can collect data of physical status of the object itself or its surroundings and also process the data by themselves, thus transferring some of the computational tasks to the edge of the network. The properties of an object can be read or tracked using for example Radio-frequency Identification (RFID) technology, but there are also other possibilities as the needs of the data transfer are different in different applications.

The objects can be of many levels of complexity: some application can be used to track the bolts used in assembling an ocean liner, and some other application can later track the ocean liner itself or its cargo. It is very likely that there will never be any universal applications due to the very diverse field of possible applications for the Internet of Things concept. [4]

3. Internet

Just having an identity or some computing power does not make things anything special. The things need to communicate with each other and outside world to transfer the gathered data for further processing. There are currently several wireless communication technologies available which differ

LAHTI, J: The Internet of Things
from each other on the grounds of range, speed, availability and technical requirements of a transmitter or transceiver.

One important factor is the possibility to generate wireless, self-configuring networks between objects and users, and purely between objects themselves. Several connection technologies can be used to transfer data between the objects, and also to Internet and back-end systems.

The transmission technologies can be divided into three groups based on their range and availability: Short range simple technologies like RFID, middle range more complex technologies like WLAN and technologies with broad geographical coverage like mobile cellular networks.

### 3.1 Simple technologies

The short range technologies are also the simplest and consume the least power. Their typical range is from some centimeters to some meters. They also have a potential to be the most deployed technologies due to low price and simplicity. Some of the common technologies in this group are Radio-frequency Identification (RFID) and Near Field Communication (NFC).

RFID is the most pervasive communication technology within the Internet of Things. This is because of its good standardized status and low price of the simplest units. International Telecommunication Union (ITU) considers RFID a key enabler of the Internet of Things concept. [3]

Main components of RFID technology are a transponders, interrogators and middleware. A transponder (tag) is attached to or implanted in an object. Data is transferred between a tag and an interrogator (reader) device, one-way or on both directions. RFID tags can be passive, semi-passive or active in nature, and information on tags can be read-only, read-write or rewritable. [5]

The communication between RFID tag and reader is performed using magnetic inductance. This mechanism can also provide enough power through tag’s coupling element, basically a coil, to allow passive tags to operate without internal power source when they are exposed to the reader’s magnetic field. Passive tags have very limited range, but some active tag types with their own power source can offer a range of up to 100 meters. As can be seen from figure 2, tags can be made very small if application needs it.

![Fig.2 Miniature size RFID chips on fingertip, adapted from [6]](image)

LAHTI, J: The Internet of Things
Information from the RFID objects is usually fed through some middleware application which forwards data to back-end systems for further processing or storage. Normally one is not going to access objects directly, but utilize the data through back-end applications.

NFC is an extension of the ISO 14443 standard, standardized in ISO 18092 and ISO 21481. NFC basically combines the operation of RFID transponder and interrogator into one unit. The range of operation is very limited, about 10 centimeters, which allows high density of objects in a given space so that they do not interfere with each other. [7] [8]

3.2 Complex technologies

Middle range technologies have a typical range of 10 – 100 meters, some spanning up to a few hundred meters in optimal conditions. Some common technologies are Bluetooth (BT) and different Wireless LAN technologies (WLAN). These technologies feature faster transmission speeds compared to short or long range technologies, but can consume more power and have more complicated interface.

3.3 Mobile cellular networks, satellite communication

There are several technologies which offer seamless coverage within a country, a continent or almost anywhere in the world. Mobile cellular networks are based on base stations with a known coverage area. A number of base stations are located so that they can offer expected coverage. GSM and WCDMA are widely used digital mobile cellular networks, which often complement each other when used with suitable wireless terminals: WCDMA offers more capacity, but GSM can offer more coverage. Satellite communication systems offer the widest geographical coverage, but cost, size and power requirements limit the deployment of devices utilizing these technologies.

3.4 Repeaters, proxies, gateways

Simple, inexpensive objects will not be equipped with large-scale active intelligence, so the data they produce and the communication methods they use to transmit the data are also relatively simple. Objects may transmit information between similar kinds of objects quite easily, but when data needs to be transferred upstream into real Internet, for example, it is not feasible to make every object handle all the burden of the TCP/IP stack.

Repeaters can be used to amplify or re-transmit wireless transmissions to allow more distance between objects. Gateways are used to transfer simple data transmissions into more complex networks or systems, or between the systems. Proxies can be used to hide a network of objects under one identity and aggregate the traffic. Middleware applications can be used to relay the object data to back-end systems or databases. [3]

3.5 Power consumption

One major field of related technologies is electrical power related technologies. Autonomous active operation needs autonomous power supply for prolonged use, or a good battery. Electrical power can be generated for example from sun, wind or object’s movement, or it can be converted from some sort of thermal, biological or chemical source. Power can also be transferred wirelessly during an active period of an object, or used to charge its battery. The main point in power technologies however is minimal power consumption. It has been made possible by advances in modern electronics, and the usage of different sleep modes and transmission power control.

LAHTI, J: The Internet of Things
4. Sensors

If the object is to have anything more than a passive identification, it needs some computational power and one or more sensors. Sensor is a transducer converting some form of physical energy into electrical energy which can be measured, processed and logged using an integrated circuit (IC). [9]

Sensors can be digital having just two states (1/0, on/off), or analogue which can be sampled for a range of values. Simple sensors need to be handled by an IC to get readings while advanced sensors come with all the needed electronics to provide the measurements in processed form.

4.1 Simple sensors

Sensor in its simplest form can be an open/closed switch or a variable resistor which change its resistance when a selected physical energy affecting it changes. Some basic digital switches include a momentary button switch and a magnetic switch. A tilt switch can provide information if the sensor is tilted or not, but not the amount of tilting.

The most common variable resistor is a potentiometer for detecting rotational or linear position changes, but there are many more sensors utilising the same idea. A thermistor changes its resistance when the ambient temperate changes. A photocell detects changes in the intensity of light and a force-sensing resistor has a variable resistance that depends on the amount of force exerted on the sensor. A flex sensor changes its resistance depending on how sharply it is bent.[9]

4.2 Advanced sensors

Many sensor types require extensive processing to control the sensor’s behaviour or manage the sensor data. It is convenient to package the sensor with the supporting electronics into one unit which can be operated using some digital protocol, often via Inter-Integrated Circuit (I²C) or Serial Peripheral Interface (SPI) bus.

Some of the sensors include motion detectors and infrared (IR) and ultrasonic sensors. These can be used to detect objects movement or position, or movement around the object. Accelerometers (often 2 dimensional (2D) or 3D) can be used to sense static or dynamic acceleration of the object. A gyroscope is used to measure the speed of rotation and an electronic compass for heading. Temperature, pressure and humidity sensors can be combined into one advanced unit for collecting environmental data.[10]

4.3 Sensor networks

A wireless sensor network (WSN) is a wireless network consisting of spatially distributed autonomous devices using sensors to cooperatively monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion or pollutants, at different locations. The base stations are one or more distinguished components of the WSN with more computational, energy and communication resources. They act as a gateway between sensor nodes and the end user.[11]

Some characteristics of the WSN are mobility of heterogenic nodes and dynamic network topology, which can cause communication failures. The network must cope with failures unattended and automatically rebuild or reroute the communication channels between the nodes.[12]
4.4 Positioning technologies

A satellite positioning system receiver can also be considered some sort of sensor in that it can be used to determine geographical location of the receiver’s antenna, and if attached to an object also the location of the object. There are several satellite navigation systems available or to be available in near future, but the Global Positioning System (GPS) is currently the only freely available system providing global coverage. Using a GPS receiver the location of an object can be logged, transmitted real-time or acted upon by the object. With GPS, several other sensors and a satellite communication system, it is possible for example to track the real-time status and location of a sea container all the way while it is been shipped from the other side of the globe. Also data from objects stored in the container can be relayed via the container’s communication channel.

5. Problems, challenges

Many of the challenges connected to the Internet of Things are not technical, but social. Security and privacy issues need to be considered when deploying objects using identification technologies and services based on those objects.

When some large-scale applications of the Internet of Things concept become available and deployed, they may soon become targets for data or identity thefts and denial of service attacks. The attacks may render the network useless, and in worst-case scenarios lead to life-threatening situations if the network is used to support life or control critical applications.

Some individuals may feel threatened by the emerging new applications of technology which allow tracking of people’s everyday life more thoroughly than before. There are already groups against the use of RFID technologies. [13]

6. Conclusion

The Internet of Things is possible to build already at this time. All the important technologies needed already exist, and many of them are also standardized to allow interconnectivity when building networks between the objects. Different applications of RFID are already used world-wide, but most of the applications are for one closed purpose only and relatively small scale. Implanting ID chips or active components to objects is no longer technically or cost-wise challenging. The challenges are more social than technical. Security and privacy need to be considered when deploying ID objects and services. Utilizing the possibilities offered by the Internet of Things concept may, for example, bring more speed and efficiency to material handling thus reducing costs, or allow easier monitoring of physical conditions of various structures thus making maintenance of large machines or buildings more efficient.

Even though many uses of the Internet of Things concept is already possible, it may take quite a while for the pervasive services to be widely deployed, and even then many objects will still be left without own electronic identity.

LAHTI, J: The Internet of Things
7. References

PERVASIVE RFID

Erkki VAAJA

Summary

This report explains the principles of RFID technology and how it can be used in pervasive networks. The report also presents some considerations that should be taken into account when designing RFID environments and also explores some of the problems that still hinder the deployment of RFID networks. There are some solutions presented in the report for the various problems. There is also few applications presented that already take advantage of the possibilities of RFID.

1. Introduction

The growing need for intellectual environments and smart housing has raised a problem of attaching information to various objects. Our technological response has been a wide use of bar codes and similar tagging options. For the future usage the bar codes will become however inadequate to mark and track objects because of the increasing data storage capacity demands. Proposed answer lies in radio frequency identification technology. It has potential to track objects easily and with greater precision as well as possibility to store more information about the tagged object. Radio frequency identification is not a new invention. It's reported to been used even as early as 1940s in the World War 2 by the British to identify their own airplanes from the enemy ones. Reader technology has improved significantly over the last two decades.[1] Widespread object identification has introduced new set of problems to consider. The reach of RFID tag reader, the authority of the tag read event and the price of one single tag are some of most important questions that inspire further research.[1][2] Researchers have been able to increase the probability to get successful tag reads and they have also found ways to combat the reader reach problem.[2] Most important issue of the technology is however yet to be answered so that everyone is satisfied. That problem is of course the security and privacy concern. People and objects carrying RFID-tag have no way to know or to stop anybody from reading the tag. This paper explores some solutions for problems mentioned and depicts the principle of RFID architecture.

2. What is RFID and how does it work?

RFID system has tree main components which are the reader sometimes also referred as the interrogator, the tag itself and the database for the tag information.[3] The reader may include additional antennas to enhance the precision.[2] Reader also powers the passive RFID-tags. Database in the other hand contains corresponding data for all the tags. That data might be something like the owner of the tag, the tagged object, manufacturing date or even history of tag owners.[1][3] The possibilities are limitless. The tag can also contain more data than merely the identification number. Tags are divided into two types that depend upon their powering method. These types are active- and passive tags.[1] Getting the information from the tag is based on radio waves that are being frequency modulated. The reader sends a carrier signal which the tag modulates and the reader then picks up the tiny modulation in the carrier signal.[4] This is how the data is being transferred between the tag and
the reader. There are some standards for RFID communication for example the EPCglobal classes 1 through 4 which define different far-field tag functionalities and provide communication parameters for the tags.

2.1 Active RFID tags

Active RFID tags [Fig.1] are powered by their own power source for example battery [1], solar panel or something similar. Active tags have much greater range than the passive ones. The range can be even up to 20 meters from the reader.[3] Active tags transmit the identification signal continuously or when instructed to. Because the active tag is more complex to produce it has significantly higher cost compared to passive tags. One active tag can costs from 7 € to 35 € depending on the features of the tag.[6] Active tags have a possibility to add sensors to the tag making them able to detect various events such as temperature rise or sudden impact.[7] Some active tags have also the ability to communicate with other tags. This tag to tag communication makes possible the communication to blind spots and places where there are obstacles for EM-wave propagation.[8]

![Fig. 1 Active RFID tag](image1)

![Fig. 2 Passive RFID tag](image2)

2.2 Passive RFID tags

Passive RFID tags [Fig.2] get their power from the reader. There are also two types of passive tags which are the near-field tags and the far-field tags. The near-field tags operate by the principle of induction from the magnetic field of the reader. The reader provides a large current trough its coil thus providing strong magnetic field. When tag is in the magnetic field the field induces a current into the tags circuitry that loads a capacitor in the tag. Then the current from the tags capacitor is used to power the chip that energizes the tags coil and the tags coil produce smaller magnetic field opposite to the readers’ field. The tags magnetic field is load modulated to transmit the tags data to reader. The reader then picks up the tiny variations in the current it provides to its own coil because of the tags varying magnetic field and receives the tags data through the current variations. [Fig 3.] The far-field tags operate by intercepting readers EM-waves. The reader sends out EM-waves through its dipole antenna. The tag has its own smaller dipole antenna that it uses to intercept the EM-waves that provide potential difference which the tag can use to store the energy to its capacitor. The energy is then used in the tag to power that tags chip that varies the tags antennas impedance hence reflecting different amount of the received EM-waves back to the reader. The readers sensitive RF-receiver then picks up the reflected waves and intercepts the tags data from them. This is referred as the back scattering effect. [Fig. 4.] The Near and far references in the type names also tell us something about the reading reach of those tags. Near-field tags have a reach from 1 centimetre to about 20 centimetres compared to the far-field tags that can be read even as far as 6 meters away. Passive tags are quite a lot cheaper to produce than active tags. A cost of single passive tag is at its lowest only 13 cents per tag. Passive tags are ideal in such perspective that they don't require replacement batteries, they have long life expectancy and they are cheap enough to produce in large amounts.[1] These types of tags are commonly used to tag
objects to keep track of their location and count. They are ideal in tagging industrial inventory and to speed up logistics.[9] There are some values compared for different types of tags in table 1.

<table>
<thead>
<tr>
<th></th>
<th><strong>Active</strong></th>
<th><strong>Passive Near-field</strong></th>
<th><strong>Passive Far-field</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost</strong></td>
<td>7€ - 35€ depending on the tag capabilities</td>
<td>14 cents or more per tag in a lot of few million</td>
<td>14 cents or more per tag in a lot of few million</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>up to 20 m</td>
<td>1 cm to 20 centimeters</td>
<td>1 cm to even up to 6 m</td>
</tr>
<tr>
<td><strong>Power supply</strong></td>
<td>Self powered e.g. battery</td>
<td>Reader powered (induction based)</td>
<td>Reader powered (EM-wave capture based)</td>
</tr>
</tbody>
</table>

Table 1. Comparison between tags

Fig. 3 The operating principle of passive near-field RFID tag [1]
3. Designing RFID systems and networks

When building large scale RFID environment it's important to consider how the tags are moving in the space and select the positions for readers and antennas based on that. The density of readers is also key element making the read events successful. [2] Most commonly RFID tags are used nowadays to speed up logistics and inventory. [9] In the case of tracking industrial object movement in certain space the reader and antenna deployment is easy because the assumed trajectory of the object is somewhat known. The situation changes quite a lot if the there is no apparent pattern in the movement of the object. This case requires lots of readers and good coverage depending of course on how precisely the objects movements want to be tracked. It's safe to say that truly pervasive RFID environment needs a lot more consideration for deployment of hardware, movements of the objects, sensitivity of tag read data, security of the data and functionality of the system compared to industrial inventory or products and materials tracking. [2] The architecture for RFID network varies quite a lot depending on used applications. The basic structure is however always the same. At the bottom there are tags that interact with the readers and readers interact with the middleware system that handles data refining and storage.

4. Problems and solutions in pervasive RFID systems

At least one large scale RFID environment has been built in the University of Washington to study is it possible to construct functioning RFID ecosystem. The study is still being conducted but preliminary results have already brought up expected and also unexpected problems related to RFID ecosystem. This study points out and addresses quite a lot of the pervasive RFID environment problems. In the
installation phase the problems varied from health concerns to aesthetics of the readers and antennas. The positioning of the antenna is important and they also studied the different probabilities that the tag was read while changing antenna and tag positions. The acceptable probability was achieved by increasing the number of readers and antennas. The readers' transmit power also needs to be adjusted so that excessive EM-waves don't propose a health risk.[2]

Challenges that were more technical included systems failures and tag read data related problems. To avoid single point of failure situations it's important to build the system fault tolerant for example duplicating servers handling and processing data. Because the RFID read event is not always successful or the tag gets read too many times or similar problems it's important to actively clean up the read information for the applications that use the data.[2]

The most important problem of the system still remains an issue of privacy and security. Because the tracked data is quite personal the problem of privacy is two dimensional. The study focused on the issue of privacy on collected tag read data that was stored in the system. If someone gets access to the data of tag reads it's easy to device the whereabouts of an object or a person just using the collected data. There is no simple solution for this problem. The data needs to be secured and access to it must be allowed only from authorized personnel. When tracking objects it's how ever vital to get location at all times and to avoid too detailed tracking it was proposed that some areas could have anonymity by default e.g. restrooms and the precision of the location for the object/person could be changed so that it could display part of the building the object or person is in or if so desired even be more precise and tell exact room.[2] The other dimension of the privacy is of course the problem that anyone who has reader can read the objects or persons tag without being noticed. This problem is still unanswered. Few solution are offered for the problem in the means of carrying a device which would scramble the RFID signal that the tag sends or simply notifies the tag carrier that the tag has been read.[3] If the tag is placed in some product and after sale the tag is no longer needed then the solution for unauthorised tag read is simply to kill the tag. [1] These solutions however weaken considerably the original idea of the RFID tracking/authentication.

One of the problems is also the situation where there are many tagged objects in same pile of goods and the whole thing is moved by the reader. If the tags start sending data simultaneously the reader can’t recognize the tag. The solution in the collocated tag problem is anti-collision protocol. The tags transmit at random intervals, but still collisions might occur so the reader has to sweep the pile until it has achieved high probability of reading all the tags in area. Another approach to solve the problem is based upon Query Tree protocol. In this solution the principle is best described as “The reader starts an interrogation cycle by asking which of the ID space’s top branches (modelled as a binary tree) contain tags. The algorithm recursively repeats for each subtree branch, but if a particular subtree doesn’t generate a reply, the reader won’t consider any of its branches and subtrees in the remaining search space. In other words, that branch is pruned from the binary tree. After a short time, all tags present will respond to the reader in depth-first-search order.”[1] This approach is used by the far-field RFID communication protocol standard EPCglobal class-1.[1]
5. Applications for pervasive RFID

Few of the mostly used pervasive RFID applications are nowadays the various asset tracking systems that logistics and different postal firms use. Global asset tracking is tracking the movements of products or materials. RFID provides an easy way to monitor packet movements across wide area. Precision also increases quite a lot when using electronic packet surveillance [Fig. 5]. United States Army is also one of the organizations that uses RFID based system to organize its logistics and to keep track of materials. The well known Wal-Mart chain has also adopted RFID tags to their products to speed up checkout. It's not wise to tag everything yet because of the price of a single tag, but the bigger and more expensive items are tagged.[1] RFID technology is also used in some anti theft systems.

Tagging different objects in all kinds of stores allow tracking their movements as customers pick them up and carry them in their carts. This can give shop owner unique perspective where he/she should place different kinds of products to get as much people as possible to go by them.[7]

One usage for RFID technology is the “digital wallet”-concept that enables people to pay for their shopping and services with RFID-chip [Fig. 6]. It is already used by some Clubs for their VIP-guests for authentication and payment. The RFID chip is implanted under person’s skin. It is similar system that can be used to identify for example pets. Visa has also published a credit card that contains RFID-chip which can be used for making payments.[7]

6. Conclusion

RFID technology is still at an early stage if we consider all the potential it holds. Many of the problems have already been solved, but some key problems such as privacy and security still remain. There is lots of promise in the research to solve these issues also, but before then the technology is not yet ready for wide spread use in smart environments and intelligent housing. RFID based logistics and asset tracking is pawing the way for truly pervasive RFID solutions and networks. RFID looks to
become dominant technology for attaching data to various objects and tracking them. Current range of
tags and radio frequency environment limitations are also holding back the implementation RFID. Still
despite these problems there are already many applications that take advantage of RFID technology.
Accurate and fast identification makes RFID the technology for future. When security issues have
been solved I predict that the rapid deployment of RFID networks will begin. I believe that when RFID
is adopted more widely there will be universal addressing scheme developed for all RFID enabled
objects and as a by-product better standards for RFID communication will also emerge.

7. References

visited 08.05.2008
Infrastructures. PERTEC 2007 Workshop on Pervasive RFID/NFC Technology and Applications,
[3] Compiled by the Society and Technology Group at the UW Dept. of Computer Science & Engi-
Electromagnetic fields and waves ECSE 352a McGill University, Nov. 30, 2004.
[6] Griebenow, A. Active RFID - on the path to deliver pervasive enterprise efficiency,
p3?template_id=218&id=8263, visited 05.05.2008
Summary

This article gives a brief look at the definitions of privacy, security and trust terms and emphasise their importance in people’s lives and organizations’ reliability. Then, a fast look at the security in general in pervasive networks, and details of risks’ entities and causes of Radio Frequency Identification (RFID) systems as part of this kind of networks. The article will end by listing some of the current issues related to this topic and some remedies and possible directions that can be taken into account for the future of RFID.

1. Introduction

This report on privacy, security and trust in pervasive networks, as it is a huge domain with unlimited boundaries and has multiple facets, will concentrate only in the RFID side. It seeks to clarify the capabilities and limitations of RFID and aims at informing the further development of new security policies. This paper is based on previous work and issues dealing with the security aspect on RFID systems and their weaknesses.

At the end of this report, some possible solutions will be proposed that would be beneficial for the future of RFID

2. Background

Organizations are aggressively deploying RFID systems, which makes activists increasingly concerned about RFID’s potential to invade user’s privacy.

As security has been always a central topic aimed by technologies’ developers, vendors and even final users/consumers, it has been always inquired by people, not only in computing, but also in “normal” and daily life such as security guards, insurances, social security numbers, PIN codes and alarms.

The biggest problem is that the growth’s rate of any kind of technology and especially the ones related to computing and information systems is always higher than the security and risks’ evaluations of it.

3. Privacy, security and trust

This section will give a quick overview about the terms privacy, security and trust which are highly related to each other. The assurance of one of them is complemented by the pledge of the other one. For instance, no trust can be guaranteed if the privacy is not asserted, and no privacy without security.

3.1 Privacy

“The state of being away from the disturbance of others like person, group or any thing else.” [1]
Privacy means keeping its own information for itself without getting known by other people or system.

3.2 Security

“Any thing providing safety, freedom, from danger or anxiety.” [1]

Security has three classical aspects: Availability, Confidentiality and Integrity.

- **Availability**: is the assurance of timely and reliable access to data services for authorised users when required.
- **Confidentiality**: is the assurance that the information is accessible only by who are authorized to access. When related to persons, if confidentiality constraint is violated, trust will be absent.
- **Integrity**: is the assurance that the available data has not been altered or lost during a transmission or data communication.

3.3 Trust

“A feeling of confidence, reliability of person or thing and strong belief. On trust, being confident without proof.” [1]

As a conclusion, to guarantee privacy, security policies should be taken into account, and as much as the security is “strong” enough, as much as trust is “high”.

4. Security in pervasive networks and RFID systems

This section will present in brief some of the pervasive networks’ weaknesses and advantages in general, and then will get into details of RFIDs’ systems security facts and risks.

4.1 Security in pervasive networks

Pervasive networks are by definition the interconnection of different devices of different sizes, provided from different vendors using distinct technologies. Thus, it is hard to maintain their interoperability and correlation. Security is a big challenge to swear by protecting pervasive network’s elements from each other and from external elements.

We can list below some of the pervasive networks’ risks:

- Small devices (can be considered as an advantage as well): the risk of being stolen or damaged is higher when we deal with small devices.
- Mobility and communication such as cars, Personal Digital Assistants (PDAs) and cell phones.
- More and more new emerging technologies.
- Inter-working and interoperability.
- Users’ ignorance.
- Quality and type of middleware.

We can list as well some of the plenty of positive points of pervasive networks:

- Interconnection of different devices at any time and any where (availability).
- Mobile prosumer (a combined producer-consumer).
- Internet connection and Internet of things.
- Portable, intelligent, programmable devices.
- Invisible in use (data process is invisible to user regarding time and space).
• Used by in many fields such as by hospitals, security agents, taxis, fire department, airports, schools, libraries and vendors.

4.2 Security in RFIDs

RFID technology has been described as the “world’s oldest new technology”. It was developed in the 1940s and was used for identification by military aircrafts. Its first commercial use was in the 1960s in the field of electronic article surveillance against theft.

The problem of security related to RFIDs has been faced from the very first applications and it is getting more and more complicated and hard to maintain.

RFIDs have two big dimensions: a technical one; RF, and a social and economic one; objects and people IDs. That is why, providing security, privacy and trust has to cover these aspects. But, there is no one-size-fits-all RFID security measure that would offer an efficient result of security strategy.

RFID is a technology that enables data collection with contactless transmitters for the purpose of identification. This tracking faculty can be seen from two different windows; anti-theft and privacy violation. These faculties will be described later on in this paper.

RFID tags have replaced bar codes technology developed in the 1970s. Those tags offer an improved enumeration system, giving each tag at least a 96-bit number that is both globally unique. But, unlike barcodes, RFID tags can be read at a distance without a person’s knowledge. As a result, tags placed in consumer items for one purpose might be covertly used to track people as they move through the world. This is especially true of RFID tags that might be embedded in items such as shoes and clothing. Nowadays, RFID tags are can be encapsulated in IDs and passports and even under human skin for health care or kidnapping purpose (as shown in figure below).

Fig.1 Under skin RFID tag [1]

In addition to tags, RFID’s readers/writers (interrogators) are part of the risk evaluation and prevention. More details are explained in the next sections.

4.3 RFIDs potential risks

RFID has many sides; one of them is tracking persons or objects. The latter can be used without the person’s knowledge and then the violation of the privacy constraint. For example, a press coverage in “Wired News” article that erroneously reported clothing giant Benetton’s plans “to weave radio frequency ID chips into its garments to track its clothes worldwide.”[3] Those tags are exposed as well, if they are “visible” to the malicious user, to so called forgery or cloning. Tags can be replicated or replaced, which can be a serious loss to the issuer. For example, duplicating hotel keys can harm the porter of the implantable ones.

RFID’s tags have the characteristic of data storage in their memories. Those data can be considered as critical for the person itself or even for the whole system/organization. Then, by intercepting data transmission, a misuse or data leaking can happen and as a result, a violation of the confidentiality constraint. Those memories can be exposed to a Denial of Service (DoS) attack which can cause a
delay of the system’s response that might be critical such as passports or access control cards used by authorities in airports or wherever. That kind of attack can lead also to a serious problem and even death of a patient by injecting erroneous codes and queries while processing medicine information in health care related domains such as hospitals, ambulances and pharmacies. Thus, data availability and integrity constraints are violated.

Data collection should be invisible to the attacker by using cryptography techniques. Data processing should be transparent and as fast as possible, depending on the type of the RFID system’s purpose. Authentication and permission initiatives should be taken into account.

In fact, many of these strategies have been implemented and already uses, but, still could be broken and hacked by “professionals”. So, enhancing security policies and privacy strategies seems urgent and so important to achieve trust. An overall security strategy and risk evaluation should be developed by considering each stage of system’s life: starting from planning, deployment, operation and maintenance, till technology’s “death”. This issue will be detailed in the next sections.

5. Causes and responsible entities in RFIDs

RFID’s risks and security related problems are related and can be caused by many entities. Those entities can be strongly related to the RFID system or considered as an external “badly influencing” agent.

5.1 End users

People or end users’ ignorance is considered as a medium risk of misusing the technology by leading to a data leak or destruction unintentionally. RFIDs are not really well understood by individuals, which makes it harder to prevent external attacks such as man in the middle, viruses, jamming or spoofing. People’s ignorance and lack of awareness and information should be treated by organizing some awareness initiatives by the government, media, issuers, or even vendors.

5.2 Open air media

The open air media forms another problem allowing “attackers” to intercept data transmission while critical data communication. Then, data accessibility should be limited by using authentication, private/shred keys or cryptography methods. Interferences with the other waves or signals travelling the air medium can cause a serious problem, so choosing the right frequency ranges, antennas’ types and signal power can reduce severely the RFID’s risks related to medium’s type.

The environment itself might cause a problem, regarding whether it is a liquid, metallic, plastic or a very hot one. It can influence on the data transmission quality and information protection from reflections, multi path, absorption and noise detection.

Person’s movement will lead to data caching and logging, which can be used in a privacy violation purpose.

5.3 Different devices and technology’s components

Potential RFID’s components requiring security can be listed as below:
Tags/transponders can differ from each other by memory capacity. A code injection, scripts or SQL attacks can damage the whole RFID system by just damaging only 1 tag. For instance, database consistency can be broken by “breaking into” one RFID tag’s memory and altering its data format, type or content.

Readers/interrogators and Writers can be intercepted while proceeding on data processing.

Tag and reader power level can be different from one tag to another. The power level can be affected by antennas’ nature (omni-directional and unidirectional). For instance, when a transmitter or a receiver is in movement, the directional antennas will not be practical. The higher is the power level, the greater is the probability to detect white noise.

Electromagnetic communication can be affected by the environment noise and signal interception, loss or modification as discussed earlier.

Databases that contain critical information related to tags and even companies or issuers, can be accessed by malicious third parties and as a result can be damaged.

So, a large number of parameters should be taken into account when developing and designing the RFID technology. But, the big challenge is to try to make a balance between personal convenience such as rich multimedia, band width, storage capacity and display capabilities, and security and potential abuse of privacy.

RFID risks can be divided into four forms of sources. The table below shows the different risks result from threats sent by vulnerabilities or weaknesses that can be exploited by malicious parties and cause adverse effects.

<table>
<thead>
<tr>
<th>Business process risks</th>
<th>Direct attacks on RFID system components potentially could damage the business processes that the RFID system was designed to enable.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business intelligence risks</td>
<td>An adversary or competitor potentially could gain unauthorized access to RFID-generated information and use to harm the interests of the organization.</td>
</tr>
<tr>
<td>Privacy risks</td>
<td>Personal privacy rights or expectations may be compromised if an RFID system uses what is considered personally identifiable information for a purpose other than originally intended or understood. The personal possession of functioning tags is also a privacy risk because it could enable tracking of those holding tagged items.</td>
</tr>
<tr>
<td>Externality risks</td>
<td>RFID technology potentially could represent a threat to non-RFID networked or collocated systems, assets and people.</td>
</tr>
</tbody>
</table>

Table 1  Risks’ typology

6. Current issues and future remedies

In this section, a quick view about some of the current issues developed concerning the RFID’s potential risks’ prevention or detection and security policies will be presented. It will be shown later in the second paragraph, that those issues developed could be broken and hacked.

6.1 Current issues: Examples

Many studies about RFIDs’ remedies are going to be detailed in the below paragraphs.

In February 14th, 2008 at Johns Hopkins University, three technologists (Ravi Pappu, Ari Juels and Bryan Parno) have developed a tag data protection process without derailing existing efforts to integrate RFID throughout the supply chain. It is based on secret-sharing cryptography, which uses a
secret key to encrypt a number, and then splits that key into multiple shares. The party attempting decryption must collect a specific number of those shares to figure out the key. [5]

In November 1st, 2007 at the University of Massachusetts Amherst (UMass); Passive RFID tags made with chips containing volatile memory should be unique. The researchers have identified methods of exploiting the tags’ uniqueness so that the tags can be authenticated to deter tag’s cloning and its data secured with cryptography. [5]

In May 2006 (by John Halamka MD, Ari Juels, Adam Stubblefield, and Jonathan Westhues), an RFID tag: VeriChipTM, was produced commercially for implantation in human beings. It includes identification of medical patients, physical access control, contactless retail payment, and even the tracing of kidnapping victims. [7]

The Auto-ID Center has designed in May 2005, Electronic Product Code (EPC) tags strengthening against cloning using the EPC kill command as a pro-privacy technology. They viewed killing EPC tags at the point-of-sale as an easy way out of the apparent privacy dilemma. [8]

In November 2005 at the Institute of Textiles and Clothing, The Hong Kong Polytechnic University, lightweight cryptography and a simple authentication method based on hash function for RFID Class 1 passive tags has been developed. [5]

An RFID blocker tag that takes a different approach to enhancing RFID privacy was created. It creates a radio frequency environment that is hostile to RFID readers. The blocker tag is a specially configured, ancillary RFID tag that prevents unauthorized scanning of consumer items; it “spams” the misbehaving of readers so they can’t locate the protected tags’ identifiers and permits authorized scanners to proceed normally. [9]

RFID tags is used also in automobile immobilizers systems, the car key incorporates a passive RFID tag that the steering column authenticates, thereby enabling vehicle operation. Payment systems introduced by the mid-90s by mobile stations and several years ago, by the European Central Bank has considered embedding RFID tags into currency.

And in many other fields such as inventory management, airports and healthcare systems (see figure 2).

![Fig.2 RFID in health care usage](image)

6.2 Lack of basic existing security

Researchers at Johns Hopkins University and RSA Laboratories recently identified a serious security weakness in the RFID tag in speed-pass devices and many automobile immobilizer systems. By demonstrating that such tags could be cloned, the researchers revealed the possibility of payment fraud.
and new modes of automobile theft. Although their discovery doesn’t directly undermine consumer privacy, it demonstrates that RFID tags could have security consequences beyond merely tracking or profiling consumers. [11]

In different types of RFID systems used in hotels, a hacker found out that none used encryption. He also found out that many systems which use encryption failed to change the default key set by the manufacturer, or that they used sample keys provided in user manuals sent with the cards. [12] A DB of key samples can be created and a dictionary of attacks can be used. Philips Electronics’ Mifare are most used as a key-access system, and could be easily cracked. A master key can also be designed, just by collecting few keys to determine the structure of the cards.

The implantable RFID "Verichip" [7] was cloned in less than 10mins by a 23 years old Canadian hardware developer for the purpose of an article in the magazine Wired. “The tag implanted in a journalist arm for the purpose of the article, featured no security at all”. [12] According to a Japanese newspaper, data about passengers stored in the Suida card can be read by basic readers, such as the one embedded in Sony Clié PDA. [12]

RFID-alarms can crash as well by hackers and an opened way for thief into building is free. For example, expensive cars can be stolen by using a simple laptop. For instance, this example shows how can that be possible: "The expert gang suspected of stealing 2 of David Beckham’s’s BMW X5 in the last 6 months did [used] SW programs on a laptop to wirelessly break into the car’s computer, open the door and start the engine". [12]

6.3 Possible threats

Inter-communication between devices and data exchange is one of the most important sources of risk and causes of security violation. So, it seems that one of the “stupidest” solutions or remedies can be proposed is not to connect to Internet at all.

But, even without Internet connections, communication with other devices is still possible (which makes it called pervasive). So, avoiding direct connection with the net will not prevent from an indirect type of attack.

Another possibility is to avoid RFID use by users that do not know at all how to use the technology and might lead to a misuse and security policies’ braking. This solution limits the market and then not so good for the economy and for the producers/vendors.

Another possibility is to increase security and safety. Well, this idea sounds better, but, increasing security means increasing the budget. Time and people’s investigations will be time and money consuming as well. Then, a good balance between that should be guaranteed.

The problem is balancing between security policies efficiencies and cost is not always evident. It could be possible, if we wait for the technology to get a bit "older”, but that means, that there is newer and better ones! Besides, by that time, attacks, leaks of weaknesses and better knowledge of it is already developed by malicious intenders.

So, what’s the best then? What about combining RFID technology with the old bar codes system (see figure below). It might increase safety and security levels and cumulate both lacks coming from both technologies.
High-level standard protocols such as Internet Protocol security (IPsec) or Secure SHell (SSH) besides to Transport Layer Security (TLS) protocol can then be implemented to small devices to enhance security levels (still a big deal).

7. Conclusion

RFID is a powerful technology that can offer potent benefits for businesses and consumers alike. But, ensuring security and privacy certificate to get people and entities’ trust seems to be a big deal. Since, a good balance between cost and high level security is the best way.

Consumers should be notified when items they purchase contain RFID tags. The latter should be disabled by default at the checkout counter (sales point) and/or should be placed on product packaging instead of on the product when possible. RFID tags should be readily visible and easily removable.

Some new researches propose to introduce high level security standards and protocols to lower level layer based devices like IP security extensions, cryptography and authentication mechanisms, using certificates or agent based mechanisms, but this seems to be really hard. Secure and stable technology has to have a secure base on which implemented new enhanced and secure high layers.

To develop a culture of privacy of RFID system, awareness raising initiatives about technology capabilities and limitations, risk reduction methodologies, initiatives to integrate security and privacy protection in the design of current RFID technologies and systems should be taken into account.

Finally, enabling security in pervasive networks is really huge topic whose boundaries are unlimited. Even its RFID side still so large and researches are incomplete or imperfect which need enhancement and more attention to the security, privacy and trust assurance branches.

8. References


FREJ, I: Privacy, security and trust in pervasive networks


SMART PHONE INTERACTIVITY AND PERSVASIVE CONNECTIVITY

Carlo IORIO

Summary

The idea described in this paper is to use the smart phone interactivity and pervasive connectivity in order to facilitate the interactions between people and smart objects of the real world. The ability to detect objects in the user’s proximity offers a natural way of interaction and strengthens the role of smart phones in a large number of application scenarios. This paper describes some applications scenarios and which are the models of interaction, the physical mobile interactions and how to choose them ones.

1. Introduction

Today interacting with the objects of the real world is simple, but tomorrow people can interact an infinity of objects; there will be problems to interact with these objects all together, so it is very important to find simple and intuitive methods in order to facilitate the interactions between the people and the smart objects of the real world. This is possible because: on the one hand there is the Internet of “Things” and in the other hand there is the physical mobile interactions. The idea is to combine these different domains in a unique domain where the smart phone is the mediator between them. This is possible thanks to the fact that physical objects are associated with digital information (Internet of Things) and to usage of the physical mobile interaction techniques as: Touching, Pointing, Scanning, User-Method object selection and Indirect-Remote control. In this paper will be described the interaction models, the technologies used to interact with real object, the physical mobile interaction techniques, advantages and disadvantages, limitations and example scenarios, that permit us to understand when a techniques is preferred in comparison to another.

2. Physical mobile interaction with Internet of “Things”

The term physical mobile interaction indicates the interaction between a user, a mobile device and a smart object in the real world, where a smart object can be a thing, a place or a person. Physical mobile interactions support the discovery and usage of services in a given context. In the Internet of Things the real objects are tagged in order to have a digital information of the object. The idea is to combine these different domains, physical mobile interaction and Internet of Things, in a unique domain where the smart phone is the mediator between them. This represents the goal of PERvasive serviCe Interaction (PERCI) [1]. This is possible because the physical objects are associated with digital information through the augmentation with visual [2] and wireless markers such as RFID tags or NFC tags and with the use of the physical mobile interactions [3] as: Touching, Pointing, Scanning, User-Metho object selection and Indirect-Remote control. The Figure 1 [4] shows a High level architecture for physical mobile interactions with the Internet of Things; the Universal Client represents the Smart Phone, which is the mediator between the Internet of Things domain and Physical mobile interaction domain [4]. It uses the Interaction Client and Service Client to interact with these different domains. The Interaction Client obtains the ID and additional data of the physical object, while the Service Client communicates with the service domain. This data will be send to Web service domain that will generate the user interface and correspondent service of physical object. Interaction Proxy is used to connect the service domain and the physical interaction domain, it provides three main functions: service composition, reasoning and user interface generation [4].
3. Smart Phone Interaction Models

A Smart phone is a mobile phone offering advanced capabilities beyond a typical mobile phone, often with PC-like functionality. A Smart Phone can be used to interact with the environment in different ways. In this paper there are four interaction models used depending on the connection types: universal remote control, dual connectivity model, gateway connectivity model and peer-to-peer model. With these models, a Smart Phone can be used to execute many type of applications, for instance to open a smart locks or to perform a payment in a restaurant.

3.1 Universal Remote Control Model

The Smart Phone can act as a universal remote control to interact with embedded systems located in its proximity. This is possible only if both Smart Phone and embedded systems have short-range wireless communication capabilities. Bluetooth is preferred because it is the more important candidate for the short-range wireless technology, due to its low-power, low-cost features and it has a free-bandwidth [5]. Figure 2 illustrates such interactions using Bluetooth.

In order to learn the identity information and the basic functionality description of the embedded system located in their proximity, the Smart Phone uses a discovery protocol. In this model it is necessary that the user has to know the interfaces, in order to interact with the embedded systems, otherwise it is not possible to use this model. To resolve this problem it is used another interaction model, that is Dual Connectivity Model.

IORIO, C: Smart Phone Interactivity and Pervasive Connectivity
3.2 Dual Connectivity Model

The dual connectivity model is based on two type of connectivity in the same Smart Phone: short range wireless connectivity (e.g., Bluetooth) and Internet connectivity (e.g. GPRS). With this model, the users can interact with the nearby environment, using the short-range wireless connectivity and with the rest of the world using the Internet connectivity. Figure 3 illustrates the Dual Connectivity model. A typical application is opening/closing Smart Locks. Supposing that a user must enter in a building protected by Smart Locks (e.g., locks that are Bluetooth-enabled and can be opened using digital door keys), dual connectivity model provides a simple way at user to open these locks in a secure manner.

![Fig.3 Dual Connectivity Model](image)

When the user is near the Smart Locks, using the Smart Phone, he can establish a connection with the lock, in order to obtain the ID of its. After the user sends a request to an Internet server and he puts in this request its own ID. The server checks in a database if the user is authorized to open the lock. If the control is successful then the server sends a response to user, where there are the digital keys for the Smart Lock and the code for the interface; in this way the user is able to open the lock.

3.3 Gateway Connectivity Model

Many pervasive applications use wireless communication by means of the IEEE 802.11 family of protocols. This family of protocols provides significant advantages in the distance and bandwidth in comparison to Bluetooth, as shown in Table 1.

<table>
<thead>
<tr>
<th>Advantage</th>
<th>Bluetooth</th>
<th>IEEE 802.11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>10 m</td>
<td>250 m</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>1Mbps</td>
<td>11/54 Mbps</td>
</tr>
</tbody>
</table>

Table 1 Advantage IEEE 802.11 relative to Bluetooth

The disadvantage of 802.11 is that it consumes too much energy, and consequently it is not possible to use it directly with the smart phone, otherwise the lifetime of the smart phone’s batteries will be very short. In order to take the advantages of the 802.11 family protocols, it could be used a gateway device that performs a change of protocol from Bluetooth to 802.11 and vice-versa. The Figure 4 shows the Gateway Connectivity Model.

![Fig.4 Gateway Connectivity Model](image)
3.4 Peer-to-Peer Model

The Smart Phones can also communicate among themselves (or with other Bluetooth-enabled devices) in a multi hop similar to mobile ad hoc networks. For instance, this model allows people to share different types of data (audio, video, picture, game, etc.) with others even if they are not in the proximity of each other. Figure 5 shows an example of this model where a group of friends can pay a check in the restaurant using their Smart Phones [6].

![Fig.5 Peer-to-Peer Model](image)

4. Physical Mobile Interaction Techniques

The term physical mobile interactions describes the interaction between the user, a mobile device and a smart objects in the real world, where a smart object can be a real world object, a person or even a location. The most important physical mobile interaction techniques are: touching, pointing, scanning, user-mediated object selection and indirect remote controls. The selection of a specific type of interaction technique depends on several factors as location of the object, the distance between object and user, the service related to the object, the capabilities of the mobile device and the preferences of the user.

4.1 Touching

With touching the user can select a real world object by touching it with a Smart Phone. In order to use touching the user must be near the object and after he has to touch the object, in this way the Smart Phone knows exactly the object and its related service. The main advantage of this technique is that it is similar to every interaction that the human does when touch an objects in the daily life; since the main disadvantage is that the user must be near the object, then the touching is convenient when the user is nearby the object. Technologies used for implementing this interaction technique are short range RFID and Near Field Communication (NFC) [7]. With the RFID or NFC the objects is not directly touched but it is sufficient a distance of about 0-5 cm for the selection. Another disadvantage is the ability of the user, because the user has to know where the tag reader is located in the device. A typical RFID system consists of a RFID tag and a RFID reader or writer. NFC is a short range data communication technology using the frequency band of 13.56 MHz, it works via magnetic field induction and is designed for simple and safe transfer of data between compatible devices. It is a combination of RFID contactless communication technology and wireless networking technology. The main difference between NFC and RFID is that in the NFC is possible to have a bidirectional transmission of information, thing that is not possible in the RFID. The touching is also an unambiguous interaction technique, because it is technically impossible to read two tags at the same time and because of the directness it is impossible to select a wrong tag or smart object.

4.2 Pointing

In order to select or control a smart object, a user points it with the Smart Phone, but the user has to know that the smart object is able to support the interaction technique pointing. The advantages of this
interaction technique are the distance between the physical object and the user and the likeness with the daily human physical interactions; in fact a person that is talking about an object usually points it with his index in order to show interest about it. The pointing can be realized by several technologies, for example: visual markers, image recognition, light beams or laser pointers, unidentified interaction or infrared technology. Today there are several visual markers techniques like QR code [8], Semacode [9] or visual codes [10]. They are bidimensional and black&white codes which can store more information than one-dimensional code, like the EAN-13 bar codes [11]. Colour codes are not used because they are more expensive to print and they are more sensitive at change of the light. An important factor, when markers are used, it is the minimal and maximal distance between the camera phone of the Smart Phone with which it is possible to capture the markers and the smart object. The distance is calculated in depending on: the size of the marker, the resolution of the camera, the viewing angle and an eventually existing optical zoom. The advantage of the usage of visual markers or image recognition based approaches is that no power supply for the augmentation of the smart object is needed and it can be used with current mobile phones without additional hardware. It is possible to use the pointing with an additional device as light sensors attached to the object [12].

4.3 Scanning

The scanning is based on the closeness of smart phone and smart object. The smart phone scans the environment for nearby smart objects in two ways: either the user starts the scan or the smart phone scans permanently the environment. The result is the same in both cases, that is a list of nearby smart objects. The advantage of this interaction technique is that the user knows all nearby smart objects and the services they offer, but the disadvantage is that there is no direct link between the user and the smart object. The main connectivity used in this technique is the Bluetooth [5], because every smart phone is equipped with Bluetooth device. Using the Bluetooth there are some disadvantages which make it not the perfect implementation of the interaction technique scanning, because the procedure of device selection, service selection and an optional password exchange uses long time. For this reason others approaches have been developed for example using: infrared [13], RFID [14] or visual markers [15]. Another possible communication technique used for the scanning is wireless local networks (WLAN) such as WI-FI [16].

4.4 User-Mediated Object Selection

With the user-mediated object selection, the user keys in information provided by the object in order to establish a link between them. In this kind of interaction the user is the sole responsible of establishment of a link between the smart object and the smart phone, so it is not necessary additional device, this represents the main advantage of this interaction technique. A practical example for the user-mediated object selection is a portable museum guides like PDAs where the visitor has to type in a number to get information about an exhibit or a URL printed on an advertisement poster to get access to the corresponding services [3]. Thus it is inappropriate when the URL printed on an advertisement poster is very long; this represents the main disadvantage, because the user has to copy all URL.

4.5 Indirect Remote Control

With this technique the Smart Phone is used as indirect remote control for interactions with remote displays. These interaction techniques represent extensions of conventional mobile interaction techniques that are based on the usage of the touchpad, joystick or keypad [3].
4.6 Physical Mobile Interaction Framework

The previous section discussed and classified different physical mobile interaction techniques. In this section it will be presented a framework that supports different implementation of the physical mobile interactions described before. This framework is the Physical Mobile Interaction Framework (PMIF), which was developed to support the rapid development of mobile applications and services based on physical mobile interactions [3]. Within a physical mobile interaction, the mobile device acts as a mediator between the physical (smart object) and the digital world (server). The Figure 6 [3] shows the PMIF architecture.

![Physical Mobile Interaction Framework](image)

The figure shows which are the technologies used for every interaction [3] and it can be understood that the PMIF supports all physical mobile interaction techniques.

5. Evaluation and Comparison of Physical Mobile Interactions

There are numerous scenarios in which mobile interaction in smart environments makes sense including the provision of additional services such as reading the manual of a microwave after touching it with the mobile device or requesting direct support for a specific device. Other examples are adding interaction functionalities for devices without an interface (e.g. information about power consumption of electronic devices) or remote control of objects (e.g. requesting the current status of the washing machine while watching TV). By means of three different tests [3], in this section it will be described that the type of application, the location of the object, the distance between object and user, the service related to the object, the capabilities of the mobile device and the preferences of the user are important factors for the selection of an interaction technique. In the first test there are three different application areas. In the first application area, the participant must take information related to an object, for example opening a web page related to a device (washing machine), or an online guide for the television or radio Figure 7 (yellow). The second application area was retrieving status information about physical appliances such as the status of the coffee machine (switched on/off) and the remaining time a washing machine needs to complete a wash Figure 7 (cyan). The last application enabled controlling a device such as the heating remotely Figure 7 (red). The participants had response about these application areas as one can see in the following Figure 7.
The second test shows when the participants would use the mobile phone for interactions with objects in smart environments. Figure 8 shows that the majority of the respondents (43%) would use such a system independently of their location. About a third of the respondents (34%) would use it only when not being at home, 13% would use it only when at home and 10% of the respondents would refuse to use such a system at all.

In end the Figure 9 shows that, in general, users preferred pointing and that they were almost equally split on the use of scanning, but generally disliked touching for the reason seen in the previous section.

6. Using Smart Phones to Access Site-Specific Services

In this section will be shown how it is possible to use the smart phones with electronic service or application allocated in specific location, that is site-specific services. By using personal information stored on smart phones, site-specific services can provide a personalized service to a particular user.
Moreover, user can store information obtained from an interaction with a service on their smart phones for future usages. This is possible using the Mobile Service Toolkit (MST).

### 6.1 Mobile Service Toolkit

The MST is a client-server software framework supporting the development of site-specific services that exploit interaction with smart phones. It is structured in two parts: MST client software and MST server software. The MST client software, called the Mobile Service Explorer or MSE, runs on smart phones, instead, MST server software runs on a device providing a site-specific service, on Linux and Windows. User in the proximity to these services can connect and access them through their MSE-enabled smart phones. The MSE performs three primary functions: Connection establishment, Personal Information management and General-purpose data entry and display [17]. The Figure 10 [17] shows an architecture of a system built using the MST:

![Fig.10 architecture of a system built using MST](image)

Two cases studio concerning the usage of the MST are described in [17], they are Virtual Queuing Service used to manage the queue in a public place and Interactive advertising that demonstrate how integrate smart phones with existing site-specific services.

### 7. Conclusion

This paper described the way in which the smart phone works as a mediator between the real world and the virtual world, in order to help the human to interact with smart object. It is possible to interact with smart object using the physical mobile interaction techniques, where the selection of them depending on the application areas used, location of the object, the distance between object and user, the service related to the object, the capabilities of the mobile device and the preferences of the user, in this way the user can take the advantages seen previously. Moreover, four interaction models could be used depending on type of connectivity. As shown in the paper, using the smart phone, different applications could be executed. In a not so far future, in order to help mainly the human life, new types of physical mobile interaction techniques could be used based on new sensor technologies[3].
8. References


IORIO, C: Smart Phone Interactivity and Pervasive Connectivity
WIRELESS PERSONAL AREA NETWORKS: ZIGBEE

Asif Zubair Bhatti

Summary

This report is written to partially fulfill the requirements to pass the course “TLT2656-Special course in networking”. The focus of this report is ZigBee, a Wireless Personal Area Network specification standard. This report provides introduction to PAN and ZigBee. Proceeding section discusses some networking aspects of the technology in detail. These networking aspects include networking topologies, addressing, properties, network formation, reliability and security. After that some issues of mobility in ZigBee network and integration with IP networks are described. In the last section some real world application are discussed and report is concluded with some direction in the current and future research aspects. Most of the material in this document is based on Jennic Ltd e-learning web based tutorial on ZigBee. [1]

1. Introduction

PANs are computer networks for communication between personal devices like computers, telephones and PDAs. These can be divided into wired and wireless PANs. Wired PANs with computer buses include USB and FireWire, while ZigBee, Bluetooth, UWB and IrDA are wireless. This report focuses on ZigBee wireless standard.

ZigBee is a standard for wireless radio networks. It has the features of Low power consumption, unlicensed radio band, easy installation and low cost. Network is extensible, flexible and automatically setup and exchange routing messages.

ZigBee Alliance defines specifications and certifies its implementation. It is a global ecosystem of companies for the creation of wireless solutions. The specification is built on physical and MAC layers and defined in IEEE 802.15.4. The specification also includes four other components: network layer, application layer, ZigBee device objects (ZDO) and manufacturer define application objects. ZDO is responsible for keeping device role, management of requests to join the network, device discovery and security.

It is mesh networking standard in its core and also supports star and tree topologies. Every network must have a coordinating device which is used to set up network parameters. Mesh and tree topologies allow to use the ZigBee routers. Tree uses frame beacons and communication is hierarchal while mesh does not use beacon frames. Beacon-enabled networks use routers to transmit beacons periodically to confirm their presence in the network. Power consumption is asymmetric (some device always active, others mostly sleeping). Non-beacon enabled networks use unslotted channel access mechanism.

2. Basic Architecture

Basic architecture of ZigBee software stack is shown in figure 1. It consists of three layers: application, ZigBee stack and physical/data link. Application layer describes the applications that can be run on the nodes. An application does conversion between input/output and digital data. Several applications can be run on one node. The ZigBee stack provides ZigBee functionality. It is concerned with structure of network, routing and security. The physical/datalink is concerned with addressing and message reception/transmission. It consists of two layers: MAC and PHY as described in figure 1. Those are based on IEEE .802.15.4 standard.
3. Networking aspects

Network layer is responsible for routing, according to network topology. It can handle the configuration of new devices to establish network. AODV (Ad hoc on demand distance vector) routing algorithm is used and routing requests are broadcast to neighbors to trace the low cost path for a destination. Each router updates its routing table with next hop and path cost.

ZigBee operates in unlicensed RF bands. It uses three different bands in different territories. These bands have center frequencies at 868, 915 and 2400 MHz with data rates 20, 40 and 250 kbps respectively. Each band is divided in channels and there are 27 channels in three bands, numbered from 0 to 26. The best frequency channel is selected at initialization. 868 and 915 bands favor fewer users, less interference, absorption and reflection of signals. 2400 band provides high data rate (250 kbps) and more channels and lower power consumption. Higher software layers have even energy detection functionality, used to avoid interference in radio transmission.

Transmission range depends on the environment. In outdoor environment range is longer than indoor. In indoor environment range is reduced due to reflection, diffraction and absorption and standing wave effects of walls and solid objects. A standard device with 0 dBm power can give 200 meters or greater range in outdoor and 30 meter in indoor. High power module can even obtain longer range. Range can also be increased through intermediate devices like routers.

3.1 Network Topologies

ZigBee can adopt three different network topologies: Tree, Star and Mesh. These are discussed in following section.

In start topology there is a central node and all the messages pass through central node called coordinating node. If the coordinating node goes offline or bottlenecked than communication is not possible and the whole network fails. For star topology network layer is not necessary, IEEE802.15.4 layer can provide star topology.
In tree topology nodes are arranged like a tree. One node is top node and other are arranged as branch/leaf of the tree. A message travels up and down in the tree as needed to reach the destination. Disadvantage of this topology is that there is no alternative route in case of link failure.

In mesh topology there is also a top node and arrangement is like a tree but leaf nodes can also be interconnected. A message travels to destination through a suitable route. It gives rise to efficient message propagation and alternative routes.

![Figure 2: ZigBee Network topologies](image)

In a single network ZigBee can support up to 65535 nodes. There are three kind of nodes in the network: coordinator, router and end device. All networks must have a coordinator node. In star topology central node while in tree and mesh top node is coordinator. In the above diagram blue color node is coordinator. Coordinator initializes network, selects the frequency of the channel and allow other devices to connect with the network.

End devices are located at extremities of the network. These are perimeter nodes in star and leaf nodes in tree and mesh networks. In the figure 2 these nodes appear in light blue color. These devices send and receive messages and can not relay messages. These are powered by batteries more often and when not active go to sleep mode to save power.

At least one router is needed in tree and mesh topologies. Its main tasks are relay messages and allow child nodes to connect. Star networks do not need routers and this function is performed by coordinator nodes. They are present mostly at position where relay is required. One important feature of router is that they can not sleep. These are represented by red nodes in figure 2.

### 3.2 Network Addresses

Each node must have unique address for identification. There are two types of addresses: IEEE addresses and Network addresses. IEEE/ MAC/ extended addresses are 64-bit addresses that uniquely identify a device. Network/ short addresses are 16-bit identifier of nodes. In network local address and these addresses are allocated by the parent node. Coordinator network address is 0x0000.

### 3.3 Network Properties

To define the size and shape of the network multiple properties can be preconfigured. These properties are taken in to consideration when a coordinator initializes a network. These properties are network depth, number of children and child routers. Depth defines the number of nodes from coordinator to the end device. Number of children parameter defines the maximum number of child devices (End devices & routers) that can be attached to a router. Thirds parameter defines how many children can be routers.
3.4 Forming a ZigBee Network

Coordinator searches a radio channel with least activity. The coordinator also assigns PAN ID to the network. This ID is either predetermined or a non-conflicting number is chosen with the neighbors by detecting neighboring networks. Once coordinator finishes the configuration it starts listening and responding the queries from other devices to join the network. Both routers and coordinator can listen and respond. The node scans available networks and selects the one it wishes to join. If node is able to see coordinator and multiple routers it selects one parent with strongest signals and sends joining request. The router or coordinator decides whether device is permitted and it has address space to allow device to join. Usually routers and coordinators have time periods to allow nodes to join the network.

Route discovery mechanism is used to find the best route to destination. It initiates with data request message. There are three options available. In SUPRESS route discovery mode message is routed along the tree. In ENABLE mode the message travels on already discovered route if exists otherwise discovers a new route. If router does not have capacity to store route it simply follows tree. In FORCE mode route discovery is always initiated. Use of this option is restricted as it generates a lot of traffic.

Router Discovery messages contains address of destination node and this is send by the parent router to all routers. The parent router of destination node sends back reply to the source nodes parent router. The hop counts and signal quality measures are recorded at each node while reply travels back to source. The best path choice depends on either less number of hopes or good signal quality.

There are two types of discoveries: device and service. In device discovery MAC or network address of destination is discovered. Service discovery involves determining the capabilities of remote node which could be type of device, its power characteristics and applications run on the nodes.

3.5 Reliability and Security

Reliability is to avoid data loss and corruption caused by radio interference or poor transmission/reception. ZigBee networks ensure reliability by taking certain measures which are described in following discussion. ZigBee devices implements CSMA-CD to listen before sending data. Acknowledgement mechanism ensures the successful reception of data. Alternate route discovery mechanism discover alternate route in case of a route failure.

ZigBee networks uses 128 bit encryption based on AES (Advance Encryption Standard) algorithm. Message timeout features rejects timed out messages. ACL (Access control list) is used to identify if the sending node is a valid or not. If node address is not ACL message is rejected.

4. Mobility in ZiBee Devices

ZigBee Networks work fine as long as all nodes are static. Detrimental packet loss can be seen when end devices becomes mobile. This becomes even worse when multiple objects move at the same time. Speed of end devices also impacts on the packet loss. Faster the speed greater is the packet loss and lower is the performance. ZigBee routers suffer less due to less mobility and because of their routing capabilities. [3]
5. Working with IPv4 & IPv6

ZigBee networks are not compatible with IP based networks. These both networks are different at their physical layer. To communicate between these two networks an intermediate device is required. That device acts as gateway to router messages between the networks. IPv4 has lot of limitations and is not usable in future due to shortage of available addresses. Therefore current research mainly focuses on inter-operatability of ZigBee with IPv6 networks. In this document two solutions for interoperability are discussed with their limitation and drawbacks.

“IPv6 over ZigBee network” solution put a stack over ZigBee NWK layer. In this all ZigBee nodes should have IPv6 address. When a gateway receives a packet from 802.3 networks it encapsulates it before forwarding to ZigBee networks. The packet size is a key problem in this approach. ZigBee supports SDU (service data unit) of 127 bytes with minimum transfer unit required for IPv6 is 1280 bytes. ZigBee does not support fragmentation of IPv6 packets to smaller ZigBee frames. The next approach 6LoWPAN creates an adaptation layer above MAC. This handles fragmentation of IPv6 packets to ZigBee frames. But this solution excludes NWK layer of ZigBee. This means that all the routing structures, address assignment and data forwarding needed to be redesigned. So this approach is also not feasible. The “translation approach” says every 802.3 node who wants to communicate with ZigBee network should also have a ZigBee address and must register itself with gateway. The ZigBee nodes send messages on ZigBee addresses and gateway node translate them before sending to 802.3 networks. Few other approaches have also been proposed and published. [4][5]

![Figure 3: Integration of ZigBee and IP networks](image)

6. Applications

ZigBee alliance has specified six application areas for ZigBee networks: consumer electronics, PC and peripherals, residential commercial controls, industrial controls, building automation and personal healthcare [6]. It also has applications in vehicle monitoring and agriculture.

In home electronics it can be used for HVAC, lightening, door locks, curtains/blinds and home entertainment systems. In commercial buildings it can be used for security checks against intruders and fire detection. In healthcare it is used in diagnostic devices such as patient monitoring during a health
fitness program. In vehicle monitoring it is useful when cabling is not possible e.g. pressure sensors in tyres. In agriculture land and environment conditions can be monitored to optimize the crop yields.

Figure 4: ZigBee applications

7. Conclusion and future directions

ZigBee is new and rapidly developing wireless PAN. It aims to provide cheap and easy solution for short range wireless communication. A lot of new devices have already built-in support for it. It has a lot of research potential especially the areas we saw in this report are integration of ZigBee Networks with IP networks and mobility issues.

8. References


[2] Figure licensed under Creative Commons Attribution-Share Alike 2.5 Spain License


WIRELESS PERSONAL AREA NETWORKS: 6LOWPAN

Vesa Salo

Summary

IP connectivity is spreading to all kinds of applications. New development IPv6 over Low-Power Wireless Personal Area Networks (6LoWPAN) refers to IPv6 which is integrated to IEEE 802.15.4-2003 – the current standard for LoWPANs. These networks are characterized by short range, low bit rate, low power and low cost. Many devices used will also be limited in computational power, memory and energy. This document will give an overview of the technology and mention some potential applications and problems as well as possibilities concerning it.

1. Introduction

Standard 802.15.4 offered framework for wireless connectivity for various types of Low-rate networks, e.g. sensor networks. When that technology is integrated to IPv6, we have a vast amount of possibilities implementing applications because IP has been used for a long time and technologies related to it already exist.

LoWPAN is a network which offers wireless connectivity in applications that have limited computational capacity, power and relaxed throughput. Some typical characteristics of 6LoWPAN are: small packet size, support for 16 bit or IEEE 64-bit extended media access control addresses, low bandwidth, two kinds of topologies (mesh and star), low power, low cost and so on. [1]

The listed characteristics and a need for IP connectivity form together a promising technology. Still there are many issues that need to be dealt with before implementing such networks. The maximum frame size on top of IEEE 802.15.4 is 127 octets and protocols like IPv6 and UDP have a big packet size for applications which means little or no space for the sensor’s data. Routing in different kinds of topologies should be implemented in such a way that computation and memory requirements are minimal. The network could consist of large number of devices and the network management should have little overhead and should enable self healing for the network. Also problems with service discovery should be considered so that controlling and maintaining services is possible. The security issues are also present but are considered only briefly in this document.

IPv6 is the basis of 6LoWPAN technology. The Internet Engineering Task Force (IETF) has succeeded eliminating the use of configuration servers by using the configuration and neighbor discovery capabilities of IPv6. Also unique stateless header compression reduces the space that IPv6 header needs to only 4 bytes. [2, 3] The header encoding and compression techniques are discussed in the following sections.

2. Technology in detail

2.1 Network and device types

The LoWPAN networks are expected to have a special function and the devices in the network have very limited capabilities. The devices are categorized into Reduced Function Devices (RFD) and Full Function Devices (FFD). FFDs are mains powered and have more resources than RFDs. The RFDs can
only communicate with FFDs, so clusters are expected to be formed between many RFDs and a single FFD [4]. One example of a network with two FFDs and multiple RFDs (e.g. sensors) are shown in Figure 1. The FFDs are considered to be network coordinators, packet forwarders and they are interfacing with other types of networks. [1] 6LoWPAN network topology can be a star or a mesh and both can be effective in current and future applications.

![Network example](image)

Figure 1. Network example.

All 802.15.4 devices have 64-bit MAC addresses but the 16-bit short addresses can be used within a PAN. The PAN coordinator assigns the short address during association and the validity expires when that association is over. [3]

IEEE 802.15.4 standard defines four types of frames. The frame types are beacon for synchronization purposes, MAC command frames, Acknowledgements and data frames. 6LoWPAN frames are sent in data frames. [3]

Because of real-time solutions and that RFDs are almost always battery operated; the RFDs are saving energy by ‘sleeping’ and those devices are intended to power their radios to receive mode during beacon slots so that time synchronization can be maintained with FFDs. [4]

There are three types of communication channels defined in IEEE specifications: 16 communication channels in the 2.4 GHz band with 256 kbps, 10 channels in 915 MHz band with 40 kbps and one channel in the 868 MHz with 20 kbps.

### 2.1 Protocol Stack

Like already mentioned the IPv6 offers solutions to these networks work without proxies. Also developers and designers can use existing standards to implement their solutions [2]. Standard protocols like Transmission Control Protocol (TCP), User Datagram Protocol (UDP), Internet Control Message Protocol (ICMP) and Domain Name Server (DNS), are potentially used.

![Protocol stacks of 6LoWPAN and Zigbee](image)

Figure 2. Protocol stacks of 6LoWPAN and Zigbee

SALO, V: Wireless Personal Area Networks: 6LoWPAN
Figure 2 shows the comparison between 6LoWPAN and Zigbee. It is easy to see that IPv6 offers many solutions in between the MAC layer and application layer. UDP is used in some applications today and many different solutions are developed using other existing protocols. The standardisation of IP networks and protocols offer at least partly usable solutions for 6LoWPAN. The Zigbee needs more adaptation and application specific adjusting to implement solutions. Proxies are also needed in Zigbee networks and it often means more complexity.

2.2 Frame structure

The maximum packet size for IPv6 over 802.15.4 is 1280 octets. As can be seen from Table 1, standard size of IPv6 packet doesn’t fit the 802.15.4 frame which is only 127 octets. If standard frame structure were to be used the worst case scenario would be as shown in Table 1.

<table>
<thead>
<tr>
<th>Max Phy layer packet</th>
<th>802.15.4</th>
<th>127</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max frame overhead</td>
<td>-</td>
<td>25</td>
</tr>
<tr>
<td>Result</td>
<td>=</td>
<td>102</td>
</tr>
<tr>
<td>Link layer security AES-CMAC-128</td>
<td>-</td>
<td>21</td>
</tr>
<tr>
<td>Result</td>
<td>=</td>
<td>81</td>
</tr>
<tr>
<td>IPv6</td>
<td>-</td>
<td>40</td>
</tr>
<tr>
<td>Result</td>
<td>=</td>
<td>41</td>
</tr>
<tr>
<td>UDP</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>Result, left for application data</td>
<td>=</td>
<td>33</td>
</tr>
</tbody>
</table>

Table 1. Calculation about the worst case for application data per frame

In addition to this calculation the need for fragmentation and reassembly layer will use even more space. These calculations show that adaptation layer is needed so that frames can fit into the 127 octets. Also compression will be needed so that the different overheads will not use 74% from the single frame.

Solution to the problems shown in calculations are the 4 basic header structures defined in RFC 4944 [3]. The first two bits define the header type and the following 6 bits indicate the options used in the headers. The basic frames include [2]:

- **Dispatch Header**: used to define what type of header follows, options show if frame is using compressed addresses or not.
- **Mesh Header**: options include information about addresses (16-bit or 64-bit) used by originator and the final destination and the hops left count.
- **Fragmentation Header**: bit pattern indicate if the frame is the first fragmented and if it is followed more by the same type.
- **HC1 Header (Header Compression)**: options about the compression of addresses.

These separate headers can be stacked for specific needs of a packet and network. For a simple case when network topology is a star and the packet is a small compressed IPv6 packet only Dispatch header and HC1 header are included as shown in Figure 3.

<table>
<thead>
<tr>
<th>Frame Hdr</th>
<th>Dispatch Hdr</th>
<th>HC1 Hdr</th>
<th>IPv6 compressed Hdr</th>
<th>UDP Hdr</th>
<th>Application Data</th>
</tr>
</thead>
</table>

Figure 3. Point to point small packet. [2]
The stacked headers approach has the benefit when designer needs a new header type. A new dispatch type can be requested from IANA (Internet Assigned Numbers Authority). Currently only 5 of the available 64 Dispatch Headers are defined so much can be designed for these networks.

3. 6LoWPAN applications

3.1 Application potentials

6LoWPAN technology has great potentials. Many solutions have been developed already and probably much more will follow. Many kinds of industrial monitoring systems and hospital systems as well as structural monitoring can be handled with 6LoWPAN.

Industrial monitoring system can be associated with methods trying to increase productivity or energy efficiency. Companies are forced to use expensive monitoring systems to predict and identify failures. Many of these systems can potentially be replaced with wireless sensor networks. Sensors can ensure product quality, control energy consumption and increase operation safety. Vibration detectors can detect equipment failure or a need for maintenance. Storage conditions can be monitored so that products will not be retained in too high temperature conditions. [5]

Same kinds of potentials are in a hospital environments. Temperature must be right in storage rooms so that blood cells or organs will not be kept in heat for too long. Then there is a possibility that the organs are moved into different parts of the hospital while still monitored that the temperature is correct or will not be exceeded for too long (blood can’t be used when held in wrong temperature approximately 30 minutes). [5]

Structural monitoring can be related to safety checks or periodic monitoring structures such as buildings or bridges. The network can be accessed from the internet to check different parameters and sensors are in sleeping mode much of the time. [5]

3.1.1 Application developer: Sensinode ltd.

Finnish company called Sensinode ltd is one of the first to develop and implement 6LoWPAN solutions. Their solutions offer enterprise-wide building automation, asset management and metering infrastructures real-time. For example the network used in asset tracking can have range up to 300 meters using the 2.4 GHz band and usage can differ from logistics to human or vehicle (location and status) tracking.

Sensinode’s sensor networks are semi-infrastructure, so network can be a star when each node communicates directly to coordinator called NanoRouter (USB connected to a computer) or alternatively nodes can have multiple hops. These networks are built upon UDP (see figure 2) and can be maintained with ICMP. [6]

Sensinode also offers a Devkit that consists of multiple components such as Devboard needed to programming and debugging, radio modules, NanoRouter and NanoSensors (programmable example node with temperature and light sensing capabilities). The developers kit also includes Java based graphical network tool. The common network is an ad hoc and the devices can find the router using ICMP. [6] Sensinode’s idea is to offer the starting point and then companies can develop their devices and capabilities further.
3.2 Issues related

The discussed applications often require integrity and confidentiality. For example very sensitive information can be transferred in hospital environment. Security functions can be provided in all of the layers but the limiting factors are small code size, low power operations and small bandwidth requirements. Also denial of service and man-in-the-middle attacks should be considered. IEEE 802.15.4 devices already support Advanced Encryption System (AES) and it can be part of the security solution. [1] IPsec would offer security in many ways but implementing it to small nodes is very problematic. Other security issues are for example how to implement secure neighbor and service (other nodes) discovery. [2]

Other issue is the use of available frame types. The RFC 4944 [3] leaves much room for future development of frames but the usage of these is still open. Developers might end up using the same frame structure for different purposes if IANA is not involved in the process.

4. Conclusion

The 6LoWPAN technology offers wireless networks and applications that have low computational power, low bandwidth, and low cost. Using standard Internet Protocol makes different technologies available and IPv6 allows developers to use existing protocols like UDP or TCP. The networks can easily be accessed and adjusted from the internet.

The headers in standard protocols are huge for these kinds of network devices so compression mechanisms and fragmentation and reassembly features are provided in the RFC 4944 [3]. Already defined headers give a starting point but additional header types can be requested from IANA.

Security issues remain and the use of IPsec is tricky because of sleeping nodes and small code sizes. The issues are currently studied and more solutions developed. Only time will show if 6LoWPAN is the winning technology over Zigbee or other radio sensor technologies but for now the possibilities are greater than in other standards.

6. References

PART 3

TRENDS IN OVERLAY NETWORKING
Summary

The end-to-end connectivity between two hosts in the internet can not be anymore taken for granted. Technologies like statefull firewall and network address translation (NAT) may break the connectivity entirely or partially. Networking techniques, as tunnelling and overlay networks offer a solution to restore the connectivity. In this report, I first shortly introduce the causes of the connectivity problems and after that the concepts of tunnelling and overlay network with their pros and cons along few examples of their usage together.

1. Introduction

The internet architecture has greatly changed since the beginning of the internet. The end-to-end principle in a sense as it was first described in a paper by Saltzer, Reed and Clark, “End-to-End Arguments in System Design” rarely applies anymore [1]. The networks, from which the internet consists of, have evolved from a dummy network into a more like middleware, consisting of statefull firewalls and network address translation (NAT) devices [2]. I'm not implying that firewalls are bad thing, they increase network security, and NAT is a result from the depletion of IPv4 addresses, but they both have their influence to the internet architecture and they prevent the possibility of end-to-end connectivity.

In a network, reachability is usually identified as a possibility to transmit packets to the destination. In a classical way this means that there is a route to the destination, but with an introduction of middlewares in networks reachability can be redefined as a state where only certain kind of communication is possible towards the destination. Earlier, if there was a route to the destination, all kind of communication over IP protocol was generally possible. Now reachability is not anymore same to all protocols. For example in case of protocols supporting multiplexing through ports, e.g. transmission control protocol (TCP) or user datagram protocol (UDP), the reachability may not be the same for all ports.

The introduction of private and special IP version 4 (IPv4) addresses [3] and IP version 6 (IPv6) have lead to a situation where not all the networks utilizing these addresses are connected to each others. The connectivity between them at the IP layer may not even be possible because intermediate networks along the path may not support theses special addresses or address families.

2. Tunnelling

In a context of network protocols, a tunnelling - in it's simplest form - is encapsulating packets of another protocol into the payload of another protocol. Because the tunnelling as a technique does not actually differ from the normal usage of protocols, where protocols at the layer below offer a carrier or transport service to upper layer, it is more a matter of taste if the encapsulation is called tunnelling or not. In terms of tunnelling, the carrier protocol is sometimes called as a tunnelling protocol. One reasoning for the term tunnelling could be that if the carrier and payload protocols do not depend on each other they in a way hide the their own network context. The tunneled protocol is not aware of actual network layer and the for the tunneling protocol the payload is just sequence of bits.
If the size of the encapsulated packet is greater than the maximum packet size of the tunnelling protocol, the tunneled packet will be split into smaller fragments (see Figure 1). The relative hierarchy of protocols does not matter, as it is as possible to tunnel OSI layer 2 packets inside OSI layer 4 packets as it is to tunnel OSI layer 3 packets inside OSI layer 3 packets. The endpoints of the tunnel just have to do the adaptation for the tunneled protocol if they forward it to a native network when the packet comes out from the tunnel.

![Fig.1 Encapsulation where tunneled packet is fragmented.](image)

The tunneled protocol can have an destination address and/or address family of it's own, independent from the address or address family of the tunnelling protocol and protocols below the tunnelling protocol. Depending on which OSI layer the tunneled protocol works on, the address family can be an general one as e.g. IEEE 802 MAC address, IPv4 or IPv6 or it can be application specific.

The OSI layer of tunneled protocol also affects the way of what kind of interface to the tunneling is offered to the operating system or the application. If the tunneled protocol is a link layer or network layer protocol, the tunnelling application usually offers a kernel level interface as e.g. network interface to use the tunnel, but if the tunneled protocol is an application layer protocol, as e.g HTTP which may not have it's own destination address or the address family, the interface is application specific and the tunnelling is handled by the application.

Payload of the tunnelling protocol may be in cleartext or it can be encrypted, as in most of the cases of virtual private networks (VPNs), secure socket layer (SSL) or transport layer security protocol (TLS). So even though, the term tunnel is often associated with the VPNs they are just a subset domain of tunnelling.

### 4. Overlay networks

Overlay networks are virtual networks, which have their own topology, addressing scheme and possibly address family independent from the topology, addressing scheme and address family of the actual network below (see Figure 2). Overlay networks can be totally or partially structured or they can be pure ad-hoc style peer-to-peer networks. Actual network can also be a part of overlay network. In that case, the overlay network and the real network extend each other (see Figure 3). Still in general, if the topology of the overlay network is at least partially structured, the same topology design rules which applies to real networks can be applied also to the overlay network.

The links between nodes in the overlay network are created by tunnelling packets of overlay network between the participating nodes. Because the address family of the overlay network is independent from the actual network layer, it is possible for example to create an overlay backbone network, connecting networks utilizing network address families as IPv4 multicast or IPv6, which may not be
supported by the networks along the normal path on the network layer between these networks. The links in the real network part usually do not use tunnels. In a structured overlay networks new nodes usually connect to specific nodes offering tunneling service. These nodes are generally called as tunnel brokers. In VPN context they are also often called VPN terminators or concentrators. In a peer-to-peer networks, the nodes offering access to the overlay network are called bootstrap nodes.

Hosts behind a statefull firewalls or NAT devices can restore the connectivity towards them by joining an overlay network with a tunnel created with TCP or UDP protocol, which assures that the transport layer connection to the tunnel broker or bootstrap node stays alive - possibly using keep-alive packets during idle times.

Fig. 2 Examples of different kind of overlay network topologies.
4.1 Problems with overlay networks

Even though, overlay networks solve some problems they also introduce new ones. Most common problems of overlay networks are overhead of protocol headers, inefficient routing of packets, low link bandwidth and low failure tolerance of tunnel brokers or bootstrap nodes.

The overhead of protocol headers is often accepted because overlay networking usually solves a greater problem. The overhead becomes a more serious problem when the transferring rate in the overlay network increases. In that case, the size of the extra headers just have to try to minimize. Depending on the size of the average packet - e.g. the overlay network is used for transferring big files and the high transferring rate is more important than the low latency - one solution could be fixing the size of the header and increasing the maximum packet size.

Inefficient routing is often an outcome of unorganized or ad-hoc style overlay networks because the topology of overlay network rarely reflects the network topology below it. The tunnel between adjacent nodes in overlay network can go through multiple routers and very unoptimal paths in a real network. The problem can be addressed by controlling and organizing the topology of overlay network. Technologies like anycast routing - where packets to a certain destination are always sent to the nearest possible destination - or source address based domain name service (DNS) replies - where the reply for the DNS query depends on the source address where the query came from - can be used to attain efficiency by directing tunnel clients always to nearest available tunnel broker and by right topology choices and adequate link bandwidth to ensure the capacity of the backbone of the overlay network connecting the tunnel brokers mutually. In a peer-to-peer networks, where it may be hard to try to organize where in the overlay network new clients will connect, changing the topology away

VARTIAINEN, T: Reachability via Overlay Networks & Tunnelling

105
from the star like topology to some other, where problem would not exist can minimize the problem. For example many peer-to-peer networks creating a distributed hash table (DHT) use a ring as a topology and an index number (a key of a bucket in the hash table) as an address of a node. So utilizing a own routing algorithm, the number of intermediate nodes between any two nodes of the overlay network can be decreased or minimized.

The low link bandwidth is especially a problem of peer-to-peer networks where two client peers communicate through other clients. The problem is not major if relatively low transfer rate between the nodes is enough, as in case of instant messaging or voice over IP (VoIP) calls. If the nodes want to transfer large amount of data in a shorter time, they could try to create a direct connection between them possibly using some NAT traversal technique to restore the connectivity if it is not possible otherwise.

In a structured overlay networks, the failure tolerance of tunnel brokers can be increased through the same high availability methods used for any network service. Duplicating the hardware and network equipments and links the tunnel broker is protected from the hardware failures. Based on a service availability and quality monitoring data, the service discovery as e.g. DNS can be used to serve addresses of tunnel brokers which have low utilization and better availability for the new clients. The availability of the overlay network heavily depends on the topology of the overlay network. E.g. peer-to-peer networks using ring topology and their own routing algorithm are more resistant for node failures than the overlay networks using start like topologies.

5. Examples of overlay networks

The tunnelling and overlay networks in general do not greatly impose any constraints, but neither offer any solid solutions in plain, so utilizing an overlay networks depends a lot on the design of the implementation. Here I introduce two different kind of overlay networks used for different purposes. The examples are just a small subset of the area and there is wide variety of different kinds of overlay networks especially in the field of peer-to-peer networking.

5.1 Akamai Technologies' content distribution network

Akamai Technologies is a U.S.-based firm operating a global content distribution network (CDN) and specialized in offering outsourcing solutions, for delivering data through their network of caches. The content distribution network consists of Akamai's cache server sites deployed on different internet service providers' (ISPs) points of present (PoPs) around the globe. The cache server sites are connected to the content providers sites through an overlay network. The quality of connectivity, as the latency and the packet loss between Akamai's cache sites and content providers' sites are monitored continuously and if the current default path from a cache site to a content provider's site becomes congested, an alternative less congested path through the overlay network will be used. [4]

5.2 Chord

The Chord is a scalable and robust distributed peer-to-peer system developed at Massachusetts Institute of Technology. It utilize a ring topology in a overlay network. The main purpose of the chord is to offer a distributed data storage and lookup service for key-value pairs through a distributed hash table. The address for a node in a ring is a hash value, calculated with SHA-1 hash algorithm from the node's IP address. Chord's own routing algorithm tries to forward a packet between any two nodes with
O(\log N) hops where \( N \) is a number of participating nodes in a ring, despite of the joining and leaving nodes. [6]

6. Conclusion

Tunnelling and overlay networking as concepts are enabler technologies to overcome reachability issues, created by the current middleware filled internet architecture, with an option to also increase security of the communication. They offer a solution to restore end-to-end connectivity with a cost of increased complexity and decreased efficiency. The key to the success in using the tunnelling and overlay networks is understanding the problems they introduce and using their ability to separate protocols from the different contexts and constraints, so the protocols can be applied elsewhere to solve other problems.

7. References

CONTENT NETWORKING AND IP MOBILITY

Florent BECATTINI

Summary

This paper covers three main issues. First of all, we will discuss about the user’s mobility on the IP network and how the protocol Mobile IP version 6 (MIPv6) deals with it. Secondly, we are going to discuss the way to serve content with a reliable and high capable transmission. We will focus on two kind of architectures called Content Delivery Networking (CDN) and Distributed Content Delivery Networking (DCDN). Some economic and efficiency issues concerning CDNs will lead to the study of the second one. This latter has provided really hopeful outcomes in condition to adjust the number of users per group of peers sharing their resources in term of storage capacity, processing power and bandwidth. Finally, we will turn our attention on the content adaption in order to fit with the capabilities of any terminal asking for content.

1. Introduction

Nowadays, the growth of the World Wide Web and new modes of Web services have triggered an exponential increase in Web content and Internet traffic. The main current topic could be sum up as serving any kind of users (computers, PDAs, mobile phone etc.), anywhere at any time. Through these terms, what is expected is to provide an adapted content which is suitable to the user’s terminal in a minimum delay and deal with the mobility by maintaining the connection with this one even if he’s moving on the network. Our report will describe the principle of MIPv6 and emphasizes how it is used in CDNs and DCDNs architectures. After that, we are going to see how terminals and content servers communicate in a way to adapt the content as suitable as possible according to the user’s capabilities.

2. Mobile IP version 6

Issues to take into consideration with mobility can be decomposed in three sub-problems: the device should be able to communicate, it should be reachable and it has to keep its current communications active [1].

The first one is resolved thanks to the auto configuration mechanism directly embedded in IPv6. MIPv6 will enhance this feature in term of delay to get a new address on the network. The second issue is more complex and could not been achieved with simple updates at the Domain Name Server level. Time constraints compel to find a new consideration which will be explained further in this report.

The last one has to take into consideration the address duality between localization and identification aspects. When a mobile device is connecting to a new network, its address will be modified to figure out its new localization and consequently, its identification will also be affected. However, each TCP connection needs 4 parameters which are the source address, the source port, the destination address and the destination port to keep alive the session.
between two nodes. If one of them changes, the connection is lost! We will consider how MIPv6 manages with this to prevent any interruptions.

First of all, each node on which is running MIPv6 will be in charge of two IP addresses. The first one, called Home Address (HoA), refers to the node in its home network. The second one, called Care of Address (CoA) is used when the node has moved in a foreign network. The equipment, which is responsible for keeping a link between these two addresses, is known as the Home Agent (HA). Located on the home network, it is in most of the cases a router. Thanks to association table updated by the binding message sent by the mobile node when he visits a new network, the HA is able to forward any packets at destination of HoA towards the current address of the mobile node, the CoA. Let’s have a look on three practical cases.

**2.1 Mobile node in its home network**

![Fig 1: Mobile Node is his home network](image)

Mobile node is in home network (Fig 1) and consequently, is communicating with its HoA. The communication takes place in a normal fashion. The HA is inactive.

**2.2 Mobile node in a foreign network**

![Fig 2: Triangular communication](image)  ![Fig 3: Optimized communication](image)
When the mobile node visits a foreign network, two kinds of communication can take place. The first one is called the triangular communication. As you can see in Fig 2, three nodes are involved in the session. The HA acts as a third part and is in charge to refresh its association table thanks to the binding updated sent by the mobile node. Thus, when a packet arrives at destination of HoA, the HA forwards this latter via a tunnel IPsec. To ensure the transparency at transport layer or in other words, to maintain the 4 crucial parameters similar along the entire communication, the HA uses a MIPv6’s extension called IP-IP in order to send the packet to the current destination of the mobile node. This one will delete this IP-IP extension and consider the packet as if it has received this one in its home network.

The second method establishes a direct communication between the two nodes in order to save some time and resources. For instance, if we consider that the mobile node is close to the network of the correspondent node, it would be completely irrelevant to use the HA potentially far away. Furthermore, if we take into consideration the fact the HA takes care of plenty of other nodes, it will be quickly overloaded during its forwarding process and thus, extra delay will be added. So, as you can see in Fig 3, any correspondent nodes which are communicating with the mobile node will now receive some binding updates in the same way that the HA will. One fundamental condition to perform this is to be sure that these correspondent nodes support the routing optimization. If not, they will have to keep their communication via the HA. Let’s consider that these nodes authorize the routing optimization. In this case, once they have refreshed their own association table, they transmit their packet with a routing header extension in which they indicate the HoA. The mobile node receiving these packets, will simply substitute the CoA used to route packets with the HoA. At the transport layer point of view, the mobile node has just received a packet in its home network and that’s why, no interruptions occur during the TCP connection.

Anyway, an important issue concerning the security level of the optimized communication has to be improved. In the triangular communication, the administrative relationship between the HA and the mobile node ensures the sending of binding messages safely (e.g. share of a secret key). Nevertheless, with the optimized communication, it is really complex to protect binding updates exchange between each peer. This issue has to be resolved if we want to take advantage of the optimized communication in a secure way.

Now, let’s turn our attention on CDNs and DCDNs architectures. In the following parts, we describe the main components of each one and how they try to serve content in an efficient way. Besides, major pitfalls which led to study carefully DCDNs will be presented.
3. CDN and DCDN architectures

3.1 Content Delivery Networks

With the rapid development of Internet, more and more users are daily requesting for any kinds of content. To serve it, the content provider can’t anymore afford to use one or even a farm of servers to store and distribute the content. That’s why, the Content Delivery Networks (CDN) have been designed. Actually, CDNs create their own network of servers around the globe to effectively deliver Web content to the end-user. Currently, we distinguish commercial and academic ones. Concerning the first ones, we can quote the company Akamai, leader in this business. For the second ones, we have some organizations such as Globule or Coral. Before focusing on advantages and drawbacks of each one, let’s have a look on the architecture and how it works [2].

![Diagram of CDN architecture]

*Fig 4: Content served by CDNs*

When a user, the correspondent node, asks for content (*Fig 4*), its request is sent to the DNS server which will then forwards this latter to the IPv6 Request Router (a) (b). This one is considered as the Home Agent and will forward the packet to the most appropriated CDN server, the mobile node (c). To perform its choice, the IPv6 Request Router evaluates each CDN server in term of proximity, bandwidth, number of requests to server in their queue, processing power available, memory used … The communication between the client and the CDN server could be achieved (d). Thanks to the routing header extension they use, the CDN server can be modified for any reasons without affecting the communication.
The user is so served by the most suitable CDN server at any time and this, in a completely transparent fashion.

Commercial CDNs relies on Client-Server architecture and provide content in a really efficient way. However, the high rental rates resulting from huge infrastructure cost make it inaccessible to medium and low profile clients. To deal with this issue, academic CDNs build their architecture with the help of voluntary groups. Unfortunately, even though it is a manner to save some money, these academic CDNs are at the mercy of peers and consequently, don’t provide a sufficient performance level in comparison with commercial CDNs.

3.2 Distributed Content Delivery Networks

The challenge of Distributed Content Delivery Networks (DCDN) is to reach the efficiency of commercial CDNs while using the concept of academic CDNs in order to save some money. Here is the hierarchy architecture of the DCDN [3]:

As we can see in Fig 5, we have two main components in DCDN architectures, DCDN servers and DCDN surrogates. DCDN servers are basically used as redirectors. They have the knowledge about the location of the content but they do not store any content as such. Master ones are the first point contact with the content provider. They control a group a local DCDN servers and distribute content among them according to its own policies. Local DCDN servers are in charge of sharing out content between different DCDN surrogates. They are the large number of Web users offering resources in terms of storage capacity, bandwidth and processing power to store and make available Web content. The main idea to encourage them to be active in this process is to give them remuneration in function of their participation. In the previous picture, we have an example where a user is downloading content. Similarly with
the CDN, the request will be forwarded by the DNS server towards the local DCDN server (1) (2). This latter has to perform a load balancing algorithm to choose the best surrogate (3). The IP address of the most suitable surrogate is sent to the user (4). This one establishes a connection directly with the surrogate to download the content expected (5).

In terms of performance, the following graph (Fig 6) shows that DCDNs could provide as much as commercial CNDs in condition to adjust some critical parameters such as for instance the number of surrogates available to serve a certain amount of users’ requests.

![Average Page Response Time (seconds)](image)

**Commercial CDN scenario:** 150 clients, 3 servers, 100 Mbps link  
**DCDN, scenario 1:** 150 clients, 6 surrogates, 10 Mbps link  
**DCDN, scenario 2:** 75 clients, 6 surrogates, 10 Mbps link  
**DCDN, scenario 3:** 30 clients, 6 surrogates, 10 Mbps link

We clearly observe that scenarios 2 and 3 are not comparable with commercial CDN results. Nevertheless, the scenario 3 (red curve) displays good outcomes. In this particular case, we notice that only 30 clients were served and 6 surrogates were storing content. This is 5 times less of users’ requests and 2 times more of servers to store content than the commercial CDN scenario. We can conclude that, roughly, we need to replace each commercial CDN server with 10 surrogates (Web users) if we want to guarantee a corresponding level of performance.
4. Content Adaptation

Nowadays, there is a proliferation of handheld devices (PDAs, smart phones, handhelds phones) which are able to connect on the IP network to request some content. Consequently, it is necessary to provide them a content which is adapted to their capabilities. To understand this issue, we will consider which types of adaptation are conceivable, what methods are available to perform this adaptation and how the terminal can inform the content provider of its capabilities [4].

4.1 Types of content adaptation

To fit with the capabilities of any terminal asking for content, there are 5 types of content adaptation: Format Adaptation, Characteristics Adaptation, Appearance Adaptation, Encapsulation Adaptation and Size Adaptation. The first one converts original content format to the format understandable by the receiver (e.g. from .jpg to .gif). The second one is used to modify properties of content in a given format (e.g. resolution, number of color used…). The third one changes the readjustment or deletes some components of the content (e.g webpage without pictures). The fourth one uses a different application protocol to transmit the content (e.g. from one Email to several SMS due to low bandwidth or size of the receiver’s buffer). The last one takes advantage of all the previous ones to decrease the size of the content.

4.2 Methods of content adaptation

To achieve the content adaptation, we usually consider two methods: the Multimedia Transcoding and the Content Selection. The first one, totally automatic, performs different modification in relation with the types of content adaptation presented previously. Although its process doesn’t need the human intervention, it may be require a lot of processing resources for a result not acceptable or usable for the terminal asking for content. Besides, some copyright issues have to be taken into consideration. It may be illegal to modify the content. That’s why, the second method was designed to solve these drawbacks. In this latter, two key technologies are used to provide content adapted to the user’s capabilities: the Infopyramid and the Customizer.

![Infopyramid of a video item]

*Fig 7: Infopyramid of a video item*
The Infopyramid (Fig 7) is a representation scheme that provides a multimodal, multiresolution representation hierarchy for multimedia content. This one is built thanks to transcoding module but also with the author’s supervision. The modality axis provides the same information under different media modes (text, video, image etc.). The resolution axis provides, for a given modality, the content at different quality levels.

Concerning the customizer, he is responsible for selecting the most suitable content according to the terminal capabilities sent in the user’s request.

### 4.3 Terminal’s capabilities

When a user is requesting for any content, he needs to provide its terminal’s capabilities and preferences profile. There are two principal way to perform that. The most common way is to insert user agent information (UA) as a header field into HTTP request. This one contains the terminal model and its software version. Thanks to them, the customizer is able to find the entry corresponding to this terminal in a capability database. While really basic to implement, through the UA, there is no support for attributes not related to the terminal such as connection bandwidth or user preferences. Moreover, dynamic attributes as for instance the current available memory on the terminal couldn’t be communicated. Last point concerns the possibility to install a couple of applications on modern terminals (PDAs, smart phones). UA can’t handle with personal tuning that everyone can achieve because of the complexity to update the capability database according to each one. Consequently, a new way to communicate the terminal’s capabilities has been established. User Agent Profile (UAProf) relies on Composite Capability / Preference Profiles (CC/PP) in order to provide an adapted content as suitable as possible with the requesting terminal. UAProf can be perceived as a plugin vocabularies which provides components and attributes relevant to mobile terminals. This one has been agreed on by all mobile terminal manufacturers. The real innovation is the simplification to deal with the transmission of capacities and preferences profile. More precisely, the UAProf header embedded in HTTP packet could refer to a repository maintained by the terminal manufacturer (static way) via URL to indicate its capabilities but also transmit some information which is likely to be different from default values.
5. Conclusion

In this report, we have covered some issues about user mobility, content delivery and content adaptation. DCDN architecture appears as an excellent way to combine all of these in a cheap way. Using the MIPv6, it permits to select the best server to deliver some content at any time and without breaking the connection with the user. It could be also imaginable to provide mobility for the user while downloading content on a surrogate by allocating some home agents in charge to communicate with local DCDN servers. However, some stakes stays problematic. The content adaptation is complex to perform due to the location of the content. Surrogates resources are not sufficient to perform relevant transcoding adaptation or store infopyramid to provide a suitable content to the requesting user. Finally, security and integrity have to been carefully taken into consideration. The content is very vulnerable to corruption given that it is located directly on peers without actual and strong security administration. The future studies have to focus on these issues without forgetting to describe a relevant business model to remunerate the active surrogate in function of their participation.

6. References

PEER-TO-PEER COMMUNICATION IN PERVASIVE NETWORKS

Jani PELTOTALO

Summary

At the moment there is a growing interest to research, develop, and use different types of peer-to-peer media streaming applications. Lots of studies how to improve the performance of a peer-to-peer file sharing survivor BitTorrent are also available. In this paper a brief overview about peer-to-peer media streaming and Clustered BitTorrent concepts are given. This paper also summarized results from experimental tests carried out by using four different peer-to-peer media streaming applications with a PC over different network connections.

1. Introduction

During previous few years peer-to-peer communication between different devices using different access network technologies has become more and more attractive. Users want to use the same tried and tested applications all the time regardless of their location. Currently the file sharing field is quite quiet and many popular applications have been shut down because of legal issues. Still one technology, BitTorrent, keeps going and lots of analysis and improvements are published every year. A new field in the peer-to-peer communication, media streaming, has become very popular research and development topic around the world. In the future, when it is possible to use peer-to-peer media streaming applications any time anywhere with a huge amount of popular channels, it would be possible that the dominant position of the traditional broadcast type of media delivery is challenged.

Several challenges exist when applying peer-to-peer communication in pervasive networks. Issues to be taken into account include firewalls, churning phenomenon, vertical handover, variations in the delay and throughput of access networks, and different CPU capabilities between devices. Application developers must somehow find a good balance between different requirements of different devices to gain the maximum number of peers possible to join the network.

Next, some background information about peer-to-peer communication is given in Section 2. An introduction to the Clustered BitTorrent and peer-to-peer streaming concepts are presented in Sections 3 and 4. After that some results of experimental tests carried out by using different peer-to-peer media streaming applications are given in Section 5. Finally, Section 6 concludes this paper.

2. Peer-to-Peer Communication

Peer-to-peer networks can be classified in many ways. Usually the classification is done based on the use case, and peer-to-peer networks are divided into six distinct categories: file sharing, instant messaging, voice over IP, storage networks, distributed computing, and media streaming. On the other hand, with file sharing, instant messaging and voice over IP, networks could be also divided based on the overlay network layout and used categories are unstructured, centralized, pure, hybrid, structured and DHT-based. Similarly with media streaming, the networks could be divided based on the media delivery network layout, and used categories are tree- and mesh-based.

At the moment many peer-to-peer networks can be used with different devices using different access network technologies. BitTorrent file sharing client can be used with Windows, Linux, MAC, Symbian...
and Nokia N800 Internet Tablet. Skype client for instant messaging and voice over IP can be used with Windows, Windows Mobile, Linux, MAC, Symbian and Nokia N800 Internet Tablet. With media streaming the range is more constrained. Octoshape client for live media streaming can be used with Windows, Linux and MAC devices, and SopCast client for Video on Demand (VoD) media streaming is available for Windows and Linux.

There are several challenges when using peer-to-peer technologies in pervasive networks. Some of the challenges, like firewalls, churning phenomenon, and vertical handover, affect to all technologies, and some affects only to one technology.

Because firewalls might block connections, bi-directional communication between two peers must be somehow allowed. This can be done by using some Virtual Private Network (VPN) solution if the application cannot handle the firewall issue by itself. Another important issue to be solved by the application developers is how to handle churning phenomenon when multiple peers rapidly attach and detach to the network within a short period of time. To be able to support seamless mobility applications must also handle vertical handover across various technology domains.

With media streaming the delay and throughput of the access network affects to the user experience, and should be taken into account when designing an application to be used with different access network technologies. The delay bears on the buffering period at the beginning of the stream, and if the used stream bit rate is too high compared to the throughput it is not possible to use the application at all. Differences in CPU capabilities might also exclude some content encoding formats, especially when dealing with mobile devices.

3. Clustered BitTorrent

In the original BitTorrent [1] peer-to-peer file sharing system, the tracker could easily be overloaded by the state information updates of constantly arriving and leaving peers. When a connection request from a peer is received, the tracker responses with random list of peers, and these randomly selected peers among the all peers in the networks might create a long delay to the file sharing between two peers. More information about the original BitTorrent protocol can be fetched from [2].

To improve the file sharing performance of original protocol, a hierarchical architecture named Clustered BitTorrent (CBT), which group peers into clusters according to their proximity in the underlying overlay network is proposed by Y. Jiadi and M. Li [3]. In this study proximity between peers is measured using Round Trip Time (RTT) and Time-to-Live (TTL) values. It would be interesting to see what happens to the performance if the proximity is measured using geographic locations when the actual data route between two peers may go through a distant router.

In the CBT two novel algorithms, a peer joining algorithm and a super-peer selection algorithm are introduced. By using these algorithms clusters in the CBT network are evenly distributed and peers in the cluster are relatively close to each other. Because peers in a peer-to-peer network are free to join and leave whenever they want cluster maintenance is also taken into account. A big cluster can be split into two separate clusters and a small cluster can be merged into some neighbour cluster if needed.

Figure 1 presents example architecture of a CBT network. The CBT network is composed of a tracker and three distinct types of peers: super-peer, seed and downloader. The tracker has normal BitTorrent functionalities and acts as a super-peer in the fundamental cluster. Periodical monitoring of the super-peer backbone network, to be able to provide precise information to new peers joining the network, is...
also tracker’s responsibility. The super-peer acts as a cluster head, and collects and maintains the state information of all peers in a cluster. Seed is a peer who has a whole file to be uploaded by other peers, and downloader is a peer who downloads the file and simultaneously uploads correctly received file pieces to other peers.

The file downloading process can be divided into cluster joining and data exchange with other peers in the cluster. Before a peer can join to the cluster a .torrent file must be fetched somewhere, usually from some WWW server (phase 1 in Figure 1). This file contains an URL of the tracker to ask the super-peer list for the file in question (phase 2). If there is only a fundamental cluster then the peer joins to it, otherwise peer joins to that cluster which super-peer is nearest. Next the peer gets a random peer list from the selected super-peer (phase 3). After this phase the file download process continues as normal BitTorrent file exchange inside the cluster. Peer requests file pieces from several peers (phase 4) and after each file piece is correctly received a download status update message is sent to the super-peer (phase 5).

Section 5 in the CBT paper presents simulation results to evaluate the proposed system against the original BitTorrent approach. These simulations showed that the CBT system can achieve faster download speed and higher file availability compared to the original system. Because the lack of CBT implementation large scale field testing is not possible and simulation results cannot be verified in the real environment.

4. Peer-to-Peer Media Streaming

Peer-to-peer media streaming is a method for multicasting or broadcasting streaming media over the Internet using a peer-to-peer network. It can be seen as a combination of traditional television or radio broadcast type of media delivery over a new kind of delivery medium. The user wishing to receive a stream needs join the network before the actual data traffic can occur. After the user has joined the
network, there is a varying warm-up/initial buffering/bootstrap time before any data can be consumed to ensure seamless viewing or listening of the streaming media.

With live streaming, users wish to receive an ongoing event, like a football match. After an initial buffering the user starts to watch the stream from a certain place and all peers consume the data in the same time frame. With VoD streaming the user searches video from some catalogue and after a certain amount of initial buffering the user starts to play the video from the start.

Peer-to-peer media streaming systems can be either push-based or pull-based. In a push-based system, peer sends the data to other peers without explicit request after it has received the data. One of the problems with this approach is that the packet forwarding decision must be based on predetermined routing algorithm, which makes the system somewhat inflexible. Loss recovery is also a big problem with push-based system. There is no request for the data, and it is not possible to re-request the missing data. Third problem is the amount of duplicate data because peers just forward the data based on the predetermined routing algorithm. All above mentioned problems are solved in a pull-based system where a peer wishing to receive a packet from another peer must request it prior to receiving. An obvious weakness with the pull-based system is the issue of dealing with free-riders.

### 4.1 Media Delivery Network Layouts

Figure 2 presents tree layout for a peer-to-peer streaming network. Each node in the network, except the original data source, is connected to a parent node and several child nodes. In this kind of network a single point of failure type of problem exists. If any of the parent nodes happen to fail the whole network originating from the failed parent node fails to receive the stream.

![Figure 2: Tree layout](image)

Figure 3 illustrates mesh layout for a peer-to-peer streaming network where each node is connected to several other nodes. In a mesh network a single point of failure type of problem does not exist except with the original data source. When a peer leaves the network it only affects those peers that are connected to it.
4.2 Media Delivery

Figure 4 presents a single stream approach with the actual media delivery. There are two major problems with this approach. Firstly, if some of the nodes happen to fail the whole network originating from the failed node fails to receive the stream. Second problem is that clients with a low throughput access network connection cannot receive the stream if the used stream bit rate is too high.

Figure 5 presents a multiple stream approach with the actual media delivery. A single stream can be divided into several sub-streams by using Multiple Description Coding (MDC) [4] or Advanced Video Coding (AVC) [5], and each of the sub-streams are then forwarded separately in the network. In this approach a peer receiving all of the sub-streams can reconstruct the full quality original stream. If a peer sending out one sub-stream leaves the network the overall stream quality does not collapse remarkably because satisfactory quality can be achieved from the sub-streams that still exist. Multiple stream approach allows low throughput access network connection clients to play the stream by receiving only some of the sub-streams, and the overall quality of the stream increases because more peers are joined to the network.
4.3 Experimental Tests

In this section a brief overview of experimental tests presented in [6] is given. In the tests a PC running Windows XP SP2 (Intel Core2Duo 6300, 2GB DDR2) was used to test four peer-to-peer media streaming applications with EDGE, UMTS, HSDPA, ADSL and LAN connections. More information about connection setups and tested applications, Octoshape [7], SopCast [8], TVAnts [9] and TVU networks TVUPlayer [10] can be obtained from the paper.

Table 1 shows measured average throughput and Round Trip Time (RTT) values for each of the selected access network technologies. The downlink throughput values (DL) are measured by downloading a 10 MB file from ftp://ftp.funet.fi/dev/ and using the Wireshark network protocol analyzer [11] to calculate the value. The uplink throughput values (UL) are measured by uploading a 10 MB file to a server and using again Wireshark to calculate the value. RTT values are measured to FICIX (http://www.ficix.fi/) to get the access network delay. These values and other available information, such as the quality of the channel and the number of peers in the channel, were used to select suitable channels for testing purposes.

<table>
<thead>
<tr>
<th>Connection</th>
<th>Throughput (DL / UL)</th>
<th>RTT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDGE</td>
<td>154 kbps / 77 kbps</td>
<td>623 ms</td>
</tr>
<tr>
<td>UMTS</td>
<td>351 kbps / 123 kbps</td>
<td>135 ms</td>
</tr>
<tr>
<td>HSDPA</td>
<td>888 kbps / 347 kbps</td>
<td>88 ms</td>
</tr>
<tr>
<td>Leased line</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>1 Mbps / 512 kbps</td>
<td>48 ms</td>
</tr>
<tr>
<td>LAN</td>
<td>100 Mbps / 100 Mbps</td>
<td>5 ms</td>
</tr>
</tbody>
</table>

Table 1: Network characteristics

Table 2 presents measured times from pressing the play button to a good user experience (bootstrap times), for different stream qualities with different network connections. ¹ in the table means that sometimes the stream works with this connection and sometimes does not and it is not recommended to use this connection with this stream. ² in the table means that the connection is not suitable to be used with this stream.
During the tests it was also measured how long time the stream continued to play without noticeable interrupts (stream buffer size) when the application was disconnected from the access network. Average stream buffer sizes for tested applications were: ~45s for TVU networks, ~45s for TVAnts, ~47s for SopCast, and ~32s for Octoshape. From these values it can be concluded that after the initial buffering is performed, delay and jitter in the access network do not have much effect to the user experience because these big buffer sizes will smooth the variation between packet arrival times.

One interesting aspect is how the current peer-to-peer media streaming applications can handle vertical handovers between different technology domains. Octoshape and SopCast were able to recover from the connection change, e.g., from UMTS to ADSL, quite easily and with the shortest interrupt but they required the new connection to be available almost instantly when the originating connection was disconnected. TVAnts was also few times able to recover from the connection change but usually with noticeable interrupts.

Table 3 presents the Quality of Experience according to the authors of [5]. **Satisfactory** means that the stream was watchable but from time to time there were some interrupts or buffering that lowered the user experience. **Good** means that the stream played as expected without errors but the experience was lowered because of the bootstrap time. **Very Good** means that the bootstrap time was also low in addition to the good signal quality.

<table>
<thead>
<tr>
<th>Stream bitrate</th>
<th>Octoshape</th>
<th>SopCast</th>
<th>TVAnts</th>
<th>TVU networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Very Good</td>
<td>Satisfactory</td>
<td>Satisfactory</td>
<td>Good</td>
</tr>
<tr>
<td>Medium</td>
<td>Very Good</td>
<td>Satisfactory</td>
<td>Satisfactory</td>
<td>Good</td>
</tr>
<tr>
<td>High</td>
<td>Very Good</td>
<td>Satisfactory</td>
<td>Satisfactory</td>
<td>Good</td>
</tr>
</tbody>
</table>

**5. Conclusion**

This paper presents some interesting topics currently studied and developed in the peer-to-peer communication field. BitTorrent has proven its strength for deploying large scale peer-to-peer file sharing network, but there are still scalability issues with the original approach. Clustered BitTorrent concept presented in this paper will improve the file downloading performance by grouping peers into clusters according to their proximity in the underlying overlay network.

At the moment peer-to-peer media streaming is securing its position among users and lots of effort has also been put to research and development. It would be very interesting to see how quickly peer-to-
peer media streaming will challenge and maybe also beat traditional broadcast type of media delivery. Results from experimental tests presented in this paper also shows that some of the currently existing peer-to-peer media streaming applications are suitable to be used with different access technologies but still there are many issues to be solved before an optimized solution for all kind of devices and access network exists.

6. References

SOCIAL NETWORKING AND VIRTUAL COMMUNITY
Pasa MAHARJAN

Summary

The potential of social networking has shortened the distance between the people around the world as millions of people from different demographic location use these social networking sites these days. Despite the diverse methods supplied for connecting with the people, social networks and development of new era web 2.0 has changed the way people connect and behave on the net, providing various features for communication.

The purpose of this report is to give brief introduction about social networks and its related web technologies that helps people to connect more easily. This report also includes a brief discussion on the new Internet era web 2.0 along with some of its services and applications. Finally, the report presents some of the problems with the current social network sites and the future prospective to overcome the problem and further developments.

1. Introduction

We are fascinated by the way social networking sites take part in blurring the boundaries of people’s network and people’s online and offline lives. People spend time on online social interactions on sites such as MySpace and Windows Live Spaces where they tend to meet new people or interact with people they know. User numbers in social networking have been increasing at the dramatic rate for the several years. For example, as of June 2007, MySpace was the most visited website in the US with more than 114 million global visitors, representing a 72% increase on 2006. Facebook increased its global unique visitor numbers by 270% in the year ending June 2007[1].

The trend of social networking on the web began with natural human desire to connect with the people, reconnect with the lost friends. Then it expanded to sharing messages, music, and videos with people sharing pieces of their culture and life interests. Today, business professionals are recognizing the profitable advantages of social networking sites, and are seeking niche networks to communicate fluidly with others in their line of work and outside their industry. Some of these advantages to business networking sites are their on demand information, quicker communication than through email, an improved culture at work, and a more personable and trustworthy professional presence with pictures, bios, and information for potential partners and clients to access. Social networking sites have become the virtual gathering spots where young, old, honest, dishonest all gather to hang out, gossip and make connection with legitimate and illegitimate. They are the virtual community center of the Internet age.

In 2004, according to Tim O’Reilly & the O’Reilly Radar Team, the Web 2.0 era started on the Internet which has changed the way people behave on the net. It seems that it is impossible to talk about current social network sites without considering Web 2.0. Due to the development of Web 2.0, new tendencies have arisen on the social networking sites which facilitate a more socially connected web where everyone is able to add and edit the information space. All these sites are technically software about networking and user participation in one way or another and are characterized as user-oriented sites.

2. Social Network: A Definition

In Internet term, social networking is web base service based on a traditional social network that
allows individuals to construct a profile within a bounded system, articulate a list of other users with whom they share a connection, and view and traverse their list of connections and those made by others within the system. One can join a social networking site and meet friends of friends analogous to going to a friend’s house for a dinner party. Some social networking sites are used to meet new friends or establish romantic relationships while others can be used to network among business peers. The basic functions that social networking sites usually offer: network of friends listings (showing a person’s “inner circle”), person surfing, private messaging, discussion forums or communities, events management, blogging, commenting (sometimes as endorsements on people’s profiles), and media uploading. With such features, social network demonstrates how the Internet continues to better connect people for various social and professional purposes.

As we look back the history, the Internet was a medium for connecting people. Email, mailing list Usenet, and bulletin boards allowed people to connect and form online social networks. However, the first recognizable social network site (SixDegree.com) was launched in 1997, which allowed user to create profiles, list their friends [2]. From 1997 to 2001, different kind of new features and a number of community tools began supporting various combinations of profiles and publicly articulated Friends [2]. People using sites like Friendster.com and Facebook.com tend to focus on sending messages, invites, and bagging to stay in touch. Another popular social network (MySpace.com) began to promote independent music and socialite scene across the globe. People have also started using MySpace for business networking, posting resume information, and uploading work samples to showcase their portfolios.

The social network sites are unique not only because they allow individuals to meet strangers, but also they enable users to articulate and make visible their social networks. This can result in connections between individuals that would not otherwise be made, but that is often not the goal, and these meetings are frequently between "latent ties" [3] who share some offline connection. On many of the large social networking sites, participants are not necessarily "networking" or looking to meet new people; instead, they are primarily communicating with people who are already a part of their extended social network. Beside these, people have joined virtual world site ‘Second Life’. This site take social networking one step with the creation of 3D virtual communities populated with 'avatars' and other real world characteristics displayed on line.

2.1 A Second Life

A second life is a three dimensional virtual community created entirely by its membership, which was developed and launched in 2003 by Linden Research Inc [4]. Second life has evolved from social networks, 3D online gaming and simulations, which provides advance level of social network service combined with general aspect of a metaverse. Members are called as 'Residents' and they assume an identity to create a motional avatars or personage to represent them which is highly customized. The residents can explore, meet other residents, socialize, and even participate in any activity happening around through second life basic features including sound, instant messaging, built in chats.

Second life exist with the residents creativity and consider as one of the distinguish characteristics. Residents create most of the content of the world and resident avatar is one example of such user generated content. With the 3D modeling tools available in second life, enables residents to create different objects such as home, landscape, vehicles, furniture and the machines which can also be sold. Residents can buy property, start businesses, create objects, join club, or just hang out. Property in second life is purchased with currency 'Linden dollar' which can also purchased with the real money through credit cards [4].

MAHARJAN, P: Social Networking and Virtual Community
Second Life residents also constantly buy and sell parcels of land where they can store their creations. With nothing to limit them but their own imaginations, players have even built entire towns on their land, filling them with houses, sports arenas, churches, nightclubs, and more [4]. Powerful scripting features enable them to trigger audio or even video when another resident wanders by. All of this serves to grease the wheels of social interaction within the game, encouraging players to work together as they build the digital world to even greater heights [4].

2.2 Threats

Social network users are often not aware of the size or nature of the audience accessing their profile data and the sense of intimacy created by being among digital ‘friends’ often leads to disclosures which are not appropriate to a public forum. As in [5], following are some of the most important threats to users of social network sites.

- Digital dossier aggregation: profiles on online social networking sites can be downloaded and stored by third parties, creating a digital dossier of personal data
- Face recognition: user-provided digital images are a very popular part of profiles on social networking sites. The photograph is, in effect, a binary identifier for the user, enabling linking across profiles, e.g. a fully identified Bebo profile and a pseudo-anonymous dating profile.
- Content-based Image Retrieval (CBIR) is an emerging technology which can match features, such as identifying aspects of a room (e.g. a painting) in very large databases, increasing the possibilities for locating users.
- Linkability from image metadata: many social networking sites now allow users to tag images with metadata, such as links to SNS profiles (even if they are not the owner/controller of that profile), or even e-mail addresses. This leads to greater possibilities for unwanted linkage to personal data.
- Difficulty of complete account deletion: users wishing to delete accounts from social networking sites find that it is almost impossible to remove secondary information linked to their profile such as public comments on other profiles.

3. Web 2.0

Web 2.0 is said to be the set of services that allow web user to share their opinion and resources easily. It is group of technologies which have become deeply associated with the term: blogs, wikis, podcasts, social networking, RSS feeds etc, which facilitates more socially connected web where every one is able to add to and edit the information space. In addition, web 2.0, is more that a set of cool technology and services, that also define the powerful ideas that has changed how we all think and use the web.

According to Dale Dougherty and Tim O'Reilly, who officially coined the term “Web 2.0” in 2004 during a team discussion on a potential future conference about web [6], “Web 2.0 is the business revolution in the computer industry caused is the move to the Internet as a platform and attempt to understand the rule for success on that new platform”. Tim O’Reilly also explained the Web2.0 era started on the Internet with the following key principles of Web 2.0 applications [6], [5].
### Key Ideas

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The web as platform</td>
</tr>
<tr>
<td>2</td>
<td>Harnessing collective intelligence</td>
</tr>
<tr>
<td>3</td>
<td>Blogging and wisdom of crowds</td>
</tr>
<tr>
<td>4</td>
<td>Data is Next Intel Inside</td>
</tr>
<tr>
<td>5</td>
<td>End of software release cycle</td>
</tr>
<tr>
<td>6</td>
<td>Lightweight programming models</td>
</tr>
</tbody>
</table>

### 3.1 Web 2.0 Service/Applications

#### 3.1.1 Blogs

Blogs or web-log refers to a simple webpage consisting of brief paragraphs of opinion, information, personal dairy entries or links, called posts, arranged chronologically with the most recent first, in the style of an online journal [6], [7]. The blogs enable users to leave comments in an interactive format and provides commentary or news on particular subjects. A typical blogs combines text, image or links to the other blogs, web pages. Some of the well known blogs are: [http://radar.oreilly.com/](http://radar.oreilly.com/), [http://www.techcrunch.com/](http://www.techcrunch.com/).

#### 3.1.2 Wikis

A Wiki is the webpage or set of web pages that can be easily edited by user who is allowed to access. In other words, it is a collaborative tool that facilitates the production of the work group and to power community website. Wikipedia is one of the best known wikis these days. Wiki supports hyperlinks and has a simple text syntax for creating new pages and crosslink between internal pages on the fly.

#### 3.1.3 Mash up

Mash up represents a powerful service that allows user to combine information from different sources into single integrated tool. A mash up, for example, could overlay traffic data from one source on the Internet over maps from Yahoo, Microsoft, Google or any content provider. The term mash up comes from the hip-hop music practice of mixing two or more songs.

#### 3.1.4 Multimedia sharing

Multimedia sharing is one of the popular services these days that facilitates the storage and sharing of multimedia content. Some of the well known example includes 'Youtube'(video), flicker (photos) and Odeo (Podcasts).

#### 3.1.5 RSS FEEDS

Really simple syndication (RSS) represents a way to receive information based on granular subscription. It is a family of formats which allows user to find out about updates to the content of RSS enabled websites, blogs or podcasts without actually having to go and visit the site. Instead, the
information from the website is collected with in a feed and piped to the user in a process known as syndication.

4. Future of social networking

Today, social networking sites are used by millions of people all over the world. The potential of social network to connect people from different demographic location has more shortened the distance between people. Current social networking is about uploading photos, inviting friends, and posting to user groups, where the user has to work with in the framework and the network monetized the traffic of the aggregated users. Social networking sites such as MySpace and Facebook has taken step further and introduced customization of pages and enabled user to monetize their traffic.

Despite the cool new features get added daily on the network such as feed aggregation, IM, data export, mobile integration, yet the lack common objectives. For example, some user connects to social networking sites just to boost the number friends they have in the profile. Many more browse other profile just for the curiosity's sake. These explicitly established connections become increasingly meaningless because they aren’t backed up by common objects or activities [8].

Another major problem is the strong competition between the social networking sites which is not compatible with each other. Most social networking sites are being developed as standalone systems forcing the user to start form the scratch every time they join a new site. Thus, user has to reenter their profile and redefine their connecting form the beginning. Actively participating in multiple sites consumes more free time than the average has available for such activities.

With reference to IEEE article 'The future of social network on the Internet: The need for Semantics', one way to provide meaning to current social network sites is to develop object-centered sociality at its core through the user's action around the content they create, link to, or for which they use similar annotations[8]. For instance, if the object is a job, it will connect to one set of people whereas a date will link to a radically different group.

![Object-centered social network.](image)

Figure 1. Object-centered social network. [8].

The above Figure1 illustrates the object -centered social network for three users. Bob and Carol are
connected via booked marked Web sites that both have annotated. Alice and Bob have matching tags on media items, and they subscribe to the same blogs. As online connections between people become increasingly intertwined with real-world interests, social networking methods are moving towards simulating real-life social interactions rather than randomly approaching each other, people meet through things they have in common[8].

![Diagram showing the social networking stack](image)

Figure 2. Making social networking a shared component across various desktop and Web applications [8].

The article also purposed a social networking stack which would allow the reuse of one's personal profile, social network connection and the content creation history across various sites and applications [8]. Thus, it requires taking account into people's relevant objects of interest and different layers to perform data portability:

1. The personal identification and authorization layer that use open ID to connect their social networks
2. The social network access layer that would utilize an individual social networking contacts across various platform
3. The content object access layer that would collect the user's relevant content objects.

5. Conclusion

We see social networking sites as phenomena that are here to stay as we consider social networking sites to be a meeting place on the Internet where people can meet and socialize in order to comply with their social need. There are an enormous demand for on line social networks which is backed up by large numbers of users to such sites as YouTube, MySpace and most successful social networks tend to have a specific purpose, for example ‘deLicio.us’ for bookmarking, ‘Flickr’ for photo sharing and ‘YouTube’ for video sharing. With such demand, and development of new era Web 2.0 has enhanced the existing functions and added cooler feature for people to get connected more easily and efficiently.

Despite the fact that web 2.0 have changed the way people behave on the net, current social networking sites still lack common objectives, as current social networks have become a competition of who has the most contacts, resulting in users adding individuals who they will never interact with. Lack of interoperability between the sites is another issue forcing the user to start form the scratch every time they join a new site. Actively participating in multiple sites consumes more free time than the average has available for such activities which might be one good reason for social networking sites to lose its popularity in future.

Object centered sociality approach appears to me a major forward in the creation and management of contact lists. Having different categories for jobs, friends, family would enable a user to cross reference the group, looking for individuals with multiple shared objects. The concept of social networking stack would allow the reuse of user profile data creating a universal open ID. One of the significant advantages of these approaches is, it is build on pre-existing standards and achieve good
balance between centralization and decentralization of contents.

Reference


PART 4

DELAY TOLERANCE VERSUS SEAMLESSNESS
INTERMITTENT CONNECTIVITY AND DELAY TOLERANCE

Vincent BATTAGLIA

Summary

People are becoming every day more mobile. At the same time, in order to follow them, network applications must communicate across difficult environments. We observe that the current Internet model is not adapted for those environments, because it’s based on assumptions like end-to-end connectivity or low-delay path. An answer to intermittent environment is Delay-Tolerant Networking (DTN). Its aim is to create systems “nomadically enabled” which allow dealing with sudden changes in conditions in a natural and transparent fashion. Those new issues imply new mechanisms in networking and routing that we are going to discuss in this document.

1. Introduction

The exponential increase in the number of mobile terminals changes the behavior of people concerning technologies. In a transparent way, you’d like to access every classical and new service of our terminal [3]. However in wireless transmissions, environment has different characteristics than in classical wire transmissions and because of those, most of the difficult/extreme environment networking didn’t use the traditional TCP/IP layers. One of the aims of DTN is to permit those different networks to communicate in order to create a multi-network service [1].

In order to understand the problematic, here are the characteristics of DTN environments:
- Intermittent connectivity
- Long and variable Delay
- Asymmetric data rates
- High error rates

What are the types of wireless networks with DTN characteristics? We can point out for example terrestrial civilian networks which interconnect mobile devices such as cell phone network, personal communicator, emergency communications and in the future intelligent highways [4]. On other field is wireless military communication, in order to interconnect troops, aircrafts, boats, satellite in battlefields. Logically, all kind of sensors networks [5] can be interested in DTN’s projects. Less obvious, but with great support, outer-space networks, such as InterPalNetary (IPN) project, supported by NASA [2] are future applications of DTNs.

This document is going to explain why the current Internet model is not adapted to those new challenges and why DTN are necessary. After this, it’ll describe the characteristics of DTN new protocol, the Bundle protocol.

2. Failure of the current Internet model

The current Internet model has been created for wire connection. The initial statement has settled assumptions about the network environment such as:
- Continuous, Bidirectional End-to-end path
- Short Round-Trips
In addition it is around those initial assumptions that the TCP/IP model has been developed. Important TCP/IP characteristics are:

- It requires a negotiation to control the flow of Data
- Transmission order necessary
- End-to-end connectivity required

However the problem in this system is its inefficiency in order to react within a difficult environment as described in the introduction. Disconnection, low latency and other, are treated as failures. As I explain in the introduction, this is the reason why TCP/IP model is not adapted to DTN environment [2]. However TCP/IP is the base of today’s Internet, and it is not possible to stop using it for obvious reasons. On the other hand, an ad-hoc patch to the current system which is rigid in their view of connect-endless will not be efficient. The solution proposed by DTN architecture is overlaying actual network to permit connectivity, even in difficult environment, as it’ll be describe.

3. Delay-Tolerant Networking architecture

3.1 Concept of DTN

The purpose of the DTN architecture is to provide flexibility in order to adapt to the environment. In this idea, it must be based on [2]:

- No explicit necessary dialog
- No timely connection
- No problems of transitions order

![Figure 1 - Bundle Layer](image-url)
The DTN architecture implements store-and-forward message switching by overlaying a new protocol, called Bundle layer, on the top of heterogeneous region-specific lower layers. The Bundle layer ties together the region-specific layers so that application program can communicate across multiple regions.

### 3.2 DTN nodes

The DTN architecture identifies different types of nodes in the DTN network. Those nodes are related to DTN regions. A DTN region is a common technological region. So, what distinguishes two different DTN regions is that layer 1 to 4 is different. Based on this fact, the DTN’s nodes are:

- Hosts, which send or receive bundle but don’t forward them. Hosts are final customers of the DTN network.
- Routers, which forward bundles within a single DTN region. The best example is actual routers of the Internet network.
- Gateway, which are really the key in DTN networks. They have the ability to forward bundles between two or more DTN regions. As it’ll be explain after, they have also a critic rule in the store-and-forward system and in the custody transfer system.

So it means that all routers have to be change? Not obligatory, we can imagine that on classic Internet network, DTN routers are not mandatory. The capacity to evolve of the router depends of the flexibility of the router software. For gateways, specific equipment should be developed.

### 3.3 Custody transfers and Delay isolation

Main mechanism of the Bundle protocol is based on custody transfer and delay isolation by storage. As you understand, some DTN regions can be distinguished by disconnection and poor conditions which didn’t permit any end-to-end connectivity. In order to face off those problems, DTN architecture uses several mechanisms [2].

![Figure 2 - DTN nodes](image-url)
The custody transfer permits to a router or a gateway to give acknowledge of the transmission to the previous nodes which send the actual packet. This technique permits to use the layer stack of the current DTN region, which often requires end-to-end connectivity. In the case of a gateway node, custody transfer is always in use, because often assumptions of reach capability cannot be made on the next DTN region [7]. Different types of custody transfer can be use that will be detailed in “Classes of Bundle Service” part.

The custody transfer is associated to store-and-forward technique. Packets are stored until acknowledgement of custody transfer arrives. This persistent storage permits eventually to resend the packet after in case there is a problem during the transmission.

This architecture, using the Bundle layer, allows reusing actual layers of a specific region, adding an end-to-end availability to be reached.

The dual use of those mechanisms is the key of DTN architecture to face disconnection is DTN regions. It also interesting to notice that DTN architecture doesn’t provide any kind of new connectivity is case of disconnection. So associated to the “network layer” provide by DTN techniques, applications itself should be ready to wait a packet for a long time without consider it a failure, because of a specific environment [1].

Figure 3 - Custody transfer
3.4 Naming and addressing

As an overlaying protocol, Bundle protocol has to use a specific addressing system. In a DTN architecture, each bundle is tagged with (<Region ID>, <Entity ID>) of both sender and destination hosts. <Region ID> characterize the DTN region of the node and <Entity ID> the specific ID within this region. An interesting point to note is that <Entity ID> is not necessary an IP because layer stack is not necessary TCP/IP in DTN.

3.5 Classes of Bundles services

According to the flexibility which is a base of DTN philosophy, DTN provide different types of bundles services to be adapted to different traffic or environments. According to the figure 4, you can find several possibilities [7]:

- Custody transfer : Custodial-acceptance acknowledgement is sent to the previous custodian
- Return receipt : confirmation to the source
- Custody-transfer notification: notification to the source when a node accept a custody transfer to the bundle
- Bundle forwarding notification : notification to the source whenever the bundle is transfer to an other node
- Priority of Delivery: Bulk, normal or Expedited
- Authentification: digital signature or other

![Figure 4 - Custody Services](image)

3.6 Routing in DTN environments

Routing protocols is a big issue in DTN’s environment. Main problem is dynamicity of the systems. The Bundle layer just provide routing between DTN region, all routing issues in region are provided by layers specific to each region, like OSPF/BGP for IP’s region.
Classic routing protocols (BGP /OSPF…) aren’t adapted to DTN environment especially because of the intermittent characteristic of this environment [2]. Main problems are:

- Most of them have timeouts
- No possibility to divide the structure
- No possibility to introduce predicted intermittent connectivity

A possible solution can be provided by new generation protocols based on tiered routing. Interesting points on tiered routing are firstly that it permits scheduled connectivity and secondly that it may be predictable based on region-specific structural awareness.

The second point is the most interesting one. It’s almost impossible to provide routing in a complete random intermittent network, but studies proof that most of the difficult environments can be predictable. It is why tiered routing should be customizable for specific environment.

3.7 Implementing DTN

Several groups of searchers are currently working on implementing DTN architectures. Even though works are still running, several conclusions on difficulties of implementing DTN networks have been discovered [6].

The first conclusion is that in challenging environments, a store-and-forward message overlay performs significantly better than traditional approach which is very important to notice.

The second conclusion is that buffer limits can be a problem in DTN environments. Basically, DTN gateway stores packets from DTN region. It’s easy to understand that if traffic is very important, the amount of data to store will be very important too.

Searchers also discovered that effects of intermittency are rarely transparent for all layers of the system which should be theoretically the case.

Finally, everything points to say that DTN architecture is the not easy to implement.

Recent attempts to simulate a DTN architecture in Helsinki under Jörg Ott direction [3] pull up that a lot of problems appear, especially in full load situations.

4. Conclusion

Delay-Tolerant Networking can be a solution for difficult environments. As we have seen, the current Internet model fails to answer to the numerous challenges of those environments. Also, we can bet that those challenges will become more and more important in the future. DTN architecture seems to perform significantly better than traditional approach in extreme environment. Nevertheless, it also seems that it may be difficult to implement. The key factor will probably be new multi-network services, which will permit to bring more and more interest in this type of architecture and also military applications, which will surely finance those research.

5. References

CONNECTIVITY IN CHALLENGED ENVIRONMENTS

Fausto GROSSI

Summary

Up to the present the concept of “network” rang around the assumption of contemporaneous end-to-end connectivity. An environment is called “challenged” if there is lack of connectivity. Solutions obtained in the military field and in the creation of underwater networks, are attracting increasing interest from researchers in challenged environments networks creation. In this paper several key aspects of military networks and underwater communications are investigated. The proposed scheme aims to give an exhaustive explanation about such existing networks and an overview of future solutions.

1. Introduction

In a world in which mobile devices are increasing and in which “Internet Anywhere” is the best wish of every user, end-to-end connectivity assumption gives way to “island of connectivity” concept. The convergence of mobile telephony and handled computing led to a finally ubiquitous developing of mobile networks. It is well known that mobile devices have one or more wireless interfaces, used to exploit wireless connectivity via access point. However, the applications commonly deployed on such devices (e.g. email, web browsing), are rarely able to fully exploit this local wireless connectivity, then a large amount of wireless bandwidth capacity remains unused, because Internet, has not been designed to take advantage of local and intermittent connectivity. During the years, different models were studied to exploit this unused wireless bandwidth capacity when connectivity is intermittent, periodic, prone to disruptions but all are based on Disruption Tolerant Network (DTN) concept. [1]

The Internet architecture and its great success during the past years was based on classic assumptions such as the use of TCP over IP (Internet Protocol) and then relatively short transmission delays, low error probability and the existence of end-to-end paths. In an environment in which there is lack of connectivity, frequent connectivity disruptions, extremely long transmission delays, the classic assumptions fail and then the Internet architecture based on TCP/IP is no more applicable. In a challenged environment there is the necessity of a new kind of network to reach connectivity anytime and anywhere; this leads to the creation of Disruption (or Delay) Tolerant Networks.

DTN is a new area of research in the field of networking that deals with extending existing protocols or inventing new ones in a coordinated, architecturally clean fashion, to improve network communication in a challenged environment. Using this type of network, the “always on” paradigm is relaxed because it would be extremely expensive or even hard to realize in challenged environments. There are two types of approach: one studied by the DTN Research Group [2]; the other aims to improve the operation of existing, interactive Internet applications and protocols in scenarios where network connectivity is intermittent. The latter approach (which is the most used today) extends the functionality of Internet by using another layer in the protocol stack and it is the Bundle Layer [3]. Thanks to this layer there is the possibility to save the message that the user would transmit, whereas there is no connectivity. Using the bundle layer, which is between the application layer and the transport layer, the DTN router keeps a copy of every packet sent until acknowledgment message comes back. It happens only when connectivity is retrieved and then the message could be sent and will be received without it was known by the receiver. This is the great power of DTN networks. However there is question about who cares for the message when it is sent by the transmitter and there is no connectivity. The answer stays in the basic concept of DTNs, which is the use of both fixed and
mobile network, keeping in mind that every device inside the DTN could be a DTN router and then could be the caretaker of the sent message. The TCP/IP model slumps, in fact every node inside the network now could be a router and the classic concept of routing is no more applicable.

With regards to this, it could be simpler to understand thinking about an user, without Internet access, that would send a message but he has the possibility to connect with another user using a wireless interface. He could exploit the connectivity of the other user sending him his message; now the latter acts as a DTN router either keeping the message in the bundle layer until (if he has not connection to Internet) or sending the message to another router (if he has Internet connectivity).

The first one to discover the power of the DTN concept was the U.S. army to provide connectivity inside a battlefield to coordinate attack and defence operations. Great results were obtained in the field of underwater sensor networks and great results will be obtained as well in the interplanetary Internet and in the everyday life thanks to the creation of communities in which all users share their connectivity with each other to reach “connectivity anywhere”.

The remainder of this paper is organized as follows. In Sections 2 there is an explanation about underwater acoustic sensor networks, which are the most developed in the field of networks in challenged environments. In Section 3 the most challenging topic of Interplanetary Internet. In section 4 an overview of the first field of application of DTNs that is MANET, a particular kind of network used to provide connectivity within a battlefield. Section 5 is the final question in which there is a little about a future transfer of technology from military field to everyday life, the Pocket Switched Network.

2. Underwater Acoustic Sensor Networks

In this section it is described what is an underwater sensor network, why it is classified as a network in challenged environment and what is the current architecture used to build such a network.

Oceanographic data collection, pollution monitoring, offshore exploration, disaster prevention, assisted navigation, tactical surveillance applications, undersea explorations; all this applications need the creation of underwater sensor networks [4]. The traditional approach for ocean-bottom or ocean-column monitoring was to deploy underwater sensors that record data during the monitoring mission, and then recover the instruments. However, in this way, there are not the best performances because there is not real time monitoring, possibility of failure detection, on-line system reconfiguration and the limited storage capacity is a problem as well [4]. All these constraints led to prefer the creation of underwater acoustic sensor networks that will enable real-time monitoring of selected ocean areas, remote configuration and interaction with onshore human operators. This can be obtained by connecting underwater instruments by means of wireless links based on acoustic communication [5].

Underwater networks can be characterized by their spatial coverage and by the density of nodes. It talks about single-hop network when all nodes are in direct contact. In networks covering larger areas, communications will require multiple hops to reach destinations and then they are called multi-hop networks. When the geographic coverage is greater than the coverage of all nodes, routing requires techniques from disruption-tolerant networking (DTN).

2.1 The physical channel

Many researchers are currently engaged in developing networking solutions for terrestrial wireless ad hoc and sensor networks. Although there exist many recently developed network protocols for wireless sensor networks, the unique characteristics of the underwater acoustic communication channel, such as
limited bandwidth capacity and variable delays [6], require very efficient and reliable new data communication protocols. In fact underwater networks are classified as in challenged environment, because the physical channel is no more the ether but the water. Light and IrDA are strongly scattered and absorbed underwater; moreover long-wave radio can be used only for short distances (1-8 kbit/s up to 10m). Finally, underwater it should be used short-wave radio in the range 1-50 kHz and then the sensor nodes cannot be light or infrared sensor, but acoustic sensors. This forced choice lead to several problem and constraints. It is well known that the speed of sound is much lower than the speed of light; speed of sound is about 1500 m/s instead of about 300.000.000 m/s of light and then it must be considered large propagation delay. Underwater obstacles and mammals cries (they interferes with the acoustic transmission) cause phase and amplitude fluctuations and then high BER. Moreover, as shown in Figure 1 [5], signal amplitude attenuation is directly proportional with increasing frequency leading a very limited bandwidth [5].

Unmanned or autonomous underwater vehicles (UUVs, AUVs), equipped with sensors, will enable the exploration of natural undersea resources and gathering of scientific data in collaborative monitoring missions. Underwater acoustic networking is the enabling technology for these applications. Under-Water Acoustic Sensor Networks (UW-ASNs) consist of a variable number of sensors and vehicles that are deployed to perform collaborative monitoring tasks over a given area [7].

The employment of mobile sensor leads to another problem. The AUVs are equipped with engine to move within the area of interest; the propulsion noise interferes with the acoustic transmission by which the surface base station tries to send message towards the fixed or mobile sensors. If the bit rate is huge there is a great Bit Error Rate (BER) because the receiver is no more able to understand the messages arriving from the surface. The solution is inside much sophisticated modulation techniques that allow to decrease very much the Bit Error Rate [5]. In transmission there is the necessity of a modulation which is very efficient in bandwidth, because they should be sent the monitoring data from the sensor to the surface, and it could be the PSK (Phase Shift Keying) modulation. It is not efficient in power because it needs large amplitude signals to be sent, but it is not a problem in transmission; it is a problem in reception because in reception the propulsion noise interferes with the receiving signal. For this reason it would have a modulation which was very efficient in power and it is the FH-FSK (Frequency Hopping-Frequency Shift Keying) modulation. It does not matter if it is not efficient in bandwidth because in reception there is not the necessity of a large bandwidth (in reception the sensor receives only control signals which do not need high bit rate). In table 1 [5] is summarized the evolution of modulation in underwater networks during the past years.
Table 1 – Evolution of modulation techniques

2.2 Three-dimensional underwater sensor networks

Three dimensional underwater networks are used to detect and observe phenomena that cannot be adequately observed by means of ocean bottom sensor nodes. In three-dimensional underwater networks, sensor nodes float at different depths in order to observe a given phenomenon. One possible solution would be to attach each uw-sensor node to a surface buoy, by means of wires whose length can be regulated so as to adjust the depth of each sensor node [8]. However, although this solution allows easy and quick deployment of the sensor network, multiple floating buoys may obstruct ships navigating on the surface, or they can be easily detected and deactivated by enemies in military settings. For these reasons, a different approach can be to anchor sensor devices to the bottom of the ocean. In this architecture, depicted in figure 2 [5], each sensor is anchored to the ocean bottom and equipped with a floating buoy that can be inflated by a pump.

Fig.2 Architecture of 3-D under water acoustic sensor network

2.2.1 DTN in three-dimensional underwater sensor networks

As already said in the previous chapter, usually mobile sensor is more used than fixed sensor because the latter is very expensive, due to the small quantities in commerce, and it covers a smaller area. Then the scenario is a certain number of mobile devices (AUVs) moving around within the network; of course the concept of DTN comes back. An interesting document explains very well the network operations in this scenario [9].

GROSSI, F: Connectivity In Challenged Environments

144
2.3 Future extensions

It is possible to work on devices in the future, in order to improve the quality of devices to reach minimum BER and then the best performance. The good results obtained in DTN field, on the other hand, allowed to reach all the modern targets and then to cover almost all the underwater applications.

3. Interplanetary Internet

In the last years the most important space agencies tried to use an Internet-like network to perform all the applications that they want in the space. These applications were previously performed used specific links and modulations; they could be either to send data in the space or to provide connectivity between planets. Using the typical protocols used in Internet it could be performed the ground satellite control and, why not, a homesick spaceman could have a chat session with his wife from Earth. The question is if it is possible with modern technology and actual knowledge and all agencies (first of all NASA that made the best efforts in this field) agree saying that it is possible. However there are two schools of thought:

The first argue that there is no other solutions but only the creation of new protocols, specific for interplanetary Internet, but it is not taken in consideration because it would take too much time and money. The second one states that it is enough extension of basic networking protocols that run the Internet to make it possible.

The latter school of thought is embodied by a group of researcher, most of them working at Goddard Space Flight Center in Maryland, USA. Thanks to this team, for the first time in history, in 2003, a file made its way back and forth from earth to Columbia Shuttle. It is possible to discover how they did the successful experiment of CANDOS in some paper; one of those is IEEE property [10].

3.1 Architecture

As mentioned earlier, the solution that now provides significant results is the extension of existing earthbound software and network architecture; however it is important to say that the starting point was the TCP [13]. It is clear because the IP is the so called “Lingua Franca”; it means that two machines built in different times and in different places, could communicate with each other only “speaking IP”. The Transmission Control Protocol provides delivery control not provided by the Internet Protocol, and then it was used as transport layer protocol, but with some extension. In fact there is the necessity to include an interfaces to another upper layer that is the bundle layer and it will be said something about it later. Moreover the RTT (Round Trip Time) should be increased to avoid connection expiration because in interplanetary transmission the distances are very large. However it is not enough to extend TCP to create an interplanetary Internet, because the greatest issue is that, in this scenario, routers are set on spacecraft, satellites, space stations. They are not fixed because they orbit the earth and then they have not always the line of sight with the earth-ground base stations.

As said before, every kind of space device, like satellites, spacecraft and space stations can be nodes and, due to their mobility, they join or leave dynamically the network. The solution is inside the development of a DTN (Delay Tolerant Network) architecture. Thanks to the introduction of the bundle layer, the DTN router keeps a copy of every packet sent until acknowledgment message comes back. As could be seen in the next figure, this schema ensures that no data gets lost.
3.2 Future extensions

Of course it is only the beginning of the developing of this kind of network, but it is sure that great enhancements were made in this field thanks, above all, to NASA and to its cooperation with other space agencies like the ESA, the European Space Agency. The great current result is sending a file from a ground station to a space station and wait for the answer, as shown in Figure 4 [10]; then it suggests that in future papers there could be only greater results of future experiments or, maybe, of interplanetary Internet real implementation.

Fig. 4 The current NASA ground network and the future interplanetary Internet

4. Connectivity in battlefields

The military field is the first field in which researchers developed DTN concepts to provide connectivity anyway and anywhere. It is important to spend a few words about this topic to understand at the best how it leads to a potential transfer of technology in the daily life. It is since the 70’s that there are great efforts to create a mobile network able to work inside a battlefield. U.S. Army was the first to understand the great potential of the DTN architecture to provide connectivity between the...
devices inside a battlefield during attack and defence missions. All the research led up to the creation of MANET (Mobile Ad-hoc NETwork).

4.1 MANET – General aspects and architecture

MANET is a network established for a special, often extemporaneous service customized to applications. The ad-hoc network is typically set up for a limited period of time, in an environment that may change from application to application. For the MANET to retain its efficiency, the ad hoc protocols at various layers may need to self-tune to adjust to environment, traffic and mission changes. Indeed, an architecture that can be deployed to monitor the habits of birds in their natural habitat, and which can be structured to launch deadly attacks onto unsuspecting enemies.

MANETs are set apart from conventional wired or wireless infrastructure type networks by a number of unique attributes and requirements which could be listed as Mobility: because inside a MANET nodes may move (it is not a fixed sensor network). Multihopping: the path from source to destination traverses one or more nodes. These two aspects hide the intrinsic aspect of Self organization because the network must autonomously determine its own configuration parameters: addressing, routing, clustering, position identification, power control. Much important are Energy conservation and security because the network must be powered for the whole mission and it must detect and fold active and passive attacks. Such a network could not be constructed following a hierarchical model, but there is the necessity of an architecture able to face with scalability problems and to ad-hoc routing; all this is provided by MBN (Mobile Backbone Network [12]), using LANMAR (LANdmark Ad-hoc Routing [12]).

5. Pocket Switched Network – PSN

In this section it is described how the results obtained in the others fields were exploited in everyday life applications. Actually there is no real transfer of technology to commercial and the reasons are different; ad-hoc networks are tuned on a specific application and their lifetime is about a mission time. In everyday life a user wants Internet access for several applications and then the way was towards “commodity” ad-hoc network; moreover military networks are too costly to deploy and too difficult to sell to the public. Due to the large developing of mobile devices and their overlapping with handled computing, it has gone towards mesh networks use. Pocket Switched Networks fall under the mesh networks but with some interesting aspect in addition [13].

The key points are the presence of several interfaces on the mobile devices, by which users can communicate with each other within the network; the Opportunistic Forwarding and the Dynamic Routing. When global connectivity is available to a node, messages can be forwarded directly to suitable nodes which are also globally connected, or to available proxies for those recipients, waiting for connectivity. Also routing is very important because in such a network every user-device could be an user agent and a router as well; the result is coverage extension.

5.1 PSN – Issues and current solution

Taking as an example the nomadic access inside a Campus, the user could walk around within the Campus and he could need to pass from a technology to another to keep his connectivity; then vertical handover protocols are necessary. Another important aspect is the security because in such a network everyone could wish to share his connectivity with other users, advertising his moral integrity and instead he is a malicious which tries to accede to the unsuspecting user’s secret data. In addition “Third
Party Forwarding Incentives” is the problem to provide incentives to users to share their connectivity with other users to reach the greatest covered area. The current unique solution is the creation of a COMMUNITY. Inside a community there is a network manager that “knows” all the users and fixed members need Registration and Authentication to the network; in this way the security is well-assured. There is also a reputation mechanism: if one is a PSN user and share more and more connectivity with other users, his reputation will increase and more and more bandwidth will be given to him; this is, maybe, the best incentive to share connectivity within a network.

5. Conclusion

Summarizing, connectivity in challenged environments groups a set of area of interest. The best results were reached in the military field and in the creation of underwater sensor networks and it could be said that almost all was made in these areas. Only some successful experiment was reached, thanks to the NASA and ESA efforts, in the creation of Internet-like network in the space. The mission cost and the great developing time (extreme distance and then long waiting time) could be overtaken and a real interplanetary Internet could be created. However there is not a real transfer of technology from these fields to the daily life, because of the costs and the uncommon necessity of people to have connectivity everywhere, anyway. The creation of communities, in which the users share connectivity with each other using the Pocket Switched Network, is the first step towards the creation of ubiquitous connectivity.

6. References

SEAMLESSNESS AND VERTICAL HANOVERS

Antti LEINONEN

Summary

Seamless mobile computing is a big challenge in a near future where we have a multiple radio access networks available and user would like to be connected to best available network. Vertical handovers are needed if seamless user experience is wanted. There are many possible protocols proposed for handling vertical handovers. Session initiation protocol (SIP) and Mobile IPv6 are most promising ones. One really interesting technique for archiving seamless mobile computing is The Internet Suspend/Resume where mobile computing is attained without people carrying their devices.

1. Introduction

Currently we have multiple wireless networks available for users. All networks have different kind of properties. Wireless local area networks (WLAN) have good bandwidth but very limited coverage where as cellular networks (e.g. Universal mobile telecommunications systems (UMTS)) have limited bandwidth but very good coverage. Devices have also multiple radios and are able to connect many networks. But when user makes a handover from one radio technology to other user will loose all connections to different services that he or she had because his/her IP address changes.

Handover between different networks with different radio access technologies is called vertical handovers (e.g. handover from WLAN to UMTS). Handover decision is traditionally based on relative signal strength (RSS) but in vertical handover decision is more complex. Handover decision can be based on RSS, cost, user preference, bandwidth, etc… That decision is really hard to make because different user will have different preferences. One could value bandwidth whereas other could value cost of usage of network. [1]

Seamlessness in vertical handovers is also big question. If vertical handovers are seamless for users it can be hard to know how big a cost for using network is. On the other hand usability suffers if vertical handovers are not seamless for user (e.g. user have to always confirm handover by clicking “ok”).

This paper is divided into two parts. In sections 2, 3 and 4 three protocols that could be used with vertical handover are presented. And in section 5 new mechanics for mobile computing called Internet Suspend/Resume is presented.

2. Session initiation protocol

Session initiation protocol (SIP) is simple, text-based protocol that includes extensibility and provision for session control. The main entities in session initiation protocol are proxy servers, redirect servers and user agents. In SIP there are many methods defined to set up sessions between the user agents, e.g. INVITE, ACK, BYE, OPTIONS, CANCEL and register. [2]

Terminal mobility requires SIP to make a new connection either during the start of a new session (pre-call mobility) or in the middle of a session (mid-call mobility). In a pre-call mobility mobile host registers its new IP address with its SIP-server by sending a REGISTER message and in a mid-call mobility mobile host needs to intimate to corresponding host by send a INVITE message about mobile
host’s new IP address. Corresponding host starts sending data to new address right after it gets INVITE message. Mobile host need also to register with the redirect server in the home network for future sessions. [2]

2.1 Vertical handovers in a WLAN-UMTS internetwork

SIP is an application-layer protocol and it relies on the protocols in the lower layers to handle the physical network connection. Additional procedures are needed in SIP mid-call mobility to get the mobile host attached to the wireless access network infrastructure before the SIP re-INVITE message can be sent. (E.g. Mobile host attaches to the UMTS network using GPRS Attach and Packet Data Protocol Content Activation procedures.) Because SIP needs additional protocols and procedures vertical handover delay consist of SIP location update and delay of network attachment. [2]

Before the mobile host (MH) can send SIP INVITE message when making a handover from WLAN to UMTS network MH needs to establish the data path to carry SIP messages. This is done by making a GPRS attach and PDP context activation. Signalling required for WLAN to UMTS handover is shown in fig 1. [2]

---

**Fig.1** Signaling for WLAN-to-UMTS handover [2]
Less signaling is needed when mobile host moves from UMTS network to WLAN network. First MH makes a DHCP registration and after that MH re-establishes connection with corresponding node with SIP message establish. Exact signaling that is needed for UMTS to WLAN handover is shown in fig. 2. [2]

![Fig.2 Signaling for UMTS-to-WLAN handover](image)

Because WLAN to UMTS handover requires so much message exchange could handover delay be as long as 1.5 seconds. That is too long delay for any delay-sensitive application like video-streaming. Soft handover techniques are needed to reduce WLAN to UMTS handover. UMTS to WLAN handover is much faster (~200 ms) because it requires much less signaling and that is not a problem for delay-sensitive applications. [2]

3. Mobile IP

Mobile IP allows a mobile host to move from one place to another without changing its home address. Home agent will route all packets to mobile host regardless of its current location. There are some drawbacks Mobile IP v4: Messages sent to mobile host in visitor region are tunneled through home agent and that could add some delay. Also handovers in Mobile IP v4 are horizontal handover because the link layer remains the same and only the point of attachment changes. [3]

There are lots of improvements in Mobile IPv6 compared to v4. One goal when designing mobile IPv6 was that roaming between different L2 technologies (WLAN, GPRS, WiMAX…) is possible. Route optimization is also possible in mobile IPv6 so that messages do not need to be routed through home agent. [4]

There have also been lots of improvements proposed for optimizing handovers in mobile IPv6. E.g. Fast Handovers for Mobile IPv6 (RFC4068) and Hierarchical MIPv6 mobility management (RFC4140). [4]
4. A ProActive Handover Enhancing protocol

A Proactive handover enhancing protocol (APACHE) is a vertical mobility management protocol that is proposed for UMTS-WLAN interworking networks. APACHE uses location information from the mobile user and proactively tries to decide when handover should be made. Location information is taken from GPS unit and information is passed to network. [5]

Because WLAN networks have very limited coverage it can be really hard for fast moving mobile user to detect WLAN networks. Fast moving user could also be very deeply inside WLAN coverage area before one could detect network and therefore loose opportunity to use extra recourses (bandwidth). Location based handovers are really good solution for domain with fast moving users and networks with limited coverage. [5]

Seamless user experience is possible because there should not be any packet loss. APACHE also avoids high network load with each handover, which is usual for a forwarding based handover techniques. One of the weaknesses of APACHE is that it generates some overhead on the wireless link because of location updates. [5]

5. Internet Suspend/Resume

Seamless mobile computing can be achieved with other kind of mechanics. Internet Suspend/Resume is a new mechanic that tries to make mobile computing possible without people carrying their laptops. Internet Suspend/Resume is inspired by suspend/resume feature of laptops. In Internet Suspend/Resume same concept is expanded to situations where you could suspend your computer at one Internet site, travel to other site and resume your work with different computer. [6]

Internet Suspend/Resume combines two technologies: Virtual machine technology and distributed file system technology. There have been three prototype implementations of Internet Suspend/Resume: ISR-1, ISR-2 and ISR-3. All prototypes have had Linux as a host operation system and VMware as the virtual machine monitor. ISR-1 (proof-of-concept prototype) used NFS for file storage but ISR-2 and ISR-3 used Coda for file storage. [6]

5.1 Internet Suspend/Resume performance.

Key performance metrics for user are:
1. How soon after resume can he/she begin useful work?
2. After he/she resume, how much is his/her work slowed down?

Performance depends mainly on two factors. First one is speed of network and second one is size of data that is transferred. Machine that had 1 GB of RAM and 8 GB disk it took about 40 minutes to resume if all data for transferred in 100 Mbps network. But Resume can be fast as 14 seconds if only the minimal state of virtual machine is fetched and rest of the data is transferred as execution proceeds. [6]

Slowdown was measured to be roughly 8 percent at 100 Mbps while machine was fetching rest of the data. 8 percent is easy accepted number because it does not slowdown too much of your usage of computer. It was also noticed that network bandwidth is critical factor if only the minimal state of
virtual machine is fetched at begin of resume. It was really sluggish to use machine if bandwidth was lower than 10 Mbps. [6]

6. Conclusion

Seamlessness is quite hard to obtain in environment of multiple networks. Handover decision is harder to make in multi radio environment and seamless handover are usually not desired if networks have different kind of costs for user. There are many protocols proposed for vertical handovers but all of them have some weaknesses. Mobile IPv6 and SIP are most promising protocols. SIP’s strengths are: simplicity and that it do not need support from network elements. Mobile IPv6’s strengths are that there are much research and studies around mobile IPv6 and that it support end-to-end IPsec.

7. References


PART 5

LOCATION AND SERVICE DISCOVERY
LOCALIZATION IN SHORT-RANGE WIRELESS NETWORKS

Ville KASEVA

Summary

In this paper, single-hop localization methods applicable in short-range wireless networks are introduced. The methods are categorized into geometric localization, proximity-based localization and scene analysis. For every category, basic techniques are discussed and examples presented. After reading this paper, the reader should have knowledge of the basic localization methods and their operation principles.

1. Introduction

The problem of localization includes determining the physical coordinates or the area of a given node [1]. The existence of location information enables a myriad of functions. At application level, asset tracking [2], location-aware monitoring, context-aware applications [3], and personal positioning [4] are made possible. Enabled protocol level functions include position-based routing algorithms [5] and geographic addressing [6], which provide scalable and energy-efficient solutions for routing in ad hoc networks [7].

This paper introduces single-hop localization methods that can be used with short-range wireless networks, which include Wireless Local Area Networks (WLANs), Wireless Personal Area Networks (WPANs), and Wireless Sensor Network (WSNs). The basic localization scenario includes location nodes who are positioned using single-hop ranging measurements to anchor nodes. The positions of the location nodes are unknown whilst the anchor nodes’ locations are known a priori. After the ranging procedure, the unknown locations are resolved using a location estimation algorithm.

The localization methods can be categorized to geometric, proximity-based, and scene analysis [8]. Geometric methods exploit standard geometry and geometric properties of a given scenario to deduce unknown locations [6,9]. They rely on distance or angle measurements between the location node and the anchor nodes. Proximity-based methods measure nearness to known points [8, 10]. More specifically, in wireless network environment, the neighborhood of a node is the information used in location determination [6]. In scene analysis, the features of the environment are observed from a particular point of view [10]. Then, the observed features are used to determine the location of the observer or the objects in the environment [10]. In static analysis, the scene is usually pre-measured. It can be comprised of any measurable physical phenomena, such as visual images or electromagnetic characteristics of Radio Frequency (RF) signals. In this paper, only scene analysis based on RF signals is considered. The principles of the introduced methods are described in 2-Dimensional (2D) space. In general, the methods are quite straightforwardly extendable to 3-Dimensional (3D) or n-Dimensional (nD) space.

The rest of the paper is organized as follows. In section 2, geometric localization methods are described. Section 3 covers proximity-based localization. Scene analysis approaches are introduced in Section 4. Section 5 explains location data sensing techniques. Finally, Section 6 concludes the paper.

2. Geometric Localization

The geometric techniques exploit either distance or angle information. The methods utilizing distance are often referred to as range-based. Similarly, methods using angles are angle-based. Range-based
methods include lateration and hyperbolic positioning. The only angle-based method is called angulation.

2.1 Lateration

Lateration [6, 10, 11] is based on distance measurements from the location node to \( n \) anchor nodes. Fig. 1 depicts an example of lateration in 2D space. In the example the distance estimates have no error. The estimated location of the location node (\( X \)) is in the intersection of the three circles centered at the locations of the anchor nodes (\( S_1, S_2, \) and \( S_3 \)). The radii of the circles (\( r_1, r_2, \) and \( r_3 \)) are the measured distances from the location node to the anchor nodes.

2.2 Hyperbolic Positioning

Hyperbolic positioning is based on measuring distance differences between pairs of anchor nodes [11]. The points having the same distance difference between a pair of anchor nodes form a hyperbola. Fig. 2 shows an example of 2D position resolution with hyperbolic positioning. It includes three anchor nodes (\( S_1, S_2, S_3 \)) forming two pairs (\( S_1, S_2 \)) and (\( S_2, S_3 \)) and corresponding hyperbola groups \( h_{12,k} \).
and $h_{23,k}$, respectively. A hyperbola group $h_{ij,k}$ shows examples of points (indexed by $k$) having the same distance difference to anchor nodes $i$ and $j$. The location of the location node is determined to be in the intersection of the hyperbolas given by measurements to different pairs of anchor nodes. As can be seen, two distance difference measurements give ambiguous estimates for the unknown location in 2D case (points $X_1$ and $X_1'$). To get an unambiguous location estimate some a priori knowledge of the unknown location or more measurements would be needed.

### 2.3 Angulation

Angulation [4, 6, 10] is based on measuring angles between the location node and the anchor nodes. An example case of angulation in 2D space is presented in Fig. 3. $S_1$ and $S_2$ are the anchor nodes used for angulating the location of the location node $X$. As the coordinates of $S_1$ and $S_2$ are known, the distance, $r_{12}$, between these nodes can be computed. Furthermore, measuring angles $\alpha_1$ and $\alpha_2$ between the nodes $S_1$, $S_2$ and $X$, enables solving angle $\alpha_x$. Now, using the law of sines $r_1$ and $r_2$ can be solved. Finally, the unknown position $X$ is given by adding distance $r_k$ to $S_k$ at angle $\alpha_k$, where $k$ denotes the anchor node.

### 3. Proximity-Based Localization

Proximity-based techniques can be divided into single-anchor and multi-anchor methods. Cell Identity (CI) and the strongest anchor node method belong to the former category. Multiple anchor nodes are exploited in the overlapping connectivity technique.

#### 3.1 Cell Identity

CI [4] is a method originating from cellular networks, where an area is covered with Base Transceiver
Stations (BTSs), each BTS covering one cell. More generally, the radio coverage area of an anchor node can be considered to form a cell. Each cell covers a known geographical area. A location node can determine that it is in the area of a cell if it hears the corresponding anchor node. Fig. 4 shows three anchor nodes $S_1$, $S_2$, and $S_3$. The location node $X$ can hear anchor node $S_1$ but not nodes $S_2$ nor $S_3$. Thus, the location node can determine that it is located in the cell of $S_1$ (solid line circle).

### 3.2 Strongest Anchor Node

In WLANs mobile devices are typically connected to the Access Point (AP) they are closest to in signal space. In the strongest base station method [12,13] the location of the mobile device is estimated to be the same as the location of the AP it is connected to. In this paper the method is named more generally as the strongest anchor node method. Fig. 4 can also be used to give an example of localization based on the strongest anchor node method. However, now the location node $X$ determines that it is located in the same location as anchor node $S_1$.

### 3.3 Overlapping Connectivity

Localization based on overlapping connectivity [1] assumes perfect spherical radio propagation and identical transmission range for all radios. A location node determines its connectivity to an anchor node using some metric, for example the percentage of successful beacon packet receptions from a fixed amount of beacon packets. Upon resolving its neighborhood, the location node can estimate to be located in the centroid of the anchor nodes it is connected to.

Fig. 5 illustrates an example having four anchor nodes and one location node. The location node calculates its location to the centroid of the anchor nodes. However, it must be noted that even with ideal radio propagation the location node can actually be located anywhere in the striped area.

### 4. Scene Analysis

Scene analysis consists of an offline training/learning phase and an online localization phase. The online phase of RF-based scene analysis includes recording Received Signal Strength (RSS) values corresponding to different anchor nodes as a function of the users location. The recorded RSS values and the known training locations are used either to construct an RF fingerprint database [13-15], or a probabilistic radio map [9, 16-20]. In the online phase the location node measures RSS values to
different anchor nodes. The location is estimated using the measured values and the data collected in the offline phase.

4.1 RF Fingerprinting

With RF fingerprinting [13-15], the location of the location node is determined by searching through the RF fingerprint database to find the recorded reference fingerprint values that are closest to the measured one (in signal space). Basically, a reference fingerprint \((r)\) consist of the training location \((l)\) and of tuples of form \(<ID_{ik}, s_{ik}>\), where \(ID_{ik}\) is the ID of anchor node \(k\) heard at location \(l_i\) and \(s_{ik}\) is the vector of RSS measurement heard from anchor node \(k\) at location \(l_i\). The unknown location is then estimated to be the one paired with the closest reference fingerprint.

If there are many equidistant closest fingerprints, the location can be estimated to the centroid of these reference locations. Also, a weighted centroid of \(k\) nearest reference fingerprints is possible. In the weighted solution the locations of the fingerprints are weighted according to their distance from the measured RSS values. For estimating distances between fingerprints, for example Euclidean and Manhattan distance metrics have been considered to the determination of distances between fingerprints [14, 15].

4.2 Probabilistic Radio Map

With techniques based on probabilistic radio maps [9, 16-20], the RSS samples collected in the offline phase are used to construct joint signal strength probability distributions for every training location used. That is, for every training location \(l_i\), a conditional probability distribution \(p(s_i|l_i)\) for the observed signal strength vector \(s_i\), containing RSS values from audible anchor nodes, can be obtained. The joint probability distribution of the whole training data is denoted by \(p(s|l)\).

In the online phase the unknown location is of interest. That is, the conditional probability distribution \(p(l|s)\) (posterior), where \(l\) is the location and \(s\) is the observed RSS vector including values from all audible anchor nodes. Furthermore, an estimator for the location, such that the conditional distribution is maximized, needs to be found. The estimator is the maximum of the likelihood function \(p(s|l)\).

5. Localization Data Sensing

The localization methods described in the previous sections require localization data sensing to enable location estimation. The measurements obtained from the physical medium need to be transformed into a form that the location estimation method can use. This requires a model, which can map the measured quantity into a localization data estimate. Geometric localization includes a multitude of localization data sensing methods. With proximity-based methods, there are fewer options, since the decision is binary. Either the location node is near an anchor node or not. RF-based scene analysis includes also a straightforward localization data sensing method: the observation data tuples and their locations need to be saved in a database during the offline training phase.

Distance between the location node and an anchor node can be measured using signal Time-of-Flight (ToF) or path loss. Commonly, ToF measurements are done to either RF or ultrasound signals. The used methods are Time-of-Arrival (ToA) or Time-Difference-of-Arrival (TDoA). Path loss measurements are typically done using RF signals. Path loss can be measured using RSS or Transmission Power based Path Loss Metering (TXPPLM). The operation of measuring distance between two points is often referred to as ranging.

KASEVA, V., Localization in Short-Range Wireless Networks
Distance difference to two anchor nodes is usually measured using Time-Difference-of-Flight (TDoF) of RF signals. TDoF can be measured using TDoA of signals transmitted from different sources.

Angles can be measured using a method called Angle-of-Arrival (AoA). AoA is typically used with RF or ultrasound signals. It can be derived either from signal phase difference or TDoA at two rigid points whose separating distance is known. Also, directional rotating antennas can be used to get more coarse grained angle measurements.

Proximity can be inferred using an RSS threshold or Packet Error Rate (PER). With RSS threshold the RSS value needs to be above a certain value before proximity is assumed. Similarly, using PER, the PER value should be bigger than some threshold.

6. Conclusions

In this paper, single-hop localization methods applicable in short-range wireless networks were introduced. The methods were categorized into geometric localization, proximity-based localization and scene analysis. The basic principles of techniques such as lateration, hyperbolic positioning, angulation, CI, strongest anchor node, overlapping connectivity, RF fingerprinting and probabilistic radio map were covered. Furthermore, the most common localization data sensing methods were introduced.

Geometric methods, including lateration, hyperbolic positioning, and angulation, exploit standard geometry. Thus, they can and achieve accurate results in the presence of accurate measurements but cannot handle measurement errors well. Proximity-based methods, including CI, strongest anchor node, and overlapping connectivity, require simple determination of node’s neighborhood, but can achieve only coarse grained location estimates. Scene analysis, that includes RF fingerprinting and probabilistic radio map techniques, is based on learning the features of the environment. Thus, the location estimation procedure can be optimized in per environment basis. However, the learning phase can be tedious and may need to be repeated regularly in order to maintain the achieved accuracy.

7. References


NAMING AND SERVICE DISCOVERY PROTOCOLS FOR PERSVasive NETWORKS

Jukka Suhonen

Summary

Service discovery allows devices to find services without prior configuration. This paper identifies the challenges in service discovery with pervasive networks. First, pervasive networks should have naming that supports locating a service easily. Second, different link layer techniques and discovery protocols should be supported, thus enabling heterogeneous networks as a device may use the most suitable protocol for its task. Last, naming and service discovery in resource constrained networks is considered. Protocols that overcome these challenges are presented.

1. Introduction

Service discovery allows devices to find services without prior configuration. Naming describes a service and allows a device to search for services that fulfill a certain criteria. Essentially, naming answers to a question “what?” while the service discovery protocol answers to a question “where?”.


The problems with the existing standards is that they are specific to a certain technology and support only simple naming. The problem with the simple naming is that it often follows names akin to the Domain Name System (DNS). Instead, a more descriptive name specifying the properties of the service or its location would be beneficial. In addition, due to different communication formats, the existing standards are not compatible with each other. As pervasive networks may consist of wide range of devices using different service discovery protocols, techniques that bridge these protocols together are required.

This paper presents issues and challenges related to the naming and service discovery in pervasive networks. Protocols that aim to overcome these obstacles are presented.

The remaining of this paper is arranged as follows. Section 2 describes location based naming, as it is of particular interest for mobile users in pervasive networks. Section 3 presents service discovery protocols and discusses how service discovery may be supported in heterogeneous networks, in which devices use different protocols. Naming and service discovery in resource constrained networks, such as WSNs, is discussed in Section 4. Finally, Section 5 concludes the paper.

2. Location Based Naming

In pervasive network, a user is often mobile and trying to discover new services in an environment that is new to the user. Thus, a common question that a naming must answer in regard of a service is “where?”.
Two types of location models are used for naming: coordinate and hierarchical. The coordinate model defines an exact coordinate in which a service can be found. This way, the distance to the service can be calculated easily based on service's coordinates and user's current coordinates. The drawback of the model is that the coordinate format is not intuitive. Hierarchical model offers more descriptive system as it may define a room, a building, a street etc. in which the service is found. While it is easy to understand, it might be hard to determine which service is closest.

### 2.1. Location Model with Aura Project

Aura project [5] at Carnegie Mellon University uses a hybrid location model that defines naming scheme with a hierarchical space tree. It aims to combine the benefits of hierarchical and coordinate models. The Aura location naming is shown in Fig. 1.

In Aura, the relationships between hierarchical locations are defined by mapping hierarchical location into coordinates. Each location is assigned with geometric attributes consisting of shape, extension (volume/area coverage of the shape) and point coordinates. Based on this information, functions like distance, within, and contains are realized. In the project, a practical implementation has been made with database. The functions are implemented in SQL. Thus, applications requiring a service make simple SQL queries to the database.

### 3. Service Discovery

Fig. 2 shows generic messaging procedure used in service discovery protocols. A device may attempt to discover services with active polling by sending service queries or by listening to service's period announcements. If multiple services are discovered, the device (or user) selects one of these with some criteria. For example, when finding a printer, the selection criterion might include distance to the printer (select nearest) or printers capabilities (e.g. prefer color printer). A service access phase begins once the service has selected. The access phase includes association and authentication to the service, the use of services, and disassociation.
3.1. Service Location Protocol

In SLP, a device or software requesting services is referred to as User Agent (UA). A device that provides services is referred to as Service Agent (SA). In addition, a network may contain an optional Directory Agent (DA) that contains the locations of available services.

Fig. 3a shows the operation of SLP without a DA. To discover services, a UA multicasts query to every SA which reply independently to the UA. When a DA is used, SAs register their services to it as shown in Fig. 3b. Instead of resource consuming multicast queries, a UA makes an unicast query to the DA that then returns a list of matching services.

SLP operates on top of User Datagram Protocol (UDP) and Transport Control Protocol (TCP). While SLP does not limit the type of services it can be used with, it commonly used only in Local Area Networks (LANs) to discover printers and servers.

3.2. Service Location in Multiparty Peering

Message bus (Mbus) [6] is a service discovery protocol that is targeted at group communication. Group communication is required for many coordination scenarios, such as multimedia conferencing. Mbus includes group and point-to-point communication services, security services (authentication and encryption of messages), and assigns unique address for each device participating in a Mbus session.

SUHONEN, J: Naming and Service Discovery Protocols for Pervasive Networks
The service discovery in Mbus follows the basic procedure presented in Fig. 2. However, the unique feature in the Mbus is that it enables devices with different link layer technologies. This is achieved by noting that after the service discovery phase, all services are contacted with the device that initiated the discovery. In Mbus, the initiator becomes a coordinator for group communications. If native multicast is not available, the coordinator acts as a hub as shown in Fig. 4.

In Mbus, devices broadcast periodic visibility reports that list their available neighbors. This way, the coordinator knows whether it needs to forward messages, and can thus react to changes in network topology.

![Fig. 4 Group communication in Mbus. If native multicast is not available, a coordinator forwards messages between devices.](image)

### 3.3. Multi-Protocol Service Discovery

A service discovery protocols are usually targeted at certain application scenarios. While certain protocol might be feature-rich, it might be too heavy for all uses, and all of its features might not be required. As pervasive networks consist of wide range of different devices and services, it would be beneficial to allow a device use a discovery protocol that is most suitable for it. However, a device should still be allowed to connect to other devices using a different service discovery protocols, even if such connection would have reduced capabilities.

Multi-Protocol Service Discovery and Access (MSDA) [7] is a middleware platform that integrates existing service discovery and service access protocols. It allows devices using different service discovery protocols to discover each others' services. The components in MSDA are show in Fig. 5. SDA plugins convert service discovery based messaging into MSDA messaging and communicate with a MSDA manager component. MSDA manager disseminates messages to the network. Separate networks can be connected with MSDA bridges between MSDA managers.
To avoid unnecessary dissemination of service requests to the whole network, MSDA manager and bridges can direct queries based on the context information. Supported contexts include static information (supported network protocols, level of security, ...), network dynamics (load, number of users, ...), admission control policies, and user information. For example, a context based query might request certain minimum bandwidth and security level, in which case the query is not routed through insecure or loaded networks.

4. Naming and Service Discovery in Resource Constrained Networks

A drawback with the traditional service discovery is that it involves considerable messaging. While it is not usually a problem in computer networks, the traditional service discovery methods are not practical in resource constrained networks, such as Wireless Sensor Networks (WSNs).

In WSNs, devices are often battery powered and deployed in harsh locations which makes replacing batteries hard or impossible. Therefore, network lifetime is a primary concern. Typically, a transceiver consumes most of the energy. A typical power consumption for a MicroController Unit (MCU) in active mode is around 2-3 mW, whereas a transceiver may consume 30-60 mW in its active mode. Thus, minimizing communication overhead is important.
Directed diffusion [8] is a routing protocol suitable for WSNs. It integrates naming and service discovery into the routing. Instead of traditional approaches, directed diffusion uses a network topology independent naming scheme that is relevant to an application. Routing is performed according to named data instead of addresses. This eliminates the typical overhead in resolving name bindings and allows application-specific, in-network processing that reduces communication overhead.

The directed diffusion is illustrated in Fig. 6. Initially, a node that requires services, referred to as a sink, floods an interest message to the network. The interest contains attribute/value pairs describing what kind of data the sink is interested in. For example, an interest might request periodic temperature updates in a certain location. A node having data that matches the interest broadcasts an exploratory data message back to the sink. As sink receives the message, it reinforces a path to the source node by sending a reinforcement message to one of its neighbors (e.g., to a neighbor that delivered the data message first). The neighbor then reinforces one of its neighbors until the whole path to the source is reinforced. The remaining data messages are sent via the reinforced path. To cope with unreliability and mobility, directed diffusion uses local messaging to repair broken links.

![Diagram of directed diffusion](image)

Fig. 6 Directed diffusion, a) sink floods an interest, b) source broadcasts an exploratory data message, c) sink reinforces one of the paths to the source, d) source transmits data via the reinforced path.

The benefit of the directed diffusion is that nodes do not need globally unique identifiers. Instead, routing is performed solely based on interests and data contents. Filters are used as a mechanism for application-specific data handling. Each filter handles certain kind of data (that is, the data that matches some interest) and is triggered when matching data enters a node. This allows manipulating the message, caching data, or generating new message. In particular, filters enable data aggregation close to the sensed phenomenon. For example, if several source nodes that are closely located match the interest, they probably share the routing path to the sink. Instead of transmitting data immediately, a filter may delay sending for awhile and thus receive data also from the other node. Then, the filter may e.g., combine the values and send averaged results to the sink.

5. Conclusions

In pervasive networks, it is often necessary to use applications and services on unfamiliar locations. Naming and service discovery are required for accessing these services easily. Instead of associating a service with a non-descriptive name, location-based naming can be used to determine service's physical location. In addition, using a name that consists of attribute/value pairs enables specifying exactly the type of required service.
As pervasive networks contain a wide range of different devices ranging from WSN nodes and mobile phones to computers, there is no service discovery protocol that is suitable for every device. Thus, protocols that bridge different link layer technologies, such as Mbus, or enable using different discovery protocols, such as MSDA, are required.

Due to their resource constraints, WSNs require their own naming and service discovery schemes. For better efficiency, naming and service discovery should be integrated to network stack. Directed diffusion uses this approach by combining naming with routing, therefore reducing messaging overhead and increasing network lifetime.

References


DEVICE & SERVICE DISCOVERY IN THE HOME NETWORKS

Petros BELIMPASAKIS

Summary

Home networks have recently evolved to advanced, broadband, multimedia networks hosting a plethora of different device types. In order to make these devices even more intuitive and simple to use, there is the need for automatic configuration, transparency of device and service discovery, as well as seamless interconnection. In this paper we describe the basic elements of a typical networked home and we present the Universal Plug-n-Play (UPnP) and Zeroconf protocols, which are the de facto standards in the area of pervasive home networks.

1. Introduction

Until recently networking at home has been meant for connecting a home PC to the Internet, via a modem and a telephone line. Local Area Networks (LAN), such as Ethernet, which have been very common in the enterprise/business domain, had been very rare in the typical homes. However, this traditional image is now being changed [1] and home networks are based on high speed data LANs. This has been enabled by a) the emergence of digital media consumed natively on various home devices and b) the availability of affordable Wireless Local Area Networks (WLAN) access points suitable of usage in the average home with a broadband connection.

Traditional “brown goods”, such as television sets and audio systems are getting exciting new features like the ability to show images and play music from digital media, connect to the Internet for streaming live content and digitally record or time shift live programs. And many of these devices are not functioning just as stand-alone entities any more. High-end models can wirelessly discover each other, negotiate their capabilities and use their combined services in a networked environment that promises to provide new levels of wellbeing and entertainment for the people possessing them [2].

The home is a special place where the computer, consumer electronics and telecommunications domains converge. And the end users are amateurs, every-day people, who do not have advanced knowledge and expertise in technological fields. In order to make the “digital home” experience pleasurable and stress-less, there is need for technology to “just-work”. That means, without device configurations, driver installations and endless settings. There have been many service discovery protocols suggested [3] for this problem, but most of them have not survived up to date. For the home domain, there have been mainly two active protocols. The dominating one, called Universal Plug-n-Play (UPnP), was created when the consumer electronics, information technology and telecommunication industries joined their forces in order to provide interoperable devices and solutions. The UPnP Forum [4] tries to address and solve all technical and usability issues in the networked home where all devices could form a transparent grid of information and media flow, without complicated configurations and maintenance actions from the users. An alternative, but no so widely used approach, has been the developed by the Internet Engineering Task Force (IETF) Zero Configuration Networking (Zeroconf) Working group [5]. Both solutions try to make sure that devices are now able to “talk the same language” and automatically configure themselves or cooperate for making more out of their combined capabilities.

In this paper, we focus our analysis on home data networks suitable for multimedia communications, meaning that we do not study home sensor, security or automation networks. In the following chapters,
we first describe a typical home network and its special characteristics, in Section 2. Then, in Section 3, we present two of the core use cases that the UPnP has been addressing, and in Section 4 we get into a deeper presentation of UPnP logic, functionality and protocols, while in Section 5, we briefly present the Zeroconf approach. Finally, in Section 6, we discuss some of the future research directions and possibilities in the area of ubiquitous home networking.

2. Typical Home Network

There are many different configurations that one can find in a home network, based both in the physical topology as well as the networking architecture. For example, connectivity in the home network could be based on technologies such as:

- Ethernet, standardized as IEEE 802.3, which connects devices in a star-topology with the usage of dedicated data cabling
- Wireless LAN, standardized as IEEE 802.11a/b/g/n, which wirelessly links devices with a dedicated access point
- Home Phoneline Networking Alliance (HomePNA), which utilizes the existing fixed phone cabling in the building, as a medium for high speed data transfers
- HomePlug, use the existing electricity cabling in the home building, for transfer for data.

However, we will not get into more details and alternatives, as the list of options could be quite long.

From the physical point of view, the most common setup is the one where the home is connected to the Internet Service Provider (ISP) via a broadband Asymmetric Digital Subscriber Line (ADSL) connection. The connection is typically terminated on a home gateway device, which is a stand-alone box that acts both as an ADSL modem, a WLAN Access Point and router (Fig. 1) - very often this box also provides an Ethernet switch. So, all home devices, either using WLAN or fixed Ethernet, can be connected to this box that would then provide the connectivity to ISP and the rest of the Internet.

From the networking point of view, most ISPs in Europe provide to their customer just one public, globally routable, Internet Protocol (IP) address. This IP address is assigned to the external interface of the home gateway. In the internal network, all devices use private IP addresses, which are assigned by the Dynamic Host Configuration Protocol (DHCP) server running on the home gateway. The gateway also acts as a Network Address Translator (NAT) router, which means that all the traffic from the internal devices to the Internet is routed via the home gateway.
A home network has some characteristics that make it special, compared to other networks, for example corporate networks. More specifically:

- There is need for “out-of-the-box” interoperability between devices, since it is not possible to customize interfaces or making special modifications to fit the specific network.
- The vendor landscape is highly non-homogenous, meaning that people purchase device from different manufacturers. The TV is not same brand as the video recorder or the audio system.
- Home devices are bought in pieces and not as a whole-package. Thus, there might be many “generations” of devices, older and newer ones, in the same network.
- Coexistence issues, as wireless radios working in the unlicensed bands may interfere.
- There is no expert administration authority, as amateur users are using these networks.

3. Pervasive Home Network Use Cases

3.1 Audio & Video Example

Imagine having multiple digital media devices, e.g. a television set, a picture frame and an audio system, connected on the home network. Also, a media server with different kinds of content, such as music, images, home videos and movies. It would be great if this content could be automatically discovered and accessible from the multimedia home devices, as shown in Fig. 2.

Those devices could be connected on the home network over multiple physical connections (IEEE 1394, Bluetooth, 802.11, Ethernet, etc.), but still they could discover each other, and the media server. Which they could then access, use its meta-data and content services to discover multimedia content and play the it from device to device within the home. Users could view their digital photos in the family room, on the TV screen or listen to their MP3s on the stereo system instead of the PC speakers. These various audio/video devices could also subscribe to web entertainment services, such as Internet radio and stream the content to multiple devices in the home [6].

3.2 Imaging Example

Sharing printers is common in home networking, and in a pervasive and ubiquitous environment it would be expected that the user does not need install support drivers on all the PCs on the home network. Also, when adding a new device to the network, such as a work laptop or a mobile phone, it should automatically discover the printer or scanner and will have access to its services without any need for any configuration. All that needs to be done is plug Printer/Scanner directly into any available network port, or associate it to the wireless home network. Currently, a user has to go through every network printer, in order to install and finalize a setup and configuration procedure, before it can be usable by other devices [6].
4. Universal Plug-n-Play Protocols

4.1 Basics of UPnP

The UPnP Forum [4] is an industry initiative, supported by 850 members, designed to enable simple and robust connectivity among stand-alone home devices and computers by different vendors. This technology tries to make networking of home devices simple and affordable. UPnP is based on open network architecture and it is independent of operating systems, programming languages or physical medium. It uses the standard TCP/IP connectivity, and higher level IP based protocols, in order to be able to work in existing computer networks.

There are three major classes of UPnP devices:

- Control Point – this is a device that is capable of discovering and controlling other devices. It can invoke the actions of a service and get its status information. For example, it could be an advanced home remote control.
- Controlled Device – any UPnP device that can be controlled or provides services to the rest of the home network.
- Bridge – an entity that connects non-UPnP devices to the UPnP home network. It could be considered as a translator between UPnP and other “proprietary” protocols.

It is important to know that a UPnP device contains services. A service is a unit of functionality implemented by a device [7]. Each service provides a set of actions that can be invoked, with a set of input parameters, and can return a value. In a similar way as programming language methods accept parameters and return a result, upon execution. UPnP Forum specifies in detail the action and parameters of services, so that they are compatible and interoperable among different vendor implementations.

4.2 Steps to UPnP Networking

In this section we will see how UPnP works, in a “bottom-up-approach”, starting from the lower level connectivity issues, going all the way up to presentation, as shown in Fig. 3.

Addressing is the very first step towards UPnP Networking. Every UPnP device needs to automatically acquire an IP address, before it can communicate with any other entity. Control points and devices get addresses for a DHCP server, if available in the network. If not, then they should select an address based on Auto IP, which is an implementation based on an IETF Draft for Automatically Choosing an IP Address in an Ad-Hoc IPv4 Network. Auto IP essentially suggests that a node picks an address in 169.254/16 range and checks to see if it is used, using the Address Resolution Protocol (ARP).
Discovery of network-based resources is the next step. When a device joins the UPnP network, the discovery protocol allows the device to advertise its services to control points across the network; and allows a control point to search for available services. This is done by sending multicast search message, using Hyper-text Transfer Protocol (HTTP) over User Datagram Protocol (UDP) over IP, using the HTTP Multicast (HTTPMU). Devices that receive the search message reply back in a unicast manner. The discovery multicast address, and the format of the advertising messages defined by the Simple Services Discovery Protocol (SSDP), which is an IETF draft. A UPnP control point can limit the search parameters and just look for devices of a particular type (such as just audio devices) particular services (such as clock services) or even a particular device.

Description covers the format of the documents used to describe UPnP devices and the services provided by them. If a control point needs to know more about the services offered by the device it needs to request the description document that is in Extensible Markup Language (XML) format. This document describes the physical device and all its logical embedded devices, services, manufacturer information, etc. Service description lists what is implemented by a service. That includes the set of actions/methods implemented, which may have zero or more arguments (parameters). Each argument is also tied to a state variable. Variables can be read by control points, and as we will see later, they can “subscribe” so that they are automatically notified when a variable changes. Note that this service description information is defined by UPnP Forum for a standard service, but if a vendor would like to create a non-standard service, then the vendor should provide this information.

Control describes the remote procedure call mechanism, used by control points, to invoke services offered by UPnP devices. This is done using the Simple Object Access Protocol (SOAP), which is an IETF draft. Through the use of SOAP a control point can query or change the elements in a service’s state table, such as turning the device on or controlling the volume of an audio system. The control point creates the XML document that contains the commands and posts it to control element of the device to be controlled. SOAP messages are delivered on top of HTTP – TCP – IP.

Eventing defines how control points are notified about changes on the state of devices. Once a device or control point subscribes to the services of a UPnP device, it can stay informed of the state of those services. They can subscribe by using the event notification service URL (provided in the service description) and are then notified by the UPnP device once the state of the services changes. For example, subscribing to the events of a television set, a control point would get a notification when the television is turned on, or off. Or, it can be notified when the temperature of a room change, if such a service is offered by a UPnP enabled home air-conditioning system. Event messages are delivered via HTTP using the General Event Notification Architecture (GENA), which is yet another IETF draft.

Presentation defines how a UPnP device may use standard web protocols to provide a typical web page that can be accessed via any web browser. Basically, it is allowed for a UPnP device to include, in its device description, a URL that points to its own web server and links to a standard web page, which can be, however, dynamically updated. This way, a user can use a standard web browser in order to access information, and maybe also control, a device.

4.3 Anatomy of UPnP Protocol Stack

As mentioned in the previous Section, UPnP utilizes many of the existing Internet protocols, such as IP, TCP, UDP, HTTP and a variant of HTTP that works on top of UDP multicast (HTTPMU). Fig. 4 shows the “anatomy” of the UPnP and its relations with the existing protocols and related IETF drafts.
The fact that UPnP protocols work on top of the very widely spread HTTP makes it easy to be deployed to devices that already have a web server, since the essential HTTP components are already there. Moreover, UPnP implementations typically mall footprint software component on the device and since it is independent of operating system, language, or physical connectivity it can be an addition to many existing class of devices.

5. The alternative: Zeroconf

5.1 Steps to Zeroconf Networking

Similarly to UPnP, Zeroconf protocols allow users to connect computers, printers, and other devices together, without the need of configurations. Zeroconf currently covers three layers/issues:

- Choose network addresses for the newly added networked devices
- Find out which device has a certain name
- Find out where to get services (e.g. printing or scanning)

Those layers are shown, in a bottom-up approach, on Fig.5.

![Fig. 5. Layers of Zeroconf Networking](image)

**Addressing**

Following pretty much the same approach as in UPnP, Zeroconf requires the usage of the IETF RFC “Dynamic Configuration of IPv4 Link-Local Addresses”, for proving an address to a device that joins the network. A small detail is that the UPnP is based on an earlier draft of the protocol, while the Zeroconf actually uses the RFC protocol.

**Naming**

Multicast Domain Name Service (mDNS) [8] is a way of using interfaces, packet formats and interaction models, similar to the typical DNS. But, it is optimized in order to be used in a small network where no typical DNS server exists, such as a home network. mDNS allows a network device to choose a domain name in the ".local" namespace and announce it using a special multicast IP
address, which is the IPv4 address 224.0.0.251. Each device on the LAN stores its own list of DNS records and when an mDNS client queries the name of a device, in order to resolve its IP address, the device with the corresponding DNS record replies with the appropriate result, which contains its IP address.

**Discovery**

DNS based Service Discovery (DNS-SD) is built on top of the standard DNS. It uses the DNS SRV (Service), TXT, (Text) and PTR (Pointer) records to advertise services. Basically, the hosts/devices offering the different services publish details of those services, for other devices to discovery and use. The service types are given informally and a full registry is maintained at the website DNS-SD.org.

### 5.2 Comparison to UPnP

If one would like to very briefly compare the two protocols, the following issues would be the most important ones:

- **Addressing**: Both protocols are very similar, basically utilizing the same solution.
- **Naming**: available only in Zeroconf, but not in UPnP.
- **Discovery**: UPnP utilizes HTTPMU, while Zeroconf relies on mDNS.
- **Application**: Zeroconf is agnostic about the application-layer protocols. Zeroconf provides only the foundation of easy to setup IP networking, via its three core layers, but does not deal with the application layers, where anything can be applied. On the other hand, the application-layer protocols are very strong in UPnP, and this is where a significant part of the standardization activities have been focused. UPnP defines the actions, methods, variables (but with space of vendor variations).
- **Underlying protocols**: Both Zeroconf and UPnP utilize existing and public IETF RFCs or Internet drafts.
- **Adoption**: UPnP is supported by more than 850 manufacturers, which belong to the UPnP Forum. On the other hand, Zeroconf is mainly supported by Apple and printer manufacturers. Thus, it has not managed to enter the consumer electronics market.

### 6. Future Directions

Both UPnP and Zeroconf have been very focused on connecting devices that are in a physical home, and they pretty much assume that those devices are static. However, with the mobile devices, cellular network and public WLANs becoming common place, most users carry some of their personal devices (such as laptops or mobile phones) even when outside the home. Currently, it is not possible for these mobile devices, or even remote static devices (such as an office PC), to remotely connect to the home network and communicate with the home devices, as they would do if they were directly connected on the home network. The problem mainly relies on the multicast-based advertising of the discovery mechanism and the private IP addresses used by intra-home devices. Thus, the next logical step would be that UPnP is enhanced with some new extensions/entity that would allow access of home resources, from remote clients. This would allow the users to enjoy their content and services from any location, without noticing any difference in the perceived experience, thus one step closer to the ultimate pervasive goal. Some solution have been already suggested, such as the [9, 10].

Moreover, the existing protocols have been criticized for their underlying technology selections. For example UPnP uses technologies such as SOAP, which has been adapted from the Web Services, for remote procedure calls, and it has been challenged as of how appropriate it really is [11]. Also, it has been suggested that mechanisms built on Representational State Transfer (REST) principles can actually deliver better performance and less overhead, in a UPnP environment [11].
7. Conclusion

It is evident that a home network, being multi-vendor, having devices being brought in pieces and lacking expert administrative authority, has some special characteristics. These make it very challenging in order to make sure that devices can work seamlessly and transparently in such a way that users would tolerate them in their every day home-life. This is why there is need for ubiquitous, device & service discovery, with far better experience than the PC (no drivers installations, control panels, etc.). UPnP and Zeroconf have been the key protocols for this, providing “invisible” networking and device communication on top of any physical network supporting IP technology, as they are is based on Internet protocols, such as HTTP, DNS and many IETF RFCs. Moreover, they are programming language & OS independent, making it possible to be deployed in any kind of networked home device.

8. References


