Bachelor's thesis

Electronics

Telecommunication Systems

2014

Tommi Tuomisto

REAL-TIME TRANSMISSION IN A WLAN

TURUN AMMATTIKORKEAKOULU
TURKU UNIVERSITY OF APPLIED SCIENCES

BACHELOR´S THESIS | ABSTRACT
TURKU UNIVERSITY OF APPLIED SCIENCES

Electronics | Telecommunication Systems

2014 | Total number of pages

Juha Nikkanen, Lic. Tech., Principal Lecturer, Jouko Leppäniemi

Tommi Tuomisto

REAL-TIME TRANSMISSION IN A WLAN

This thesis focuses on specifying the most notable factors of a wireless local area network impact on real-time, IP –based data transmission.

The subject for this thesis was assigned by Kantio Oy, a company providing WLAN solutions and consultation regarding network performance analysis and optimization.

This thesis first introduces the basic WLAN architectures and channel topologies and examines the current Ethernet and WLAN techniques, such as CSMA and QoS, and their impact on the disturbances, latencies and general performance of the network. Voice over IP serves as an example of a real-time application with which the problems of WLAN as a communication medium arise. The thesis also studies the characteristics of a voice call and the effects that cause the quality degradation of a voice-call, such as jitter and packet loss.

Lastly, the thesis examines the theory behind a WLAN handover and its effect on real-time transmission. The examination is conducted by first studying how the handover mechanism differs between WLAN channel architectures and then by measuring the handover delay differences of various types of mobile phones.

As a result, this thesis gives a general presentation of the problems behind deploying real-time data through a WLAN and how the effect of those problems can be decreased. The DiffServ QoS prioritization framework and the time-slot division -based Airtime Fairness method were found to be effective tools for optimizing network performance and guaranteeing high-quality service. In addition, based on the handover comparison made between channel architectures, and the handover delay measurements made with different mobile devices suggest, that a beneficial decrease in handover delay can be achieved by deploying a WLAN with a single channel architecture.

KEYWORDS:

WLAN, real-time, VoIP, Ethernet, single channel, multi-channel.

OPINNÄYTETYÖ (AMK) | TIIVISTELMÄ
TURUN AMMATTIKORKEAKOULU

Elektroniikka | Tietoliikennejärjestelmät

2014 | Sivumäärä

TKL Juha Nikkanen, Jouko Leppäniemi

Tommi Tuomisto

REAALIAIKAINEN TIEDONSIIRTO WLAN -VERKOSSA

Tämän opinnäytetyön tavoitteena oli tuoda esille ne langattoman lähiverkon osatekijät ja menetelmät, jotka vaikuttavat merkittävästi reaaliaikaisen, IP-pohjaisen datan kulkemiseen verkon läpi.

Aiheen opinnäytetyöhön antoi Kantio Oy, WLAN ratkaisuja sekä tietoliikenneverkkojen analysointiin ja optimointiin liittyvää konsultointia tarjoava yritys.

Työssä perehdyttiin tämän hetkisiin WLAN- ja Ethernet -tekniikoihin, joiden avulla on mahdollista vähentää langattoman verkon häiriöitilanteita ja viiveitä sekä yleisesti parantaa verkon suorituskykyä reaaliaikaisessa tiedonsiirrossa. Reaaliaikaisista sovelluksista kiinnitettiin huomiota erityisesti Voice over IP eli VoIP-tekniikkaan ja sen laatua heikentäviin tekijöihin. Työssä käytiin läpi äänipuhelun toimintaperiaatetta ja tutkittiin mm. viiveiden, jitterin ja pakettien hävittämisen merkitystä hyvän puhelun laadun takaamisen knnalta.

Lopuksi tutkittiin päätelaitteiden toteuttamaa solunvaihto -menetelmää ja sitä miten se vaikuttaa reaaliaikaiseen tiedonsiirtoon. Käytännön mittauksia apuna käyttäen tarkasteltiin myös kuinka sekä WLAN:in kanava-arkkitehtuurin valinta, että ero päätelaitteiden välillä vaikuttavat solunvaihtoviiveisiin.

Lopputuloksena työssä saatiin kattava käsitys langattomassa lähiverkossa tapahtuvan reaaliaikaisen tiedonsiirron problematiikasta. Verkon palvelun laadun takaamisen kannalta tärkeimmiksi työkaluiksi osoittautuivat datan QoS -priorisointiin käytettävä DiffServ -menetelmä ja aikaan pohjautuva Airtime Fairness -lähetyksenjakotekniikka. Käytännön mittauksien avulla huomioitiin solunvaihtoviivettä pienentävä vaikutus, joka olisi saavutettavissa siirtymällä langattoman verkon monikanavaisesta mallista yksikanavaiseen arkkitehtuuriin.

ASIASANAT:

Reaaliaika, VoIP, WLAN, Ethernet, monikanava, yksikanava

CONTENT

APPENDICES

PICTURES

FIGURES

TABLES

LIST OF ABBREVIATIONS (OR) SYMBOLS

IP	Internet Protocol
TDM	Time Division Multiplexing
Wi-Fi	Synonym for 802.11 compatible wireless devices
LAN	Local Area Network
IBSS	Independent Basic Service Set
AP	Access Point
WLAN	Wireless Local Area Network
STA	Station, wireless client
DS	Distribution System
BSS	Basic Service Set
SSID	Service Set ID
MAN	Metropolitan Area Network
WAN	Wide Area Network
CSMA/CD	Carrier Sense Multiple Access with Collision Detection
CAT	Category (cables)
RF	Radio Frequency (3 kHz to 300 GHz)
DSSS	Direct Sequence Spread Spectrum
OFDM	Orthogonal Frequency-Division Multiplexing
3G	3rd Generation Mobile Communication Technology
4G	4th Generation Mobile Communication Technology
IFS	Inter frame space
ACK	Acknowledgement
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
RTS	Request To Send

CTS	Clear To Send
QoS	Quality of Service
RTP	Real-time Transport Protocol
RTCP	RTP Control Protocol
FEC	Forward Error Correction
IntServ	Integrated Services
DiffServ	Differentiated Services
RSVP	Resource Reservation Protocol
PHB	Per-hop Behavior
DSCP	Differentiated services code point
ESS	Extended Service Set
ITU	International Telegraph Union
ITU-T	ITU Telecommunication Standardization Section
DSP	Digital Signal Processing
TCP	Transmission Control Protocol
HO	Handover
UDP	User Datagram Protocol
DS-domain	DiffServ -domain
CQ	Communication Quality
EF	Expedited Forwarding
AF	Assured Forwarding
CoS	Class of Service
PDV	Packet Delay Variation
PCP	Priority Code Point
IEEE	Institute of Electrical and Electronics Engineers
RJ45	Modem connector type
VLAN	Virtual Local Area Network

1 INTRODUCTION

This thesis will examine the characteristics of a wireless local area network and focus on specifying the factors, techniques and workarounds that have the biggest impact on real-time, IP -based data transmission.

Although many similarities can be found between the characteristics of different wireless technologies, this thesis concentrates solely on WLAN and leaves out for example 3G, LTE, WiMAX and other mobile broadband or wireless network solutions. As a WLAN usually consists of both Ethernet 802.3 and WLAN 802.11 techniques, this thesis examines both while having a higher emphasis in the wireless area. This thesis uses Voice over IP to serve as an example of a real-time application with which the problems of WLAN as a communication medium are brought up.

A variety of theses, such as Joni Nikkari's 'VoIP over WLAN' (2007), had already been done about WLAN and VoIP, but rather than studying the lower layers (network, data link) the emphasis on these was directed closer to the upper layer, with examination of subjects such as WLAN encryption, VoIP codecs and VoIP applications.

Apart from a few direct Internet links, the references used in this thesis comprised mostly of literature and publications made by respected organizations and companies of the IT field, such as IEEE, ITU and Cisco. For more detailed information about each topic mentioned in this thesis, the reader is advised to turn to these source materials listed in the references.

2 BACKGROUND

Every year, a growing number of services are being provided through wireless networks and the Internet. Dedicated networks that once contributed to only one service, are now being made obsolete by the possibilities of IP -technology. The transition to a collective, packet-switched data service is an on-going process, driven mainly by the growing need for cost efficient networking. Organizations aim for savings by concentrating on building and maintaining a one-for-all type of network.

Meanwhile, consumers are moving from stationary devices to laptops, smartphones and tablets, each of which are sold on the premise of freeing the user from the wired communication medium, and enabling them to use the Internet and its services wirelessly, either through WLAN or a mobile broadband connection. As the use of Skype, Facetime and other real-time applications with these mobile devices increase, in order to guarantee a good quality of service to the end user, the networks have to be able to work around some of the problems specific to the packet-switched and wireless network.

Although IP networks are considered an efficient platform for data transmission, for certain types of applications the cost of this efficiency has meant a decrease in user-end quality. Due to its non-deterministic principle of operation the packet-switched network is not considered as an ideal method for real-time transmission. "The Internet traditionally provides service that is commonly characterized as a best-effort service. Many applications run very well using this service model but some new interactive applications such as telephony or video conferencing impose stringent demands to the network."[1]

3 WIRELESS LOCAL AREA NETWORK

To better understand the difficulties that real-time data transmission may face in a wireless network, it is good to first understand the basic technology that enables the use of WLAN and to visualize the most common ways of how wireless LANs are implemented.

3.1 Ethernet 802.3

Ethernet is the most used LAN technology in modern computer networking. By a set of standards and protocols, Ethernet provides an efficient way for multiple devices to access and use a single, wired transmission medium.

On a wireless LAN, Ethernet usually handles the traffic between the various networking infrastructural devices. Controllers, access points, hubs, switches and routers, depending on the desired setup, are interconnected with an Ethernet cable. This is to ensure maximum service quality and a minimum of interference to the backbone data of the network and also to the data that is forwarded to-and-from a wider area network.

The most common transmission medium used in an Ethernet LAN is a twisted-pair cable coupled with RJ45 -connectors. These types of cables are usually simply referred to as Ethernet cables. Optical fiber cables are also becoming increasingly popular as fiber gives more bandwidth and makes covering longer distance with less attenuation easier.

802.3 is a collection of standards designed and modified by the IEEE 802.3 working group. 802.3 standards define the very basis of Ethernet and provide a way of regulating how the physical layer and the data link layer access the wired medium.

In addition to controlling the media access, 802.3 also defines the standard data rates used over optical fiber and twisted-pair cables. The following are the most current data rates in use[2]:

Table 1. IEEE 802.3 defined data rates.

Name	Bit rate
10Base-T Ethernet	10 Mb/s
Fast Ethernet	100 Mb/s
Gigabit Ethernet	1 Gb/s
10-Gigabit Ethernet	10 Gb/s

3.2 WLAN 802.11

In a Wireless LAN the cables connecting the router or Access Point (AP) to the end station are replaced with a wireless medium. Communication between an AP and a client station (STA) is done via electromagnetic radio-frequency (RF) waves.

The fundamental technique behind WLAN technology is the spread spectrum method. In the spread spectrum method the generated power of a signal is spread in to the frequency domain resulting in a wider bandwidth signal. The main reason for using spread spectrum technique is to minimize the interference that the WLAN might have towards surrounding RF solutions. A wider band is also more resilient to interference, making the WLAN traffic more robust and stable in different circumstances.[3]

The three modulation techniques used in spread spectrum are:[3]

1. Frequency hopping: The transmitter and receiver change frequencies in conjunction.
2. Direct-sequence Spread Spectrum (DSSS): The signal is multiplied with a continuous, higher rate pseudonoise which results in a wide-band uniform frequency distribution.
3. OFDM (Orthogonal Frequency Division Multiplexing): The signal is divided in to several orthogonal carrier-waves that together form a wide-band distribution.

Table 2. IEEE 802.11 Working groups and focus areas.[4]

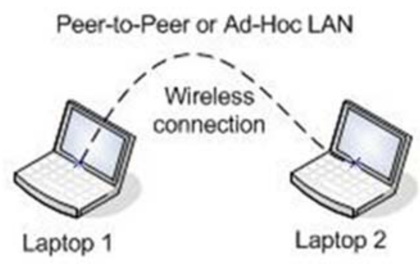
Task group	Bitrate / Description	Frequency	Modulation
802.11	1 and 2 Mbit/s	2,4 GHz	DSSS
802.11b	10 Mbit/s	2,4 GHz	DSSS
802.11g	54 Mbit/s	2,4 GHz	OFDM
802.11ª	54 Mbit/s	5 GHz	OFDM
802.11h	Added interference correction to 802.11ª	5 GHz	OFDM
802.11n	600 Mb/s	5 GHz	OFDM
802.11ac	1,3 Gb/s	2,4 & 5 GHz	OFDM
802.11ad	*6,75 Gb/s*	*60 GHz*	*OFDM*
802.11i	*Network security*		
802.11e	*Service quality*		
802.11f	*Roaming within a WLAN*		
802.11r	*Fast roaming within a WLAN*		

The IEEE 802.11 series of standards listed in Table 2. control the majority of the current applied WLAN solutions. WLAN products that are based on the 802.11 standards are more commonly referred to as Wi-Fi devices.[5]

3.3 Network Topology

The most basic wireless network setup is called an IBSS or Independent Basic Service Set. A service set of this kind is based on an ad-hoc or peer-to-peer topology.[3] In an ad-hoc network the wireless stations communicate directly with each other without using an AP as seen in Picture 1.

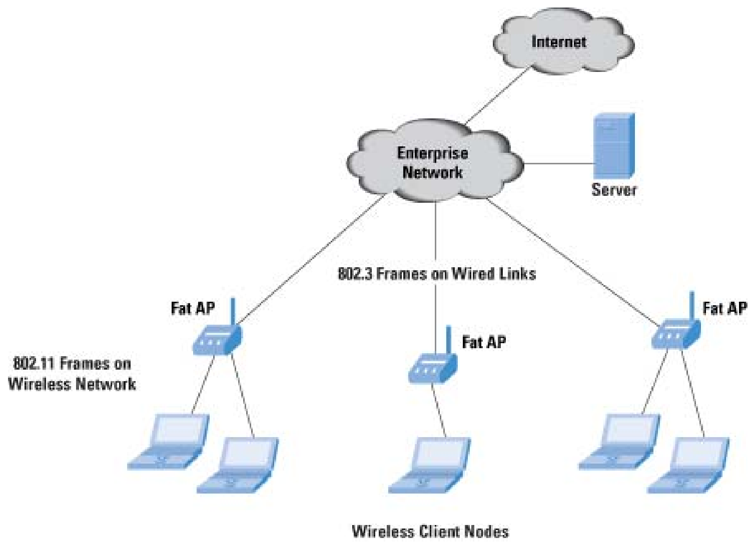
Because of its simplistic principle of operation, an ad-hoc network can't provide the same stable and dynamic functionality that can be achieved with an infrastructural WLAN setup. An infrastructural WLAN consists of one or more APs which are used as a bridge to connect wireless stations or STAs to each other or the wider area network. The APs are interconnected by a DS (Distribution System) which is usually provided via Ethernet, but can also be implemented with WLAN. The area or cell covered by one AP is called a BSS (Basic Service Set) and together every BSS on the same network form an ESS (Extended Service Set). To differentiate the network from others, an SSID (Service Set Identification) is assigned to the network. Wireless Client Stations (STAs) trying to connect to the network have to know the SSID in order to gain access.[3]



Picture 1. Ad-hoc WLAN.

The main component in setting up a infrastructural WLAN is the AP. The AP is the link through which the wireless client stations or STAs communicate with the wider area network or Internet. In an infrastructural- topology, an AP can be deployed in many ways, of which the autonomous and centralized architecture models are the most used.

In an autonomous architecture, each AP works on its own, which means that the AP is given full control of handling the data between the wired medium and the wireless STA without any communication with the other APs on the network. APs in an autonomous architecture are called 'Fat APs', as more complexity is required from them in order to implement all of the 802.11 WLAN functionality. An example of an autonomous architecture is visualized in Picture 2.[6]

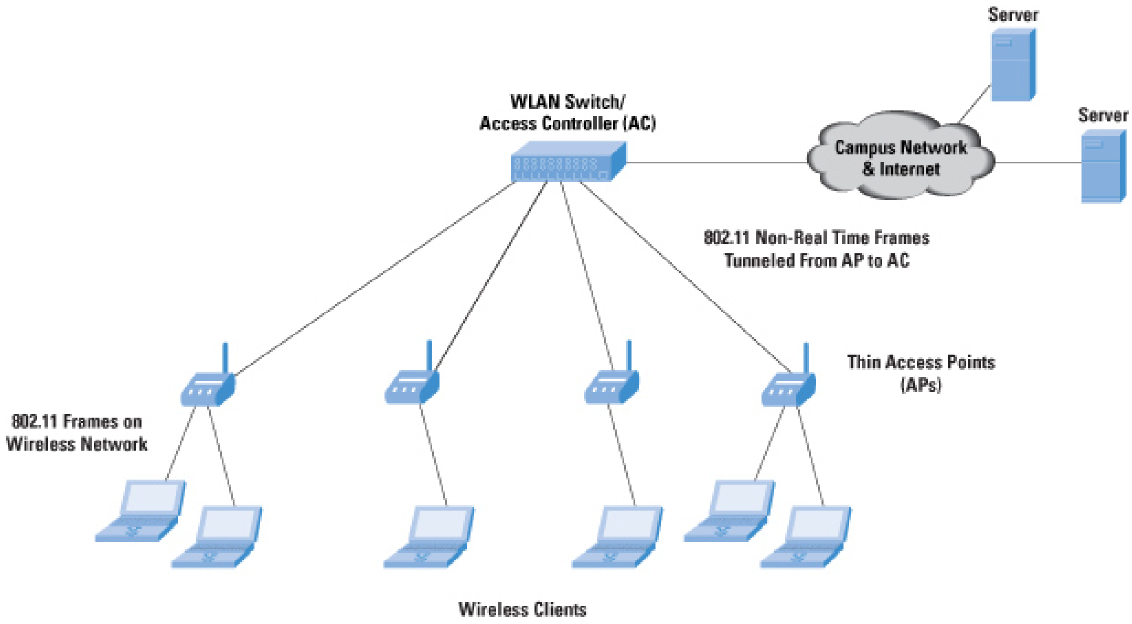


Picture 2. Autonomous WLAN architecture.[6]

In a centralized architecture, the network APs are connected to the wider area network and to each other through a WLAN Switch or Access Controller. The controller configures and controls the APs and also manages the data they forward. As the controller handles most of the tasks and processing, less complexity is needed from the APs, thus making the APs on a controller-based network 'Thin'. Thin APs are considered more lightweight, as they lack the complexity and functionality of their Fat AP counterparts. An example of a centralized architecture is visualized in Image 3.[6]

As Fat APs are more diverse in both hardware and software, they are often more expensive to install and maintain.[6] This is why an autonomous architecture is preferred only in smaller scale set ups such as home, or small-office networks. In a centralized network the functionality and the price is concentrated on the controller, making it a more efficient architecture to cover larger areas.

In both of these architectures, usually only the path from STA to AP is wireless, making the rest of the traffic traverse through a wired medium. This is why both WLAN 802.11 and Ethernet 802.3 functionality is required from modern Access Points.

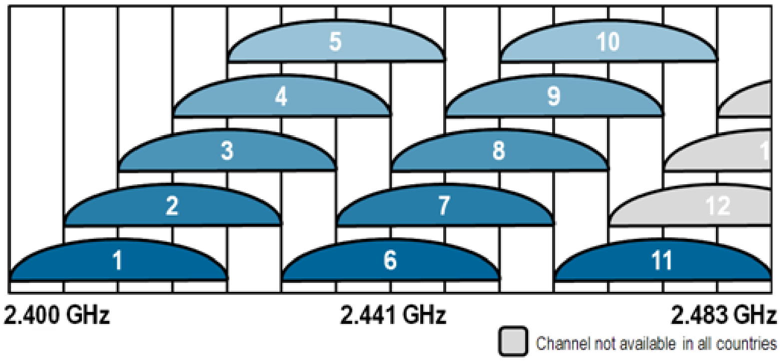


Picture 3. Centralized WLAN architecture.[6]

3.4 Channel Architecture

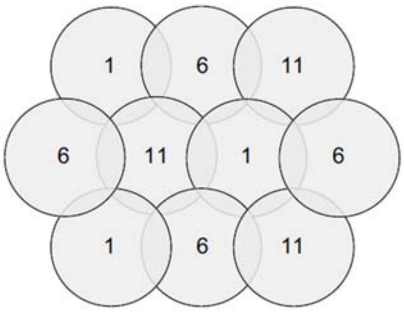
When a wireless network is designed to cover a wider area, multiple Access Points are usually implemented. Network designers have to provide good mobility for STAs moving within the network and minimize the interference that APs might cause to each other and other surrounding networks.

The most common way to design the channel layout of a WLAN is to make use of the different channels that are available within the frequency range. As mentioned before, WLAN operates on non-regulated 2,4 GHz and 5 GHz frequencies. The total bandwidth of WLAN is divided into 14 channels, from which 13 are used in Europe and 11 in the US. Because of the 2 channel difference between these two big markets, 11 channels are mostly used to maximize commonality. The 11 channels in use are allocated partially on top of each other on the frequency domain, in the way that is shown in Image 4.[7]



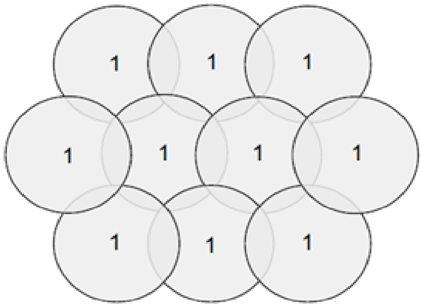
Picture 4. 2,4 GHz channel frequencies.[7]

Due to this allocation, only 3 channels that do not interfere with each other, 1, 6 and 11, can be used when designing a multiple-AP WLAN.[7] As can be seen from image 5, to eliminate the possibility of two cells interfering with each other, the channels of the APs have to be assigned so that two adjacent cells don't operate on the same channel. A wireless networking approach that uses multiple channels like this when forming the whole ESS is usually called a multi channel architecture.



Picture 5. Multi-Channel cell architecture.

The other approach is called the Single Channel Architecture. As can be seen in image 6, Instead of using 3 channels, the single channel approach uses only 1. Making use of the centralized network architecture, the problem of cellular interference is solved by using a controller which coordinates the RF decision making of the APs. Rather than being able to distuingish cells from eachother, the STA connected to the BSS can only see one continuous cell.[8]



Picture 6. Single Channel cell architecture.

From a network design perspective, a single-channel approach removes the complexity of normal channel planning. As all of the APs of the network work within the same RF channel, the network designer is left with only the task of deciding one channel for the entire network.[8]

4 VOICE CALL PROBLEMATICS

In the original circuit-switched phone line, when a phone call was dialed, a user reserved a capacity of 64kb/s from the network. This capacity was fixed, meaning that it was all reserved for one phone call, whether it was used or not. Although not as efficient capacity-wise as many modern approaches, a network like this provided a deterministic and predictable transfer medium, making it easier to assure a stable, high quality user experience.

With IP -technology, the circuit-switched model changes into a packet-switched network, where a voice-packet reserves the network only for the time it takes to transmit the packet from source to destination. A packet-switched network like this, relies on statistical availability when accessing the transfer medium and therefore is considered as a more efficient but less predictable service than the deterministic model, given that more than one user uses the network simultaneously.

Although both of these networks face similar difficulties when trying to assure a good quality user-experience, some of the characteristics explained below are unique to the packet-switched system.

4.1 Latency

In computer networking, latency is the time that it takes for a packet to travel through the network, from source to destination. In telephony, latency means the time that it takes for the sound that is made in the speakers mouth to travel to the listeners ear.[9] Latency doesn't directly affect the quality of a voice signal, but can have a major impact on the synchronization between the two participants of a phone call.[10]

The ITU-T G.114 document covering one-way transmission time recommends a 150 ms maximum latency for high-quality, end-to-end voice transmission. The same document, however, presents a user-experience study, where participants of a phone call have expressed their satisfaction towards the quality of a call while the latency between speakers has been gradually increased, see Image 13. In the study, the amount of delay is based on the developed E-model estimate (ITU-T Rec. G.107). The E-model is a transmission planning tool, that provides a prediction of the expected quality of a phone call.[12] The rating R indicates the quality of speech transmission, 100 being the best, 50 worst.

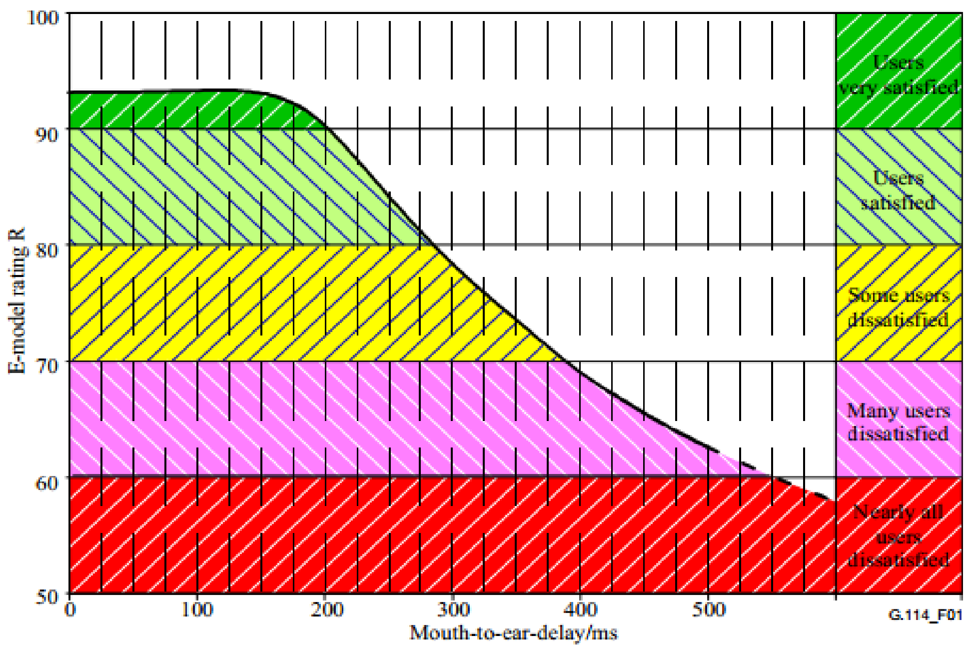


Figure 1. Effects of mouth-to-ear delay on user satisfaction.[13]

According to the graph in Figure 1, users with the rating R being over 90 have been very satisfied even up to a 200 ms latency. Therefore, 200 ms can be considered as the maximum one-way latency that shouldn't be exceeded when trying to guarantee a good user-experience.

The primary causes for latency are the following:

- Packetization delay
- Propagation delay
- Link delay

Packetization delay is the time it takes to fill a packet with data. Generally, the larger the packet size, the bigger the latency.[10] The amount of packetization delay varies between manufacturers and is affected by the design of the packeting algorithm in use. For example, the DSP of a Cisco VoIP -product using G.729 (a voice packeting algorithm) uses a codec sample interval of 10 ms. Two of these samples, each with a 10 ms delay, are made into one packet. With G.729, an added 5ms of delay is caused in every packet formation, making the total amount of delay 25 ms for one packet. [9]

Propagation delay is the time that it takes the signal to traverse through the cable. The approximate time-delay caused by propagation is calculated with the following equation:

$$T_{pd} = \frac{d}{c * 0{,}56}$$

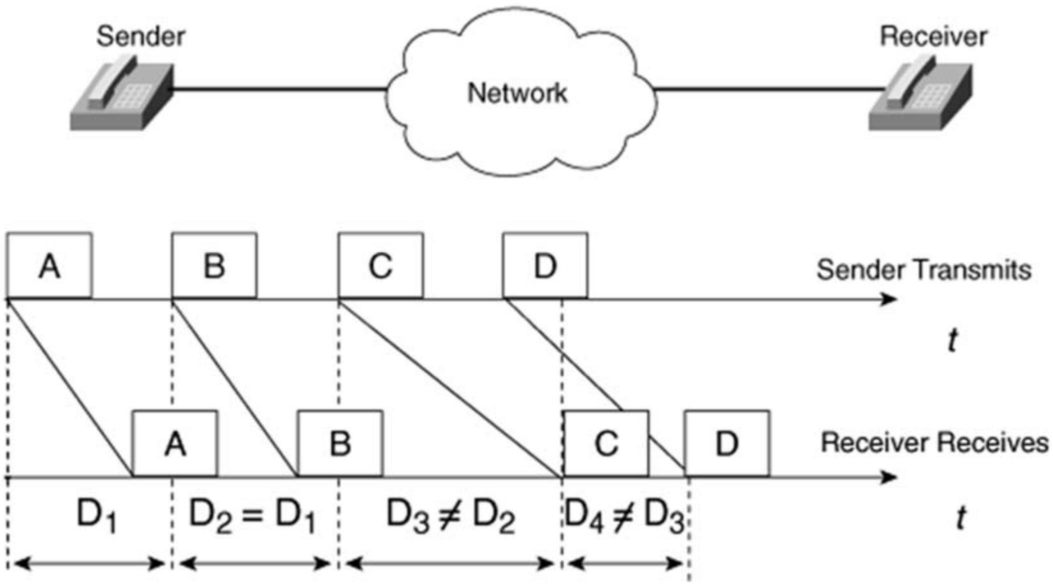
where d is the distance for which delay is calculated, c is the speed of light and 0,56 is the coefficient approximating the speed of electrons in a copper-wire. Propagation delay is unavoidable, but has a significant effect only with longer distances.[10]

Link delay is caused in the interface that controls the medium access. Link delay consists of the time that it takes to forward the data to the interface and the time that the packet has to wait in queue before transmission. The latter is also known as queue delay. Queue delay is usually a result of a situation, where the amount of forwarded packets exceed the processing power of the interface which means that the delay is inversely proportional to the links processing speed, i.e. the faster the interface, the lower the delay.[9] External factors like high traffic at the link can increase queue delay and cause network congestion, a situation that results in low network throughput.

Throughput is measurement that indicates the rate at which a communication channel is able to succesfully deliver messages. It is synonymous with digital bandwidth consumption as the momentary usage of the channel (bit/s) is compared to the maximum throughput of the network node in question.[25]

4.2 Jitter

Jitter, also known as packet delay variation or PDV, is the measure of time between the expected moment of arrival and the actual arrival time of a packet. For example, if packets are forwarded to the network with a constant rate of 20 ms, but the 2nd packet arrives to its destination 25 ms after the first packet, the network has caused 5ms of jitter to the transmission of the 2nd packet. Visualized in Picture 7.

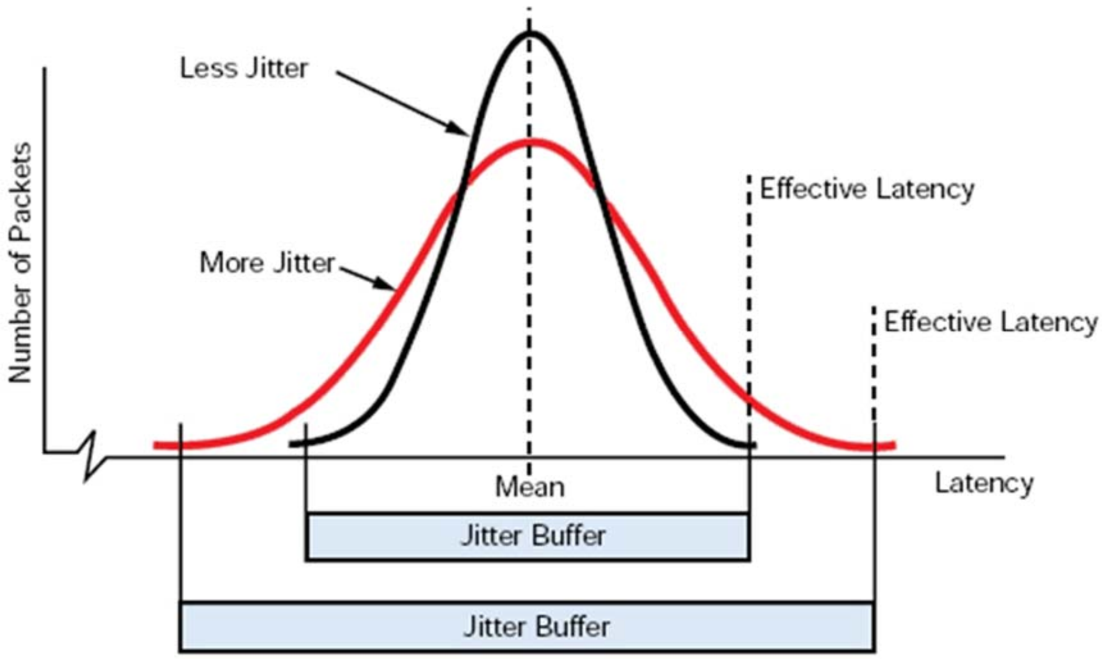


Picture 7. Jitter accumulation example.

Jitter is a packet-switched network -specific problem and it's mainly caused by queuing variations that are the result of dynamic changes in the network traffic loads.[10]

Some packets may also traverse a different path to the destination, through more or less 'hops', making the physical length of the path vary between packets. A different route can significantly deviate the packet from the expected arrival time. When sending a stream of packets in a complex network, there are no guarantees for each packet to travel the same route.[10]

A jitter-buffer is the main tool in minimizing the effect that jitter may have on a voice call. The buffer eliminates the jitter, by stacking the incoming packets into a queue and forwarding them again with the right frequency. The buffer size has to be scaled according to the variation of jitter in the network. A larger buffer can decrease the amount of packet loss, but causes more latency to the forwarded signal, whereas a smaller buffer size can cause more packet loss but decreases the latency. This scenario is visualized in Picture 8.



Picture 8. Buffer size effect on packet loss.[11]

As can be seen in from Picture 8., a compromise has to be made based on the latency fluctuations of the network. The network designer must carefully study the networks jitter and accordingly scale the jitter buffer so that the overall end-to-end latency is not exceeded.

4.3 Packet Loss

Packet loss is a situation where a packet of data fails to reach its destination. Losing packets in a network is both common and expected and can happen for many reasons. Underestimating the networks need for bandwidth may cause the queue buffers of routers and switches to overflow in high-traffic, channel congestion situations. In the event of an overflow, the network device will have to discard packets in order to maintain network operability.[10] Packet loss may also be caused by signal degradation where interference, physical impairments or errors made at network installation cause the transmission medium to be inoperative.

Many protocols use the information of packet loss to assess the networks current 'condition' and decide whether to increase or decrease the sending rate of packets. Protocols such as TCP have packet retransmission capabilities, making the applications that use these kinds of protocols more tolerant against packet loss.[10]

Real-time applications on the other hand are commonly based on the UDP protocol which makes them less tolerant to packet loss, as UDP lacks the retransmission capabilities. And even with retransmission, the RTP session would discard every packet that has arrived too late. This is why packet loss should be avoided when possible, to prevent voice quality or service disruptions from occurring.[10]

5 DATA PRIORITIZATION

The first step to make sure that certain types of data have the best possible basis for traversing through the network, with a minimal amount of disruptions, is to use data prioritization.

QoS or Quality of Service is a way of providing a consistent and predictable data delivery system by separating and prioritizing the data traffic of a network [14]. In a wireless network that is designed for asynchronous data transmission (i.e. not dedicated to one application), QoS is an important tool, given that some of the applications using the network are considered more important than others. This is especially important when considering the real-time aspect. As mentioned before, keeping the delays of real-time traffic on a minimum is crucial, thus implementing a high-priority QoS for the application in use can help guarantee a good service for voice traffic, without significantly limiting other usage of the network.

QoS takes factors such as delay, bandwidth and packet loss into account. Depending on the data characteristics and prioritization, networking devices manage the routing of individual data-flows, bandwidth resources, admission control, packet forwarding mechanisms and policy control.[14]

5.1 Integrated Services

Integrated Services (IntServ) is described as a fine-grained, flow-based QoS framework that can guarantee an uninterrupted path for a stream of data. By using a signaling system, the network protocol RSVP (Resource Reservation Protocol) used with IntServ, reserves needed resources from a network before data transmission. When an application uses the IntServ -service, it adds flow (RSVP session) specifications, parameters that describe the data flow and requested QoS -level, to the packet it is about to forward. In this case a data flow means a stream of packets that has the same source and destination address and port numbers [14]. The RSVP protocol then reads the flow ‘specs’ of each packet, in order to decide how much network resources it has to reserve for the flow and signals the details to the networking devices.

Depending on the given criteria for data transmission, an IntServ -architecture can provide 3 different reservation services:

- Guaranteed — Delays are kept on a desired amount. No packet loss.
- Controlled Load — Occasional glitches. Delays and packet loss rate kept constant.
- (Best effort — no reservation, normal networking procedure.)

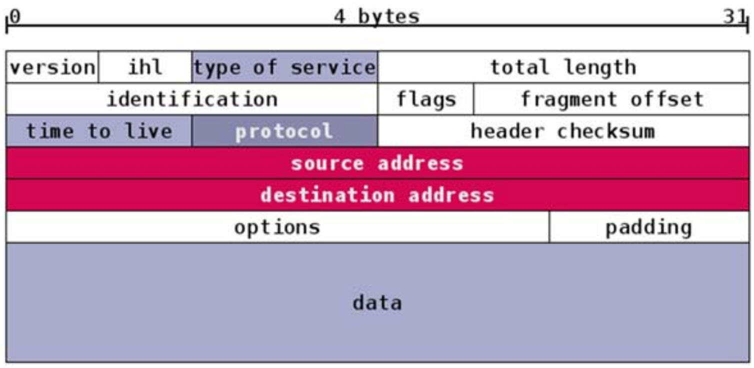
The Guaranteed -service gives the data flow the best possible circumstances to be transmitted without interruption. A best effort situation on the other hand is not particularly a service provided by RSVP but rather a weak-link on the network, as it happens only when the flow traverses through a non-RSVP router.[14]

The problem with IntServ is that it requires a lot from the networking devices. Every router and end device on the path of the packet has to be compatible to handle RSVP and be able to send QoS signals forward. Also, the reservations made in every device are 'soft' which means that they have to be periodically refreshed in order to keep them from timing out. In a smaller scale setup this is not a problem, but on a bigger scale the reservation signaling traffic on a single router builds up and adds more complexity to the routers tasks. When a routers processing limit is met, the traffic that goes through it will eventually slow down.[14]

The ‘pre-determined’ route that is made with the RSVP protocol, is also the main factor that makes IntServ less suitable for a WLAN environment. For example, everytime a mobile client station attaches to another AP, the RSVP route has to be updated, which results in unnecessary latencies. [15]

5.2 Differentiated Services

A DiffServ architecture is based on dividing a network to smaller scale parts called DS -domains. A DS -domain usually covers one or two networks that are run under the same administration, like a company intranet. A domain consists of DS -nodes which all work under a common service provisioning policy. A DS -node implements prioritization to data by dividing it to PHB or per hop behavior -groups. Every packet that enters a DiffServ network is marked with a DSCP (or Diffserv Code Point). The DSCP is written to the 8bit Differentiated Services Field (Type of Service field in Picture 9.) in the packets IP -header and it indicates the PHB -group the data is assigned to. As the packet traverses through the network, every node in its path reads the DSCP info, and depending on the PHB, applies the appointed forwarding measures.[14]



Picture 9. ToS fields location in the IPv4 packet.[16]

As the DSCP is a 6-bit value, in theory a network could have 64 different PHB -groups. However in practice the following PHBs are mostly used:[17]

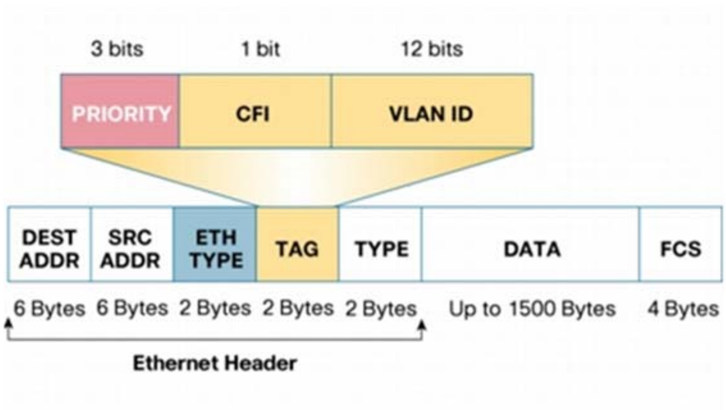
- Default PHB — best-effort traffic.
- Expedited Forwarding (EF) — low-loss, low-latency traffic.
- Assured Forwarding (AF) — assurance of delivery under prescribed conditions.
- Class Selector PHBs — guarantees backward compatibility with the 3bit IP Precedence classifications field (similar to the PCP field) that was used prior to DiffServ.

From these, PHB EF is the best Diffserv method in providing an assured bandwidth with low-latency, low-loss and low-jitter properties.

When considering the effects to STA mobility, DiffServ can be considered as the better alternative for deploying QoS measures in a WLAN. The only requirement from the DiffServ framework is that the packet is marked with the DSCP parameters, so that it can be forwarded accordingly. So in case of a handover from AP to another, the service should not be affected, if the route providided by the new AP works within the same PHB rules as the previous AP.

5.3 Class of Service

As every packet that traverses through a network does not necessarily consist of am IP –packet, CoS is used to implement packet differentiation and traffic prioritization at the data link layer. The priority values are assigned to the 3bit long PCP (Priority Code Point) that is located in the VLAN tag of the Ethernet header as seen in Picture 10.



Picture 10. PCP fields location in Ethernet packet header.[23]

With CoS, data can be treated differently by the network disciplines (routers, switches) based on the packets importance. By using the 3 bit PCP field, Ethernet packets can be categorized into 8 different priority levels listed in Table 3.

Table 3. PCP field priority levels.[18]

Priority	PCP	Acronym	Traffic Type
0	1	BK	Background
1	0	BE	Best Effort
2	2	EE	Excellent Effort
3	3	CA	Critical Applications
4	4	VI	Video
5	5	VO	Voice
6	6	IC	Internetwork Control
7	7	NC	Network Control

From Table 3. it can be seen that both video and voice have higher priority levels compared to other types data.

6 CARRIER ACCESS

In a WLAN the main types of data carriers are the fiber or copper Ethernet cables and the RF waves that propagate between Wi-Fi devices. The usage of these carriers is limited, which is why accessing them has to be controlled with access methods. By using various algorhitms, these methods try to distribute the usage of the carrier evenly, while trying to maximize the efficiency. However, in some situations the functionalities of these basic methods have been considered inadequate, which is why alternative approaches have been developed.

6.1 CSMA/CD

Carrier Sense Multiple Access with Collision Detect (CSMA/CD) is an access method that defines the rules of how devices on an Ethernet network use and access the transmission medium/carrier, for example a twisted-pair CAT cable.

CSMA is a 'listen-before-talk' transmission method which means that before transmission can happen, the device has to monitor the transfer medium for an opening. At the moment there are three different algorithms that are used to enable this access method:[19]

1. 1-persistent. The node that is ready to transmit continuously monitors the medium, and immediately starts sending data when the medium is sensed to be idle.

2. Non-persistent. If the medium is sensed busy, the node that is ready to transmit counts down a random waiting time before checking whether or not the channel is idle. If after this time the channel is sensed idle, data can be sent. Otherwise the process of counting down is repeated.

3. p-persistent. If the medium is idle, the node starts transmitting. Otherwise the medium is continuously monitored until it becomes idle again, after which the transmission starts with a probability p.

When multiple devices are connected to the same medium, simultaneous data transmission may occur and result in a data collision. As the collided data will never reach its destination, the packet has to be sent again. Instead of sending it right away, the CD or collision detection method applies a random waiting time between the receiving of collision information and the re-transmission process. This lowers the probability of the same collision from happening again, as both participants of the collision apply their own randomized time-delay.[19]

When the amount of devices on a single carrier grows, the probability of data collision rises. This eventually results in higher network latencies and delays that slow down the flow of data and decrease the quality of real-time communication.

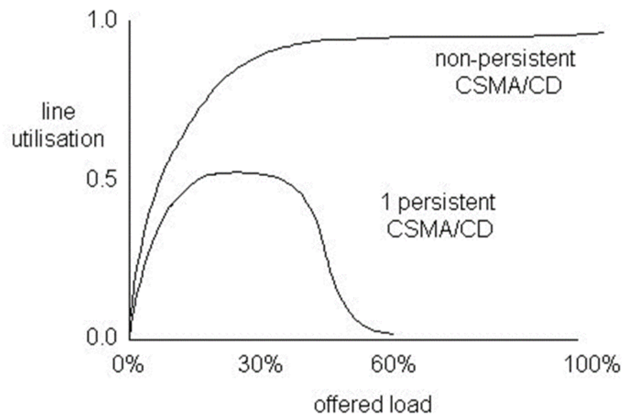


Figure 2. Traffic load effect on CSMA throughput.[20]

This situation is documented in Figure 2., where the effects of traffic (offered load) on throughput (liune utilisation) are studied. According to the graph, Non-persistent CSMA/CD performs better than the 1-persistent, but after an approximate 30% traffic load the channel becomes fully utilized which results in higher latencies for the participants.

6.2 CSMA/CA

CSMA/CA, Carrier Sense Multiple Access with Collision Avoidance, is the medium access sharing mechanism used in IEEE 802.11. Just like with CSMA/CD the transmitting node (STA/AP) first listens to the channel whether or not it is free for transmission. The node waits for a period of time called IFS (inter frame space) for the channel to be idle before transmitting. If the channel is not idle after IFS, the node waits for a random back-off time before checking for channel availability again [21]. This process is repeated until the channel has become available, after which the packet can be sent. The receiving end sends an ACK back to the transmitter to indicate that the packet has been successfully delivered. If no ACK arrives, a collision is assumed to have happened and the re-transmission of the lost packet takes place. The back-off -factor minimizes the risk of two separate stations noticing an opening and sending their packets at the same time, thus reducing the occurrence of collisions.

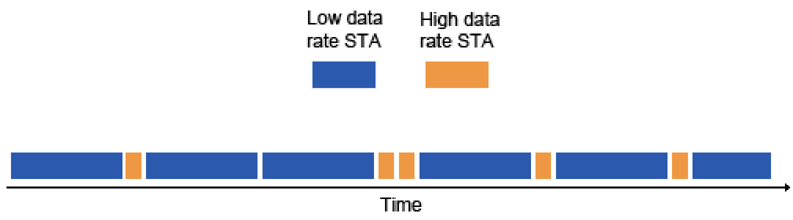
The main difference between 802.3 CSMA/CD and 802.11 CSMA/CA is that the 802.3 transceiver can listen to the medium and transmit at the same time. There are two main factors that explain why a modern 802.11 transceiver is unable to do this. First off, making a RF transceiver that could listen and transmit at the same time would be expensive. A rise in the prices of portable devices would be likely. The second reason is also known as the hidden node problem; even if the stations could listen and send at the same time, the distance between the two stations could be so long that the stations would be unable to hear each other's signals. The space between the stations could also be blocked, resulting in a similar situation. This is why the 802.11 uses collision avoidance algorithms, rather than letting stations try to detect collisions.[22]

The 802.11 standard also supports a RTS/CTS –protocol (Request to Send – Clear To Send) where the transmission of data is negotiated between the stations on a DSS. This protocol will guarantee access to the medium but doesn't improve the networks efficiency. RTS/CTS is mostly meant to be used when sending large data frames and in situations where the probability of collisions, caused by silmultanous data traffic, is high.[22]

6.3 Airtime Fairness

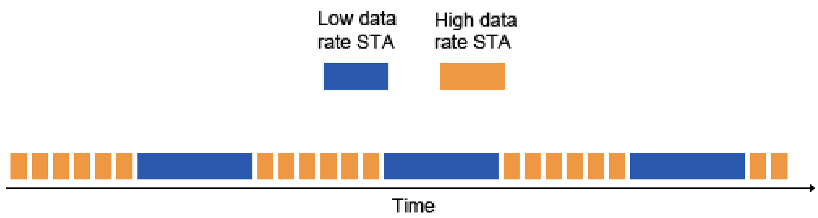
In a common WLAN usage -scenario, multiple stations will try to access the wireless carrier and send data forward through one Access Point. With CSMA/CA, when considering the distribution of the carrier usage, the ideal situation would be that each device connected to this AP would get ‘equal’ treatment. However, this is rarely the case.

With CSMA/CA and its random -based functionality, there is nothing to prevent one client from reserving the carrier multiple times in a row. Also, depending on the manufacturer, the size and the age of the device, laptops, tablets and smartphones, which are connected to the same AP, might all have different data rates depending on the properties of their Wireless Network Interface Cards or WNICs. Combined with the fact that only one of these devices can send or receive data at a time results in an unfair situation where low data rate devices take disproportionate amounts of airtime when compared to the devices using high data rates [24]. For example you may have an 802.11n AP that supports client stations operating at 300 Mbps and below, and within the cell of this AP there are two kinds of STAs: faster ones with a 300 Mbps data rate and slower ones with a 10 Mbps data rate. Assuming that CSMA/CA allocates both devices with an equal amount of access to the medium and that with each window of access the same data amount is sent, the example results in a situation where both devices achieve the same throughput but the device operating at 10 Mbps takes up 30x more airtime than the faster one.[24] In other words, the faster device doesn't benefit from having superior hardware as the queue delay caused by the slower device eliminates the advantage of a higher data rate. This situation is visualized in Image 1., where the airtime is divided into data -slots, making the low data rate station dominate the airtime over the high data rate STA. To address this problem, network vendors have started implementing a solution that works as an extension to CSMA/CA labeled Airtime Fairness.



Picture 11. Visualization of airtime distribution between two STAs using only CSMA/CA

With Airtime Fairness, deciding on the basis of either the STAs physical layer properties or data rate, APs use additional scheduling and queuing to balance the airtime usage between the stations. Since the high data rate STAs use the network more efficiently, they are given more opportunities to transmit [24]. On a basic level, each station is assigned to a fixed length timeslot or airtime, as seen in Image 2., on which they can transmit as much data as they are capable of.



Picture 12. Visualization of airtime distribution between two STAs using Airtime Fairness

Although this method doesn't eliminate the link delay that STAs face while queuing for their slot - in fact, slower devices might suffer from a slight increase - it compensates the faster device with either more airtime or less queue delay. It also makes the delays more predictable for every STA and lowers the possibility of slower legacy devices (e.g. 802.11 b/g stations) from having a crippling effect on the faster devices.

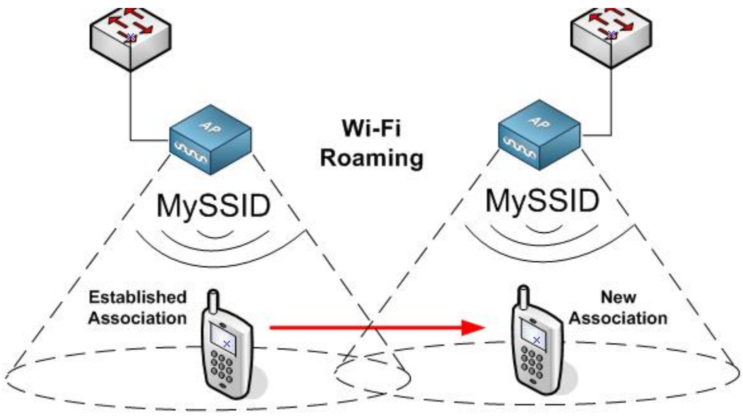
It is important to note, that the term ‘Airtime Fairness’ is by itself loosely defined and various techniques and algorithms used in its deployment may vary significantly from vendor to another.

For example, Meru Networks has standardized a time-based fairness algorithm (similar to the one explained above), which uses a token-ring type of queuing system. Each client is assigned with a fixed slot, which the AP goes through in a logical order, making the carrier access more equal and predictable[26]

7 HANDOVER DELAY

Unlike with channel access methods and data prioritization, the choices made by choosing the right channel architecture/layout can have a beneficial effect on a problem that is specific to wireless communication: handover delay.

Handover or HO is the action that takes place when a STA transfers the responsibility of handling its data transmission to another AP. It is called a mobility management process that, in a well designed network, allows mobile STAs to move around the whole ESS (from cell to another) without losing connection.[14]



Picture 13. Handover visualization.

The decision that leads to a handover is calculated by employing several algorithms that take into account factors such as received signal strength (RSS) or received power P.[14] Less common algorithms can also monitor the carrier-to-interference ratio (CIR), bit-error rate (BER) and block error rate (BLER).[14] The APs send beacon messages at regular intervals from which the STAs measure these values. From these receive condition values, a running average is calculated. This results in value known as CQ or Communication Quality. Based on the value of CQ, the STA decides whether or not it needs to expend more effort into finding a new AP with a better CQ.

Handovers can be divided into hard handovers and soft handovers. When a handover is categorized as hard, the transceiving of data between the STA and the destination AP starts right after the STA has closed the connection with the source AP. Although the transition is designed to be instantaneous, this might result in a short break in service. With a soft handover this break is prevented by having both the source and the destination AP transmit the same data in parallel. The data transmission by the source AP is stopped only after the new connection is made which also results in closing the old connection.[14]

Whether one is better than the other can be decided by comparing them to the characteristics of a good handover:

1. It's fast: The mobile STA starts receiving packets at its new location with a minimal packet delay.
2. It's seamless: Packet loss rate caused by handover should be zero or near zero.
3. Minimal signaling traffic: Keeping control data load at a required minimum.

Although less efficient processing-wise with more signaling traffic than with a hard handover, a soft handover has a major advantage when the emphasis is on minimizing the delays and packet loss of a network. The fact that a soft handover is less likely to cause a break in service makes it a more beneficial approach when considering real-time data transmission.

7.1 Delay characteristics

The factors that increase delay in a handover are detection of the need for handover, active or passive scan, re-authentication and re-association.[14]

With a multichannel deployment the main delay is caused at step 3, seen in Table 4., where the data transmission is paused for the time of an active scan. Both the probe request being sent by the STA and the probe responses sent by APs in range of the scan, are highly dependent on the amount of STAs using the same channel. As mentioned before, more traffic means longer access queues, which in return results to increased delays to actual data transmission.[14]

Table 4. Multi-Channel handover procedure.[14]

Step	Multi Channel
1	Beacon transmitted by APs
2	STA analyzes CQ from beacons
3	When CQ < Star Cell Search Threshold ► 1x active scan. Incoming data traffic stopped and buffered in AP.
4	If AP with better CQ is found ► start cell switch procedure go to 7.
5	If not ► Start emptying AP data buffer. ► Repeat 3. on a regular basis.
6	If CQ < Fast Cell Search (FCS) ► 1x active scan. Incoming data traffic stopped and buffered in AP. Initiate cell switch procedure when AP with CQ > FCS found.
7	Reassociate to new AP, disassociate from old AP. Update path for data in old AP.

With a single channel, there is no delay caused by active or passive scanning. The STA will switch to another AP as soon as it finds one that is within the handover range, and has a better CQ.[14]

7.2 Delay variations between different mobile devices

When the STA is in charge of the handover, the handover delay is also affected by the data processing capabilities of the STA. These capabilities are basically dependent on the hardware and software setups that naturally vary between different manufacturers devices.

To visualize this effect, a measurement was made in the laboratory of Kantio Oy, which compared three different mobile devices and the duration of their handover mechanisms.

In this measurement a test environment was made that included one Aruba Controller 620 and two Aruba 105 access points. The Aps were connected to the controller and mounted to opposite walls at the office. Both Aps used the 2,4 Ghz frequency band, from which channel 11 was given to the first (A) and channel 1 to the second (B) AP. The measurement data produced by the handover event was captured with the Riverbed SteelCentral Packet Analyzer and Wireshark.

The measurement was done by having each device make the same test call with Skype. The call was made from the device to the AP-A operating at channel 11. First, the device was located near AP-A. After the call was established, the device was manually moved next to the AP-B operating at channel 1, which resulted in a handover from A to B. During the test, all other data traffic was blocked.

Table 5. Handover delay measurement results.

Device	Model	HO Delay
Samsung Galaxy Xcover	GT-S5690	285,9 ms
Samsung Galaxy Tab	GT-N8010	29,7 ms
Apple iPad, 3rd generation, iOS 8.1	MD3238KS/A	4350 ms

It should be noted, that each device was tested only once, which means that the actual mean delay value might be lower than what these results indicate. The big variation of handover delays between the different devices can still be observed from the results shown in Table 5. Considering that the maximum end-to-end delay for a good quality voice call was set to 200 ms, both the Xcover phone and the iPad exceed this with only one handover.

A more detailed graphical representation of the iPad handover scenario can be seen in Appendix 1.

8 CONCLUSIONS

This thesis tried to give an answer to the question of how and by which means can a wireless local area network adjust to the needs of sensitive, real-time data, without drastically limiting the quality perceived by the end-user.

By studying the specifications of a good quality voice call the main network afflicted causes for user-end quality degradation were narrowed down to latency, jitter and packet loss.

The main tool for decreasing the effect of these three was found to be the prioritization of data. For example, by deploying a DiffServ -based QoS framework with the desired settings and by marking the sensitive data with the PHB EF -group DSCP, the latencies caused by link and queue delay can be significantly reduced. DiffServ was found to be the better framework in a WLAN, as its effectiveness doesn't rely on a pre-determined route, which in the case of IntServ causes unnecessary handover delays.

In a WLAN with a diverse set of stations, a time-based version of the CSMA/CA extension Airtime Fairness was also seen as a viable option for improving the performance of the entire network. Switching from a randomized access model to a time based model, can make the accessing of the wireless carrier more predictable and remove the crippling effect that slower devices might otherwise have towards faster ones.

Lastly, deploying a single channel cell-architecture within an ESS was considered as the best way to minimize the effect of handover delay in real-time transmission. By removing the handover decision making from the STA and reducing the need for passive or active scanning, a single channel architecture can decrease the amount of handover delay, thus improving the overall service for mobile client stations.

For the author, most of the topics discussed above were relatively unknown beforehand, which meant that conveying them in this thesis required a lot of research. The vast amount of material available was seen as an asset but in some points also increased the workload. Deciding between different sources and assessing their viability, while also keeping the content of the thesis within a graspable framework was considered as the hardest part in making this thesis.

Based on the mostly theoretical approaches this thesis had towards WLAN networking and real-time transmission, future studies could concentrate more on measurements and practical implementations. For example, the quickly ventured topic of measuring mobile devices and their effect on handover delay could be reproduced with more repetition and with a larger set of mobile devices.

REFERENCES

[1] Kos, A. Klepec, B. Tomazic, S., "Real-time application performance in differentiated services network," *TENCON 2001. Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology,* (7161447), 2001.

[2] Cisco Systems, Inc., "Cisco Active Network Abstraction Reference Guide, 3.7," [online]. Available: http://www.cisco.com/c/en/us/td/docs/net_mgmt/active_network_abstraction/3-7/reference /guide/ANARefGuide37/ethrnt.html (Accessed: 9.4.2014)

[3] Saarelainen, Kari, *IP-viestintäjärjestelmät.* Helsinki: Readme.fi, 2011.

[4] IEEE, "Quick Guide to 802.11 Working Groups" [online]. Available: http://ieee802.org/11/QuickGuide_IEEE_802_WG_and_Activities.htm (Accessed: 2.11.2014)

[5] Webopedia, "What is Wi-Fi" [online]. Available: http://webopedia.com/TERM/W/Wi-Fi.html sivu (Accessed: 4.11.2014)

[6] Shridhar, T., Flextronics, "Wireless LAN Switches — Functions and Deployment," [online]. Available: http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_9-3/wireless_lan_switches.html (Accessed 2.5.2014)

[7] National Instruments, "Wireless Data Acquisition. Range vs. Throughput" [online]. Available: http://www.ni.com/white-paper/8566/en/ (Accessed: 20.10.2014)

[8] Aruba Networks, Inc., "WLAN RF Architecture Primer: Single-Channel and Adaptive Multi-Channel Models" *WLAN RF Architecture Primer,* 2013.

[9] Davidson, Jonathan Peters, James, *Voice over IP*. Helsinki: Edita Publishing Oy, 2002

[10] Kantio Oy, "Voice call service considerations", *Study report, 2*013.

[11] Legg, Gary, Embedded, "VoIP Meets Wi-Fi: Internet Phone Calls Go Wireless" [online]. Available: http://www.embedded.com/design/communications-design/4017837/VoIP-Meets-Wi-Fi-Internet-Phone-Calls-Go-Wireless

[12] ITU-T, "E-model tutorial," [online]. Available: https://www.itu.int/ITU-T/studygroups/com12/emodelv1/tut.htm (Accessed 8.5.2014)

[13] *G.114, One-way transmission time.* ITU-T, Geneva, Switzerland, 2003.

[14] Prasad, Neeli Prasad, Anand, *802.11 WLANs and IP Networking: Security, QoS, and Mobility*. Norwood, MA, USA: Artech House, 2005.

[15] Salkintzis, A., Passas, N. *Emerging Wireless Multimedia: Services and Technologies.* West Sussex, England: John Wiley & Sons 2005

[16] IEEE, ”RFC 791” [online] Available: http://www.faqs.org/rfcs/rfc791.html (Accessed: 21.10.2014)

[17] Wikipedia, "Differentiated services" [online]. Available: http://en.wikipedia.org/wiki/Differentiated_services (Accessed: 17.4.2014)

[18] Wikipedia, ”IEEE 802.1p” [online]. Available: http://en.wikipedia.org/wiki/IEEE_P802.1p (Accessed: 25.10.2014)

[19] Shinde, S.S., *Computer Network*. Daryaganj, Delhi, India: New Age International, 2009.

[20] Bramer, Brian, “Local Area Networks” [online]. Available: http://www.cse.dmu.ac.uk/~bb/Teaching/NetWorks/LANS/LANS.htm (Accessed 15.11.2014)

[21] Ho, Tien-Shin Chen, Kwang-Cheng, "Performance Analysis of IEEE 802.11 CSMA/CA Medium Access Control Protocol," *Personal, Indoor and Mobile Radio Communications, 1996. PIMRC'96., Seventh IEEE International Symposium*, Volume 2, 407-411, 1996.

[22] Goransson, Paul Greenlaw, Raymond, *Secure Roaming in 802.11 Networks*. Burlington, MA, USA: Newnes, 2007.

[23] Cisco, ”Quality of Service on Cisco Catalyst 6500” [online]. Available: http://www.cisco.com/c/en/us/products/collateral/switches/catalyst-6500-series-switches/prod_white_paper0900aecd803e5269.html (Accessed: 19.10.2014)

[24] Burton, Marcus, Head, Tom Jackman, Shawn M., *CWDP : Certified Wireless Design Professional Official Study Guide.* Hoboken, NJ, USA: Sybex, 2011.

[25] Wikipedia, "Throughput" [online]. Available: http://en.wikipedia.org/wiki/Throughput (Accessed: 25.11.2014)

[26] Meru Networks Techspot, "Meru's Air Traffic Control: Advanced RF Architecture," [online video] Available: https://www.youtube.com/watch?v=Cco2yrTbBcs (Accessed 8.5.2014)

Handover delay measurements for iPad

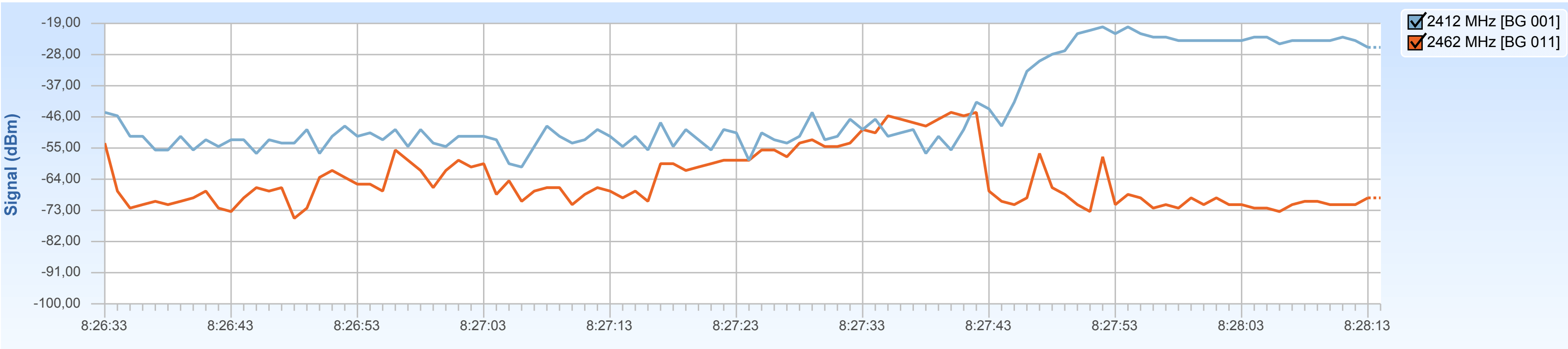


Figure 1. Indication of handover after decrease in communication quality.

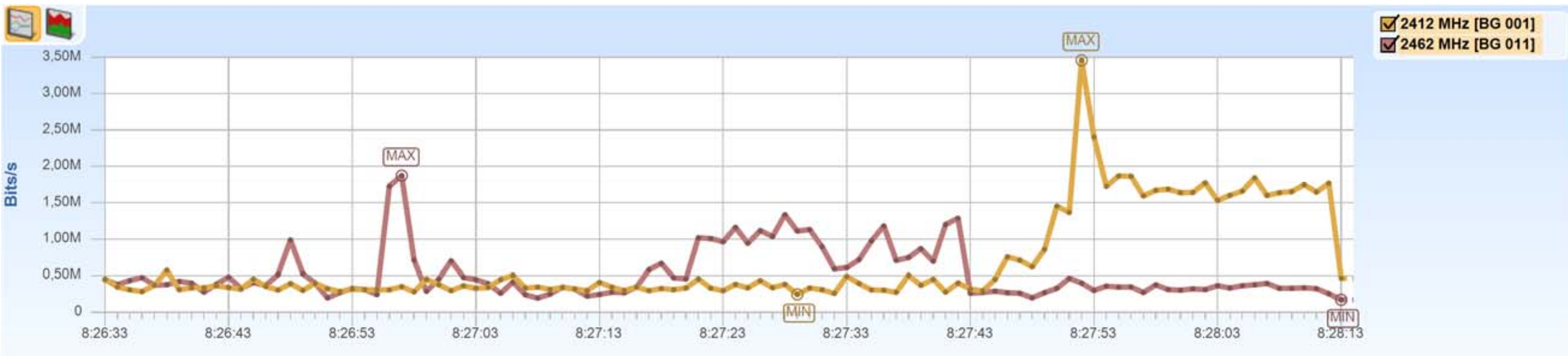


Figure 2. Unloading the data-buffer of the new channel causes a spike in traffic.

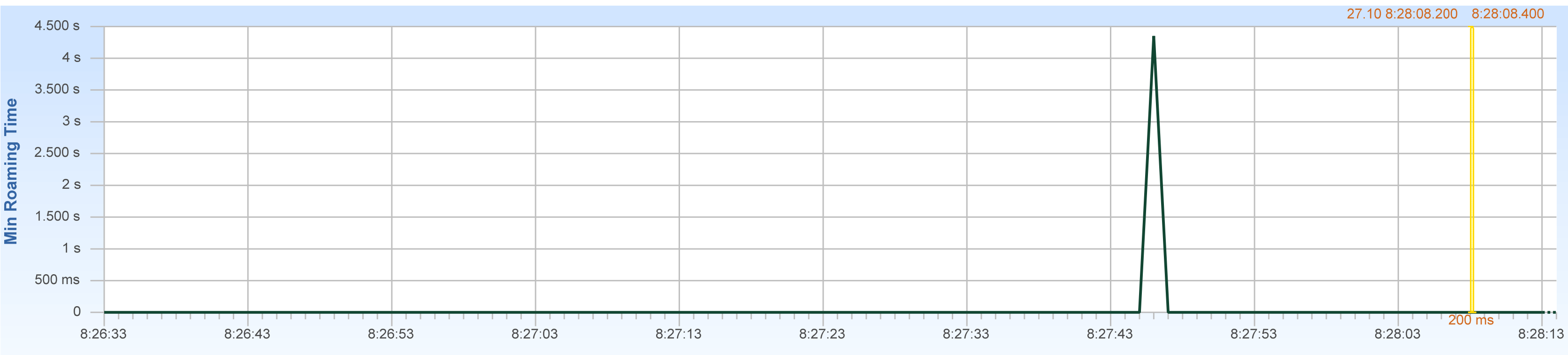


Figure 3. Spike indicates the handover moment and duration.

Roaming Station MAC

Hierarchy (Roaming Station MAC)	Roaming Time	ESSID	Start Channel	End Channel	Start AP BSSID	End AP BSSID	Start Time	End Time	Start Pack
Roaming Station MAC: 98:fe:94:bd:7a:ce	4.350 s	[1]	[1]	[1]	[1]	[1]	27.10.2014 8:27:42	27.10.2014 8:27:47	
	4.350 s	demo-emplo	2462 MHz [BG 011]	2412 MHz [BG 001]	d8:c7:c8:6d:4b:51	d8:c7:c8:6d:4a:f1	27.10.2014 8:27:42	27.10.2014 8:27:47	
	4.350 s	[1]	[1]	[1]	[1]	[1]	27.10.2014 8:27:42	27.10.2014 8:27:47	[1]

Figure 4. Total time used by the iPad for the handover sequence