Characterisation of Internet Traffic in Wireless Networks

Lei Liang, Yang Chen*, and Zhili Sun

Centre for Communication System Research (CCSR),
University of Surrey, Guildford, Surrey, United Kingdom
`{L.Liang,Z.Sun}@Surrey.ac.uk`

Abstract. Wireless connection technologies provide users (Internet Protocol) IP network access without the physical hardware connection of the wired networks. One of the applications of these technologies is the Wireless Local Area Network (WLAN), which is based on the IEEE802.11x Wireless Fidelity (WiFi) standard and is widely deployed as a flexible extension to data network or an alternative for the wired Local Area Network (LAN). In this context, the design, control and performance analysis of future wireless networks requires the study and credible characterisation of WLAN traffic. This tutorial presents measurements and analytic studies of IP traffic in a WLAN environment. Moreover, an investigation is reported into the characterisation on protocol distribution and modelling of IP packet inter-arrival times.

Keywords: WLAN, IP traffic measurement, traffic modelling and packet inter-arrival time.

1 Introduction

Traffic measurement is crucial to traffic engineering (TE) functions [1]. It provides insight of network traffic, network operation state and problem anticipation. It is also crucial for optimising the network resources to meet the requirements of traffic condition and quality of service (QoS) requirements. It can also provide the feedback data for the engineer to adaptively optimise network performance in response to events and stimuli originating within and outside the network. It is essential to determine the QoS in the network and to evaluate the effectiveness of traffic engineering policies. And experience indicates that measurement is most effective when acquired and applied systematically.

Measurement in support of the TE functions can occur at different levels of abstraction. For example, measurement can be used to derive packet level characteristics, flow level characteristics, user or customer level characteristics, traffic aggregate characteristics, component level characteristics, and network wide characteristics [1]. The measurement presented in this paper has been carried out to study the packet level characteristics of aggregate WLAN traffic.

---

* Yang Chen has left the University of Surrey.

D. Kouvatsos (Ed.): Next Generation Internet, LNCS 5233, pp. 191–202, 2011.

The history of wireless networking can stretch back over fifty years ago, during World War II, when the United States Army first used radio signals for data transmission. They developed a radio data transmission technology, which was heavily encrypted. A group of researchers in 1971 at the University of Hawaii created the first packet based radio communications network named ALOHNET with inspiration from that army system. It is essentially the very first WLAN.

In the last few years, the wireless connection technology is improving, making it easier and cheaper from companies to set up WiFi access point for access to the Internet. That resulted in a rise of WLAN usage worldwide to set up hotspots, including university campuses, airports, hospitals, companies, and warehouses. The fast development of the WLAN brings new challenges to researchers. One of them is the need to understand the characteristics of the WLAN traffic. This paper presented a measurement experiment in WLAN using packet monitor software.

The paper is organized as following: the second section gives background knowledge of WiFi including the relevant standards, and the third section presents the measurement methodologies, parameters and environment, measurement results and analysis will be presented for traffic characterisation in the forth section, and finally the last section draws conclusions based on the measurement and analytical results.

2 Background of WiFi

WLANs are currently the most common form of wireless networking for day-to-day business operations. An industry standard has to be developed to enable the WLAN widely accepted and ensure the compatibility and reliability among all manufacturers of the devices. The first standard for WLAN IEEE 802.11 [2] was defined in 1997 by the Institute of Electrical and Electronics Engineers (IEEE) as a part of a family of IEEE 802.x standards for local and metropolitan area networks. It operates at a radio frequency (RF) band between 2.4GHz to 2.5GHz with data rates of 1Mbps and 2Mbps. It also provided a set of fundamental signalling methods and services. It addressed the difference between wireless LAN and wired LAN and provided the method to integrate both LANs together. There are new main amendments defined in IEEE following IEEE 802.11. They are IEEE 802.11a, IEEE 802.11b, IEEE 802.11d and IEEE 802.11g. All of them have commercial products in the market now.

The IEEE 802.11a [3], defined in September 1999, gives method to provide a WLAN with data payload communication capabilities up to 54 Mbps in a frequency band around 5 GHz. IEEE 802.11b [4] in 1999 and IEEE 802.11g [5] in 2003 provide methods to increase the data payload rate up to 11Mbps and 33 Mbps respectively in the 2.4GHz frequency band defined in IEEE 802.11. IEEE 802.11d [6] in 2001 provided specifications for conformant operation beyond the original six regulatory domains of IEEE 802.11 and enabled an IEEE 802.11 mobile station to roam between regulatory domains. 802.11n is a recent amendment which improves upon the previous 802.11 standards by adding multiple-input multiple-output (MIMO) and many other newer features. The maximum bit rate can reach 600 Mbps. Table 1 shows the list of IEEE 802.11 families.

Table 1. List of IEEE 802.11 families

Name	Time of Definition	Document Type
IEEE 802.11	1997	Original Standard
IEEE 802.11a	1999	Amendment
IEEE 802.11b	1999	Amendment
IEEE 802.11d	2001	Amendment
IEEE 802.11g	2003	Amendment
IEEE 802.11g	2009	Amendment

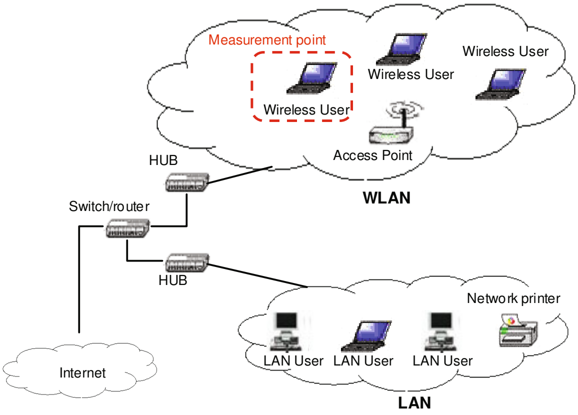

Fig. 1. Basic WLAN layout example

WLAN is normally deployed within a single building or a campus area. Fig. 1 illustrates the basic layout of a WLAN with connection to wired LAN, where several laptops are connected with the core wired network through an Access Point (AP), and the wired network has server, switch, hub, workstation, network printer, and so on. It can be seen from the Fig. 1 that the WLAN is an extension to the wired LAN.

A WLAN may consist of three network components. They are wireless network card, wireless access point and wireless bridge.

The wireless network cards include PCI cards for workstations and PC cards for laptops and other mobile devices. The card can work in an ad-hoc mode, as in client-to-client scenario (ad hoc), or in a pure client-to-AP mode (infrastructure).

The wireless network cards connect to an AP. An AP is essentially a hub that gives wireless clients the ability to attach to the wired LAN backbone. With the use of the cell structures, more than one AP can be set up in a given area. This is similar to the cell phone coverage in mobile communication systems. Wireless bridges can provide high speed longer range outdoor links between buildings.

The IEEE 802.11 standard gives two different ways to configure a WLAN using these network components: ad-hoc and infrastructure. In the ad-hoc WLAN, computers are brought together to form a network "on the fly." As shown in Fig. 2, there is no structure to the network; there are no fixed points; and usually every node is able to communicate with every other node. A good example of this is the aforementioned meeting where employees bring laptop computers together to communicate and share design or financial information. Although it seems that order would be difficult to maintain in this type of network, algorithms such as the spokesman election algorithm (SEA) have been designed to "elect" one machine as the base station (master) of the network with the rest being slaves. Another algorithm in ad-hoc network architectures uses a broadcast and flooding method to all other nodes to establish their identifications and connections.

The second type of network structure used in WLANs is the infrastructure as shown in Fig. 3. This architecture uses fixed network AP with which mobile nodes can communicate. These AP are sometime connected to landlines to widen the LAN's capability by bridging wireless nodes to other wired nodes. If service areas overlap, handovers can occur. This structure is very similar to the present day cellular networks around the world.

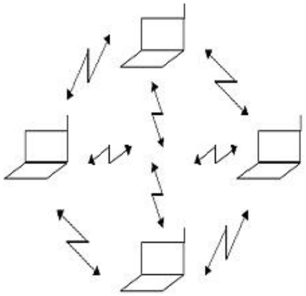

Fig. 2. Ad-hoc WLAN architecture

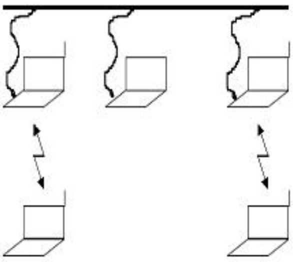

Fig. 3. Infrastructure WLAN architecture

3 Measurement Environment and Methodologies

A series of WLAN measurements have been carried out in an academic building in the University of Surrey. The network is a typical infrastructure WLAN having around 20 to 50 users in different time of a day. The users have laptop computers with wireless cards and using Microsoft Windows XP operation system. Fig. 1 shows the measurement environment. The laptops are connected to the wired core network through an AP. The Measurement Point (MP) is one of the laptops.

To measure WLAN traffic, one has to address an appropriate measurement methodology. Many measurement methodologies have been used for traffic measurement in the history, i.e. using LOG files and capturing packets form the networks using some software and hardware. The measurement methodologies can be divided into two main groups: passive approach and active approach. Both have their values and should be regarded as complementary, in fact they can be used in conjunction with one another. The MP can be used as a normal laptop while it measures the traffic situation of the whole WLAN. In principal, this is a passive measurement.

Then we have to decide what tools we can use to measure the WLAN traffic. There are so many traffic monitoring and measurement software in the market and on the Internet. Some of them are very powerful with hardware equipments. However, commercial software can be very expensive. Much open source software is distributed on the Internet free of charge. But most of them are programmed for common use or some other special purposes that don't exactly fit in a particular measurement of our interests. As a Windows version of the famous UNIX packet capture tool TCPDump, WinDump was chosen as the measurement tool in our measurement for its powerful functions and easy-handling output.

Measurement parameters are the other important factor to be considered. Some performance and QoS parameters are very important for traffic engineering such as delay, jitter and packet loss. The IETF IP Performance Metrics (IPPM) working group has carried out studies and defined metrics of these parameters. They are very useful to evaluate the performance of a network. There is a group of parameters that can reflect the network traffic status. These parameters at packet level include throughput, packet length, packet inter-arrival time, packet burstness and so on. They are useful for capacity management, queue management and traffic prediction. This paper is going only present the packet inter-arrival time, which is a key factor in our measurement and analysis. It can be calculated using the following formula:

$$IA = AT(i) - AT(i-1)` \qquad (1)$$

where IA stands for inter-arrival time, AT(i) and AT(i-1) are the arrival time of the i^{th} packet and its previous packet respectively.

During the measurement, the MP kept running WinDump software for thirty minutes as one measurement interval to capture WLAN packets from all the active users. Totally there were more than ten measurements taken at different time periods of a day. The IP packets were monitored and captured. These packets include different applications of WWW surfing, FTP, real-time online video and online games, etc.

4 Measurement Results and Analysis

Over ten sets of measurement results are stored for analysis. Results including packet length, protocol type, packet arrival time, number and volume of captured packets, and so on. All results are stored in pure text files. Every measurement interval of thirty minutes generated a text file with size ranged from 7Mb to 40Mb corresponding to the network usage at different times of a day. The number of packets captured during each measurement interval varies from 100,000 to 600,000.

4.1 Packet Type Analysis

Software was developed using Borland C++ Builder. It handled the raw data of the measurement output and generate summary in the Table 2.

Ten sets of data are shown in Table 2. Each of them presents results of one measurement interval. For each of the measurement interval, the number of the captured packets ranged from 93,744 to 598,819. This variance is mainly due to the difference

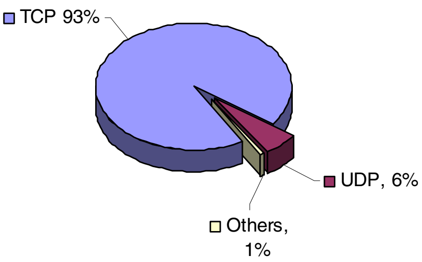

Fig. 4. Protocol percentage pie for WLAN traffic

Table 2. Summary of WLAN Measurement Output

Interval	No. of packets captured	No. of TCP packets	No. of UDP packets	Percentage of TCP packets (%)	Percentage of UDP packets (%)	TCP packets total length (Bytes)	UDP packets total length (Bytes)
1	164,280	154,549	6,888	94.08	4.19	76,400,952	579,987
2	598,819	593,464	2,514	99.11	0.4	354,190,737	623,560
3	547,050	541,934	2,636	99.06	0.5	329,917,033	307,035
4	468,306	463,013	2,861	98.87	0.7	252,739,983	290,285
5	238,616	234,931	1,589	98.46	0.7	127,104,315	164,625
6	93,744	89,455	1,897	95.42	2.02	60,215,251	184,117
7	520,495	465,212	51,526	89.38	9.9	271,855,736	9,759,117
8	576,802	522,682	50,407	90.62	8.74	350,224,666	9,719,882
9	546,570	427,324	110,400	78.18	20.2	225,362,090	21,303,531
10	205,778	199,625	3,441	97.01	1.07	79,533,200	351,498
Total	3,960,460	3,692,189	234,159	93.23	5.91	2,127,543,963	43,283,637

of the number of users in the WLAN and the applications running during each interval. Taking the interval 6 as an example, it was taken in a midnight when there were only less than ten users in the whole WLAN, which leading to the least number of packets captured among all measurement intervals.

Applications also affected the number of packets generated in each measurement interval. For instance, web surfing users generate much less packets than online video watching users because they did not generate packets continuously when they read web pages Totally, 3,960,460 packets were captured during these ten measurement intervals, which include 3,692,189 TCP packets and 234,159 UDP packets. That means TCP protocol is still counted as the majority used by Internet applications, i.e. around 93.23% of the captured packets are TCP packets as shown in Fig. 4.

The results showed that all of the online video services captured in the measurement used TCP to transport data rather using UDP. This is conflicting with the basic understanding that real-time applications are very sensitive to network latency and should use UDP as transport protocol to avoid retransmission and latency caused by congestion control function of TCP.

However, it is also understandable that online video service prefers TCP in the current best-effort Internet because it can provide reliable delivery with less packet loss and less distortion of the video services. It is also because online video service is relevantly immune to latency because it is not an interactive and real time service and, thus, network latency can be compensated by local playback buffer.

4.2 Packet Inter-arrival Time Analysis

The self-developed software also calculated the packet inter-arrival times for each measurement interval using formula (1). The objective was to find out what distribution the packet inter-arrival times of WLAN traffic follows. To establish the most suitable statistic distribution to model the measured curves, many distributions were tested by varying relevant parameters. These distributions include Chi-squared distribution, exponential distribution, inverse Gaussian (Wald) distribution, lognormal distribution, Pareto distribution and Rayleigh distribution [7]. Both Probability Density Functions (PDFs) and Cumulative Distribution Functions (CDFs) of these distributions were used to compare with the measurement plots of each of the traces. However, we found the CDF is a better way to present all of the theoretical curves and the measurement curves as well as being easier to use.

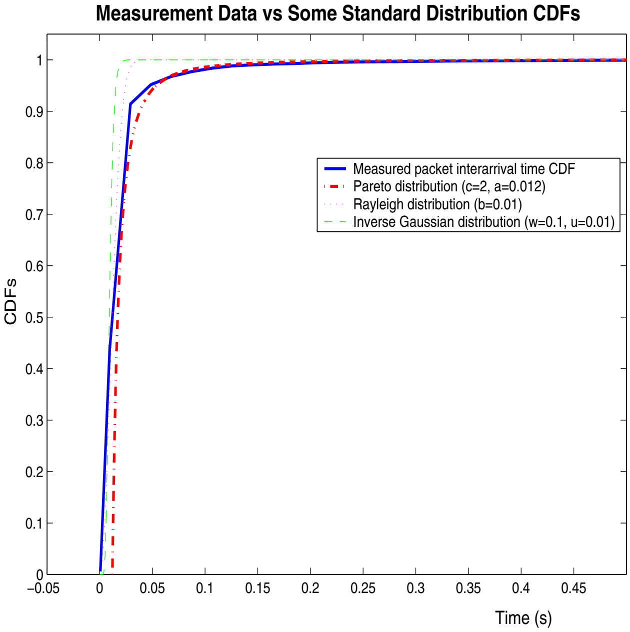

Fig. 5. Packet inter-arrival time CDF fitting

The fitting results show that we can easily tell which distribution is the best fit by the CDF of the data for the very clear fitting difference of all the experimented standard distributions. Fig. 5 shows the Inverse Gaussian (green dashed line), Rayleigh (magenta dotted line), Pareto (red dash-dot line) and one measured packet interarrival time (blue solid line) CDFs. It's clear that the different distributions have notably different CDF that can be distinguished by human eyes.

From all of these distributions, it was seen that the Pareto distribution is the best fit to nine measurement results. One measurement result can be fitted by using Inverse Gaussian distribution. Table 3 shows the fitting result.

Table 3. Packet Inter-arrival Time Fitting Results

Measurement	Max. Inter-arrival Time	Min. Inter-arrival Time	Best fit distribution
1	1.9402	5×10-6	Pareto
2	0.4004	5×10-6	Pareto
3	0.3642	3×10-6	Pareto
4	0.1736	3×10-6	Pareto
5	0.3925	5×10-6	Pareto
6	0.5594	5×10-6	Inverse Gaussian
7	0.1644	4×10-6	Pareto
8	0.2081	1.2×10-5	Pareto
9	0.3057	4×10-6	Pareto with cut-off
10	2.9733	3×10-6	Pareto

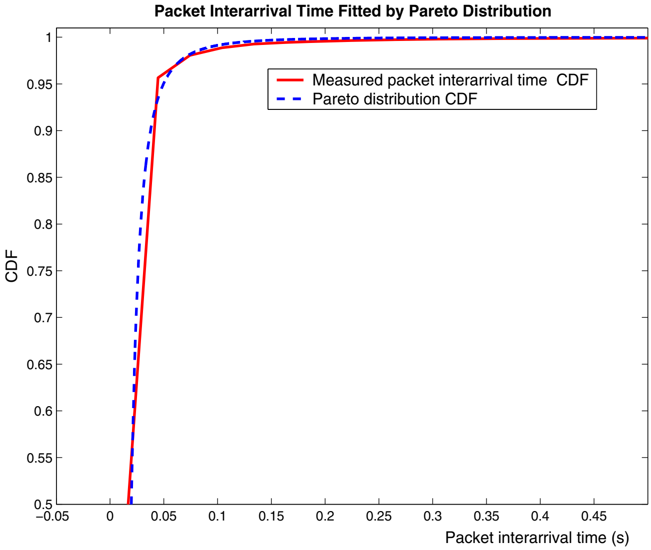

Fig. 6. Packet inter-arrival time fitted by Pareto distribution

For those measurements that can be fitted by Pareto distribution, Fig. 6 shows one of them as an example in this paper. The solid line is the measured packet inter-arrival time CDF and the dashed line is the CDF of a Pareto distribution.

The mathematic presentation of the Pareto distribution PDF is:

$$f_T(t) = \frac{ca^c}{t^{c+1}} \tag{2}$$

where $c > 0$ is the shape parameters of Pareto distribution and $a > 0$ is the location parameter. Fig. 7 shows the fitting of the measured packet inter-arrival time in measurement 6 using Inverse Gaussian distribution.

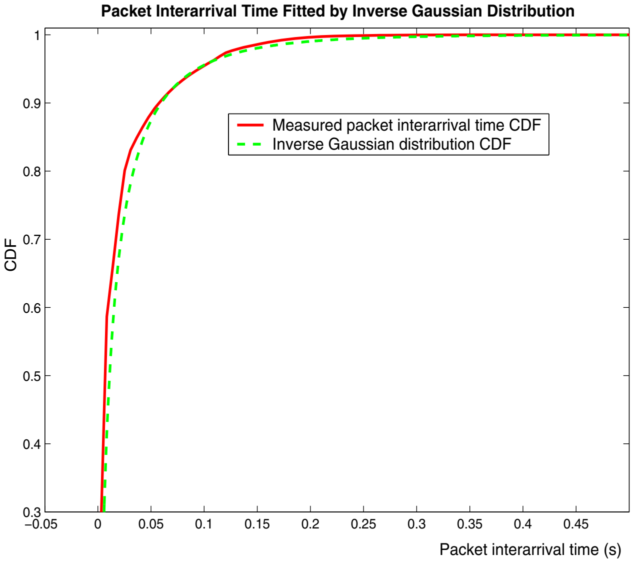

Fig. 7. Packet inter-arrival time fitted by Pareto distribution

The mathematic presentation of the Inverse Gaussian distribution PDF is:

$$f_T(t) = \left[\frac{\lambda}{2\pi t^3}\right]^{\frac{1}{2}} \exp\left\{\frac{-\lambda(t-\mu)^2}{2\mu^2 t}\right\} \tag{3}$$

where the parameters $\lambda > 0$ and $\mu > 0$ are the scale parameter and location parameter respectively for the Inverse Gaussian distribution. The mean of the distribution is μ and the variance μ^3 / λ . For different measurements, we found slightly various values of λ and μ. The characters w and u in the legend of Fig. 5 correspond to the variables λ and μ in equation (3).

The main reason that this measurement is fitted with a different distribution from the other measurements is the less packets number. The measurement was taken in the midnight and the number of users in the WLAN was far less than the peak day hours, leading to much less packet captured in the measurement interval. Statistically, the time between each packet carried in the WLAN became larger and it finally resulted in an Inverse Gaussian distributed packet inter-arrival time.

Therefore, we can use Pareto distribution to model the packet inter-arrival time for WLAN traffic during the day when many users are active and Inverse Gaussian distribution to model it at late night with few active users.

There is an exception among all the nine Pareto distribution fitted measurements. In the measurement 9, we found a sharp rise cut off the measured packet inter-arrival time CDF around the point of 0.09 second. The sharp cut-off in the CDF plotting makes it much less accurate if we still trying to use one Pareto distribution to fit the measured curves. To model this cut-off distribution, we have to consider the distribution and the cut-off separately. For example, if a distribution has a PDF function without cut-off $Y = f_X(x)$ where $x \geq k$, the function with cut-off occurring at $x=m$ can be expressed as:

$$Y' = \begin{cases} f_X(x), & k \leq x < m \\ \beta, & x = m \\ 0, & x > m \end{cases} \tag{4}$$

where $\beta = 1 - \int_k^m f_X(x)dx$.

This method was introduced by [8]. Thus the mathematic expression of the Pareto distribution for situation happened in measurement 9 should be modified to:

$$Y' = \begin{cases} \dfrac{cb^c}{x^{c+1}}, & T_{\min} \leq x < T_{cut} \\ \beta, & x = T_{cut} \\ 0, & x > T_{cut} \end{cases} \tag{5}$$

where $\beta = 1 - \int_{T_{\min}}^{T_{cut}} \frac{cb^c}{x^{c+1}} dx$, T_{min} is the minimum packet inter-arrival time and T_{cut} is the cut-off point.

5 Conclusions

A series of measurements have been presented for the study of IP traffic characteristics in a WLAN environment. The software called WinDump was chosen to be the measurement tool and the whole measurement campaign was scheduled into a set of 30 minutes intervals covering different time of a day in different user behavioural conditions. The packets captured include applications of web surfing, FTP, online gaming and video streaming, etc. Totally, the results of ten measurement intervals were studied and presented.

The results show that TCP is still the main transport protocol for network services. A total of 3,692,189 TCP packets were captured in the ten measurement intervals out of 3,960,460 IP packets, i.e. 93.23% packets are TCP packet. The UDP protocol was mainly used by online card game application. Online video streaming also used TCP protocol although it is a real-time application. This is because TCP can provide reliable packet delivery without packet loss over the best-effort Internet, where the network latency can be compensated using local buffering technologies.

The Pareto distribution was found to be the best fitting to the packet inter-arrival times in nine (out of ten) measurement intervals. The much less usage of the WLAN at midnight, made the Inverse Gaussian distribution to fit better the packet inter-arrival times in one measurement interval. Thus, it is better to model the traffic of a WLAN using the Pareto distribution during the busy day time and using Inverse Gaussian distribution when few users are active during the night. The cut-off phenomenon was observed during measurements and modelled using method expressed in equation (5).

Future works may focus on the study of particular applications using either TCP or UDP protocol and may include web surfing, FTP, online gaming and online video streaming and so on. In this context, new results will contribute towards a better understanding of the network requirements for different applications and provide input to the network management and design configuration.

References

1. Awduche, D., Chiu, A., Elwalid, A., Widjaja, I., Xiao, X.: Overview and Principles of Internet Traffic Engineering. IETF RFC 3272, pp. 4–21 (2002)
2. IEEE Std 802.11-1997 Information Technology - Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks - Specific Requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE (1997)
3. Supplement to IEEE Standard for Information Technology - Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks - Specific Requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-speed Physical Layer in the 5 GHz band. IEEE (1999)
4. Supplement to IEEE Standard For Information Technology - Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks - Specific Requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-speed Physical Layer Extension in the 2.4 GHz Band. IEEE (1999)
5. IEEE Standard for Information Technology - Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks - Specific Requirements, Part II: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE (2003).
6. IEEE Standard for Information Technology - Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks - Specific Requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Amendment 3: Specifications for Operation in Additional Regulatory Domains. IEEE (2001)

7. Evans, M., Hastings, N., Peacock, B.: Statistical Distributions, pp. 45–136. John Wiley & Sons, Inc., Chichester (1993)
8. Universal Mobile Telecommunications System (UMTS) - Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS (UMTS 30.03 version 3.2.0). TR 101 112, v3.2.0, Universal Mobile Telecommunications Systems, European Telecommunications Standards Institute, pp. 34–35 (April 1998)