Chapter

Telecommunications Protocols Fundamentals

Amer Al-Canaan

Abstract

The need for communication amongst people and electrical systems motivated the emergence of a large number of telecommunications protocols. The advances in digital networks and the internet have contributed to the evolution of telecommunications worldwide. The purpose of this chapter is to provide students and researchers with a clear presentation of telecommunications core protocols that are utilised in different research domains including telephony, brain-computer interface (BCI) and voice and digital telecommunications. Indeed, BCI involves different electrical signals, communications concepts and telecommunications protocols. This chapter introduces the reader to the core concepts in communications including analogue and digital telecommunications protocols that are utilised generally in communications and in particular in BCI systems. The topics covered in this chapter include telecommunications protocols, communications media, electrical signals, analogue and digital modulation techniques in digital communications, software-defined radio, overview on 10-Mbps Ethernet protocol and Session Initiation Protocol (SIP).

Keywords: BCI, SDR, protocols, Ethernet, analogue modulation, digital modulation

1. Introduction

Telecommunications protocols play an important role in the advanced modern communication systems that convey information, signals and messages over short and long distances. Telecommunications protocols were developed for data (digital) and voice (analogue) messages.

In a typical brain-computer interface (BCI) [1] application, the electroencephalography (EEG) [2] signals are acquired from the brain, encoded and sent over wireless protocols, such as Bluetooth or Wi-Fi data channels, to a control module. However, in a basic BCI system, signals may be sent through wires between signal acquisition and control modules through a certain serial data communications protocol. BCI is one of several vital engineering domains where researchers and students have to understand and deal with telecommunications protocols.

The need for data communications has inspired researchers and led to the emergence of digital communications, integration of Voice over Internet Protocol (VoIP) or IP telephony with multimedia services offered on IP networks over public switched telephone network (PSTN). Modern telecommunications through VoIP software are common on personal computers and portable devices including smart phones and handheld devices. VoIP systems employ packet switching protocols, which have numerous advantages over circuit switching upon which is based on the traditional PSTN.

VoIP applications for local area network (LAN), wide area networks (WAN), wireless local area network (WLAN) and mobile telephone networks offer better availability, scalability, flexibility, minimum hardware and low cost than PSTN. On the other hand, Internet-related problems such as delay and congestion causing jitter and packet loss are inherent in VoIP.

However, circuit switching is compelling in many applications where real-time, low delay and high QoS are desired, where each customer of modern PSTN profits from dedicated analogue or digital circuits. This implies that a communication channel is reserved during a call or a data session. Due to the limited number of circuits and control units in PSTN, only a fraction of customers can perform simultaneous calls within a switch.

One of the main protocols that has been developed for IP telephony is SIP, which is inspired from establishing and ending a call session and for changing parameters of an established session. The simplicity of SIP and the emergence of Java application interfaces for integrated networks (JAIN)-SIP which is a Java-based API for SIP have reinforced the development and implementation of platform-independent IP telephony services.

In this chapter, core concepts in telecommunications protocols, as well as other related topics including communications media, analogue and digital modulation techniques in digital communications, software-defined radio, overview on 10-Mbps Ethernet protocol and SIP protocol, are presented in an easy and simple style with a number of figures to explain the basic principles of telecommunications protocols.

1.1 Telecommunications core concepts

This section introduces the reader to selected core concepts in telecommunications including telecommunications media and digital encoding.

1.1.1 Twisted pair

Twisted pairs are utilised to carry analogue and digital signals. Depending on distance, analogue signals may be limited to 250 kHz, and digital signals are limited to 10 Mbps for distances around 100 m [3]. At the onset of electrical telecommunication systems, copper was the main transmission medium because of its electrical characteristics such as low resistivity to electric current.

1.1.2 Morse code

The Morse code is a variable-length code, where each character is given a series of dots and dashes. Some letters have one dot and others have one dash. The code length varies from 1 to 5, covering 36 symbols. The telegraph signals were carried using copper twisted pairs. Signal wires are twisted in order to cancel out unwanted noise and reduce the effective inductance of the transmission line. At the sending side, a switch is used to open and close the electric circuit in a certain pattern in order to produce Morse code at the receiving side.

1.1.3 Coaxial cable

A coaxial cable consists of a core wire and a cylindrical shield separated by insulation material. It provides better noise rejection and baud rate over longer distances than the twisted pair. Analogue signal frequency can exceed 500 MHz, and baud rate can reach 500 Mbps depending on distance.

1.1.4 Optical fibre

Optical fibre systems consist of a laser diode transmitter and receiver separated by transparent optical fibre. The signals are transmitted as light pulses that propagate inside the optical fibre. The optical fibre has small diameter and consists of three components: the core (pure glass or plastic), the cladding and the protective cover. The cladding material (glass or plastic) is less optically dense, which allows the light to travel easier through the core. The optical fibre can be used on longer distances with attenuation.

1.1.5 Wireless transmission

Radio and TV broadcasting was made possible through various modulation techniques of electrical signals over different carrier frequencies. For example, the short waves (SW) include frequencies from 3 up to 30 MHz, very high frequencies (VHF) range from 30 to 300 MHz and ultra-high frequency (UHF) cover frequency spectra from 0.3 to 3 GHz. Lower frequencies have longer propagation distances, while higher frequencies suffer from reflections and attenuation over long distances. On the other hand, radio frequency (RF) and high-frequency (HF) transmissions require small antennas since their wavelengths are much shorter.

1.1.6 Microwave transmission

With shorter wavelengths in the range 4–6 GHz, microwave signals travel in straight lines and do not penetrate solid objects. They are affected by clouds, rain and obstacles blocking the line of sight between the transmitter and receiver. Usually parabolic antennas are used for large systems. The received signal is focused at the focal point of the parabola.

1.1.7 Very small aperture terminal (VSAT)

In the 1980s, the very small aperture terminal devices made it possible to telecommunicate, utilising small dish dimensions between remote areas by means of highly directional parabolic antennas [4].

1.1.8 Telephone systems

The microphone in a telephone set converts sound into analogue electric signals that are conveyed traditionally through copper wires and reproduced back at the receiver into sound waves through the speaker. The first telephone systems were analogue, while today's telephone systems are completely digital with tone dialling, voice and data services. Telephone networks have profited from advancements in wireless communications by the implementation of the mobile [5, 6] communications. Old telephone networks were designed mainly to convey voice before the emergence of digital data networks and the Internet.

1.1.9 Analogue and digital signals

Digital signals are characterised by two discrete levels, high and low (1 or 0), while analogue signals have continuous forms. Digital and analogue signals are both utilised in modern telecommunications [7] systems and computer networks. Popular digital codes include American Standard Code for Information Interchange (ASCII) and binary-coded decimal (BCD). ASCII is used in basic character symbols for computer systems, while BCD is mainly used for seven-segment displays.

1.1.10 Non-return to zero (NRZ)

Non-return to zero is the simplest digital encoding as shown in **Figure 1**, where a logic one corresponds to a positive high signal level and the logic zero is simply at ground potential or zero voltage. The NRZ encoding is inconvenient for data transmission specially when data contain a long series of zeros or ones.

1.1.11 Return to zero (RZ)

Return to zero is an improved digital encoding over the NRZ encoding, where logic one signals return to zero as shown in **Figure 2**. The RZ encoding is inconvenient for data transmission when data contain a long series of zeros.

1.1.12 Manchester encoding

To assure reliable transmission of digital data (such as Ethernet and IP), the Manchester encoding (refers to **Figures 3** and **4** with clock signal) is convenient to solve the issue of sending a long series of zeros or ones through a data communication line. The Manchester encoding encodes logic 1 as a transition from level high to low signal, while a 0 is a transition from low to high. The needed bandwidth is twice as the original signal, and there is always a change in the middle of each bit.

An improved version of this encoding is called the differential Manchester encoding, where a 0 causes the signal to change at the start of the interval (refer to **Figure 5**). On the other hand, a 1 causes a change at the end of the interval. A 1 keeps the signal level unchanged as in the previous bit and changes to high at the middle. This is advantageous and permits interchanging the wiring of a differential pair without any issue.

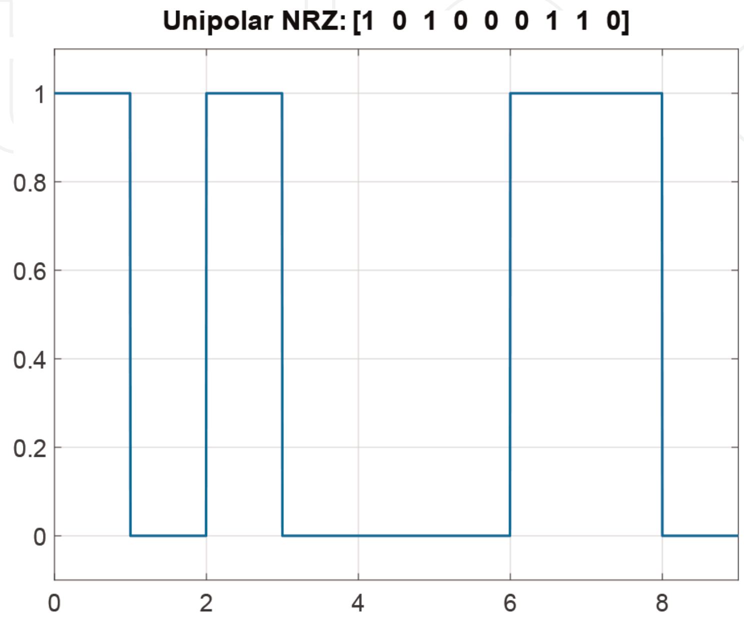

Figure 1.
Unipolar non-return to zero (NRZ).

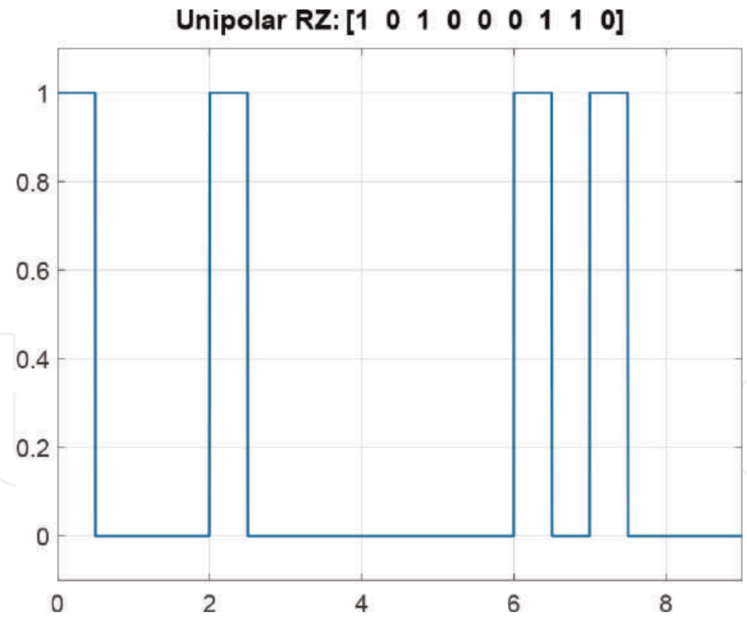

Figure 2.
Unipolar return to zero (RZ).

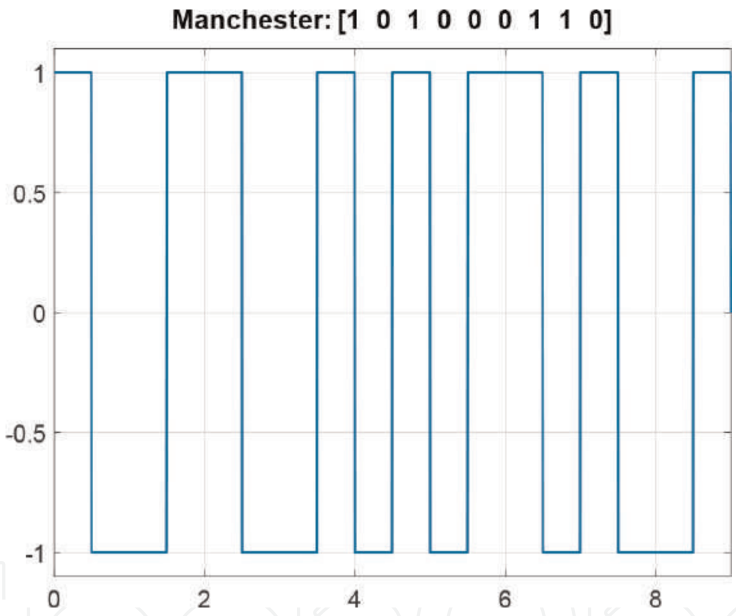

Figure 3.
Manchester encoding.

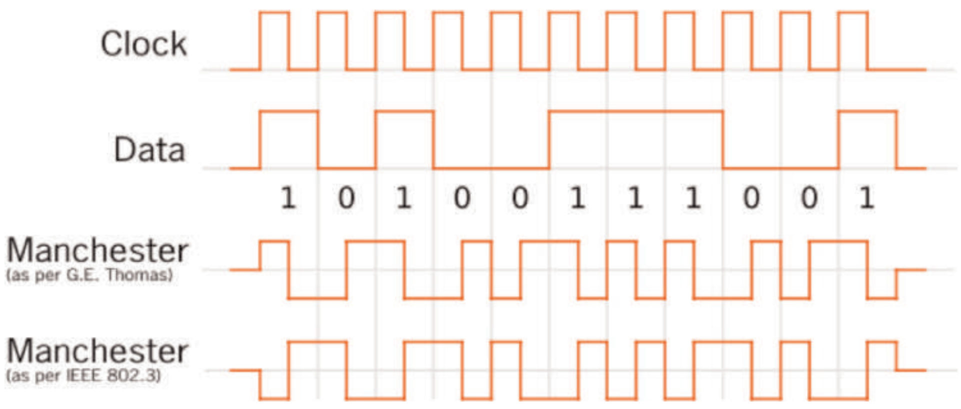

Figure 4.
Manchester encoding example.

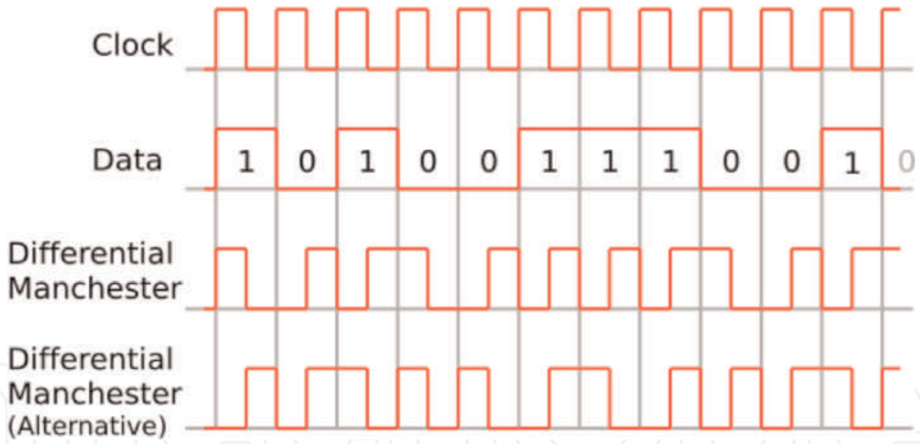

Figure 5.
Manchester differential encoding example.

1.1.13 Shannon's theory

Shannon studied noisy channels, and his theory is based upon the fact that a signal has to have high signal-to-noise (S/N) ratio in order to be successfully distinguished. This influences the maximum bit rate that can be used as follows:

$$Data\ rate\ in\ bps = bandwidth \times log_2(1 + S/N) \tag{1}$$

To increase the data rate, a channel with high S/N should be used. Other means that can increase the bit rate is data compression.

1.1.14 Sampling theory

To convert a continuous signal $x(t)$ into a digital form [8], it is first sampled at equal intervals of time. To be able to reconstruct a sampled signal, $x_\delta(t)$ is defined as

$$x_\delta(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\delta(t - nT_s) \tag{2}$$

The sampling interval T_s is $1/f_s$, where the sampling frequency f_s should be at least twice the highest frequency component f_{max} of the original signal $x(t)$. The frequency $2f_{max}$ is called the Nyquist frequency.

1.1.15 Analogue-to-digital (A/D) conversion

An analogue signal with a given frequency f_1 can be converted into a digital form by sampling it at a constant frequency f_s, where $f_1 < f_s$. A sampled signal has the form of pulses with different amplitudes called pulse amplitude modulation (PAM). The PAM signal is then quantised, and every level is given a binary code number. This process is called pulse-code modulation (PCM). The sampling frequency f_s has to be at least twice as much as the signal frequency being sampled f_1 in order to produce a good approximation of the original signal that can be reproduced and converted back to analogue form. In telephony systems the 8-kHz frequency is used to sample voice that is encoded using 8-bit code. The bit rate in this case is $8000 \times 8 = 64$ kbps. In compact disc (CD) technology, the audio is sampled at 44.1 kHz.

1.1.16 Multiplexing

Multiplexing occurs when data are collected from different sources and are transmitted into one common communication channel. Three types of multiplexing are utilised, namely:

1. Frequency-division multiplexing (FDM). This type of multiplexing employs subcarriers to transmit different message signals.

2. Time-division multiplexing (TDM). This type of multiplexing employs time slots to transmit different message signals.

3. Quadrature multiplexing (QM). This type of multiplexing employs quadrature carriers to transmit different message signals. This type of multiplexing can be distinguished from FDM by the fact that they have overlapped frequency spectra. QM represents double-sideband (DSB) and single-sideband modulations (SSB).

2. Modulation techniques

In the past, digital networks were connected through telephone networks via the modem (modulation/demodulation). Modern telecommunications systems utilise optical fibres that carry many digital channels, which can be translated into voice signals in a telephone by using a codec (coder/decoder). This involves digital-to-analogue (D/A) and analogue-to-digital (A/D) conversions. When a signal $m(t) = A_m \, cos \left(2\pi f_m t + \phi_m(t)\right)$ is transmitted, it is normally modulated using a carrier $c(t) = A_c \, sin \left(2\pi f_c t + \phi_c(t)\right)$ signal, which can be changed or modulated in amplitude (A_c), phase shift (ϕ_c) or frequency (f_c) [9]. The carrier signal can be generalised as $c(t) = A_c(t)\left[sin \left(2\pi f_c t + \phi_c(t)\right)\right]$.

2.1 Analogue modulation

To transmit analogue signals over long distances, analogue modulation techniques are used by changing either the amplitude, phase or frequency of analogue signals.

2.1.1 Amplitude modulation (AM)

Amplitude modulation (AM) takes place when $A_c(t)$ is linearly related to the modulating signals (message). In this modulation technique, the carrier frequency is kept constant, and its amplitude is varied according to the amplitude of the transmitted analogue signal as shown in **Figure 6**. An AM signal $y(t)$ is the result of multiplying the message $m(t)$ and carrier $c(t)$ functions. Assuming a sinusoidal carrier signal defined as $c(t) = A_c \, sin \left(2\pi f_c t\right)$ is used to modulate the message signal $m(t) = A_m \, cos \left(2\pi f_m t + \phi(t)\right)$:

$$\begin{aligned} y(t) &= [1 + m(t)/A_c]c(t) \\ y(t) &= [1 + \mathbf{m} \, cos \, (2\pi f_m t + \phi)]A_c \, sin \left(2\pi f_c t\right) \end{aligned} \tag{3}$$

In the above equation, **m** is the modulation index, which is the ratio of the amplitude of the message signal A_m to the amplitude A_c of the carrier signal.

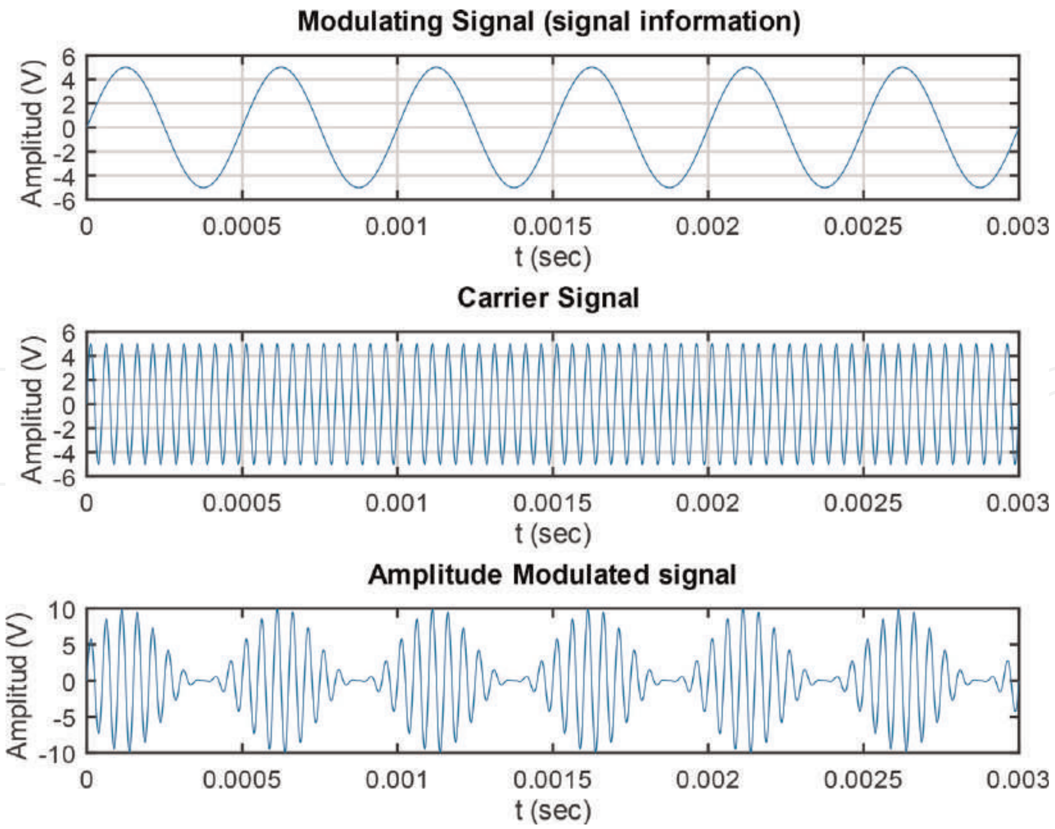

Figure 6.
AM modulation.

To be able to recover the message, **m** should be less than 1, i.e., $1 < \mathbf{m} > 0$. The resulting product function $y(t)$ is composed of three frequencies:

$$y(t) = A_c\, sin\,(2\pi f_c t) + \frac{1}{2}\mathbf{m}A_c\left[sin\,(2\pi[f_c + f_m]t + \phi) + sin\,(2\pi[f_c - f_m]t - \phi)\right] \tag{4}$$

The equation above shows three frequencies:

1. The carrier frequency f_c.
2. The sum of the carrier and modulated frequencies $f_c + f_m + \phi$ with the same phase shift of the message signal.
3. The difference between the carrier and modulated frequencies $f_c - f_m - \phi$ with the negative phase shift of the message signal.

2.1.2 Frequency modulation (FM)

Frequency modulation (FM) takes place when the time derivative of $\phi(t)$ is linearly related to the modulating signal. In this modulation technique, the amplitude of the carrier signal is kept constant, and its frequency is varied according to the amplitude of the transmitted analogue signal as shown in **Figure 7**. Frequency and phase modulations are considered as special cases of angle modulation $s(t) = A_c\, cos\,(2\pi f_c t + \phi(t))$. The carrier frequency is changed such that the frequency f_c depends on the message signal. Since the frequency is the derivative of the phase, the relation between the input signal and frequency can be written as [9] $\phi'(t) = \mathbf{m_f} m(t)$. The FM signal $y(t)$ can be written as

$$y(t) = A_c \left[cos \left(2\pi f_c t + \frac{A_m f_\Delta}{f_m} sin\ (2\pi f_m t) \right) \right] \tag{5}$$

In the above equation, A_m is the amplitude of the message signal, f_m is the frequency of the message signal and f_Δ is the maximum frequency that corresponds to the maximum amplitude A_m value. The frequency modulation index $\mathbf{m_f}$ describes the variation in carrier frequency compared [10]:

$$\mathbf{m_f} = \frac{f_\Delta}{f_m} \tag{6}$$

The frequency modulation index can be less than 1 (for narrowband FM) or much greater than 1 (for wideband FM).

2.1.3 Phase modulation (PM)

Phase modulation (PM) takes place when $\phi(t)$ is linearly related to the modulating signal. In this modulation technique, the amplitude of the carrier signal is kept constant, and its phase is varied according to the amplitude of the transmitted analogue signal as shown in **Figure 8**. The phase of the PM signal can be written in terms of the phase modulation index $\mathbf{m_p}$ as $\phi(t) = \mathbf{m_p} m(t)$.

2.2 Digital modulation

Transmission of digital signals involves modulation of amplitude, frequency or phase of carrier signals. The difference between analogue and digital modulation is that in digital modulation, the changes are at discrete intervals. For example, the

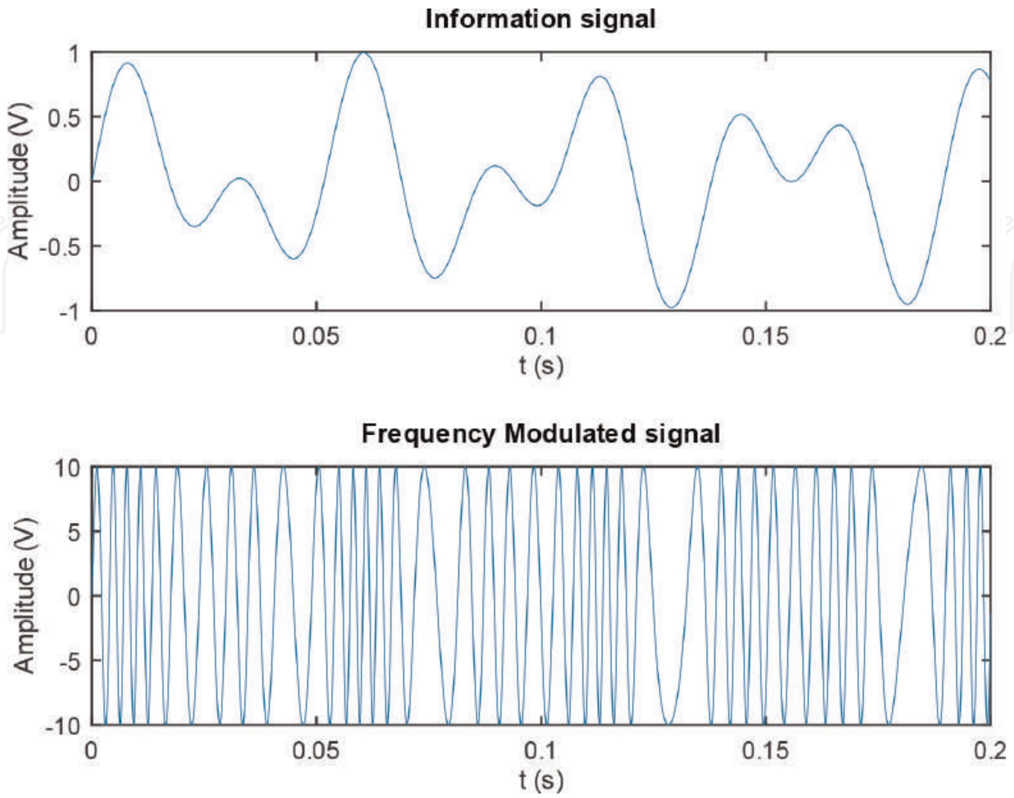

Figure 7.
FM modulation.

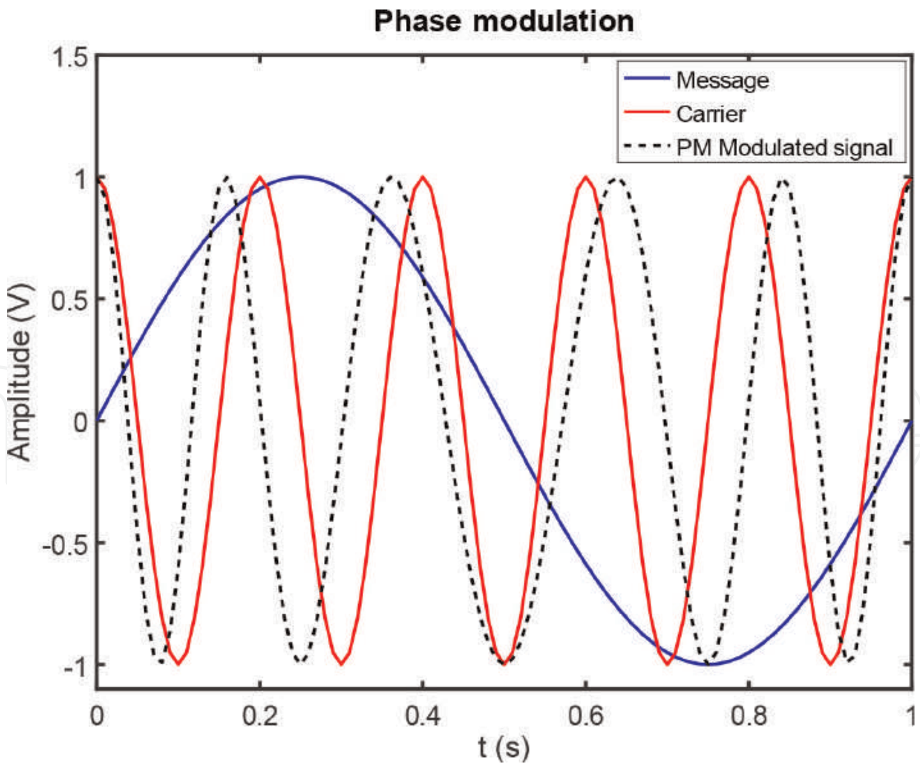

Figure 8.
PM modulation.

amplitude of the carrier signal can be assigned to a maximum value or zero to represent the binary data 1 and 0.

2.2.1 Frequency-shift keying (FSK)

Frequency-shift keying is called also frequency modulation (FM). A bit 0 corresponds to low frequency, and a 1 corresponds to high frequency as shown in **Figure 9**. An FSK signal $s(t)$ can be written as

$$s(t) = \begin{cases} A_c \, cos\left(2\pi\left(f_c + k\right)t\right), & \text{if } \ bit = 1 \\ A_c \, cos\left(2\pi\left(f_c - k\right)t\right), & \text{if } \ bit = 0 \end{cases} \tag{7}$$

In the equation above, k is a constant shift in frequency. Obviously, the FSK uses two frequencies ($f_c + k$ and $f_c - k$) for logic 0 and 1, respectively. This type of FSK is called binary FSK (BFSK).

In case k and $3k$ are used to shift the carrier frequency, the resulting FSK signal has four different frequencies and can be utilised to encode the binary codes $00, 01, 10$ *and* 11, as follows:

$$s(t) = \begin{cases} A_c \, cos\left(2\pi\left(f_c + 3k\right)t\right), & \text{if } \ bits = 00 \\ A_c \, cos\left(2\pi\left(f_c - k\right)t\right), & \text{if } \ bits = 01 \\ A_c \, cos\left(2\pi\left(f_c + k\right)t\right), & \text{if } \ bits = 10 \\ A_c \, cos\left(2\pi\left(f_c - 3k\right)t\right), & \text{if } \ bits = 11 \end{cases} \tag{8}$$

2.2.2 Amplitude-shift keying (ASK)

Amplitude-shift keying is similar to amplitude modulation (AM) as shown in **Figure 10**. Each signal amplitude is assigned to a sequence of bits. If four amplitudes

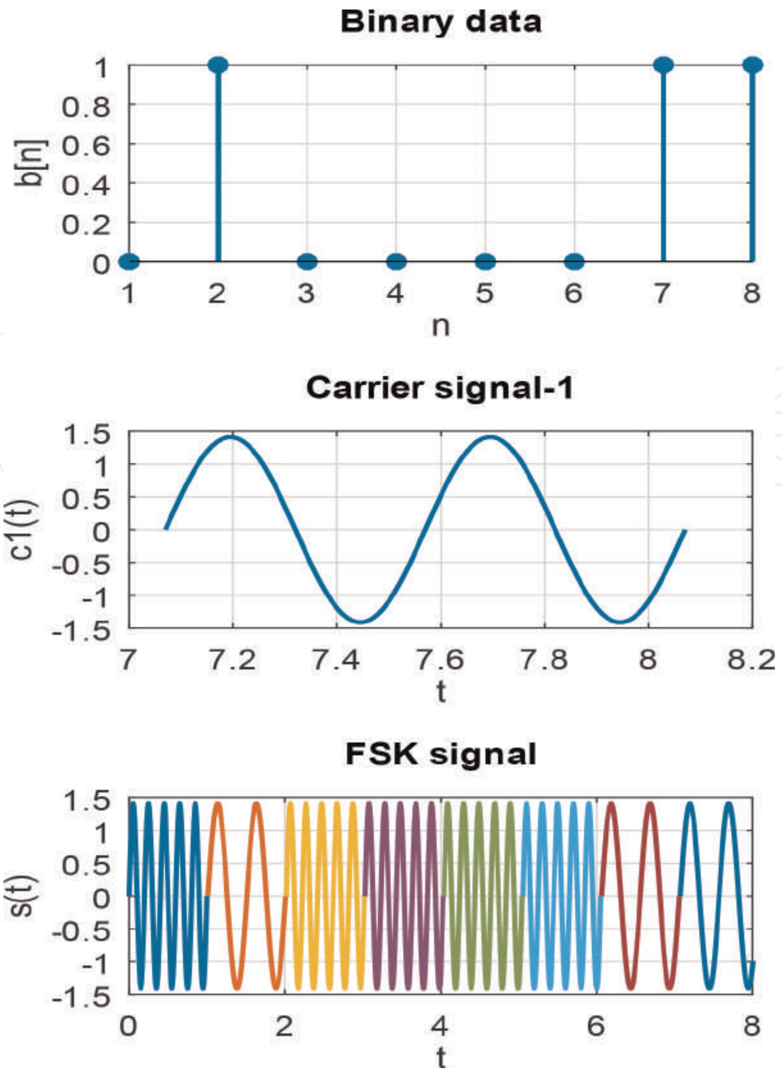

Figure 9.
FSK modulation.

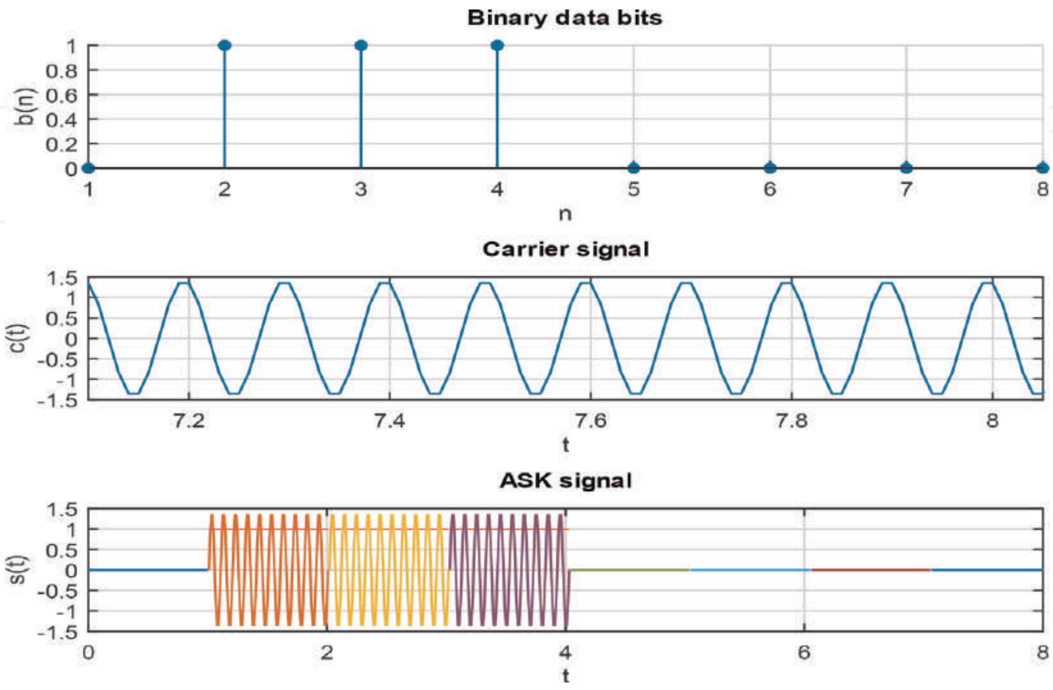

Figure 10.
ASK modulation.

are considered, the following bit code sequences can be defined as 00, 01, 10 and 11. A ASK signal $s(t)$ can be written as

$$s(t) = \begin{cases} A_c\,cos\,(2\pi f_c t), & \text{if } bit = 1 \\ 0, & \text{if } bit = 0 \end{cases} \tag{9}$$

2.2.3 Phase-shift keying (PSK)

Phase-shift keying (PSK) is also called phase modulation (PM). The signal can have a variable phase as shown in **Figure 11**. If the signal is compared with its predecessor, this technique is called differential phase-shift keying (DPSK). Each phase shift can be assigned to a given binary code [11]. A PSK signal $s(t)$ can be written as

$$s(t) = \begin{cases} A_c\,cos\,(2\pi f_c t + \pi), & \text{if } bit = 1 \\ A_c\,cos\,(2\pi f_c t) & \text{if } bit = 0 \end{cases} \tag{10}$$

Since the above equation contains two distinct phases, this type is called binary phase-shift keying (BPSK). If the number of phase variations is increased to 4, the quadrature PSK (QPSK) ca be defined as follows:

$$s(t) = \begin{cases} A_c\,cos\left(2\pi f_c t + \frac{\pi}{4}\right), & \text{if } bits = 00 \\ A_c\,cos\left(2\pi f_c t + \frac{3\pi}{4}\right), & \text{if } bits = 01 \\ A_c\,cos\left(2\pi f_c t + \frac{5\pi}{4}\right), & \text{if } bits = 10 \\ A_c\,cos\left(2\pi f_c t + \frac{7\pi}{4}\right), & \text{if } bits = 11 \end{cases} \tag{11}$$

2.2.4 Quadrature amplitude modulation (QAM)

Though the above three approaches can be used with any number of signals, they tend to be difficult to implement due to the fact that special hardware will be needed to distinguish between adjacent amplitudes, phases and frequencies. To overcome this limitation, a combination of bits can be assigned to groups of signals that can be different in amplitude and phase, for example. For example, using signals with two amplitudes and two phase shifts produces four different signals.

2.2.5 Analogue pulse modulation

Pulse modulation can be achieved by modifying either amplitude, width or position of a pulse signal:

- Pulse amplitude modulation (PAM): The PAM signal (as shown in **Figure 12**) is similar to the sampled signal. The pulses in PAM can have a finite width unlike the sampling delta pulses. The PAM-modulated signal $y(t)$ can be written as

$$y(t) = \sum_{k=-\infty}^{+\infty} x(k)\delta(t-k) \tag{12}$$

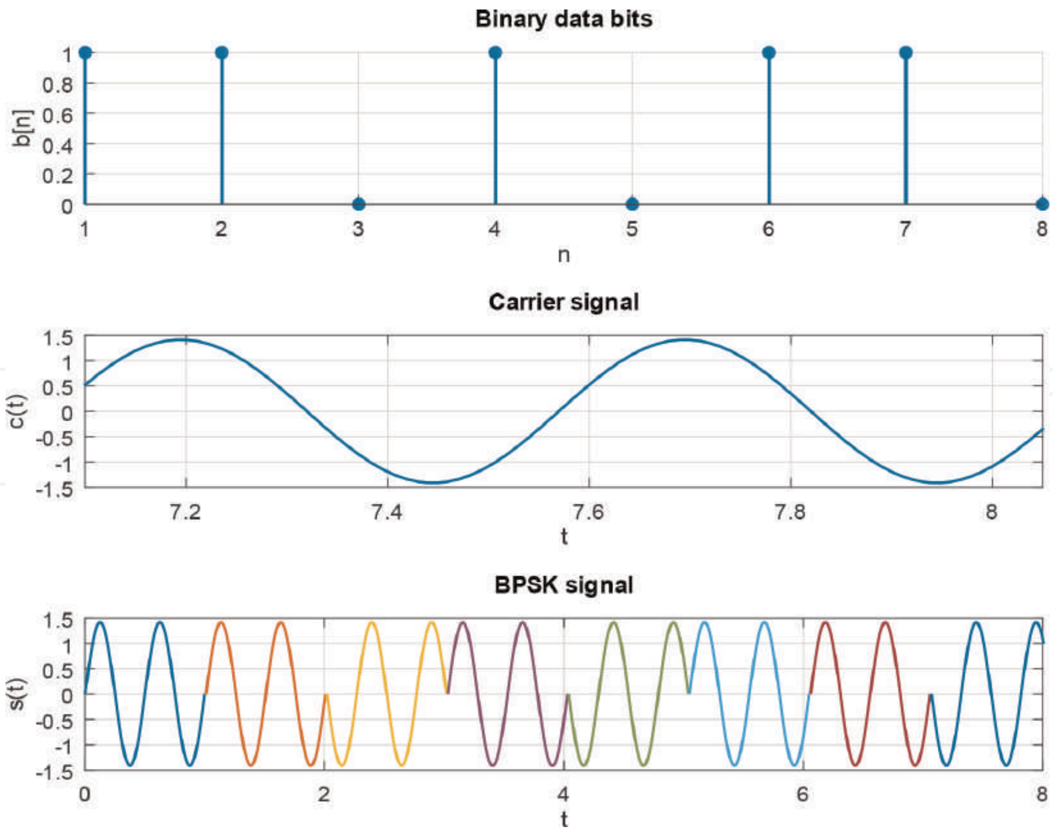

Figure 11.
PSK modulation.

- Pulse width modulation (PWM): In PWM as shown in **Figure 13**, the width of each pulse is related to the modulating signal. This type of modulation is used in DC motor control applications.

- Pulse position modulation (PPM): In PPM as shown in **Figure 14**, the position of each pulse is related to the modulating signal.

2.2.6 Digital pulse modulation

Digital pulse modulation includes two types:

1. PCM: This modulation technique is achieved by sampling the message signal and assigning a digital code (quantisation) to each pulse. The level of the signal is not transmitted; instead the quantised code is assigned according to the available bits for encoding. For example, in 8-bit PCM (with $n = 8$), each level is assigned to a discrete value between 0 and 255. For a signal that has a bandwidth (BW) and a sampling rate of $2BW$, the number of transmitted pulses becomes $2nBW$.

2. Delta modulation: In delta modulation, only the difference between the previous and following codes is sent, as shown in **Figure 15**. For a reference signal $m_s(t)$ and a message signal $m(t)$, the difference $\Delta(t)$ is computed and fed to a pulse generator in order to produce the delta-modulated signal to be transmitted:

$$y(t) = \Delta(t) \sum_{n=-\infty}^{+\infty} \delta(t - nT_s) \tag{13}$$

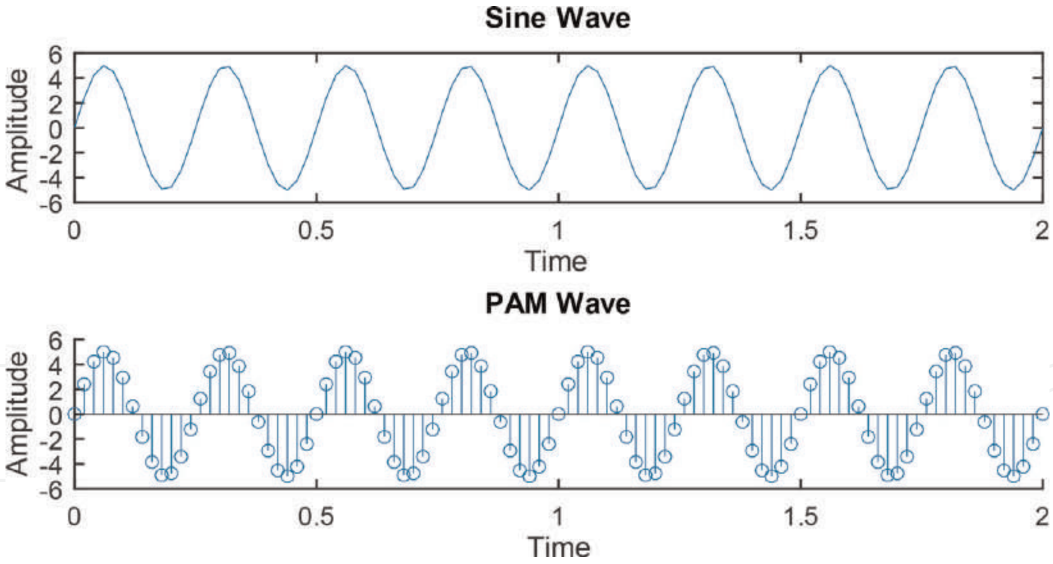

Figure 12.
PAM modulation.

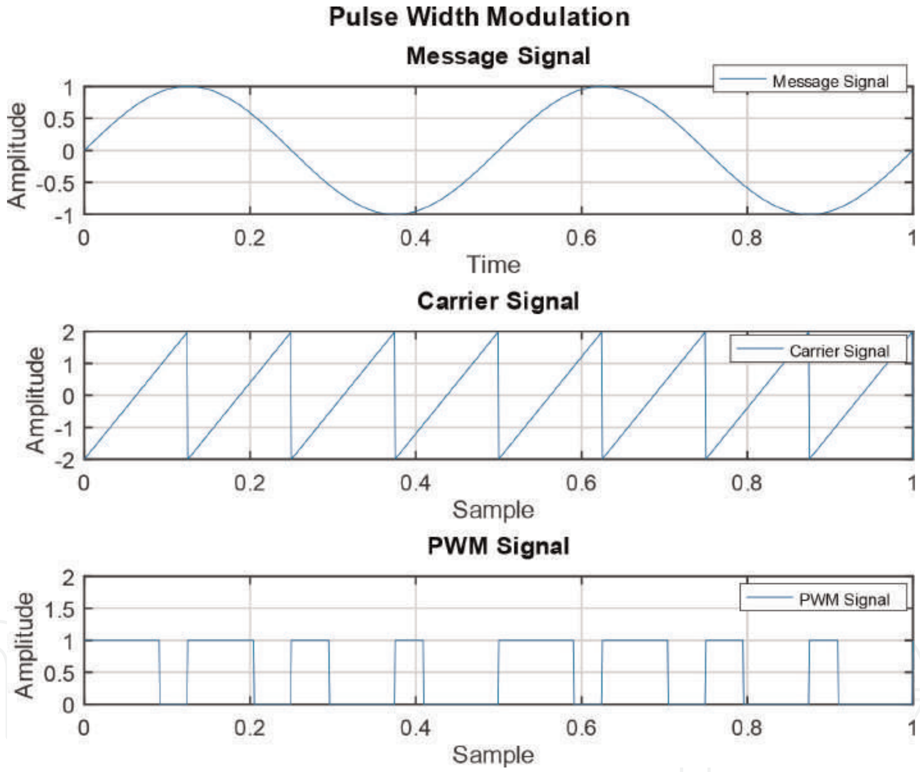

Figure 13.
PWM modulation.

In **Figure 15**, the reference signal $m_s(t)$ is the signal with rectangular edges superimposed on the smooth sine wave message signal. The reference signal is obtained by integrating $y(t)$ as follows [8]:

$$m_s(t) = \sum_{n=-\infty}^{+\infty} \Delta(nT_s) \int^{t} \delta(t - nT_s) d\tau \tag{14}$$

The difference value $\Delta(nT_s)$ is calculated at the nth sampling instant. The reference signal $m_s(t)$ is a stair-step approximation of $m(t)$ as shown in **Figure 15**.

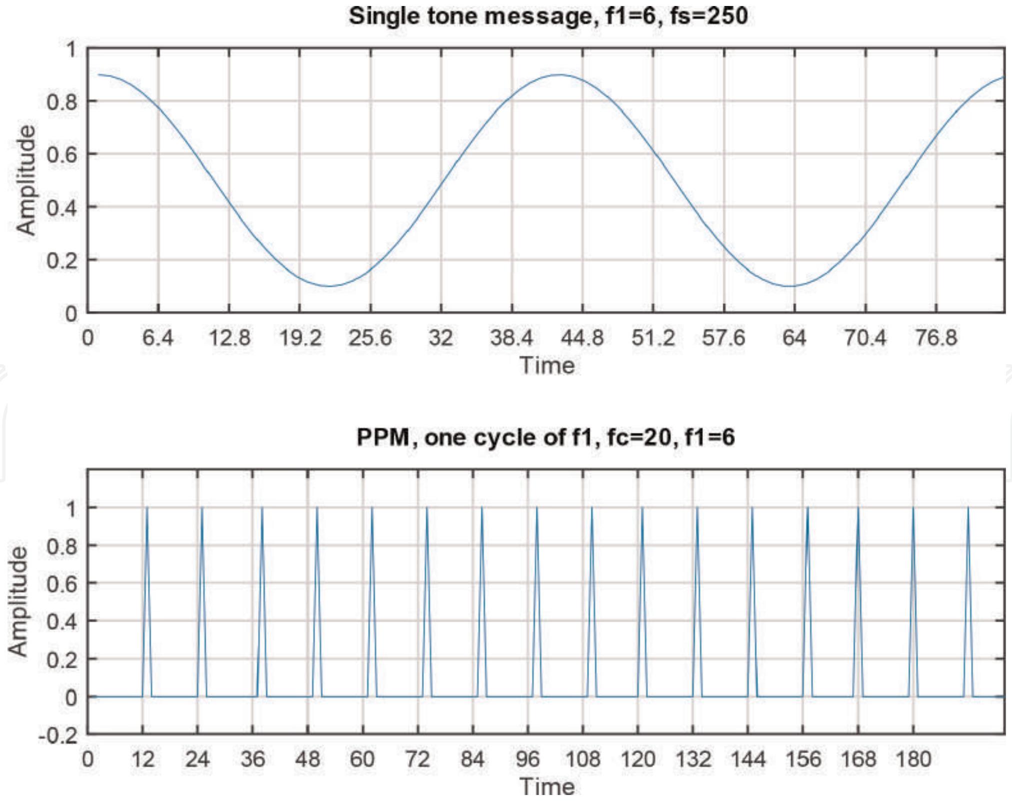

Figure 14.
PPM modulation.

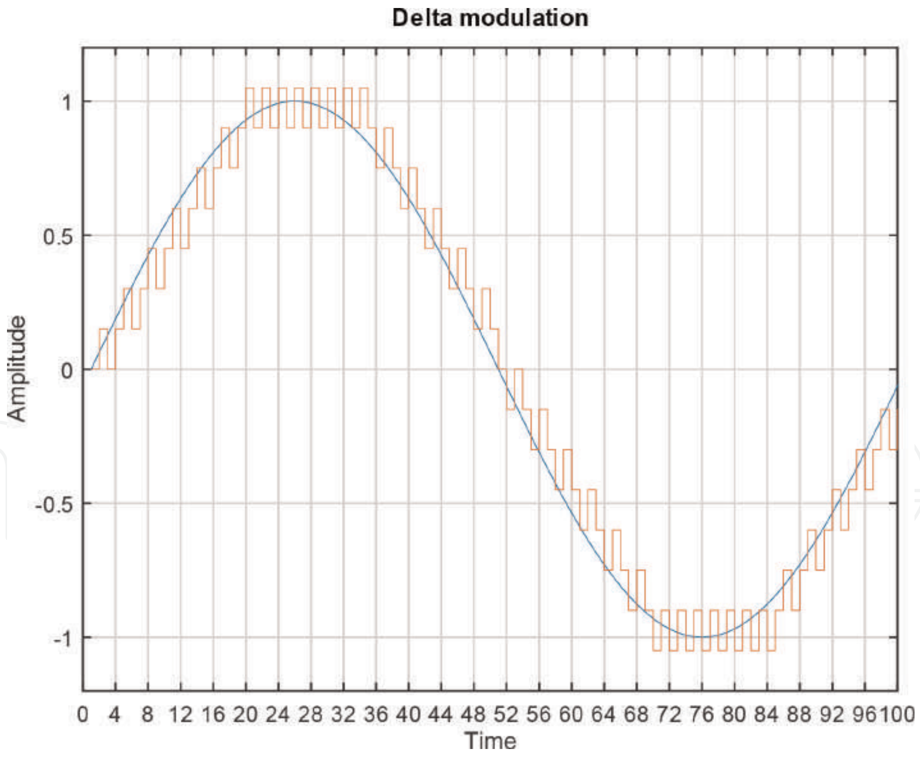

Figure 15.
Delta modulation.

3. Session initiation protocol (SIP)

Modern telephony systems are based upon the Voice over IP (VoIP) protocols, such as SIP, which is a call control and signalling protocol adopted by the 3GPP in order to deliver IP multimedia services [12] to the mobile network [6]. The design of SIP was inspired from HTTP protocol and standardised by the Internet Engineering Task Force (IETF). The purpose of SIP is to enable initiating, terminating interactive call sessions and changing parameters of ongoing sessions. The simplicity of SIP and the emergence of JAIN-SIP [13] have facilitated the development and implementation of platform-independent IP telephony services. Multimedia sessions enable communicating via voice, video and text. SIP messages are either requests or responses and use Session Description Protocol (SDP) in order to determine and negotiate session parameters at either endpoint. SIP supports name mapping and redirection functionalities and, thus, permits user mobility. A typical SIP architecture consists of SIP user agents (UAs) and servers.

4. Software-defined radio (SDR)

Software-defined radio (SDR) [14] is a wireless communication device that employs software to perform most of the operations that are traditionally done by hardware in conventional radio circuits. Similar to the first radio receivers, SDR uses the same hardware for antenna and RF amplifiers. Unlike traditional radios that are based upon hardware to perform modulation and demodulation, software-defined radios are dependent on software to achieve filtering, modulation and demodulation. The IF signal is sampled and converted to digital signal that can be manipulated using software. Common modules between traditional radio and SDR include the antenna and the D/A and A/D converters. Some SDR implementations are freely available using field-programmable gate arrays (FPGA) [15].

5. Overview of 10-Mbps Ethernet

The core protocol of the Internet is the Ethernet protocol, which is based upon serial digital communications. This section provides an overview on the 10-Mbps Ethernet standard. The composition of Ethernet frames (at the MAC sub-layer) and the generation of differential signals at the physical interface (Phy) layer can be implemented on different hardware types as well as FPGA through hardware description language (HDL) code. For 10-Mbps Ethernet, Manchester encoding is utilised, where every bit of information is encoded as a transition from 1 to 0 or from 0 to 1.This is advantageous for the synchronisation between the sender and the receiver and for the recovery of the transmission clock. This encoding method prohibits sending consecutive zeros or ones, which appear as constant DC signal in a conventional RZ encoding. Since every bit of information is composed of two voltage levels, the reference clock is at 20 MHz (double the baud rate).

To identify the beginning of an Ethernet frame, a special pattern of bits is sent, which consists of preamble and a start of frame delimiter (SFD). The preamble and SFD are sent prior to the actual data. The pattern '10' is repeatedly sent, such that a total of 62 bits of 101010 are followed by 11. The last byte (SFD) is 10101011. In hexadecimal, the preamble is 7 bytes of 0x55 followed by a single SFD byte of 0xD5. The first byte that is sent is 0x55, whereas the byte 0x*D*5 is sent last. The leftmost bytes are sent first, of which the rightmost bits (LSB) are sent first. This is why the first byte in the preamble 10101010 is sent from right to left, as 0x55, i.e., the first bit to be transmitted, is 0. Data are usually transferred from an FPGA to the Ethernet port through a physical interface. Taking into consideration the media-independent interface (MII) standard, where the Phy interface communicates nibbles (4 bits) at a time, the SFD 10101011 byte is sent as 0xD and 0x5, since the lower nibble 0xD (in binary, 1011) is sent first starting by 1 (rightmost bit). The reference clocks are 2.5 and 25 MHz for 10-Mbps and 100-Mbps Ethernet, respectively. Reduced MII (RMII) and serial MII (SMII) are two reduced versions of MII, where 2-bit and 1-bit bus widths are used for the Phy, respectively. Compared to the 10-Mbps MII, the gigabit MII (GMII) communicates through 8-bit width bus with a reference clock of 125 MHz. However, the 10-Gbit MII (XGMII) standard deals with 32 bits of data at a time.

Some implementations of Ethernet on FPGA depend upon finite state machines (FSM) programmed in HDL, such as VHDL. Several open-source codes [13] offer Ethernet implementations in VHDL or Verilog.

6. Conclusion

This review chapter contains an overview of telecommunications protocols that are part of modern telecommunications systems. This chapter also provides an overview on analogue and digital signal modulation techniques that are currently used in many research fields including BCI. The researcher in BCI domain as well as the electrical engineering student may find the flow of information smooth and convenient.

The information in this chapter are intended to introduce the reader as well as the researcher in BCI to the core concepts in communications and to analogue and digital telecommunications protocols in an easy-to-follow approach supported with multiple figures and mathematical expression.

The topics covered in this chapter include core concepts in electrical signals, communications, telecommunications protocols as well as other related topics including communications media, analogue and modulation techniques, software-defined radio, 10-Mbps Ethernet protocol and SIP protocol. The topics in this chapter are presented in an easy and simple style with a number of figures to explain the basic principles and fundamentals of telecommunications protocols.

Author details

Amer Al-Canaan
Department of Electrical Engineering, Islamic University of Al-Madinah,
Al-Madinah Al-Munnawara, KSA

*Address all correspondence to: amerc@iu.edu.sa

References

[1] Ramadan RA, Vasilakos AV. Brain computer interface: Control signals review. Neurocomputing. 2017;**223**: 26-44

[2] Jiang X, Bian GB, Tian Z. Removal of artifacts from EEG signals: A review. Sensors. 2019;**19**(5)

[3] Shay WA. Understanding Data Communications and Networks. Boston, USA: PWS Publishing Company; 1994

[4] Manohar V, Kovitz JM, Rahmat-Samii Y. Synthesis and analysis of low profile, metal-only stepped parabolic reflector antenna. IEEE Transactions on Antennas and Propagation. June 2018; **66**(6):2788-2798

[5] Sauter M. From GSM to LTE-Advanced Pro and 5G: An Introduction to Mobile Networks and Mobile Broadband. Hoboken, NJ, USA: Wiley; 2017

[6] Lee W. Mobile Cellular Telecommunications: Analog and Digital Systems. Columbus, OH, USA: McGraw Hill Education; 2017

[7] Frenzel LE. Principles of Electronic Communication Systems. 4th ed. Columbus, OH, USA: McGraw-Hill; 2016

[8] Ziemer RE, Tranter WH. Communications Systems, Modulation, and Noise. 6th ed. Hoboken, NJ, USA: John Wiley; 2010

[9] Siva C, Murthy R, Manoj BS. Ad Hoc Wireless Networks. Upper Saddler River, NJ, USA: Printice Hall; 2004

[10] Wikipedia. Frequency Modulation, 2019

[11] Swierczynski P, Fyrbiak M, Koppe P, Paar C. FPGA Trojans through detecting and weakening of cryptographic primitives. IEEE Transactions on IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 2015;**34**(8): 1236-1249

[12] Al-Canaan A, Khoumsi A. Cross-platform approach to advanced IP-telephony services using JAIN-SIP. Journal of Networks. 2010;**5**(7): 8080-8814

[13] Al-Canaan A, Khoumsi A. Advanced IP-Telephony Service Creation using JAIN-SIP API: Crossplatform approach. In: Mosharaka International Conference on Communications, Networking and Information Technology (MIC-CNET 2008); December 2008; Amman, Jordan. pp. 46-51

[14] Machado-Fernandez J. Software defined radio: Basic principles and applications. Revista Facultad de Ingéniera. 2015;**24**(01):79-96

[15] Opencores.org. OpenCores; 2015

Chapter

Optimum Efficiency on Broadcasting Communications

Juan Manuel Velazquez Arcos,
Ricardo Teodoro Paez Hernandez,
Tomas David Navarrete Gonzalez
and Jaime Granados Samaniego

Abstract

This chapter is devoted to review a set of new technologies that we have developed and to show how they can improve the process of broadcasting in two principal ways: that is, one of these avoiding the loss of transmission signals due to abrupt changes in sign of the diffraction index and the other, preventing the mutual perturbation between signals generating information leak. In this manner, we propose the join of several of the mentioned technologies to get an optimum efficiency on the process of broadcasting communications showing the theoretical foundations and discussing some experiments that bring us to create the plasma sandwich model and others. Despite our very innovative technology, we underline that a complete recipe must include other currently in use like multiple-input multiple-output (MIMO) simultaneously. We include some mathematical proofs and also give an academic example.

Keywords: wireless communications, optimal broadcasting, information packs, negative refraction index, communication theory, wave propagation through plasma

1. Introduction

Nowadays, one of the most innovative procedures to improve communications is the random scattering of microwave or radio signals that may enhance the amount of information that can be transmitted over a channel. This fact, from a mathematical point of view, is due to the growth of the phase space available for that channel, which provides a more rich mathematical base to define every single signal. In many recent papers, a common subject is the use of a broad range of base functions to span each signal. The hope is that every single collision of the initial signals will be scattered and reaches another phase space region providing additional information, but the increase of phase space involves a more complicated set of describing functions. A multiple scattering of the obstacles enlarges the effective aperture in a time-reversed process for acoustic or electromagnetic signals when they are placed in random manner.

Another current tool is time reversal, or phase conjugation in the frequency domain, where a source at one location transmits sound or electromagnetic waves, which are received at another place, time reversed (or phase conjugated), and retransmitted. The effect is to eliminate noise pollution.

Despite the existence of the mentioned resources and others like multiple-input multiple-output (MIMO), many problems survive, but fortunately, we have proposed some additional ways to improve the broadcasting by diminishing the information loss. Some of our results are based on communication theory and others in the mathematical properties of particular integral equations and their solutions.

Through the present chapter, we introduce for convenience a hypothetical discrete system in order to write finite matrices. But we can certainly extend the validity of our expressions as we will see, even for both discrete and continuum systems provided the involved potentials fulfill very general conditions not discussed in the present work.

In the same manner, because the formalism we have developed for the study of time reversibility refers to acoustic systems, we recall that the scalar wave equation for acoustic signals can be written as:

$$k(\mathbf{r})\frac{\partial^2 f(\mathbf{r},t)}{\partial t^2} = \nabla^2\left(f(\mathbf{r},t)/\rho(\mathbf{r})\right) \tag{1}$$

We now describe the quantities appearing in Eq. (1), $\rho(\mathbf{r})$ represents the mass density and $k(\mathbf{r})$ the compressibility of the propagation medium, while $f(\mathbf{r},t)$ is the acoustic signal.

Because the wave equation is of second order in time, we can talk about time reversibility, and then allows solutions, which travel toward the future or the past. An efficient time reversal requires to ensure that the system be ergodic, making possible that the signal may travel both senses in time. To improve focusing, we must describe the signal propagation towards the future or past by means of equations of the same type [18, 22, 27] that is both directions inhomogeneous or both homogeneous. Linearity permits that a signal traveling toward the past can be written with the aid of the integral equation:

$$f(\mathbf{r};T-t) = f^{(\circ)}(\mathbf{r};T-t) + \int\limits_{V}\int\limits_{-\infty}^{\infty} U^*(\mathbf{r}')G^{(\circ)*}(\mathbf{r}',\mathbf{r};T-t',t)f(\mathbf{r}';T-t')dt'dV' \tag{2}$$

In Eq. (2), $G^{(\circ)*}(\mathbf{r}',\mathbf{r};T-t',t)$ is the free Green function, $U^*(\mathbf{r}')$ depicts the complex dispersion coefficients, and $f(\mathbf{r};T-t)$ is the returning signal that has traveled toward the past. The inhomogeneous term $f^{(\circ)}(\mathbf{r};T-t)$ is known as a sink term and makes both the outgoing and returning equations inhomogeneous integral equations. In Eq. (2), the parameter T represents the time during which the outgoing signal (the one traveling toward the future) is being considered and recording. It is observed experimentally [9] that the time-reversed signal has a definition of a 14th of , the wavelength of the used signal for acoustic signals but this is also true for electromagnetic waves. On several experiments [9, 10], Lerosey, de Rosny, Tourin, and Fink have shown that when such a source term is included, the apparent cross section is increased in two ways: first, the multiple scattering also multiplies the available phase space so when the time is reversed, the information is increased, and second, in the electromagnetic case, the sink term stimulates and triggers the braking of the confinement of the evanescent waves that also raise the information and in consequence the definition to level of about $^{\lambda}\!/_{14}$. In acoustics, the sink term consists in the operation of the source in reverse order; in the electromagnetic case, the sink term can be implemented with a crest of fine wires around the antennas.

2. Recovering the matrix equations

As we have said above and considering that from a strictly mathematical point of view, both the acoustic and electromagnetic waves achieve the same wave equation type (with a vector version in the electromagnetic case). Then, we can regain, without further ado, the vector matrix formalism [1–7, 11–14] which generalizes the discrete scalar time reversal acoustic model and includes an original model for discrete broadcasting systems that we have called the plasma sandwich model (PSM) [8, 16–18] and we put some associated parameters appeared on it into the named vector matrix formalism (VMF) [8, 20, 24]. But we must underline that is the resonant behavior the one must be considered for increasing efficiency on communications and to achieve extraordinary resolution. To this end, we remember that a three-dimensional version of Eq. (1) can be written as the Fourier transform of an integral generalized homogeneous Fredholm's equation (GHFE) [21–24] for resonances, and does not matter if for acoustic or electromagnetic ones. To analyze the resonant behavior, we must eliminate the inhomogeneous term so we can write the following algebraic equation satisfied by the Fourier transform of the resonant waves:

$$[\mathbf{1}-\eta_R(\omega)\mathbf{K}^{(\circ)}(\omega)]_n^m\,\mathbf{w}_R^n(\omega)=0 \tag{3}$$

where the kernel $\mathbf{K}^{(\circ)}(\omega)$ is the product of the Fourier transform of the free Green function $\mathbf{G}^{(\circ)}(\omega)$ with the interaction U (without loss of generality we can suppose that U does not depend on ω), so this can be written explicitly as:

$$\left[\mathbf{1}-\eta_R(\omega)\mathbf{G}^{(\circ)}(\omega)\boldsymbol{U}\right]_n^m\mathbf{w}_R^n(\omega)=0 \tag{4}$$

At this point, we must say that we could obtain a transfer matrix description [16–18] instead Eq. (4), but our last equation represents the core of the VMF version. The fact is there are important differences between the two formalisms; for example, VMF makes the time-reversal process easy. Of course, we are moving over a frequency domain and not over a time-dependent one, the former the appropriate in agreement with information theory applications. And certainly, the most important difference is that VMF formalism includes the concept of the resonant solutions.

3. Introducing the PSM parameters

One of the methods we have proposed is based on experiments executed by Xiang-kun Kong, Shao-bin Liu, Hai-feng Zhang, Bo-rui Bian, Hai-ming Li et al. [8] in which they put three layers of plasma joined and alternated with one of them magnetized in the core and the other two unmagnetized in the extremes of the device; when this plasma sandwich is submitted to an external electric potential, it is observed that for a range of values of the external potential, the refraction index is negative [15, 19]. When we analyzed those experiments, we conclude that for this range of the electric potential, the plasma sandwich brakes the confinement of the evanescent waves as occurs in a left-hand material and we proposed a model named the plasma sandwich model for the behavior of the propagation media. Depending on the particular conditions of the propagation media, that is, depending of the values of the plasma sandwich parameters, and for particular conditions of the external electric potential, the propagation media may behave like the plasma sandwich and acquire a negative refraction index. In this section, we introduce the PSM parameters and find the resonant frequencies for a specific problem, underlying that resonant frequencies can be used only to associate an interval of frequencies of a real signal to a device that could be an antenna and not to a single emitted frequency by them; this is because resonant waves are released evanescent waves that vanish in the resource sites and not precisely information carriers. The frequency bands we can build from the resonant frequencies can be considered as convenient highways for the transit of information. Every kernel depends on the response of the media in circumstances that can vary for different time intervals. In this manner, we present an example very easy to work but in which is not relevant the particular behavior of the signal we used to get it. Next, we can find the resonant frequencies for an academic example. First, we choose an appropriate discrete kernel $\mathbf{K}^{(\circ)}(\omega)$, for convenience; in this particular kernel, we do not take into account the three components of the electromagnetic field (usually represented for the indices n and m). However, we propose a system constituted by two emitting antennas. One possible may be written [1, 3–7]:

$$\mathbf{K}^{(\circ)}(\omega)=\begin{pmatrix} \frac{\sin(\omega-\omega_p)\delta}{(\omega-\omega_p)\delta} & -i\frac{\cos(\omega-\omega_p)\delta}{(\omega-\omega_p)\delta} \\ i\frac{\cos(\omega-\omega_p)\delta}{(\omega-\omega_p)\delta} & \frac{\sin(\omega-\omega_p)\delta}{(\omega-\omega_p)\delta} \end{pmatrix} \tag{5}$$

In kernel (5), we have introduced the plasma sandwich model (PSM) parameter δ, which is defined as:

$$\delta \;..\kappa \bar{d}_M \tag{6}$$

Definition (6) involves κ with the physical meaning of the wave number of an incident beam that interacts with the magnetic and electric fields in a way that the whole kernel is the expressed in Eq. (5); $\bar{d}_M$ is the average thickness of a plasma-magnetized layer that generates this interaction; parameter ω_p is the average value for the plasma frequency in the magnetized plasma layer which can be written in terms of the local electron concentration in the layer as:

$$\omega_p = \tfrac{1}{2\pi}\left(\frac{Ne^2}{m\varepsilon_0}\right)^{\frac{1}{2}} \tag{7}$$

In this definition, N is the electron concentration, e is the electronic charge, and ε_0 is the permittivity of vacuum.

It is possible to note that any change in the parameter values gives different broadcasting conditions [5]. PSM suggests that there is not a single stationary set of iterated layers but a bunch of sets evolving in time and in consequence with different effects for each frequency. We must remember that the equation to solve is Eq. (3) where,

$$\mathbf{K}_m^{n(\circ)}\left(\mathbf{r}',\mathbf{r};\omega\right)=\begin{cases} 0 \textit{ if } \mathbf{r}'=\mathbf{r} \\ U^{nm}\left(\mathbf{r}'\right)\mathbf{G}_\omega^{nm(\circ)}\left(\mathbf{r}',\mathbf{r}\right) \textit{ if } \mathbf{r}'\neq\mathbf{r} \end{cases} \tag{8}$$

The last two ubiquitous conditions to achieve resonance are the vanishing of Fredholm's determinant for Eq. (4), and that Fredholm's eigenvalue λ equals to 1 [6, 11, 22, 23]. The last two conditions give us the expected resonant frequencies for the system constituted by two antennas dependent on the PSM parameters. Now, we must remember that resonances have a special behavior that can be represented by a complex frequency:

$$\omega = \mathrm{K} - i\Lambda \tag{9}$$

The transformation of the evanescent waves for traveling ones is due precisely to the imaginary part Λ. In addition, the relation between ω and the wave number κ is:

$$\kappa = \sqrt{\mu\varepsilon\omega} \tag{10}$$

Substituting expressions (9) and (10) into Eq. (3), we can write the resonance condition as:

$$\Delta\begin{pmatrix} \mathcal{M} & \mathcal{N} \\ \mathcal{N} & -\mathcal{M} \end{pmatrix} = 0 \tag{11}$$

The abbreviated components of the matrix in (11) are explicitly

$$\begin{aligned} \mathcal{M} &= \rho_p[\sin(\rho_p)ch(\gamma_p) - \lambda_p] + \gamma_p sh(\gamma_p)\cos(\rho_p) \\ &+ i[\rho_p sh(\gamma_p)\cos(\rho_p) + \gamma_p\lambda_p] \quad (12) \end{aligned} \tag{12}$$

and

$$\begin{aligned} \mathcal{N} &= \gamma_p\cos(\rho_p)ch(\gamma_p) + \rho_p\sin(\rho_p)sh(\gamma_p) \\ &+ i[\rho_p\cos(\rho_p)ch(\gamma_p) - \gamma_p\sin(\rho_p)sh(\gamma_p)] \end{aligned} \tag{13}$$

In Eqs. (12) and (13), we have used the following definitions:

$$\sigma_M = \bar{d}_M\sqrt{\mu\varepsilon} \tag{14}$$

$$\rho_p = \sigma_M(\mathrm{K}^2 - \Lambda^2 - \omega_p K) \tag{15}$$

$$\gamma_p = \sigma_M\Lambda(\omega_p - 2\mathrm{K}) \tag{16}$$

$$\lambda_p = \lambda(\rho_p^2 + \gamma_p^2) \tag{17}$$

To have an image of the solutions of Eq. (11) (see **Figure 1**), we can make $\mathrm{K} = x$ and $\Lambda = y$ those are the real and imaginary parts of ω, and fix the value for the plasma frequency ω_p so we have the following image:

We obtain for the particular conditions:

$$\mathrm{K} = \Lambda \tag{18}$$

$$\omega_p = 10^6 \, Hz \tag{19}$$

The solutions (resonances):

$$x_1 = 5.009 \times 10^5 \, Hz \tag{20}$$

$$x_2 = -985.99 \, Hz \tag{21}$$

In this case only, x_1 is properly a resonance and x_2 has not physical meaning but maintain their orthogonality properties.

4. Communication theory measurement of information loss

Because we have now a wide vision of the loss of information and we know that this is the reason that the images are not perfect, we can use the results of Shannon, Nyquist, Wiener, Hartley, Hopf [25–29], and other authors that have formulated a measure of the loss of information in communication systems. We support our mathematical results on related works [6, 11, 24, 26, 28], which give us a solid theoretical frame to our present and future papers. Indeed, because the *capacity* of a channel and *entropy* are very close concepts, we can use some of the results we have cited above to answer the problem for TRT and LHM.

Basically, we recall two theorems:

Theorem I.

If the signal and noise are independent and the received signal is the sum of the transmitted signal and the noise, then the rate of transmission is:

$$R = H(y) - H(n) \tag{22}$$

This means that the rate of transmission is the entropy of the received signal less the entropy of the noise. The channel capacity is:

$$C = \underset{P(x)}{Max} H(y) - H(n) \tag{23}$$

Theorem II.

The capacity of a channel of band Θ perturbed by white thermal noise power N when the average transmitter power is limited to P is given by:

$$C = \Theta \log \left(\frac{P+N}{N} \right) \tag{24}$$

In this expression, P is the average power of the transmitted signal and N is the average noise power.

From these two theorems, we make our proposal for a channel where we have lost information in three ways. That is, we have limitations on the maximum frequency Θ (band), the presence of different classes of noise, and on a limited time

T available for a time-reversal process. Then, defining a joint average for the power $Q(n,T)$, the channel capacity is:

$$C_T = \Theta \log\left(\frac{P + Q(n,T)}{Q(n,T)}\right) \tag{25}$$

This remains equal to zero when $P = 0$. The very significant feature of this proposal is the explicit dependence on T, in both the joint average power and the channel capacity, as opposed to the conventional treatment of the signal time duration that is considered as a limit process which tends to infinity. This is a consequence of the explicit form of the Fourier transform of the time-reversed Green function that changes with a factor $e^{i\omega T}$, so even if we are not forced to do so, we can think of it as a parameter that defines the channel. We can think of an arbitrary channel but, when we use it to reverse any signal in time, we follow a different process depending on the time T we decide to fit. Then, we can label the channel with each T as a different one and of course with a different capacity with those corresponding to other values of T. Because of the arguments expressed previously in this work, we can use this measure to the same extent on LHM, ATR, and TRT. For a related discussion of the equivalence of the time-reversal methods and the employment of left-hand materials, we can see ref. [30], and for the use of time reversal on antennas, we can see also ref. [16].

5. An academic example

In order to give an insight into information measurement applied to TR, let us propose that our system behaves like a filter. So, in this particular example, we have no loss if we select $t < T$. We also propose that we have a signal like [12]:

$$\frac{\sin(2\pi\Theta t)}{2\pi\Theta t} \tag{26}$$

And, that we have instead of the incoming signal in Eq. (15) another like [10]

$$\frac{1}{2}\frac{\sin^2(\pi\Theta t)}{(\pi\Theta t)^2} \tag{27}$$

The input function Eq. (26) is a sample of a more general function generated by the sum of a series of shifted functions

$$a\frac{\sin(2\pi\Theta t)}{2\pi\Theta t} \tag{28}$$

where a, the amplitude of the sample is no greater than $\sqrt{S}$ (S is the peak allowed transmitter power).

The channel capacity would be [23] approximately (provided that S/N is small)

$$C_T = \Theta \log\left(\frac{S + Q(n,T)}{Q(n,T)}\right) \tag{29}$$

In the time-reversal process, we have shown that for each Fourier component, we should add a complex exponential factor dependent on T. But we know now that the tool is the same and that only the numerical value of channel capacity C_T

changes. We see how in practice the time-reversal parameter T appears explicitly but also that when we cut the time duration of reversed signal, it is possible to consider them as an additive contribution to $Q(n,T)$. But the form of Eq. (25) suggests a generalized measure of a blend or mix channel capacity when sharing the same band W and differ only by the recording time $T_1,T_2,\cdots,T_n$

$$C_{T_1,T_2,\cdots,T_n} = \Theta \log\left(\frac{S+Q(n,T_1,T_2,\cdots,T_n)}{Q(n,T_1,T_2,\cdots,T_n)}\right) \tag{30}$$

The fact that we are using the same band but different cutting limits $T_1,T_2,\cdots,T_n$ also suggests that we can design an appropriate filter that can distinguish between signals according to the recording time that is we can superpose signals with the same frequency range but with different recording times. In a previous work, we have sketched a filter, but now we give a better-defined device, so we propose (see **Figure 2**) as a hint to get the filter, the following steps for both the transmitter and the receiver:

5.1 Transmitter

First, increase the n frequencies on the unique entrance band $B(\omega_0)$ (that is centered in frequency ω_0) incoming from the inverse of $T_1,T_2,\cdots,T_n$, then the n new top frequencies $\omega_1,\omega_2,\omega_Q$ are used to create n transformed signals with the rule suggested by communication theory and these last signals enter a blender. Then, the mixed signal is taken by a band generator and projected in Q new bands centered at the frequencies $\square_1,\square_2,\square_Q$ (each corresponds to a resonant frequency). Finally, each band enters this signal transmitter.

5.2 Receiver

The Q traveling signals enter the mirror band amplifier, so called because it knows that there are Q resonant frequencies and then can create (or separate the signal in Q resonant bands) Q sub-bands and amplify the signal in each band (at this moment, each band carries a piece of the original n different signals); after this,

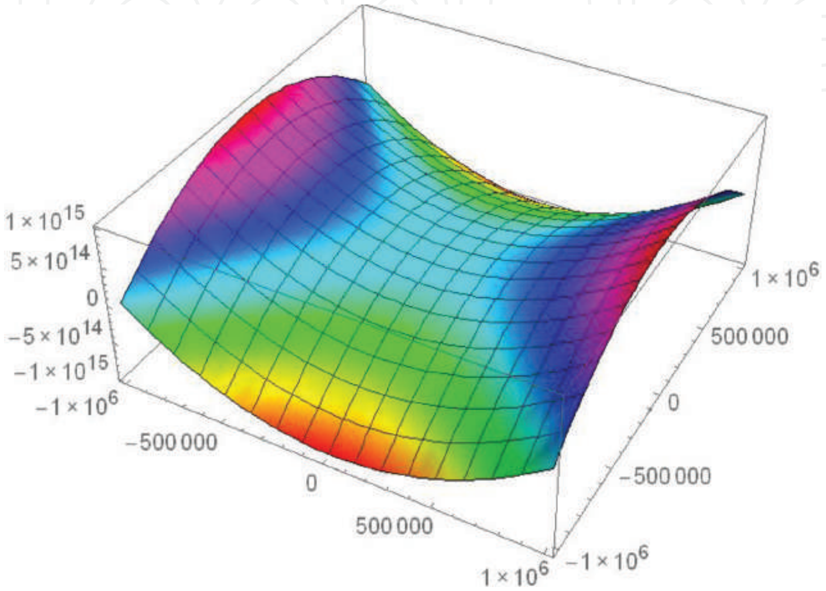

Figure 1.
Image of the solutions of Eq. (11) when the related equation is 987.93(x^2-y^2–10^6) = y (10^6-2x).

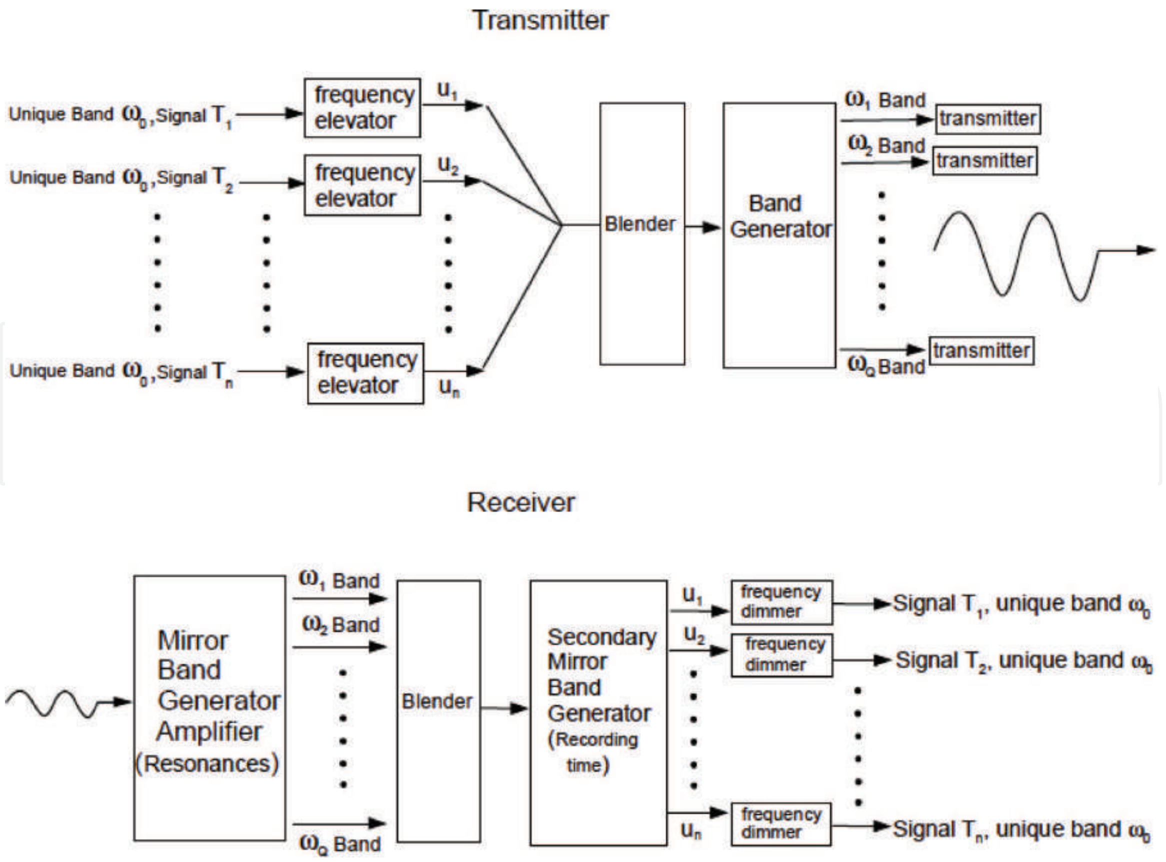
Transmitter
Unique Band ω0, Signal T1
Unique Band ω0, Signal T2
Unique Band ω0, Signal Tn
frequency elevator
frequency elevator
frequency elevator
u1
u2
un
Blender
Band Generator
ω1 Band
ω2 Band
ωQ Band
transmitter
transmitter
transmitter
Receiver
Mirror Band Generator Amplifier (Resonances)
ω1 Band
ω2 Band
ωQ Band
Blender
Secondary Mirror Band Generator (Recording time)
u1
u2
un
frequency dimmer
frequency dimmer
frequency dimmer
Signal T1, unique band ω0
Signal T2, unique band ω0
Signal Tn, unique band ω0

Figure 2.
Flow chart for a proposal device. This can emit and read the blended messages with recording times $T_1, T_2, \cdots, T_n$ *beneath to the unique band* ω_0

the Q signals are blended and then sending to a secondary mirror band generator which knows that there are n recording times $T_1, T_2, \cdots, T_n$ and because of that it can create n bands with the higher central frequencies $u_1, u_2, \cdots, u_n$ (these last signals could be amplitude-modulated signals) and distribute the blended signal among them. Then, every signal on each band enters a frequency dimmer (the inverse operation performed by the frequency elevators in the transmitter), so we retrieve the n original signals corresponding to the unique band $B(\omega_0)$. For example, in Section 3, we have that the total number of resonances is $Q = 2$, and the two resonant frequencies are $\omega_1 = 5.009 \times 10^5\, Hz$ and $\omega_2 = -985.99\, Hz$.

At this point, it is important to say that the key point on the use of the proposed device is the build of information packs described in another place in order to diminish mutual interference between different signals.

6. Error in time reversing and a related theorem

Based on the equivalence of the TRT and the properties of the Green function, we can trust that any discussion about the interaction of metamaterials with electromagnetic field can be done through this function and simultaneously observe the effect of a time reversal. For this reason, we can now describe the error in terms of the Green function by the hypothesis that LHM can be put to test by forward and backward in time signals and read the results with two points of view: first, the direct effect of the loss of information because of the limited record time T or second, how the negative refraction index helps to preserve information. Now, we can review our previous results and generalize using the kernels, so we can characterize the capacity of a channel in many different circumstances. So, we have made use of the analogies [30] between the TRT and the employment of LHM to propose that we can express the capacity of any of these negative refraction index materials in the same terms or procedures as those of TRT. Also, we can propose an identical description for the channel capacity that is Eq. (24) and its generalization Eqs. (25) and (30). Then, the matrix formalism for discrete systems can be used to characterize the channel capacity of transmission of information in a process of time reversibility using the Fourier transforms of the Green functions (properly we use the kernels with the interaction matrix $\mathbf{V} = \mathbf{1}$) forward and backward. That is, by the first step, the signal transforms like (in the following equations *I* and *F* stand for initial and final places):

$$\mathbf{Y}_F = [\mathbf{1} + \mathbf{R}(\omega)]\mathbf{X}_I \tag{31}$$

then in the second step, it returns to the initial place by means of the operation.

$$\mathbf{Z}_I(\omega) = [\mathbf{1} - \mathbf{K}^{(\circ)}(\omega)]\mathbf{Y}_F(\omega) \tag{32}$$

Then, the complete signal trip would be:

$$\mathbf{Z}_I(\omega) = [\mathbf{1} - \mathbf{K}^{(\circ)}(\omega)][\mathbf{1} + \mathbf{R}(\omega)]\mathbf{X}_I(\omega) \tag{33}$$

So that by defining the error in the time-reversing process by:

$$\delta\mathbf{X}_I = \mathbf{X}_I - \mathbf{Z}_I \tag{34}$$

We can write this like:

$$\delta\mathbf{X}_I(\omega) = \mathbf{X}_I(\omega) - [\mathbf{1} - \mathbf{K}^{(\circ)}(\omega)][\mathbf{1} + \mathbf{R}(\omega)]\mathbf{X}_I \tag{35}$$

or

$$\delta\mathbf{X}_I(\omega) = -[\mathbf{R}(\omega) - \mathbf{K}^{(\circ)}(\omega) - \mathbf{K}^{(\circ)}(\omega)\mathbf{R}(\omega)]\mathbf{X}_I(\omega) \tag{36}$$

Eq. (36) is a corollary that shows explicitly the role of both the forward and backward Fourier transforms of the Green function (we have done $\mathbf{V} = \mathbf{1}$ on Eq. (8) for convenience and also for the complete kernels $\mathbf{K}(\omega)$ and $\mathbf{R}(\omega)$). Eq. (36) is very clear about the origin of the errors because we can see, for example, that in the case that the forward and backward Green functions are mathematically one the transpose conjugated of the other for a perfect time reversal (when acting the first on a column vector and on a row vector the other), we get that the error is zero and that the error increases as the differences of both functions also increase. In a very special case, we can then propose that $\mathbf{K}(\omega)$ and $\mathbf{R}(\omega)$ only differ by the factor $e^{i\omega T}$ or $e^{2\pi i \frac{\omega}{\omega_T}}$ when the only source of error is the recording time T, so that we obtain from Eq. (36) that:

$$\delta\mathbf{X}_I(\omega) = -\left[e^{-2\pi i \frac{\omega}{\omega_T}}\mathbf{K}(\omega) - \mathbf{K}^{(\circ)}(\omega) - \mathbf{K}^{(\circ)}(\omega)e^{-2\pi i \frac{\omega}{\omega_T}}\mathbf{K}(\omega)\right]\mathbf{X}_I(\omega) \tag{37}$$

In Eq. (37), the function $e^{-2\pi i\frac{\omega}{\omega_T}}\mathbf{K}(\omega)$ has the form of the Fourier transform of the Green function but with the argument translated by an amount equal to the recording time T that appears explicitly in Eq. (19) that is the Fourier transform of:

$$\mathbf{K}(t-T) \tag{38}$$

But with the time running backward, so, as we will show in a moment, if T is very short, the error will be very huge. On the contrary, if the time goes to infinity, the error will go to zero. Resuming, the new Eqs. (33)–(38), make possible a characterization of the lost information in left-hand materials not only for microwave range, but also for visible frequencies because we have extended recently the time-reversal techniques (see ref. [3, 12]).

Now, we can define:

$$\mathcal{K} = e^{-2\pi i\frac{\omega}{\omega_T}}\mathbf{K} \tag{39}$$

So, we can write Eq. (37) like:

$$\delta\mathbf{X}_I(\omega) = -[\mathcal{K}(\omega) - \mathbf{K}^{(\circ)}(\omega) - \mathbf{K}^{(\circ)}(\omega)\mathcal{K}(\omega)]\mathbf{X}_I(\omega) \tag{40}$$

and because the kernel of the Fourier transform of the generalized inhomogeneous Fredholm's equation (GIFE) satisfies the following integral equations:

$$\mathcal{K} = \mathbf{K}^{(\circ)} + \mathbf{K}^{(\circ)}\mathcal{K} \tag{41}$$

$$\mathcal{K} = \mathcal{K}^{(\circ)} + \mathcal{K}^{(\circ)}\mathcal{K} \tag{42}$$

While Eq. (41) exactly represents the problem with a finite recording time T, Eq. (42) represents a hypothetical problem in which the recording time is infinite.

Substituting Eq. (41) into Eq. (40), we have:

$$\delta\mathbf{X}_I(\omega) = -[\mathcal{K}(\omega) - \mathcal{K}(\omega)]\mathbf{X}_I(\omega) \tag{43}$$

Then, we can suppose that the two kernels in Eq. (40) represent the real and the hypothetical problem described above. Of course, we see that if real conditions approximate the ideal ones, the error is clearly zero. But we can factorize the interaction matrix in Eq. (43):

$$\delta\mathbf{X}_I(\omega) = -\mathbf{V}[\mathcal{G}(\omega) - \mathcal{G}(\omega)]\mathbf{X}_I(\omega) \tag{44}$$

But Eq. (44) says clearly that the error does not depend on the form of the interaction, only depends on the recording time T. Even we have supposed that the only source of error was the recording time, we do not suppose any particular behavior for the interaction. So, we have enunciated and proved a theorem:

Theorem III.

In the time-reversal problem and for left-hand material conditions, the normalized error:

$$\frac{\delta \mathbf{X}_{I}(\omega)}{\left\|\mathbf{V}\right\|} \tag{45}$$

is independent of the explicit form of the interaction provided the last is isotropic.

$$(\mathbf{V}^{-1} = \mathbf{V}^{t*})$$

Returning to the time representation, for the time-dependent retarded isotropic (remember that in the following expression, the indices m and n indicates components of the field and can be omitted), free Green function related to $\mathbf{K}^{(\circ)}(\omega)$, we can write explicitly.

$$G^{mn(\circ)+}(\mathbf{r},t;\mathbf{r}',t') \equiv G^{(+)}(\mathbf{r},t;\mathbf{r}',t') = \frac{\delta[t'-(t-\frac{|\mathbf{r}-\mathbf{r}'|}{c})]}{|\mathbf{r}-\mathbf{r}'|} \tag{46}$$

and for the advanced time-dependent free Green function related to $\mathbf{R}^{(\circ)}(\omega)$:

$$\mathcal{G}^{mn(\circ)-}(\mathbf{r},t;\mathbf{r}',t') \equiv \mathcal{G}^{(-)}(\mathbf{r},t;\mathbf{r}',t') = \frac{\delta[t'-(t+\frac{|\mathbf{r}-\mathbf{r}'|}{c}-T)]}{|\mathbf{r}-\mathbf{r}'|} \tag{47}$$

That is the recording time appears explicitly in the advanced Green function and we can show that its value makes possible to blend many signals on the same channel without interference. It is important to note that for resonances, the relevant Green functions are precisely the free ones and not the complete ones as we can see in Eqs. (5) and (6).

7. Information packs

In this section, we present the support and the definition of the information packs that are required for the adequate performance of the device shown in Section 6 and that by him constitute a method to improve the broadcasting efficiency. To this end, we must remember that on communication theory [9, 10] are defined the so-called ensembles of functions dependent on time. One of their properties is really a group one from the mathematical point of view and lies in that any ensemble transforms into another member of the same ensemble when we change the function at any certain amount of time. To illustrate this property, we shift by an amount t_1 the argument of all the members of the ensemble defined as follows:

$$F_{\theta}(t) = \sin(t+\theta) \tag{48}$$

where θ is distributed uniformly from 0 to 2π.

Then, we have:

$$F_{\theta}(t+t_1) = \sin(t+t_1+\theta) = \sin(t+\varphi) \tag{49}$$

where φ is distributed uniformly from 0 to 2π.

Then, each function has changed individually, but the ensemble as a whole is invariant under the transformation. Also, if we apply the operator T which gives for each member

$$S_\alpha(t) = TF_\alpha(t) \tag{50}$$

It implies that

$$S_\alpha(t + t_1) = TF_\alpha(t + t_1) \tag{51}$$

It is possible to prove that if T is an invariant operator and the input ensemble $F_\alpha(t)$ is stationary, the output ensemble $S_\alpha(t)$ is also stationary. Now, for communication purposes, the operator T, which could be a modulation process, is not invariant because of the phase carrier that gives certain time structure, but if the translations are multiples of the periods of the carrier, then the modulation will be invariant. At this stage, it is important to remember that Wiener [6] has pointed out that if a device is linear as well as invariant (in the sense of the last definition), then the Fourier analysis is the appropriate mathematical tool for dealing with the problem. Now, suppose in addition that we are interested on functions that are limited to the band from 0 to Θ cycles per second, then we have the following theorem [10]:

Let $F(t)$ contain no frequencies over Θ. Then:

$$F(t) = \sum_{-\infty}^{\infty} X_n \frac{\sin \pi(2\Theta t - n)}{\pi(2\Theta t - n)} \tag{52}$$

where,

$$X_n = F\left(\frac{n}{2W}\right) \tag{53}$$

In this expansion, $F(t)$ is represented as a sum of orthogonal (basis) functions. The coefficients X_n of the various terms can be considered as coordinates in an infinite dimensional "functions space." We will take the last theorem (Eqs. (52) and (53)) as a very suggestive rule to consider the recently obtained resonant frequencies. If we use physical arguments about the reasons of the presence of a resonance, we can be sure that channels available for broadcasting are also limited in number. Indeed, in a recent paper, we have generalized the procedure for electromagnetic scalar and vector potentials [30] and we have established that we can use either the electromagnetic field or the potentials for obtaining the resonances and also for the use of the recording time as a resource to optimize communications. And now, we can build information packs (IP) that are functions, which represent a part of the signal we want to send with the minimum loss of information. The resultant expression is:

$$F_e(t) = \sum_{-\infty}^{\infty} X_{n,e} \frac{\sin\left[\pi(2\omega_e t - n)\right]}{\pi(2\omega_e t - n)} \tag{54}$$

where,

$$X_{n,e} = F_e\left(\frac{n}{2\omega_e}\right) \tag{55}$$

Every ω_e allows us to build a decomposition like (54) but we expect that only a few terms are necessary for a well representation of $F_e(t)$. Next, we send separately each $F_e(t)$ by its own device and it is all we need for broadcasting. To receive the signal, we need a separate device for each ω_e.

A very important feature is that because of the properties of the modulation process stated in Eqs. (50) and (51), we can recover, for any arbitrary signal, the behavior under spectral representation and under separated pack representation. So we can either talk about $F_e(t)$ in Eq. (54) as the representation of some element of the basis function for the spectral representation or directly as the e component of an arbitrary signal $S(t) = TF(t)$. Now, we recall the two resonances founded in another work [3]:

$$\omega_1 = \frac{\pi}{4d} + \omega_0 \tag{56}$$

and

$$\omega_2 = \frac{3\pi}{4d} + \omega_0 \tag{57}$$

Suppose that $S(t)$ is the signal

$$S(t) = \frac{\sin\left[\pi(2\Theta t)\right]}{\pi(2\Theta t)} \tag{58}$$

Then, we have the first pack:

$$S_1(t) = \sum_{-\infty}^{\infty} X_{n,1} \frac{\sin\left[\pi(2\omega_1 t - n)\right]}{\pi(2\omega_1 t - n)} \tag{59}$$

with

$$X_{n,1} = S\left(\frac{n}{2\omega_1}\right) \tag{60}$$

And, we have the second pack

$$S_2(t) = \sum_{-\infty}^{\infty} X_{n,2} \frac{\sin\left[\pi(2\omega_2 t - n)\right]}{\pi(2\omega_2 t - n)} \tag{61}$$

with

$$X_{n,2} = S\left(\frac{n}{2\omega_2}\right) \tag{62}$$

We can see that if $\Theta = \omega_1$, the only coordinate distinct to zero is $X_{0,1} = 1$ and if $\Theta = \omega_2$, only survives the term $X_{0,2} = 1$. So, we remark self-consistency of the method.

Even VMF has a broad application on the microwave range, maybe it would be more useful to apply for larger frequencies. But even the great technological boom, there is not any device that could manipulate visible light at length as happens with microwaves. Whatever we can recall some of the basic early ideas on radio broadcasting when the option was sending information by means of modulating the wave's amplitude as appears in **Figure 3**. However, we can take our definition of information packs and put it in a modulated visible-light signal taking the enveloping of the signal we name the wrapping signal (WS) as the information that can be injected inside Eq. (54). Technically, we rewrite Eqs. (50) and (51) in the form:

$$H_\beta(t) = \Omega S_\beta(t) \tag{63}$$

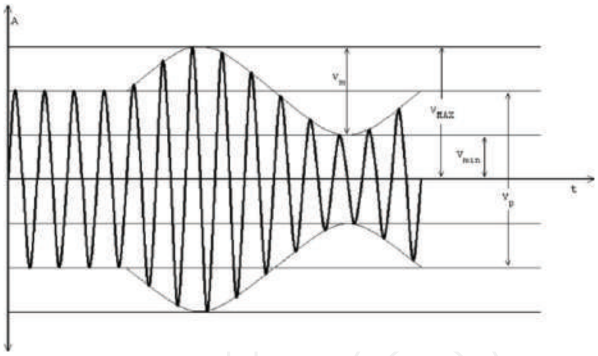

Figure 3.
The former radio broadcasting procedure: modulated amplitude. Image given by Pérez-Martinez [31].

It implies that

$$H_\beta(t + t_1) = \Omega S_\beta(t + t_1) \tag{64}$$

Now, the operator Ω is a generic operator like T but acting over the ensemble $S_\beta(t)$. Some care must be taken when reading the WS information, because the translations stated in Eqs. (63) and (64) were multiples of the periods of the carrier, and then as we said above, the modulation will be invariant. The resonant frequencies will be obtained by the same procedure.

In order to complete the methodology, we recall the concept of group velocity $c_g(t)$ and construct this inherent quotient between them and the enveloping frequency ω_g which results in the wave number κ_g, so we associate them with the resonance frequencies in a similar form as we styled with microwaves, but now these last signals come from the measured properties of the Green's function associated with the modulated signal. In this way, in Eq. (54), we put directly the WS first for a non modulated beam:

$$S_{e'}(t) = \sum_{-\infty}^{\infty} X_{n,\,e'} \frac{\sin\left[\pi(2\omega_{e'}t - n)\right]}{\pi(2\omega_{e'}t - n)} \tag{65}$$

in which the coefficients are given by:

$$X_{n,\,e'} = S_{e'}\left(\frac{n}{2\omega_{e'}}\right) \tag{66}$$

The signal $S_{e'}(t)$ in (65) can be viewed as the representation of some element of the new basis functions or as the $e^{'}$ component of an arbitrary amplitude-modulated signal $H_{e'}(t)$. Now, we can give an example where we use the same values for the resonances on Eqs. (56) and (57) and where we propose an arbitrary amplitude modulated or WS (for a modulated visible light beam) signal given as follows:

$$H(t) = a\cos\left(\Theta_A t + \delta\right) \tag{67}$$

In Eq. (67), $\Theta_A = \Theta_p \pm \Theta_m$ is an arbitrary frequency, and in a same manner, a and δ are preconceived constants but otherwise arbitrary.

With these preliminaries, we can build the first IP:

$$H_1(t) = \sum_{-\infty}^{\infty} X_{n,\,1} \frac{\sin\left[\pi(2\omega_1 t - n)\right]}{\pi(2\omega_1 t - n)} \tag{68}$$

where explicitly the coefficients are:

$$X_{n,1} = H\left(\frac{n}{2\omega_1}\right) \tag{69}$$

And taking expression (67)

$$X_{n,1} = a\cos\left[\Theta_A\left(\frac{n}{2\omega_1}\right) + \delta\right] \tag{70}$$

In a similar manner, the second IP will be:

$$H_2(t) = \sum_{-\infty}^{\infty} X_{n,2}\frac{\sin\left[\pi(2\omega_2 t - n)\right]}{\pi(2\omega_2 t - n)} \tag{71}$$

in which

$$X_{n,2} = H\left(\frac{n}{2\omega_2}\right) \tag{72}$$

Also, by taking Eq. (67):

$$X_{n,2} = a\cos\left[\Theta_A\left(\frac{n}{2\omega_2}\right)\delta\right] \tag{73}$$

As we said above, the resonances must come also for the WS. By this procedure, we have enlarged the scope of the formalism we named vector-matrix or VMF [1–3].

In order to complete our example, we put explicit values of the resonances for the two visible light IP:

$$H_1(t) = \sum_{-\infty}^{\infty} X_{n,1}\frac{\sin\left[\pi\left(2\left[\frac{\pi}{4d} + \omega_0\right]t - n\right)\right]}{\pi\left(2\left[\frac{\pi}{4d} + \omega_0\right]t - n\right)} \tag{74}$$

And explicitly

$$X_{n,1} = a\cos\left[\Theta_A\left(\frac{n}{2\left[\frac{\pi}{4d} + \omega_0\right]}\right) + \delta\right] \tag{75}$$

For the second IP

$$H_2(t) = \sum_{-\infty}^{\infty} X_{n,2}\frac{\sin\left[\pi\left(2\left[\frac{3\pi}{4d} + \omega_0\right]t - n\right)\right]}{\pi\left(2\left[\frac{3\pi}{4d} + \omega_0\right]t - n\right)} \tag{76}$$

in which

$$X_{n,2} = a\cos\left[\Theta_A\left(\frac{n}{2\left[\frac{3\pi}{4d} + \omega_0\right]}\right) + \delta\right] \tag{77}$$

8. Conclusions

In Eqs. (25), (29), (30), (34)–(40), we have shown that it is possible to use an operator language and the properties of the Green function to define the capacity of a channel, the loss of information, and finally, the error in the time-reversal process. Therefore, we can use our results to describe the behavior of LHM interacting with electromagnetic field whether forward or backward in time. Thanks to our interpretation of a resonance in the broadcasting problem with the left-hand material conditions, and the application of the model PSM, we make up a broadcasting system that has the power for distinguishes between signals according to their recording time, and allows to superpose signals in the same frequency range having different recording times with the minor loss because of resonance technology; to this end, we have presented a detailed support and definition of the information packs (IP) and the possibility of application for visible light. In addition, we have enunciated and proved a theorem (theorem III) that establishes: for the TRT and LHM, the normalized error is independent of the particular behavior of the interaction. Summarizing, we give a complete recipe for optimizing communications efficiency.

Author details

Juan Manuel Velazquez Arcos*, Ricardo Teodoro Paez Hernandez,
Tomas David Navarrete Gonzalez and Jaime Granados Samaniego
Universidad Autónoma Metropolitana, Mexico City, Mexico

*Address all correspondence to: jmva@correo.azc.uam.mx

References

[1] Velázquez-Arcos JM, Granados-Samaniego J. Wave propagation under confinement break. IOSR Journal of Electronics and Communication Engineering (IOSR-JECE). 2016;**11**(2. Ver. I):42-48. e-ISSN: 2278-2834, p-ISSN: 2278-8735. Available from: www.iosrjournals.org

[2] Velázquez-Arcos JM, Granados-Samaniego J, Vargas CA. Electromagnetic scalar and vector potentials on the problem of left-hand materials conditions. Electromagnetic in Advanced Applications. 2016;**1**:1. edited by IEEE

[3] Granados-Samaniego J, Velázquez-Arcos JM, Vargas CA, Tavera Romero F, Hernández López RT. Resonant technology and electromagnetic packaging. In: International Conference on Electromagnetic Advanced Applications 2015; Torino, Italia; 2015

[4] Velázquez-Arcos JM, Granados-Samaniego J, Vargas CA. Fredholm and Maxwell equations in the confinement of electromagnetic field. In: International Conference on Electromagnetic Advanced Applications 2015; Torino, Italia; 2015

[5] Velázquez-Arcos JM, Granados-Samaniego J, Vargas CA. Resonance regime on a plasma sandwich simulates a left-hand material broadcasting condition. In: Electromagnetics in Advanced Applications (ICEAA), 2014 International Conference; 2014. pp. 137-140. DOI: 10.1109/IC EAA.2014.6903842

[6] Velázquez-Arcos JM. Fredholm's alternative breaks the confinement of electromagnetic waves. AIP Advances. 2013;**3**:092114. DOI: 10.1063/1.4821336

[7] Velázquez-Arcos JM, Granados-Samaniego J, Vargas CA. The confinement of electromagnetic waves and Fredholm's alternative, Electromagnetics in Advanced Applications (ICEAA). In: 2013 International Conference; 2013. pp. 411-414. DOI: 10.1109/ICEAA. 2013.6632268

[8] Kong X-k, Liu S-b, Zhang H-f, Bian B-r, Li H-m, et al. Evanescent wave decomposition in a novel resonator comprising unmagnetized and magnetized plasma layers. Physics of Plasmas. 2013;**20**:043515. DOI: 10.1063/ 1.4802807

[9] de Rosny J, Fink M. Overcoming the diffraction limit in wave physics using a time reversal Mirror and a novel acoustic sink. Physical Review Letters. 2002;**89**(12):124301

[10] Lerosey G, de Rosny J, Tourin A, Fink M. Focusing beyond the diffraction limit with far-field time reversal. Science. 2007;**315**:1120-1122

[11] Velázquez-Arcos JM. Fredholm's equations for subwavelength focusing. Journal of Mathematical Physics. 2012; **53**(10):103520. DOI: 10.1063/1.4759502

[12] Velázquez-Arcos JM, Pérez-Martínez F, Rivera-Salamanca CA, Granados-Samaniego J. On the application of a recently discovered electromagnetic resonances to communication systems. IJETAE. 2013; **3**(1):466-471. Website: www.ijetae.com. ISSN: 2250-2459

[13] Velázquez-Arcos JM, Granados-Samaniego J, Vargas CA. Communication theory and resonances on electromagnetic systems, Electromagnetics in Advanced Applications (ICEAA). In: 2012 International Conference; 2012. (IEEE Cape Town). pp. 392-395, DOI: 10.1109/ ICEAA.2012.6328657

[14] Velázquez-Arcos JM, Granados-Samaniego J. Recent technologies and

recording time detection on time reversal signals. International Journal on Recent and Innovation Trends in Computing and Communication. 2014;**2**(11): 3447-3950

[15] Grbic A, Eleftheriades GV. Negative refraction, growing evanescent waves, and sub-diffraction imaging in loaded transmission-line Metamaterials. IEEE Transactions on Microwave Theory and Techniques. 2003;**51**(12):2297-20305

[16] Xu H-X, Wang G-M, Lv Y-Y, Qi M-Q, Gao X, Ge S. Multyfrequency monopole antennas by loading metamaterial transmission lines with dual-shunt branch circuit. Progress In Electromagnetics Research. 2013;**137**: 703-725

[17] Kato H, Inoue M. Reflection-mode operation of one-dimensional magnetophotonic crystals for use in film-based magneto-optical isolator devices. Journal of Applied Physics. 2002;**91**:7017-7019

[18] Kato H, Matsushita T, Takayama A, Egawa M, Nishimura K, Inoue M. Theoretical analysis of optical and magneto-optical properties of one-dimensional magnetophotonic crystals. Journal of Applied Physics. 2003;**93**: 3906

[19] Hernández-Bautista F, Vargas CA, Velázquez-Arcos JM. Negative refractive index in split ring resonators. Revista Mexicana de Fisica. 2013;**59**(1): 139-144. ISSN: 0035-00IX

[20] Velázquez-Arcos JM. Nanotechnology can be helped by a new Technology for Electromagnetic Waves. Nanoscience and Nanotechnology. 2012; **2**(5):139-143. DOI: 10.5923/j.nn.20120205.02

[21] de la Madrid R. The decay widths, the decay constants, and the branching fractions of a resonant state. Nuclear Physics A. 2015;**940**:297-310

[22] Velázquez-Arcos JM, Vargas CA, Fernández-Chapou JL, Salas-Brito AL. On computing the trace of the kernel of the homogeneous Fredholm's equation. Journal of Mathematical Physics. 2008; **49**:103508. DOI: 10.1063/1.3003062

[23] de la Madrid R. The rigged Hilbert space approach to the Gamow states. Journal of Mathematical Physics. 2012; **53**(10):102113. DOI: 10.1063/1.4758925

[24] Velázquez-Arcos JM. Fredholm's equation and Fourier transform on discrete electromagnetic systems. IJRRAS. 2012;**11**(3):456-469. Available from: www.arpapress.com/Volumes/Vol11Issue3/IJRRAS_11_3_11.pdf

[25] Shannon CE. A mathematical theory of communication. The Bell System Technical Journal. 1948;**27**:379-423, 623-656

[26] Nyquist H. Certain factors affecting telegraph speed. Bell System Technical Journal. 1924:324. Certain Topics in Telegraph Transmission Theory, A.I.E.E. Trans., vol. 47, April 1928, pp. 617

[27] Hartley RVL. The Interpolation, Extrapolation and Smoothing of Stationary Time Series, NDRC Report. New York-London: Wiley; 1949

[28] Wiener N. The Ergodic theorem. Duke Mathematical Journal. 1939;**5**:1-18

[29] Hopf E. On causality statistics and probability. Journal of Mathematical Physics. 1934;**13**(1):51-102

[30] Velázquez-Arcos JM, Granados-Samaniego J, Vargas CA. Resonances and different recording times on electromagnetic scalar and vector potentials. In: 2017 Progress in Electromagnetics Research Symposium-Fall (PIERS-Fall), Singapore; 2017. pp. 229-235

[31] Pérez-Martinez F. Fundamentos de Sistemas de Televisión. México, D.F.: Universidad Autónoma Metropolitana; 2009

Chapter

Android Application Security Scanning Process

Iman Almomani and Mamdouh Alenezi

Abstract

This chapter presents the security scanning process for Android applications. The aim is to guide researchers and developers to the core phases/steps required to analyze Android applications, check their trustworthiness, and protect Android users and their devices from being victims to different malware attacks. The scanning process is comprehensive, explaining the main phases and how they are conducted including (a) the download of the apps themselves; (b) Android application package (APK) reverse engineering; (c) app feature extraction, considering both static and dynamic analysis; (d) dataset creation and/or utilization; and (e) data analysis and data mining that result in producing detection systems, classification systems, and ranking systems. Furthermore, this chapter highlights the app features, evaluation metrics, mechanisms and tools, and datasets that are frequently used during the app's security scanning process.

Keywords: Android, application, scanning, security, malware

1. Introduction

This section introduces the Android operation system and its applications. Moreover, it defines Android malware and shares its recent statistics. Android permissions and security model are also presented. This section ends with discussing the security scanning framework for Android applications.

1.1 Android and application definition

Android is one of the most popular operating systems that provide open-source development environment based on Linux. It allows the development for mobile, tablets, smartwatches, and smart TVs. Android was established by Open Handset Alliance that started working in 2003, while Google released its first Software Development Kit (SDK) in 2007, but the first commercial version was released in September 2008 called as Android 1.0 [1] with the first device executed being HTC Dream. The sale of the Android phone was increased from 75% in 2013 [2] to 88% in 2018 [3]. **Table 1** lists the sales of smartphones from 2011 to 2018 which show a clear capture of the market over the years. This market penetration reveals the successful implementation of features as well as cheap price.

The Android system is composed of five important layers:

- **Applications** refer to the software stack of native as well as user-based applications.

- **Android runtime** allows the application to run on mobile devices by converting the Android code into DEX format or byte code. The conversion of DEX code into device-related code is done before compilation, and this kind of technique is referred to as ahead of time (AOT).

- **Application framework** manages and runs the applications using the services such as activity manager, content providers, telephony manager, package manager, location manager, etc.

- **Android libraries** are a set of Java-based development application programming interfaces (APIs) that can help in performing general purpose tasks, as well as location-based and string handling.

- **Android kernel** is based on the Linux 2.6 kernel and is used to provide abstraction between device hardware and other software layers [4, 5].

The efforts for making each of the component secure have been made. However, still there are issues due to open-source development, and every vendor and company following their own standards has led to serious security issues [6].

The Android application contains four types of components shown in **Figure 1** [7]:

- **Activities:** each activity represents a single screen with a user interface.

- **Services:** a service operates in the background to execute long-running operations. Services could be initiated by other components like activity or broadcast receiver.

- **Content providers:** to share data between different applications.

- **Broadcast receivers:** to listen for specific system-wide broadcast announcements and react to them.

Android applications are written in Java programming language and distributed as .apk files. Android application package (APK) file is a ZIP compressed file that includes the following files:

- **AndroidManifest.xml file**: it describes the application's capabilities and informs the OS about the other components of the application. It identifies the needed hardware and software features such as the camera, in addition to, the minimum API level required by the application. The permissions requested by the app and the permissions required to access the application's interfaces/data are defined in its manifest file.

- **Dalvik executable or classes.dex file**: the Java classes and methods defined in the application code are grouped into one single file (classes.dex).

- **.xml files**: which are used to define the user interface of the application.

- **Resources**: the external resources that are associated with the application (e.g., images).

Android applications run in a virtual environment to improve security. However, they can be downloaded from any source whether trusted or not. After an application is initiated, it grants its own virtual environment, so the code will be isolated

Year	Android share	iOS shares	Other OS shares
2011	46.66	18.87	34.45
2012	66.34	19.11	14.53
2013	78.50	15.54	5.94
2014	80.70	15.37	3.91
2015	81.60	15.88	2.50
2016	84.79	14.44	0.75
2017	85.91	13.98	0.09
2018	88	11.75	0.03

Table 1.
The detail of the Android phone compared to iOS and other smartphone sales shares from 2011 to 2018 retrieved from Statista.com [3].

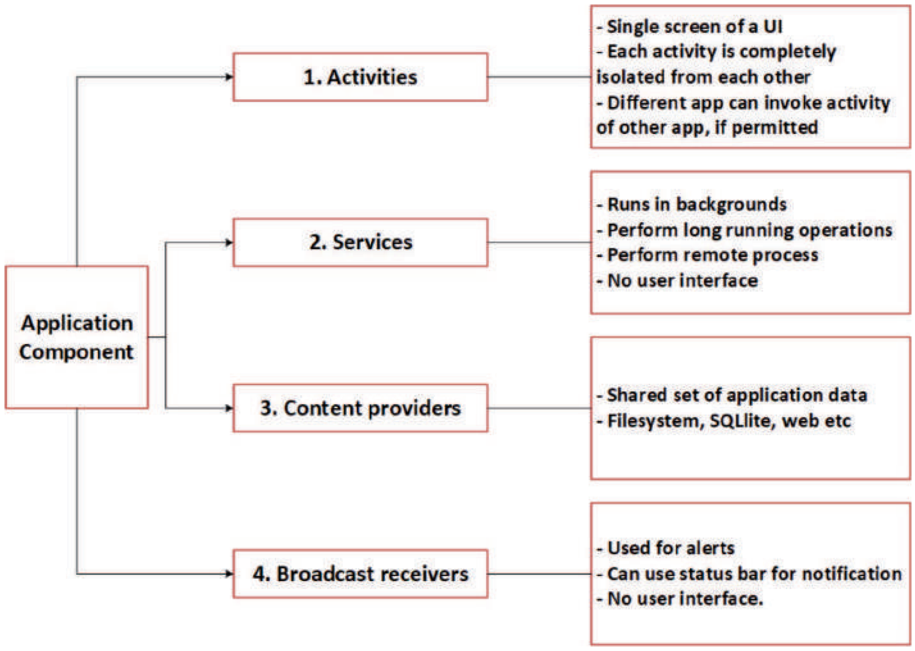
Application Component
1. Activities
- Single screen of a UI
- Each activity is completely isolated from each other
- Different app can invoke activity of other app, if permitted
2. Services
- Runs in backgrounds
- Perform long running operations
- Perform remote process
- No user interface
3. Content providers
- Shared set of application data
- Filesystem, SQLlite, web etc
4. Broadcast receivers
- Used for alerts
- Can use status bar for notification
- No user interface.

Figure 1.
The Android application components.

from other apps. Although the applications are isolated, still they can interact with the system and other applications through APIs. Meanwhile, Android assigns Linux user ID for each application.

API stands for application programming interface that refers to the set of tools providing interfaces for communication between different software components. APIs are used to access data and key features within Android devices. API framework consists of a set of API packages that include specific classes and methods. Additionally, it contains a set of XML elements and attributes for declaring a manifest file and accessing resources besides permissions and intents. Looking into API component calls in the executable file may allow exploring the behavior of an app and reporting its capabilities. However, in many cases, the attackers hide the API calls using cryptography, reflection, or dynamic code loading techniques to increase the difficulty of analysis.

1.2 Malware definition and statistics

Presently, mobile device apps are distributed through online marketplaces such as Google Play Store. Such marketplaces are considered hubs to allow developers to publish their apps and distribute them as well. Today, there are more than 2.5 million applications available in the Google Play Store [8].

When downloading apps from unofficial markets, the user is usually at risk because there is no centralized control like official markets. As more users shift to Android devices, cybercriminals are also turning to Android to inflate their gain. However, many Android apps turn out to be malicious. The number of malicious software (malware) samples in the Android market has surged to an alarming number reaching over 5.49 million by the end of 2018 [9, 10].

A recent report from F-Secure [11] showed that over 99% of all malware programs that target mobile devices are designed for Android devices. Another report from the security firm G DATA shows that a new instance of Android malware pops up nearly every 10 seconds. Another report from AV-TEST [12] states very clearly that anyone seeking to make money by attacking mobile devices will choose Android devices as targets.

Malware is an umbrella term used to stand for an assortment of types of hostile or intrusive software, including viruses, worms, Trojan horses, ransomware, spyware, adware, and different malicious programs [13].

Ransomware is considered one of the most threating malwares nowadays. There are two types of ransomware: crypto ransomware and lock screen ransomware. The crypto ransomware encrypts the information, while the locker ransomware hinders users from gaining access to their data by locking the device's screen. For both types, the attack demands a payment (ransom) to recover the files or access to the device. It is worth mentioning that paying the ransom money does not guarantee that the files will be back or that the ransomware will be removed from the device [14, 15].

According to Kaspersky, ransomware has taken place in most of the majority of notorious security attacks for the past decade. Also, 116.5 million attacks were noted in 2018, compared to 66.4 million in 2017, an increase of twofold in just 1 year [16].

Malicious apps, in general, are distributed mostly through phishing, drive-by attacks, and app stores. Phishing messages might comprise links to malicious apps and are sent over SMS or WhatsApp. Drive-by attacks are carried out by Web page exploits. When the user has a vulnerable browser, the exploit is able to execute a code. To infect users through app stores, malwares are submitted to them hiding as some legitimate app. In fact, in some cases some popular apps are modified to include malicious actions while keeping the app's main functionalities [17].

Therefore, a reliable tool is needed to test the trustworthiness of these apps before being installed. App risk scoring or rating should be empirically calculated according to different risk scoring techniques. The visualization of these risks should be easy enough for a normal user to recognize the risk associated with a specific app.

1.3 Android permissions and security model

Android platform is very popular due to its available and comprehensive API framework [18]. Android API offers the developers of mobile apps the ability to gain access to hardware information, accessing user's data, knowing phone state, changing phone settings, etc. The developers are impacted by the permission model while developing mobile applications. To develop a mobile app, the engineers are required to determine, for each API functionality, what permissions are needed and how they are correctly activated. Android asks the developers to list publicly what permissions are used by the app; however, there are no mechanisms to know the exact purpose of such permissions and what kind of sensitive data they could use.

Android permissions mainly fall under four categories [19]:

- **Normal:** minimal risk permission is assigned automatically by the system and does not require an explicit declaration.

- **Dangerous:** the permission to private data, system process, and other hardware is referred to as dangerous and should be assigned explicitly at the time of installation or usage of the application.

- **Signature:** the applications get the same ID and the same access rights if the two application certificates are the same.

- **SignatureOrSystem:** the applications that are signed with the same certificate will get the same permission as the base system automatically.

Take the camera permission as an example; it belongs to the dangerous category. These permissions also ensure the safety of the system by keeping the user aware of what he is trying to do and what permissions have been requested. The issue with Android permission is that they are coarse-grained and violate the principle of least privileges (PoLP) that ensure that the only required thing is permitted. In contrast, Android allows overall permission about most of the features such as phone contact permission can allow checking phone state and other details.

The Android permission system obliges app's developers to state which security critical resources are needed. At runtime, the access requests are controlled by the permission checker component in order to secure the critical resources and operations.

In general, the security policy for the phones is delegated to their users. The lists of permissions will be shown to the users where they can accept or reject. It is essential and challenging to make sure that these apps appropriately deal with great value sensitive data [18].

Since Android 6.0, dangerous permissions are now asked explicitly to the user when requested the first time and then granted automatically. Android 6.0 changed some areas with regard to permissions. Two major changes were introduced. (1) Apps targeting SDK 23 (Software Development Kit) or higher can request permissions at run-time. (2) If an app requests a dangerous permission, with another permission from the same group that has been already granted, the system immediately grants it without any further interaction with the user.

The Android permission system received several criticisms [20]. The system is considered to be too coarse-grained since the user has to choose whether to accept all of the permissions declared by an app or to refuse to install the app. Users are usually not sure to determine if an app can be trusted or not. Actually, how Android is showing the required permissions is not very user-friendly and quite difficult to understand the risks associated with these permissions.

Android apps are allowed to define new custom permissions on Android. These permissions are used to protect an app's own resources from others. To define new custom permission, a permission name is needed, optionally including a permission group and a description regarding the permission purpose. Sixty-five percent of Google Play Store apps define their custom permissions, whereas 70% of these apps request them for their operation [21].

Mobile app history has shown that the users' privacy and security must be protected against benign applications not to mention malware ones. Actually, lots of widely used apps have been reported as requiring too many permissions or leaking user information to their servers intentionally [22].

Android uses both discretionary access control (DAC) and mandatory access control (MAC) to form a multilayer security model [23]. The model implements a kernel-level application sandbox that uses Linux user identifiers (UIDs) and UNIX-style file permissions. Since version 4.3, Security-Enhanced Linux (SELinux) was introduced, and from version 4.4 it started being deployed in enforcing mode.

Android security has seen other improvements as well. In version 5.0, Google introduced smart lock, which allows users to unlock the phone using a trusted device, such as Bluetooth/NFC beacon, smartwatch, or facial recognition. In version 6.0, they introduced a fingerprint API. All these features are an extra step to make security easier for the average user. However, Android's security model is still based on a set of coarse-grained permissions.

Android builds its security basis on multiuser capabilities of Linux by assigning a unique ID to each application that will manage its own processes [24]. The runtime manager runs the applications in its sandbox that provides security as it does not allow:

- Inter-process communication

- Data access to other processes

- Hardware access such as camera, GPS, or network

- Access to local data of the phone such as media libraries and contacts

As a contrast to other OS platforms, the sandbox facility is provided by runtime manager for direct access to resources and hardware, while other operating systems provide sandboxing based on their kernel. This is based on features such as all kinds of requests outside the applications are by default denied and have to be permitted explicitly. When an application is installed, the permissions are allocated in addition to a unique, permanent identification (ID) that is also assigned to this app. This application ID is used to enforce the permissions for application, processes, and file system [25–27].

The files in the application are always private unless they are explicitly set to be shared using two modes, (1) readable and (2) writable. In order for two applications to share other's files, then the application ID must be the same for both applications, as well as the user ID. Additionally the public key infrastructure (PKI) certificate value must be shared to be considered as one application [27]. The paranoid network security mechanism is used to protect the network access by keeping all kinds of network access in separate groups such as WiFi, the Internet, and Bluetooth. Thus, if the application or process gets the permission to access a Bluetooth, then its application ID will be added to the group access list for Bluetooth and similarly for others. Consequently, one application can be assigned to one or more access groups [28].

Before any application distribution, Google that manages the main play store requires to sign the application using developers' personal certificate to make sure that the distributed copy is done through the right developer and no modification can be made to the application. If the two application matches the same certificate, Android will assign the same application ID to both applications and will access to private files for each application [27, 28].

Relying only on the current Android security model and permission levels to secure Android app is inefficient. Other more comprehensive security systems need to be considered and implemented to ensure efficient detection of malware apps. Consequently, the following sections present a reference model for Android application's security scanning process.

1.4 Android application scanning framework

A reference model for Android scanning process is shown in **Figure 2**. This model provides the core steps/phases vital to analyze Android apps and malware detection. The following sections highlight each one of these phases, starting from allocating the source of Android apps, downloading mechanisms, app's source code generation process, app's features extraction, applying static and dynamic analysis, generating datasets, detecting and classifying the app into benign or malware, and ranking its risk if it is detected as a malware app. Moreover, the mostly used mechanisms and tools utilized by researchers and developers at each process's phase are also presented.

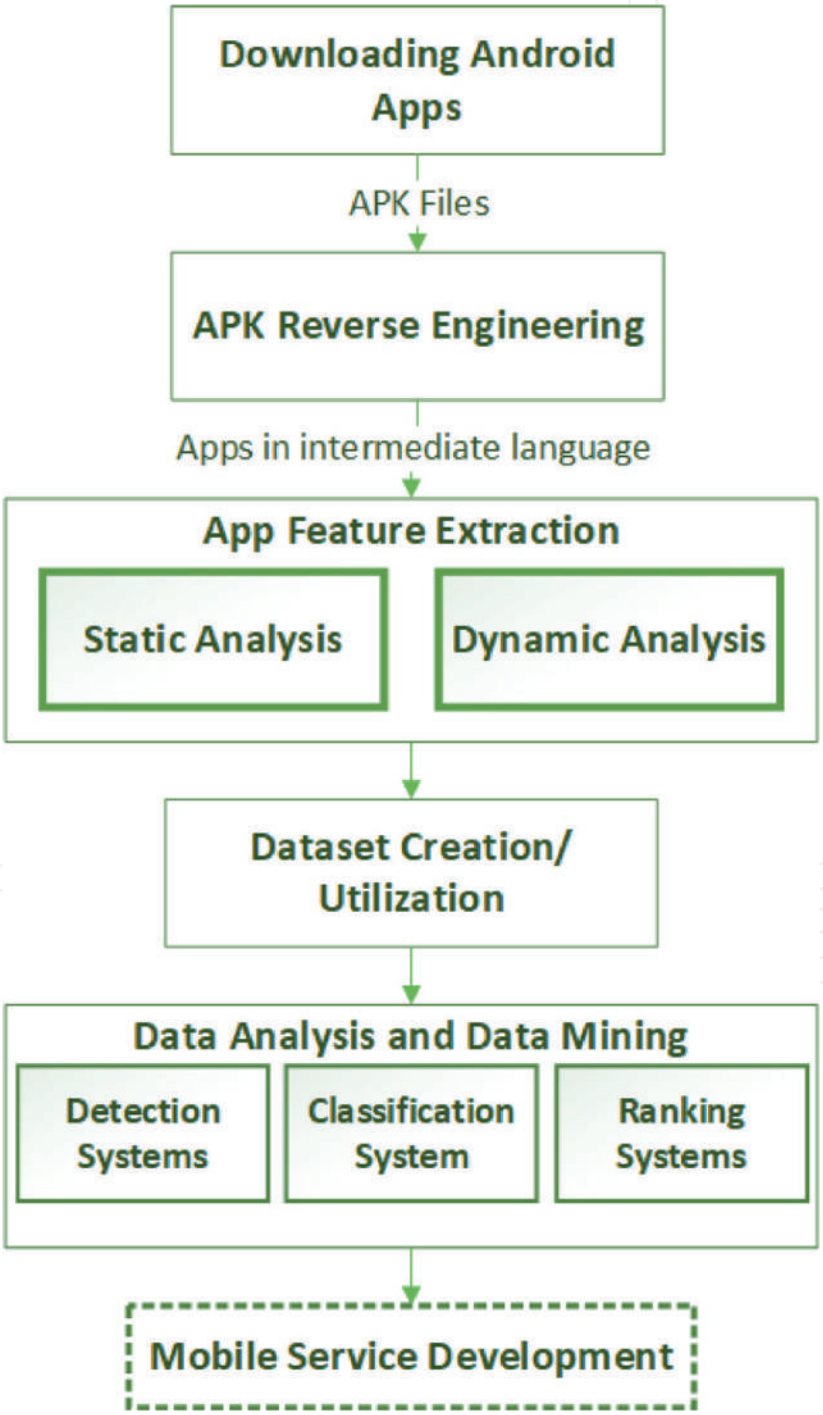

Figure 2.
Android application security scanning model [29].

2. Android application

The app's source and how they are downloaded are presented in this section, in addition to the source code generation for Android applications.

2.1 App source and download

Android application collection process includes gathering APK files from different Android marketplaces. The main application sources include Google Play, Anzhi, and AppChina. For every potential free app, the crawler script must ensure that the app has not been downloaded before and then calculate the app's hash using the SHA256 algorithm [30]. Once the app is downloaded, it can be archived for future use. Chrome APK Downloader, a desktop version of APK downloader tool, can be used to download the APK files of the free Android applications into desktop from Google Play marketplace [31]. For the paid applications, the Raccoon APK Downloader can be used to download APK files from Google Play Store [32].

2.2 Source code generation

After downloading the apps, they need to be analyzed. In order to do that, APK reverse engineering process is required to decompile, rebuild, and convert the Android executable code (.apk file) into an intermediate language such as Smali, Jimple, and Jasmin [33]. The aim of reverse engineering is to retrieve the source file from the executable files in order to apply program analysis. Unzipping the APK files generates .dex files. By reassembling the dex files using an APK reverse engineering tool, the Java files can be retrieved. Three of the most popular tools that have been used in Android APK reverse engineering are Apktool, Dex2jar, and Soot. A comparison of Android reverse engineering tools was conducted in [33]. The results showed that Apktool which uses Smali reassembled 97% of the original code, whereas Soot which uses Jimple and Dex2jar which uses Jasmin preserve 73% and 69% of the app's original code, respectively.

3. Android application analysis

The process of analyzing Android apps to detect different types of malwares and the result of such analysis in terms of datasets are illustrated in this section.

3.1 Feature extraction

Once the app's source code is retrieved, the feature extraction process starts. The features that are usually extracted depend on the type of malware and the analysis mode whether static or dynamic. This will be explained in the following two subsections. **Table 2** lists the most commonly used static and dynamic features [34].

3.2 Static analysis

The static analysis aims to check the existence of malware by disassembling the source code without executing the application. Tools which perform static analysis are mainly categorized into three approaches as shown in **Figure 3**: (1) signature-based detection, (2) permission-based detection, and (3) Dalvik Bytecode detection. There is some limitation which is related to each static detection approach. The

Static features	Dynamic features
Permission	Network, SMS, power usage, CPU, process info, native and Dalvik memory
API calls	Data packets being sent, IP address, no. of active communications, system calls
String extracted	Process ID, system calls collected by strace, returned values, times between consecutive calls
Native commands	Network traffic, destination IP address
XML elements	System calls collected by strace, logs of system activities
Meta data	Data collected by logger, Internet traffic, battery percentage, temperature collected every minute
Opcodes from .dex file	
Task intents	

Table 2.
Most commonly used features in static and dynamic analyses [34].

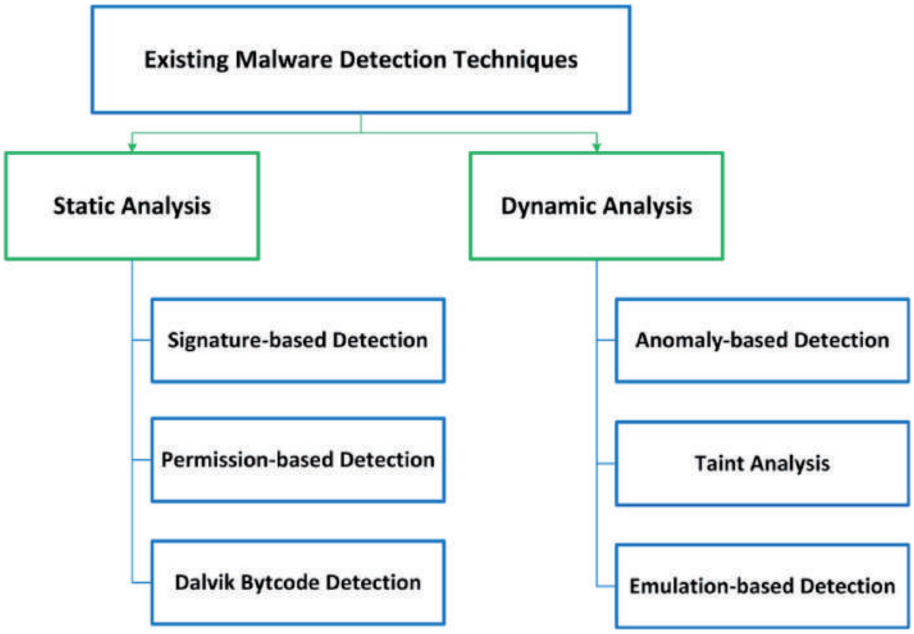
Existing Malware Detection Techniques
Static Analysis
Dynamic Analysis
Signature-based Detection
Permission-based Detection
Dalvik Bytcode Detection
Anomaly-based Detection
Taint Analysis
Emulation-based Detection

Figure 3.
Existing malware detection techniques.

signature-based detection which relies on stored signatures for known malwares does not have the ability to detect unidentified malware signatures [35]. In the permission-based detection, the benign app could be considered, incorrectly, as a malware due to the minor variation of the requested permissions from the original and the malware application [30]. Finally, the Dalvik Bytecode detection, which assists in evaluating the applications actions, consumes more resources [36]. Several tools were discussed and analyzed in [35–41] such as FlowDroid [39], PScout [40], and ApkAnalyser [41]. Each of the aforementioned tools focuses on one or more features. The most generally extracted static features are the permissions [18], API calls [42], and source code metrics [43].

Years ago, different rival static approaches have been proposed like TaintDroid [43], DroidRanger [45], and RiskRanker [46] to detect malicious malware features. But all of them rely on manually crafted detection patterns which may not be able to detect new malware and come with significant device performance cost [47].

Authors in [48] proposed DREBIN as the first approach which provides detection of Android malicious code directly on the mobile device. They used static analysis in a machine learning system to distinguish malware from trusted applications. They considered linear support vector machines for classification. This approach, however, cannot detect runtime loaded and obfuscated malicious applications because it relies on static analysis [47].

Yang [49] developed a prototype (AppContext) that detects malicious apps using static analysis. They mined 633 benign apps from Google Play and 202 malware apps from various malware datasets. AppContext identifies applications using machine learning based on the contexts that trigger security-sensitive behaviors (e.g., the conditions and events that cause the security-sensitive behaviors to occur). But this approach can be evaded by dynamic code loading and consumes huge human efforts in labeling each security-sensitive behavior [50].

Akhuseyinoglu and Akhuseyinoglu [51] have proposed an automated feature-based static malware detection system called AntiWare for Android devices. It is automated since it engages the machine learning method for detecting malicious applications by using the extracted apps' features. They took into consideration the requested permissions and Google market data, including developer name, download time, and user ratings from Google Play as a feature set. AntiWare is designed to predict the rank of an application inquired by the user as malicious or benign and then report the results to the user. The main disadvantages are primarily depending on market data on Google Play and the requested permissions. The market data is not reliable since a lot of applications are invented by different new developers every second. Additionally, the permissions by its own are not sufficient to assess the malicious behavior of an application.

3.3 Dynamic analysis

In dynamic analysis, the application actions are dynamically analyzed and monitored during the execution time. The unexecuted code might be missed by this approach, but it can effectively detect the malware behaviors which are not detectable by the static analysis. Since this approach occurs during runtime, it can be performed in a controlled environment to avoid damaging the device [52].

Android dynamic malware analysis detection techniques (see **Figure 3**) can be classified into [53, 54]:

- **Anomaly detection:** the anomaly-based detection has the ability to identify suspicious behaviors to indicate the presence of malware. A drawback for this technique is that it can sometimes flag a benign application as malware because it displayed similar behaviors of malware.

- **Taint analysis:** it is an efficient technique that checks and monitors sensitive information; however, a limitation is that the performance becomes very slow rendering it useless to be applied in real time.

- **Emulation-based detection:** it is a detection technique, where it scans the application behavior by simulating the conditions of its execution environment to determine if the application is a benign or malware application from the behavior. Similar to this technique is sandbox-based detection, but the main difference originates from the details of designing each approach. A major drawback for this approach is that it requires more resources.

Tam [55] applied dynamic analysis method and machine language to detect malware. They capture real-time system calls performed by the application as key information to discriminate between ransomware, malware, and trusted files and called it CopperDroid. CopperDroid runs the Android application in the sandbox and records all system calls, in particular inter-process communications (IPC) and remote procedure call (RPC) interactions which are essential to understanding an application maliciousness behavior. However, some types of malware can detect the virtual environment and act differently (as a benign) which gives false positives.

Recent research [56] dynamically classifies Android applications to malicious or benign in the first launching of the app. The classification is applied based on the frequency of system calls as an indicator of suspicious behavior. They have built a syscall-capture system to capture and analyze the behavior of system call traces made by each application during their runtime. They have achieved an accuracy level of 85% and 88% using the decision tree algorithm and the random forest algorithm, respectively.

Also, Wang [57] proposed a dynamic analysis to analyze an application on the fly to detect malicious behavior. They developed a prototype called Droid-AntiRM to identify malware applications that employ anti-analysis techniques. The prototype identifies the condition statements in applications that could trigger the malicious acts of malware, which are unable to be recognized by static analysis. However, their prototype cannot handle dynamic code loading, encryption, or other various obfuscation techniques.

Many tools have been developed based on the dynamic perspective such as TaintDroid [44], Droidbox [58], and MobSF [59]. Additionally, some tools are considering both static and dynamic analysis in their solutions such as VirusTotal tool [60].

3.4 Ransomware detection

Unfortunately, there were very few researches studying ransomware where the malicious app blocks access to the Android device or/and its data. In [61] the authors presented a tool called Cryptolock that focuses on detecting ransomware by tracking the changes in real-time user data. They have implemented the tool on Windows platform. However, Cryptolock may send a false-positive alert because it cannot differentiate whether the user or the ransomware is encrypting a set of files [62]. They focus on changes on user's data rather than trying to discover ransomware by investigating its execution (e.g., API call monitoring and access permissions).

HelDroid tool [63] was developed to analyze Android ransomware and to detect both crypto and locking ransomwares. The tool includes a text classifier that uses natural language processing (NLP) features, a lightweight Smali emulation technique to detect the locking scheme, and the application of taint tracking for detecting file-encrypting flows. The primary disadvantage of this approach is that it highly depends on a text classifier as it assumes the availability of text. Also, it cannot be applied to some languages that have no specific phase structure like Chinese, Korean, and Japanese. This approach can be easily avoided by ransomware by applying techniques such as encryption and code obfuscation [63]. Moreover, like whatever machine learning approach, HelDroid trains the classifier in order to label an app as a ransomware. The detection capability of the model depends on the training dataset [64–66].

Another work in literature exploring the ransomware detection in Android mobiles was presented in [67]. The authors presented R-PackDroid as a static analysis approach that classifies Android applications into ransomware, malware, or benign using random forest classifier. The classification employed was based on information taken from the system API packages. An advantage over the previous approach (HelDroid) is its ability to detect ransomware regardless of the application language. Also, it flags the applications that were recognized as ransomware with very low confidence by the VirusTotal service. However, R-PackDroid cannot analyze applications with a feature code that is dynamically loaded at runtime or classes that are fully encrypted because it relies on static analysis.

Likewise, Mercaldo [68] focused on ransomware detection specifically in Android. They tested a dataset composed of 2477 samples with real-world ransomware and benign applications. The main issue of this approach is that it is manual and requires a lot of effort to analyze and build logic rules used for the classification [69].

Another automated detection approach was introduced in [70] to analyze and penetrate the malicious ransomware. They have introduced some features of static and dynamic analysis of malware. In static analysis, malicious features can be discovered with permissions, API calls, and APK structure, while malicious features in the dynamic analysis may include access to sensitive data or sensitive paths, access to the HTTP server, and user charge without notification and bypass permissions. The aim was to produce a better performance apparatus that supports ransomware detection in Android mobiles which they have designed but did not implement. The authors analyzed one malware and listed the steps of APK analysis as a concept but did not implement the proposed design. Therefore, there are no results that can prove the effectiveness of their approach.

In [71], the authors experimentally presented a new framework called DNADroid which is a hybrid of static and dynamic techniques. This framework employs a static analysis approach to classify apps into suspicious, malware, or trusted. Only suspicious classified applications are then inspected by dynamic analysis to determine if it is ransomware or not. The main weakness is that dynamic analysis is only applied to suspicious applications leaving the possibility of having malware not successfully recognized by static analysis.

3.5 Dataset creation and utilization

Datasets are mainly in two types. The first type is the Android application datasets. These include both benign apps and malicious apps. For the benign apps, the majority of researchers are collecting them from the app stores like Google Play Store [30, 37, 60]. For malicious apps, it depends on malicious behavior under study. For example, for malware Android apps, VirusTotal was one of the main sources for many researchers [38, 60]. For ransomware apps, HelDroid project [63] and RansomProper project [38] were also used.

The second type is the datasets generated after extracting the app’s features. The researchers can either use existing constructed datasets considering the features under study or build new ones by screening the apps and extracting their features. The main concerns regarding the use of existed datasets are (1) absence of up-to-date apps and operating system version (2) including many duplicate samples (3) and not being accessible. These reasons could motivate researchers to build their own up-to-date and labeled datasets.

4. Android malware application detection and ranking

Many previous works have considered the problem of ranking Android apps and classify them to either malware or benign apps. The majority of these solutions have relied mainly only on the permission model and what types of permissions are requested/used by the application. They used different ways and depth of analysis in this regard.

The work presented in [72] studied the permission occurrences in the market apps and the malware apps. Also, the authors analyzed the rules (a combination of permissions) defined in Kirin [73] in order to calculate the risk signals and to reduce the warning rates. Gates et al. in [74] have compared work presented in [72, 73], naïve-based algorithms and two proposed methods for risk scoring. These methods consider the rarity of permissions as the primary indicator that contributes to raising a warning. The performance comparison was in terms of the detection rate.

The authors in [75] used similar hypotheses of listing the permissions in each app and count occurrences of permissions in similar apps (a game category in their case) excluding the user-defined permissions that are not affecting the privacy. In their solution, they gave the user the choice to turn off the permission(s). In [76], the authors used the combination of features (permissions) to compare the clean app values with the malware values to come up with thresholds that will be used to classify new Android apps. Within the same context, the idea presented in [77] was to construct a standard permission vector model for each category, which can be used as a baseline to measure and assess the risk of applications within the same category. For each downloaded app, the permission vector will be extracted and compared with the standard one; the amount of deviation from the baseline will calculate the app's risk.

While discussing the approaches in the existing risk scoring systems and their main dependency on the Android permissions, it is worth asking how many of them have considered the involvement of the user with the scoring results and, if they decided to involve the user, how the risk was displayed and communicated to the user. The empirical study conducted in [78], which implemented an intensive study on top, used permissions with a high-risk level. They calculated the risk level based on the type of permissions and the probability they will be requested by the app. The risk value for each permission in addition to its technical name and description was transmitted to the users. Although a coloring code was used to indicate the level of risk, still users are involved in technical details which will not help them to take proper decisions regarding the apps' installation. The work presented in [79] has utilized fuzzy logic to measure the risk score. Also, in addition to the permissions and their categories, they took input from different antivirus tools to calculate the score. Their system allowed the user to upload the app's APK through the browser and provided them with a risk report. This report showed the risk score, permission usage rate, and unnecessary permission usage rate in addition to the list of permissions, their categories, and risk level. On the other hand, the authors in [80] have considered the statistical distribution of the Android permissions in addition to the probabilistic functions. The declared but not exploited permissions and vice versa were all considered in their risk analysis. Machine learning was also utilized to measure risk.

In terms of visualizing the permissions and their risks, the authors in [81] introduced Papilio to visualize Android application permissions graphically. This helped them to find the relations among the applications and applications' permissions as well. Papilio was able to find the permissions requested frequently by applications and permissions that either never requested or requested infrequently. The authors in [82] stressed the importance of visualizing the statistical information related to Android permissions. Having graphical representation for the permissions' statics within a specific category encouraged the users to choose more often apps with a lower number of permissions. A privacy meter was used in [83] to visualize the permissions' statistics through a slider bar which outperformed the existing warning system like Google's permission screens. Visualizing app activities enhances user's awareness and sensitivity to the privacy intrusiveness of mobile applications [84]. Another attempt to visualize the permissions statistics was also introduced in [85] using lifelog analysis views in terms of risk history and app's risk view.

From the above-related work, we can observe that the majority of the previous solutions have mainly relied on permissions either statistically or based on probability to analyze Android apps, to classify them as malwares, and to measure their risk level. Although permissions are important to analyze and classify Android applications. However, these permissions should be up-to-date. Also, other important static and dynamic metrics need to be considered to guarantee a comprehensive evaluation and consequently an accurate detection of malware apps.

There have been many types of research on designing malicious detection approaches. Such approaches resort to static analysis of the malware, and others use dynamic analysis, while some methods utilize both static and dynamic analyses to get better detection of a malicious incident. Moreover, the generated datasets will be analyzed in order to detect any potential security threats, regardless whether these datasets were constructed based on static or dynamic tests or even both. Usually, data mining techniques could be used for the purpose of detecting and classifying attacks [42, 52]. Moreover, intelligence techniques could be utilized to even rank the risk by assigning the attack a risk score [42, 86].

The scanning service might fruit in developing a mobile application that is installed on user's devices to examine the Android application and discriminate, if it is a clean app or a malicious app to warn the user and protect her/his Android device. DREBIN [87] is one of the malware detection systems available for smartphones. One of the major features that DREBIN provides is instantaneous malware detection. When a new application is downloaded, DREBIN starts the analyzing process directly. As a result, the user is protected against any unreliable sources. Another example of anti-malware software is HinDroid [88] which has been integrated as one of Comodo's mobile security scanning tools. HinDroid structures the APIs based on heterogeneous information network in order to make predictions about the tested application. Consequently, HinDroid can reduce the time and cost of analyzing Android apps.

5. Statistical analysis

This section presents a statistical study to show the frequency of the used approaches, methods, datasets, and tools in the current systems. Various, related, recent, published solutions in 2017–2018 were considered in this study.

In regard to reverse engineering tools utilized by researchers, APKtool was heavily used by 54% in comparison with other tools (see **Figure 4**). Soot was next with 20% of usage.

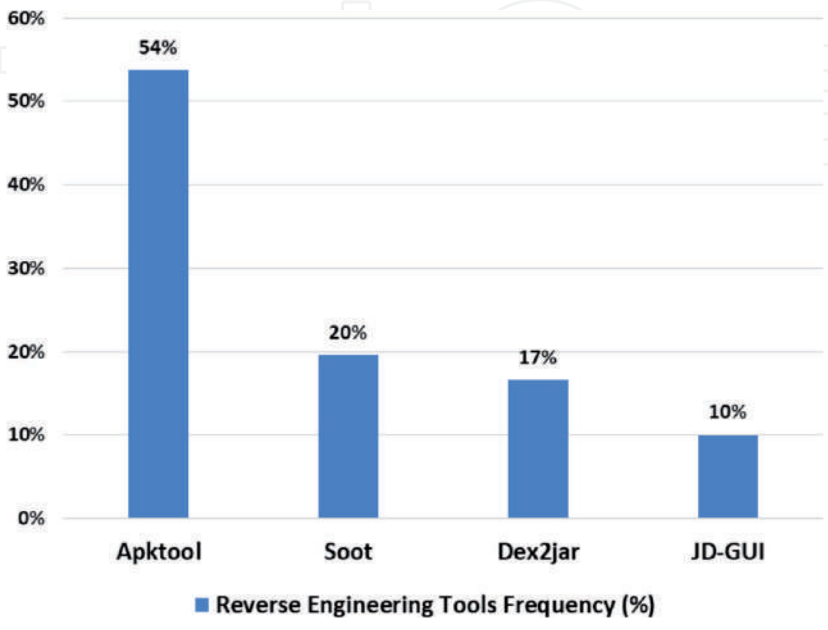

Figure 4.
Reverse engineering tool usage in 2017–2018 research.

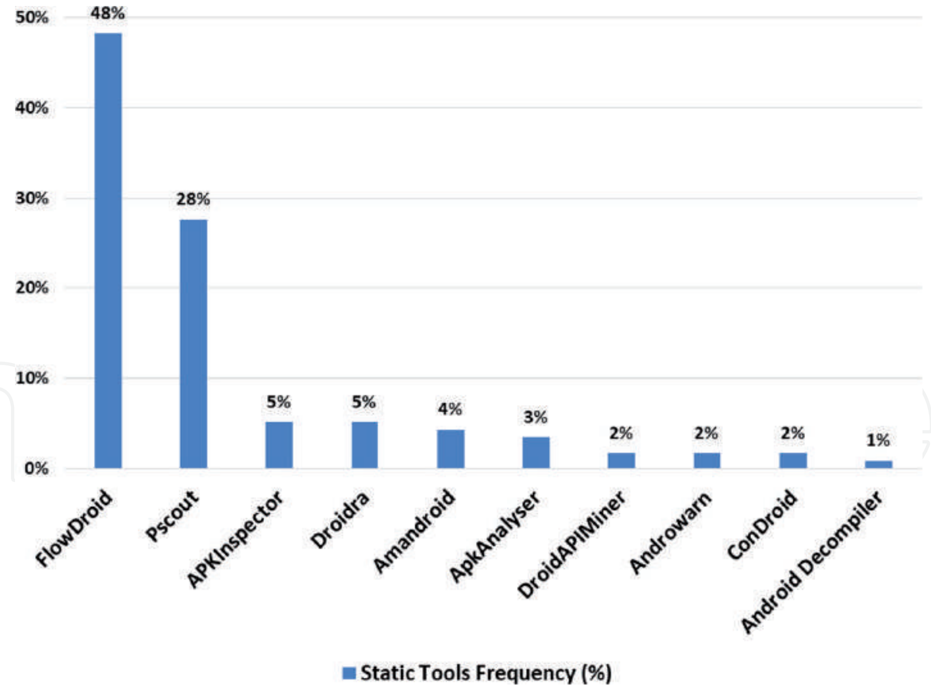

Figure 5.
Static tool frequency in 2017–2018 research work.

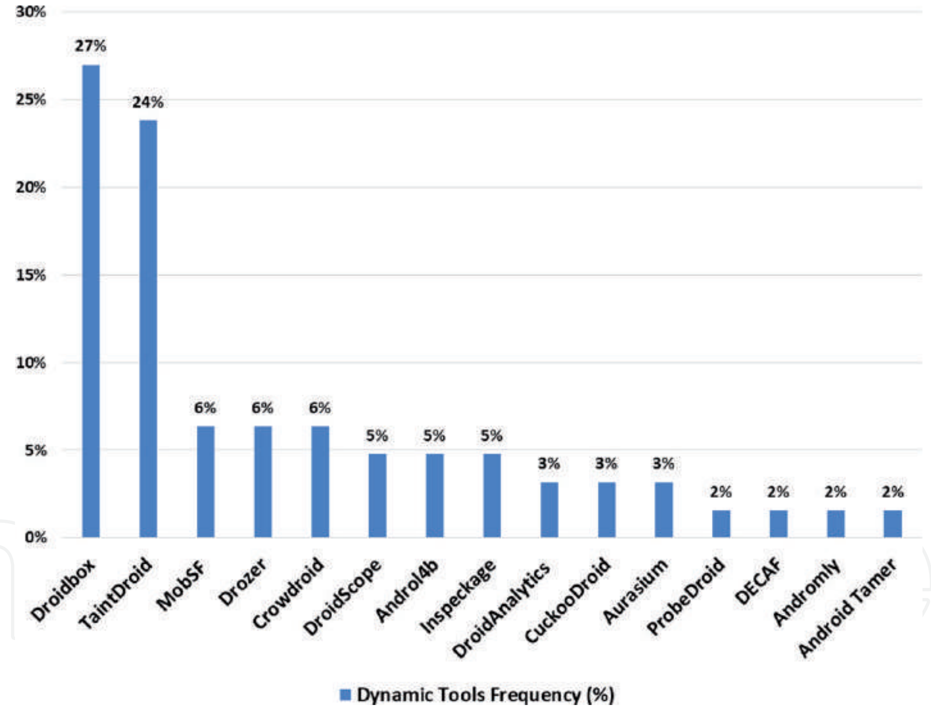

Figure 6.
Dynamic tool frequency in 2017–2018 research work.

Figure 5 shows a comparison among the static tools which were utilized by researchers. It can be observed that 48% of the static-based systems used FlowDroid tool in their solutions. PScout was the second most used with percentage reaching around 28%.

Dynamic analysis tool usage is illustrated in **Figure 6**. The majority of existing solutions used Droidbox with 27% and TaintDroid with 24% in comparison with other approaches. The rest of the results are shown in **Figure 6**.

The results in **Figure 7** reveal that AndroZoo was the most used dataset in 2017–2018. The percentage of usage reached 43%. Genome and DREBIN datasets came next with frequencies 30 and 16%, respectively.

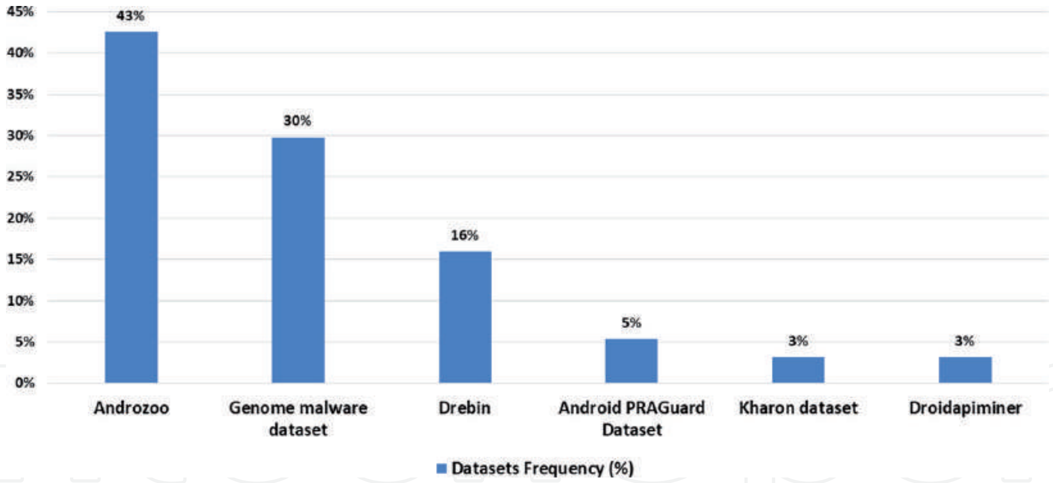

Figure 7.
Dataset frequency in 2017–2018 research work.

6. Conclusions

This chapter highlighted the booming of Android technologies and their applications which make them more attractive to security attackers. Recent statistics of Android malwares and their impact were presented. Additionally, this chapter has provided the main phases required to apply security scanning to Android applications. The purpose is to protect Android users and their devices from the threats of different security attacks. These phases include the way of downloading Android apps, decoding them to generate the source code, and how this code is screened to extract the required features to apply either static analysis or dynamic analysis or both. The feature extraction process resulted in constructing different datasets. Proper data analysis and data mining techniques could be applied to examine the app and classify it as benign or malware with high accuracy. The malware detection service could be implemented and provided in terms of a mobile application that will communicate the scanning results to the user in a friendly way. The chapter was concluded by presenting a statistical study that showed the most used tools and datasets throughout the scanning process for the last 2 years 2017 and 2018.

Acknowledgements

We would like to acknowledge the Security Engineering Lab (sel.psu.edu.sa) team and Prince Sultan University for supporting this work. Special thanks go to Ms. Samah Alsoghyer and Ms. Aala Khayer.

Conflict of interest

The authors declare that there is no “conflict of interest” in regard to publishing this book chapter.

Author details

Iman Almomani[1,2]* and Mamdouh Alenezi[1]

1 Prince Sultan University, Riyadh, KSA

2 The University of Jordan, Amman, Jordan

*Address all correspondence to: imomani@psu.edu.sa

References

[1] Alliance OH. Android overview. Open Handset Alliance. 2011;**8**:88-91

[2] Faruki P, Bharmal A, Laxmi V, Ganmoor GM, Conti M. Android security: A survey of issues, malware penetration, and defenses. IEEE Communications Surveys & Tutorials. 2015;**17**(2):998-1022

[3] Global smartphone sales to end users from 1st quarter 2009 to 2nd quarter 2018 [Internet]. Available from: https://www.statista.com/statistics/266219/global-smartphone-sales-since-1st-quarter-2009-by-operating-system/ [Accessed: 2019-03-22]

[4] Nimodia C, Deshmukh HR. Android operating system. Software Engineering. 2012;**3**(1):10

[5] Brahler S. Analysis of the android architecture. Karlsruhe Institute for Technology. 2010;**7**(8):3-64

[6] Drake JJ, Lanier Z, Mulliner C, Fora PO, Ridley SA, Wicherski G. Android Hacker's Handbook. New Jersey, USA: John Wiley & Sons; 2014

[7] Mithilesh Joshi BlogSpot. What is android application components and how we use it? 2015. https://mithileshjoshi.blogspot.com/2015/06/CITATIONS 74 what-is-android-application-components.html [Accessed November 27, 2017]

[8] AppBrain. Number of Android applications on the Google Play store | AppBrain. 2019. Available from: https://www.appbrain.com/stats/number-of-android-apps [Accessed: 2019-03-02]

[9] AV-TEST The IT-Security Institute. Malware Statistics and Trends Report | AV-TEST. The AV-TEST Institute; 2018. Available from: https://www.av-test.org/en/statistics/malware/ [Accessed: 2019-03-02]

[10] GData. Cyber attacks on Android devices on the rise. 2018. Available from: https://www.gdatasoftware.com/blog/2018/11/31255-cyber-attacks-on-android-devices-on-the-rise [Accessed: 2019-04-15]

[11] F-Secure State of cyber security. Available from: https://www.f-secure.com/documents/996508/1030743/cyber-security-report-2017 [Accessed: 2019-04-15]

[12] Security Report 2016/17 [Internet]. Available from: https://www.avtest.org/fileadmin/pdf/security_report/AV-TEST_Security_Report_2016-2017.pdf [Accessed: 2019-04-15]

[13] Delac G, Silic M, Krolo J. Emerging security threats for mobile platforms, MIPRO. In: 2011 Proc. 34th Int. Conv. 2011. pp. 1468-1473

[14] Savage K, Coogan P, Lau H. Security Response – The Evolution of Ransomware. Mountain View (CA): Symantec Corporation; 2015

[15] Liska A, Gallo T. Ransomware: Defending Against Digital Extortion. California, USA: O'Reilly Media, Inc; 2016

[16] (KasperSky) Chebyshev V. Mobile Malware Evolution 2018 [Internet]. Available from: https://securelist.com/mobile-malware-evolution-2018/89689/ [Accessed: 2019-04-16]

[17] Grégio A, Abed R, Afonso V, Filho D, Geus P, Jino M. Toward a taxonomy of malware behaviors. The Computer Journal. 2015;**58**(10):2758-2777

[18] Alenezi M, Almomani I. Abusing android permissions: A security perspective. IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT); Amman, Jordan. 2017

[19] Developer.Android. "Permission Types" [Internet]. Available from: https://developer.android.com/guide/topics/manifest/permission-element [Accessed: 2019-04-15]

[20] Gianluca D, Martinelli F, Matteucci I, Petrocchi M, Saracino A, Sgandurra D. Risk analysis of android applications: A user-centric solution. Future Generation Computer Systems. 2018;**80**:505-518

[21] Seray T, Demetriou S, Ganju K, Gunter C. Resolving the predicament of android custom permissions. In: ISOC Network and Distributed Systems Security Symposium (NDSS). 2018

[22] Arstechnica.com. Your iPhone Calendar isn't Private. 2012. Available from: http://arstechnica.com/apple/2012/06/your-iphone-c

[23] Haining C, Li N, Enck W, Aafer Y, Zhang X. Analysis of SEAndroid policies: Combining MAC and DAC in android. In: Proceedings of the 33rd Annual Computer Security Applications Conference. Orlando, FL, USA: ACM; 2017. pp. 553-565

[24] Ratazzi EP. Understanding and Improving Security of the Android Operating System. No. AFRL-RI-RS-TP-2018-001. New York, USA: Air Force Research Laboratory/Information Directorate Rome United States; 2016

[25] Backes M, Bugiel S, Hammer C, Schranz O, von Styp-Rekowsky P. Boxify: Full-fledged app sandboxing for stock android. In: Proceedings of the 24th {USENIX} Security Symposium ({USENIX} Security 15). 2015. pp. 691-706

[26] Bennet Y, Sehr D, Dardyk G, Chen J, Muth R, Ormandy T, et al. Native client: A sandbox for portable, untrusted x86 native code. In: Proceedings of the 2009 30th IEEE Symposium on Security and Privacy. Berkeley, CA, USA: IEEE; 2009. pp. 79-93

[27] Elenkov N. Android Security Internals: An in-Depth Guide to Android's Security Architecture. San Francisco, California, USA: No Starch Press; 2014

[28] Georgios P, Homburg P, Anagnostakis K, Bos H. Paranoid android: Versatile protection for smartphones. In: Proceedings of the 26th Annual Computer Security Applications Conference. Austin, Texas, USA: ACM; 2010. pp. 347-356

[29] Almomani I, Alkhayer A. Android applications scanning: The guide. In: Proceedings of the IEEE International Conference on Computer and Information Sciences (ICCIS); 3-4 April 2019; Saudi Arabia. Jouf: IEEE; 2019

[30] Allix K, Bissyandé F, Klein J, Traon, L. AndroZoo. In: Proceedings of the 13th International Workshop on Mining Software Repositories—MSR 16. 2016. DOI: 10.1145/2901739.2903508

[31] Abubaker H. Analytics on malicious android applications. International Journal of Advanced Software Computer. 2018;10:106-118. ISSN 2074-8523

[32] Ikram M, Kaafar M. A first look at mobile ad-blocking apps. In: Proceedings of the 2017 IEEE 16th International Symposium on Network Computing and Applications (NCA). 2017. DOI: 10.1109/nca.2017.8171376

[33] Arnatovich L, Wang L, Ngo N, Soh C. A comparison of android reverse engineering tools via program behaviors validation based on intermediate languages transformation. IEEE Access. 2018:12382-12394. DOI: 10.1109/access.2018.2808340

[34] Baskaran B, Ralescu AA. Study of android malware detection techniques and machine learning. In: Proceedings of the Mod. Artif. Intell. Cogn. Sci. Conf. 2016. pp. 15-23

[35] Arshad S, Ali M, Khan A, Ahmed M. Android malware detection & protection: A survey. International Journal of Advanced Computer Science and Applications. 2016;**7**(2):463-475. DOI: 10.14569/ijacsa.2016.070262

[36] Zachariah R, Yousef M, Chacko A. Android malware detection and prevention. International Journal of Recent Trends in Engineering and Research. 2017;**3**(2):213-217. DOI: 10.23883/ijrter.2017.3028.0uhbl

[37] Baskaran B, Ralescu A. A study of android malware detection techniques and machine learning. In: Proceedings of the IEEE International Conference on Smart Internet of Things (SmartIoT). 2018. DOI: 10.1109/smartiot.2018.00034

[38] Chen J, Wang C, Zhao Z, Chen K, Du R, Ahn G. Uncovering the face of android ransomware: Characterization and real-time detection. IEEE Transactions on Information Forensics and Security. 2018;**13**(5):1286-1300

[39] Arzt S, Rasthofer S, Fritz C, Bodden E, Bartel A, Klein J, et al. Flowdroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for android apps. Acm Sigplan Notices. 2014;**49**(6):259-269

[40] Au K, Zhou Y, Huang Z, Lie D. Pscout: Analyzing the android permission specification. In: Proceedings of the 2012 ACM Conference on Computer and Communications Security; October 2012. Raleigh, North Carolina, USA: ACM; 2012. p. 217-228

[41] Mujahid S, Abdalkareem R, Shihab E. Studying permission related issues in android wearable apps. In: Proceedings of the IEEE International Conference on Software Maintenance and Evolution (ICSME); September 2018. Madrid, Spain: IEEE; 2018. pp. 345-356

[42] Tao G, Zheng Z, Guo Z, Lyu M. MalPat: Mining patterns of malicious and benign android apps via permission-related APIs. IEEE Transactions on Reliability. 2018;**67**(1):355-369. DOI: 10.1109/tr.2017.2778147

[43] Alenezi M, Almomani I. Empirical analysis of static code metrics for predicting risk scores in android applications. In: Proceedings of the 5th Symposium on Data Mining Applications (SDMA2018); 21-22 March, 2018; Riyadh, KSA

[44] Enck W. TaintDroid: An information-flow tracking system for realtime privacy monitoring on smartphones. In: Proceedings of the ACM Transactions on Computer Systems (TOCS) 32.2. 2014. p. 5

[45] Zhou Y"H. You, get o_ of my market: Detecting malicious apps in official and alternative android markets. In: Proceedings of the NDSS. 2012. pp. 50-52

[46] Grace M. Riskranker: Scalable and accurate zero-day android malware detection. In: Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services. ACM; 2012. pp. 281-294

[47] Security Affairs. DREBIN Android app detects 94 percent of mobile malware [Internet]. Available from: http://securityaffairs.co/wordpress/29020/malware/drebin-android-av.html [Accessed: 2017-12-01]

[48] Arp D. DREBIN: Effective and explainable detection of android malware in your pocket. Proceedings of the NDSS. 2014

[49] Yang W. Appcontext: Differentiating malicious and benign mobile app behaviors using context. In: Proceedings of the Software Engineering (ICSE); 2015 IEEE/ACM 37th IEEE International Conference. IEEE. 2015. pp. 303-313

[50] Yang W. Malware detection in adversarial settings: Exploiting feature evolutions and confusions in android apps. Proceedings of the Proc. ACSAC. 2017

[51] Akhuseyinoglu N, Akhuseyinoglu A. AntiWare: An automated Android malware detection tool based on machine learning approach and official market metadata. In: Proceedings of the 2016 IEEE 7th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON). October 2016. pp. 1-7. DOI: 10.1109 /UEMCON.2016.7777867

[52] Allix K, Bissyandé T, Klein J, Traon Y. AndroZoo: Collecting millions of android apps for the research community. In: Proceedings of the 13th International Workshop on Mining Software Repositories—MSR. 2016. DOI: 10.1145/2901739.2903508

[53] Zachariah R, Akash K, Yousef M, Chacko A. Android malware detection a survey. In: Proceedings of the 2017 IEEE International Conference on Circuits and Systems (ICCS). 2017. pp. 238-244

[54] Kaspersky Lab. "Emulator." [Internet]. Available from: https://www.kaspersky.com/enterprise-security/wiki-section/products/emulator [Accessed: 2019-04-01]

[55] Tam K. CopperDroid: Automatic reconstruction of android malware behaviors. Proceedings of the NDSS. 2015

[56] Bhatia T, Kaushal R. Malware detection in android based on dynamic analysis. In: Proceedings of the 2017 International Conference on Cyber Security And Protection Of Digital Services (Cyber Security). 2017. pp. 1-6. DOI: 10.1109/CyberSecPODS.2017.8074847

[57] Wang ZD-ARM. Taming control flow anti-analysis to support automated dynamic analysis of android malware. In: Proceedings of the 33rd Annual Conference on Computer Security Applications (ACSAC'17). 2017

[58] Huang H, Zheng C, Zeng J, Zhou W, Zhu S, Liu P, et al. A large-scale study of android malware development phenomenon on public malware submission and scanning platform. IEEE Transactions on Big Data. 2018: 15-23. DOI: 10.1109/tbdata.2018.2790439

[59] Kaur R, Li Y, Iqbal J, Gonzalez H, Stakhanova NA. Security assessment of HCE-NFC enabled E-wallet banking android apps. In: Proceedings of the 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC); July 2018. Tokyo, Japan: IEEE; 2018. pp. 492-497

[60] VirusTotal Malware Intelligence Services [Internet]. n.d. Available from: https://www.virustotal.com/learn/ [Accessed 2018-12-15]

[61] Scaife N. Cryptolock (and drop it): Stopping ransomware attacks on user data. In: Proceedings of the Distributed Computing Systems (ICDCS), 2016 IEEE 36th International Conference on. Nara, Japan: IEEE; 2016. pp. 303-312

[62] Tripwire. Early-Warning Ransomware Detection Tool Could Help Protect Users Despite Drawbacks [Internet]. Available from: https://www.tripwire.com/state-of-security/security-data-protection/cyber-security/early-warning-ransomware-detection-tool-could-help-protect-users-despite-drawbacks/ [Accessed 2017-12-02]

[63] Mercaldo F, Nardone V, Santone A. Ransomware inside out. In: Proceedings of the 2016 11th International Conference on Availability, Reliability and Security. ARES; 2016. pp. 628-637. DOI: 10.1109/ARES.2016. 35

[64] Li J, Sun L, Yan Q, Li Z, Srisa-an W, Ye H. Significant Permission

Identification for Machine-Learning-Based Android Malware Detection, IEEE Transactions on Industrial Informatics; July 2018;**14**(7):3216-3225

[65] Mercaldo F, Nardone V, Santone A. Ransomware inside out. Proceedings of the 11th International Conference on Availability, Reliability and Security (ARES); Augest. 2016:628-637. DOI: 10.1109/ARES.2016.35

[66] Sun L, Wei X, Zhang J, He L, Yu PS, Srisa-an W. Contaminant removal for Android malware detection systems. Boston, MA, USA: 017 IEEE International Conference on Big Data (Big Data); 11-14 Dec 2017. pp. 1053-1062

[67] Maiorca D. R-PackDroid: API package-based characterization and detection of mobile ransomware. In: Proceedings of the Symposium on Applied Computing. Marrakech, Morocco: ACM; 2017. pp. 1718-1723

[68] Mercaldo F. Ransomware steals your phone. Formal methods rescue it. In: Proceedings of the International Conference on Formal Techniques for Distributed Objects, Components, and Systems. Heraklion, Greece: Springer; 2016. pp. 212-221

[69] Mercaldo F. Extinguishing ransomware-a hybrid approach to android ransomware detection. In: Proceedings of the 10th International Symposium on Foundations Practice of Security. 2017

[70] Yang T. Automated detection and analysis for android ransomware. In: Proceedings of the High Performance Computing and Communications (HPCC), 2015 IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), 2015 IEEE 12th International Conference on Embedded Software and Systems (ICESS), 2015 IEEE 17th International Conference on. IEEE; 2015. pp. 1338-1343

[71] Gharib A, Ghorbani A. DNA-droid: A real-time android ransomware detection framework. In: Proceedings of the International Conference on Network and System Security. Helsinki, Finland: Springer; 2017. pp. 184-198

[72] Sarma BP, Li N, Chris Gates C, Potharaju R, Nita-Rotaru C, Molloy I. Android permissions: a perspective combining risks and benefits. In: Proceedings of the 17th ACM symposium on Access Control Models and Technologies. Newark, New Jersey, USA; 20-22 June 2012. pp. 13-22

[73] Enck W, Ongtang M, McDaniel P. On lightweight mobile phone application certification. In: Proceedings of the 16th ACM Conference on Computer and Communications Security. Vol. 2009. pp. 235-245

[74] Gates C, Li N, Peng H, Sarma B, Qi Y, Potharaju R, et al. Generating summary risk scores for mobile applications. IEEE Transactions on Dependable and Secure Computing. 2014;**11**(3):238-251

[75] Mathew J, Joy M. Efficient risk analysis for android applications. In: Proceedings of the IEEE Recent Advances in Intelligent Computational Systems (RAICS); 10-12 December 2015. Trivandrum, India: IEEE; 2015. pp. 382-387

[76] Guo C, Xu G, Liu L, Xu S. Using association statistics to rank risk of android application. In: Proceedings of the IEEE International Conference on Computer and Communications (ICCC); 10-11 October 2015. IEEE; 2015. pp. 1-5

[77] Hao H, Li Z, Yu H. An effective approach to measuring and assessing the Risk of android application. Proceedings of the International Symposium on Theoretical Aspects of Software Engineering (TASE); 12-14 Sept. 2015:31-38

[78] Wang Y, Zheng Y, Sun C, Mukkamala S. Quantitative security risk assessment of android permissions and applications. In: Proceedings of the Lecture Notes in Computer Science, LNCS-7964. Newark, NJ, USA: Springer; 2013. pp. 226-241

[79] Yuksel A, Yuksel E, Sertbasa A, Zaim A. Implementation of a web-based service for mobile application risk assessment. Turkish Journal of Electrical Engineering & Computer Sciences. 2017;**25**(2):976-994

[80] Merlo A, Georgiu G. RiskInDroid: Machine learning-based risk analysis on android. In: Proceedings of the IFIP International Conference on ICT Systems Security and Privacy Protection (SEC). 2017. pp. 538-552

[81] Hosseinkhani M, Fong P, Papilio CS. Visualizing android application permissions. In: Proceedings of the Eurographics Conference on Visualization. 2014. pp. 1-10

[82] Kraus L, Wechsung I, Möller S. Using statistical information to communicate android permission risks to users. Workshop on Socio-Technical Aspects in Security and Trust (STAST). Vienna, Austria: Co-located with 27th ieee computer security foundations symposium (CSF); 18 July 2014. pp. 1-9

[83] Kang J, Kim H, Cheong YG, Huh JH. Visualizing privacy risks of mobile applications through a privacy meter. In: Information Security Practice and Experience. Lecture Notes in Computer Science. Vol. 9065. Beijing, China: Springer; 2015. pp. 548-558

[84] Eze C, Nurse J, Happa J. Using visualizations to enhance users' understanding of app activities on android devices. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications. 2016:39-57

[85] Yoo S, Ryu H, Yeon H, Kwon T, Jang Y. Personal visual analytics for android security risk lifelog. In: Proceedings of the 10th International Symposium on Visual Information Communication and Interaction. 2017. pp. 29-36

[86] Dash S, Suarez-Tangil G, Khan S, Tam K, Ahmadi M, Kinder J, et al. DroidScribe: Classifying android malware based on runtime behavior. In: 2016 IEEE Security and Privacy Workshops (SPW). 2016. DOI: 10.1109/spw.2016.25

[87] Arp D. Drebin: Effective and explainable detection of android malware in your pocket. In: Proceedings of the 2014 Network and Distributed System Security Symposium. 2014. DOI: 10.14722/ndss.2014.23247

[88] Hou S. HinDroid. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD 17. Halifax, NS, Canada: ACM; 2017. DOI: 10.1145/3097983.3098026

Chapter

Deployment of PON in Europe and Deep Data Analysis of GPON

Tomas Horvath, Petr Munster and Josef Vojtech

Abstract

This chapter discusses the extensibility of fiber to the x (FTTx) households, specifically in the territory of the European Union. The Czech Republic has made a commitment to other member states to provide connectivity of at least 100 Mbit/s for half of the households by 2020. Although Internet access in the Czech Republic is mostly dominated by wireless fidelity (WiFi), this technology is not capable of meeting the demanding current demands at a reasonable price. As a result, passive optical networks are on the rise in access networks and in mobile cell networks by fiber to the antenna (FTTA). Passive optical networks use much more complex networks. In cooperation with Orange Slovakia, the analysis of the transmitted data was conducted. The optical network unit management and control interface (OMCI) channel data, as well as the activation data associated with specific end units, were analyzed. We propose a complete analysis of the end-unit-related activation process, download, and initialization of the data image for setting the end units and voice over Internet protocol (VoIP) parameters. Finally, we performed an analysis of the transmission of dying gasp messages.

Keywords: dying gasp, GPONxpert, OMCI channel analysis, ONU activation process analysis, PON deployment, transmission convergence layer

1. Introduction

The optical infrastructure is essential for current applications that demand a high bandwidth [1–3]. The International Telecommunication Union (ITU) has been developing standards for passive optical networks (PONs) for over 20 years [4, 5]. The second most active organization in this area is the Institute of Electrical and Electronics Engineers (IEEE) [6, 7].

Passive optical networks are currently expanding, as the European Union (EU) has allocated budget to extend the coverage of these networks [8]. Today, the access network is not only about transferring data streams from/to the Internet. The popularity of Amazon TV, Netflix, and so on puts increased demands on bandwidth. Current transmission speeds are not sufficient, and a bandwidth of at least 100 Mbit/s in every household is still under consideration. In the Czech Republic, the utilization of gigabit PON (GPON) standard still dominates. However, such standard was in its first version approved back in 2003 [9]. This standard makes it possible to achieve a bandwidth of up to 2.5 Gbit/s in full duplex mode, but the disadvantage is that the bandwidth is fully shared by all end users (in theory, up to 128 customers per port). The available bandwidth can be operatively changed in time and according to the requirements using dynamic bandwidth allocation (DBA) algorithms [10–13]. The decreasing cost of the necessary devices allows GPON optical line termination (OLT) to be used more often for service providers; on the other hand, the standard in use may not be sufficient for the future. The cost of the next-generation PON (XG-PON) terminal units is still quite high, regardless of the OLT unit price. The price of the technology itself is determined by the price of the optical network unit (ONU) terminal units. The advantage of deploying next-generation networks would be the ability to a share the transfer rate of up to 10 Gbit/s. Together with appropriate DBA algorithms, the full bandwidth utilization or its adequate distribution between endpoints would be efficiently used. GPON networks theoretically allow us to transfer data up to 19 Mbit/s for each ONU (considered for the maximum transfer rate and a split ratio of 1:128). XG-PON networks are limited by higher split ratios but have higher transfer rates available. Theoretically, 39 Mbit/s can be achieved for each ONU. In other words, the transfer rates are the maximum possible in both GPON and XG-PON networks. Usually, the guaranteed transfer rates are several times lower according to the use of a transmission container (T-CONT) [14].

2. Current state of the access networks in the Czech Republic

The Czech Republic has committed itself within the European Union to ensuring a transmission rate of at least 30 Mbit/s toward users by 2018 [15]. In 2020, the next milestone is going to be to increase the downlink speed up to 100 Mbit/s for approximately half of all households [15]. Both variables account for asymmetric transmission rates (usually higher transmission rates in the downstream direction). Based on [16], this “scarcity” should be eliminated by 2030. At that time, only a symmetric variant of Internet access will be considered.

Current technologies such as asymmetric digital subscriber line (ADSL) are no longer able to meet the previously mentioned bandwidth requirements. The plans of the Czech Republic include a GPON or a variant of an active optical network. As presented in [17], the formal definition for next-generation networks can be defined as follows: next-generation networks (NGNs) are networks based on data packet transfer technologies capable of providing electronic communications services, allowing for the use of various high technologies that are able to manage and control the quality of the provided services, and whose functions related to these services are independent of basic transmission technologies. The network provides subscribers with unlimited access to various providers of publicly available electronic communications services and consistently supports the provision of services to subscribers at any point in the network. Additionally, next-generation networks can be split into backbone and access networks. This work, however, deals exclusively with access networks.

On the other hand, the Czech Republic is not entirely prepared to satisfy the high demands on the connection speed in all locations. Based on [18], the dominant transmission rates were mostly up to 10 Mbit/s. No significant growth of higher transmission rates has been recorded.

In 2016, the Czech Telecommunication Office published an annual report summarizing current technologies for Internet access. The associated graph can be seen in **Figure 1**. **Figure 1** clearly shows that the dominant technology in this area is wireless fidelity (WiFi) (26.8%), i.e., wireless transmission of information. The annual report does not include the frequencies used; however, the basic frequencies in the

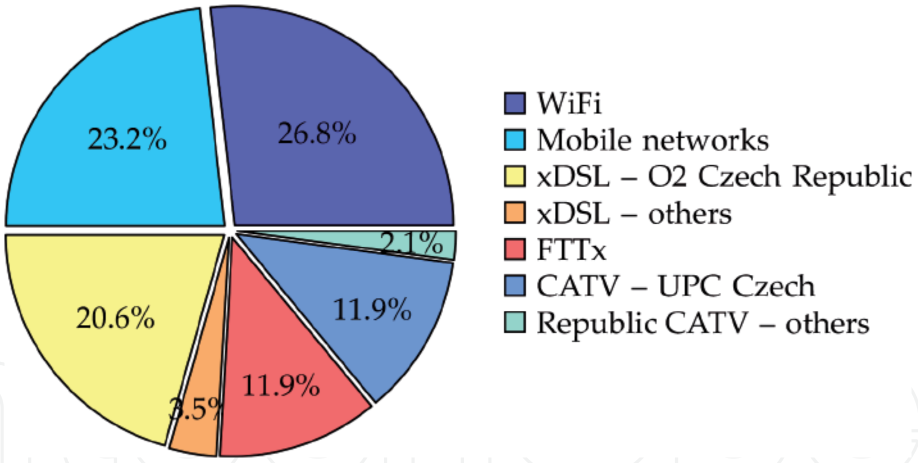

Figure 1.
Access technologies market share in the Czech Republic [15].

license-free band (2.5/5 GHz) can be assumed. The second technology with the highest penetration is represented by mobile networks (23.2%). The third technology combines all types of xDSL technologies. This area is dominated by Telefonica O2 Czech Republic, a.s., with a penetration of 20.6%. Other xDSL technologies only reach 3.5%. According to [16], fiber to the x (FTTx) connections at the same value of penetration (11.9%) as cable operator UPC Czech Republic, s.r.o., currently offers the fastest connection speed of 500/30 Mbit/s (depending on the location). Conversely, FTTx connections depend only on the selected standard as with the fiber to the home (FTTH) variant. FTTx connections can support transmission rates up to 10/10 Gbit/s (depending on the number of end units connected to the OLT control unit).

The properties of the next-generation access networks can be summarized as follows [16]:

- providing high transmission rates for subscribers and providing reliable services through optical networks or other comparable technologies,
- supporting a variety of advanced digital and converged services based on Internet protocol (IP),
- providing significantly higher transmission rates in the downstream direction, i.e., toward the user.

3. Household penetration

The current state of fiber to the building (FTTB) or FTTH connections is generally problematic to analyze. These data are usually not freely available, and the cost of these documents is high (on the order of thousands of dollars). A company named IDATE has published its market research for the FTTH Council Europe conference [19]. The outcome of the analysis for Europe clearly shows that Latvia has the best FTTB/H (households) connection (see **Figure 2**). Their household penetration is approximately 50.6% (25.3% are FTTH connections). Another dominant country is Sweden, with a total penetration of 43.3% (only 8.5% are FTTH connections). The total penetration for the Czech Republic is very low compared to other countries, with a total penetration of 3.7% (only 2% are FTTH connections). Compared to the neighboring state, Slovakia has an overall penetration of 17.7%

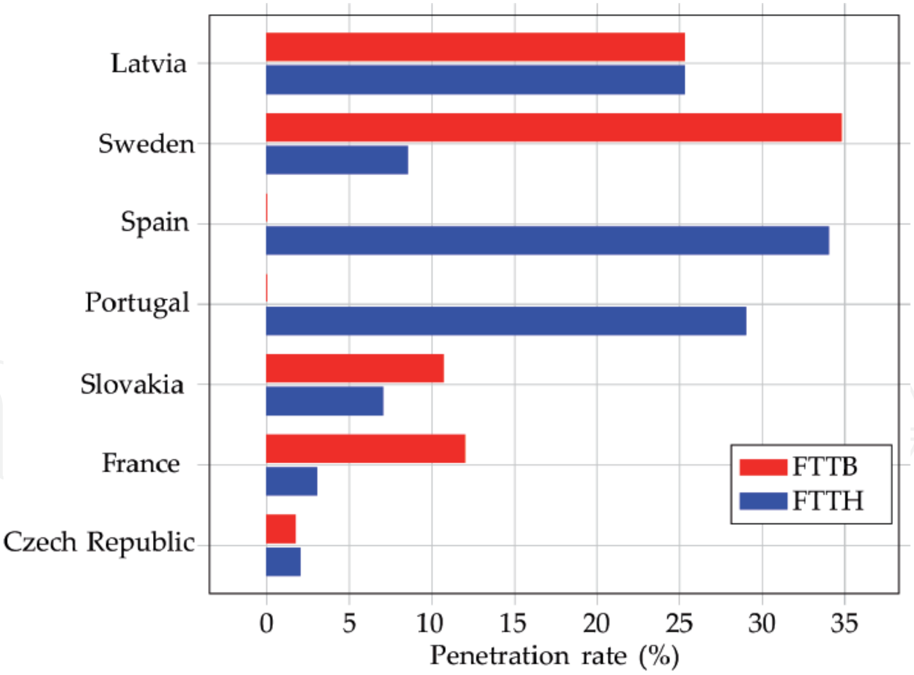

Figure 2.
Selected EU states with FTTH/B penetration [19].

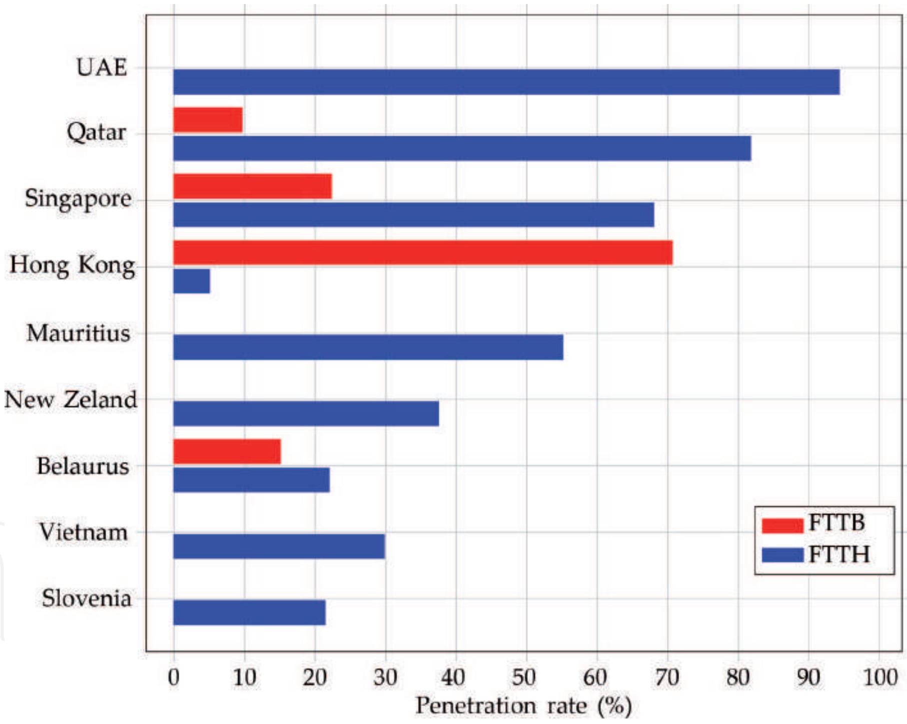

Figure 3.
Selected global states with FTTH/B penetration [19].

(7% are FTTH connections, and the remaining 10.7% are FTTB connections). This is mainly because in Slovakia, there is a very strong operator, Orange SK. Orange SK may test the use of new technologies in this relatively small market, and if this technology stands, it can be deployed, for example, in Orange home (formally France Telecom) in France.

Another objective of the current FTTB/H connection analysis is to focus on the global market (see **Figure 3**). Globally, the United Arab Emirates (UAE) has a total penetration of 94.3%. This penetration is completely composed of FTTH

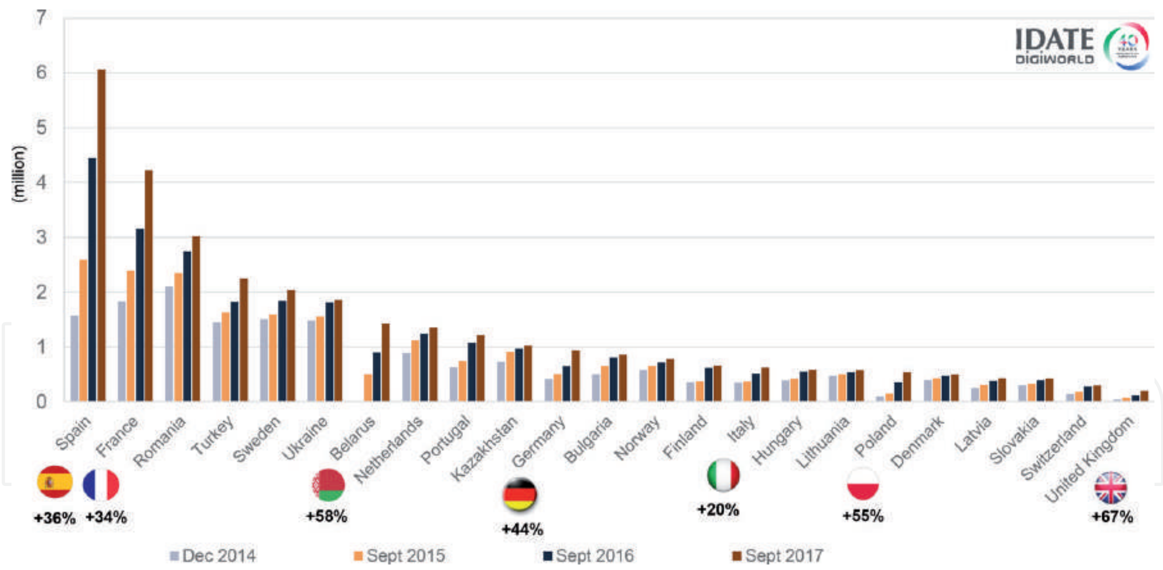

Figure 4.
Progress in FTTH/B deployments according to IDATE [19].

connections. Strong competitors for the UAE are Qatar and Singapore. Qatar has a penetration of 81% for FTTH connections and of 9.4% for FTTB connections. Singapore has a similar total penetration, but in a different ratio, 68% for FTTH connections and 22.3% for FTTB connections. FTTH-only countries are Mauritius, New Zealand, Spain, Vietnam, Portugal, Slovenia, Jamaica, Saudi Arabia, Australia, Macedonia, Switzerland, Oman, Kuwait, Chile, Ecuador, Colombia and Angola. The total penetration of the last 4 states does not exceed 5% [19].

The total penetration is strongly dependent on the number of individual connections. IDATE focused on the analysis of the global market and the comparison of the state of connections in buildings in four stages, December 2014, September 2015, September 2016, and September 2017 (see **Figure 4**). The largest increase in connections was in Poland, with a total difference of 46%. Italy was the second country with the largest increase in connections (35%), followed by Great Britain and France (31%), Spain (24%), and Portugal (22%). Unfortunately, the Czech Republic was not included in this analysis because the number of connections is not as significant. In other words, the trend of building connections is greater in Belarus, Norway, Lithuania, and Hungary.

A detailed view of the number of FTTB/H customers can be seen in report [20]. The report shows that at the end of 2010, the total number of customers was balanced across the EU28 and the commonwealth of independent states (CIS). From a wider perspective, the EU39 reached approximately 8 million customers. However, this difference must be attributed, in particular, to 11 other countries that are counted in the EU39. The aligned trend between the EU28 and the CIS was maintained until 2015. Later, the number of customers increased in the EU28, and the previous dominance of the CIS was diminished. In September 2017, the total number of customers was approximately 25 million, while for the CIS "only," it was 20.5 million. Most places for customers are connected to the provider's network, but there have also been new locations for housing, new towns, and satellite residences created. During the preparation of the work, developers are working hard to build a data infrastructure and negotiations are taking place between Internet services providers (ISPs) and developers. EXFO defines these connections as home passed: premises to which an operator has the capability to connect in a service area, but the premises may or may not be connected to the network [21].

4. Access networks and 5G networks

The primary determination of all technologies for xPON is evident from their name, a passive optical (access) network. This trend continues from the original asynchronous transfer mode PON (APON), broadband PON (BPON), GPON, XG-PON, and the latest approved next-generation PON stage 2 (NG-PON2) recommendations. The latest recommendation has become the pioneer of extending the passive optical network to mobile customers as well. However, residential customers with a fixed connection (flat or house) still remain the priority. With the onset of 5G technology in mobile communications, it will be necessary to reduce the area of cells to ensure coverage of the entire territory by radio signals. This is mainly due to the increasing permeability and diminishing cell size, so it is necessary to build more cells that cover the same area. It is possible to divide the area according to its antenna density into low density (<20 small cells/km^2), medium density (<75 small cells/km^2), dense (<200 small cell/km^2), and ultrahigh density (>200 small cells/km^2). Current long-term evolution (LTE) technology has been providing broadband data services; however, these technologies seem to be inadequate for certain services (virtual reality or generally the most sensitive services for low latency, such as access to data networks of the Internet of things devices). Current customer needs may include gigabit transmissions per second, smart home/buildings, self-driving car, working and playing in the cloud, and 3D or UHD video. Minimal latency requirements will be determined mainly based on data transmission within the national network (10–200 km). The transmission delay in the current networks ranges from 5 to 41 ms, and the delay for the access part of the network (1–10 km) is approx. 7–12 ms. Another key factor that affects the delay is the time it takes to process incoming requests from a data center (approximately 8 ms). The round-trip time (RTT) of current networks is approximately 106 + 8 ms. 5G networks aim to limit this value to 14 + 8 ms. The major merit of RTT depreciation will be to move cloud services closer to the user. Then, the RTT will be reduced to 14 ms, which will primarily generate a delay (7 ms) on the access technology. However, the question remains how the operators will move the data centers closer to the customer, since until now, a distance of 200 km a data center from the customer has been enough. Such a distance is not sufficient for 5G networks.

Among the available technologies covering the 5G signal area, there are technologies for access networks: G.fast, data over cable service interface specification (DOCSIS) and NG-PON2. G.fast technology offers symmetric transmission speeds of up to 500 Mb/s over a short distance (up to 100 m). This speed can be increased to 10 Gb/s, but the overall system reach will be shortened. In theory, G.fast can only be deployed in special cases, such as brownfield scenarios, to ensure connectivity of very small cells in buildings. The basic prerequisite is the combination of functions within the baseband unit (BBU) and remote radio unit (RRU). DOCSIS 3.1 offers bandwidth of 10/1–2 Gb/s share per coaxial segment (192 MHz orthogonal frequency-division multiplexing (OFDM) channels). Full-duplex communication (current downstream and upstream) can take up to 10 Gb/s per coaxial segment. However, neither of these methods is capable of fully serving the 5G network because the available bandwidth is shared and the common public radio interface (CPRI) does not support the lowest possible latency for transmission.

The basic idea behind the NG-PON2 network is to provide all end stations with sufficient bandwidth. The station shares the total bandwidth that the associated OLT unit is able to handle properly. NG-PON2 network parameters such as distribution ratios, power levels, transfer rates, etc. are described in [22–25]. In 5G network areas, there is ultradense deployment of basic radio stations required,

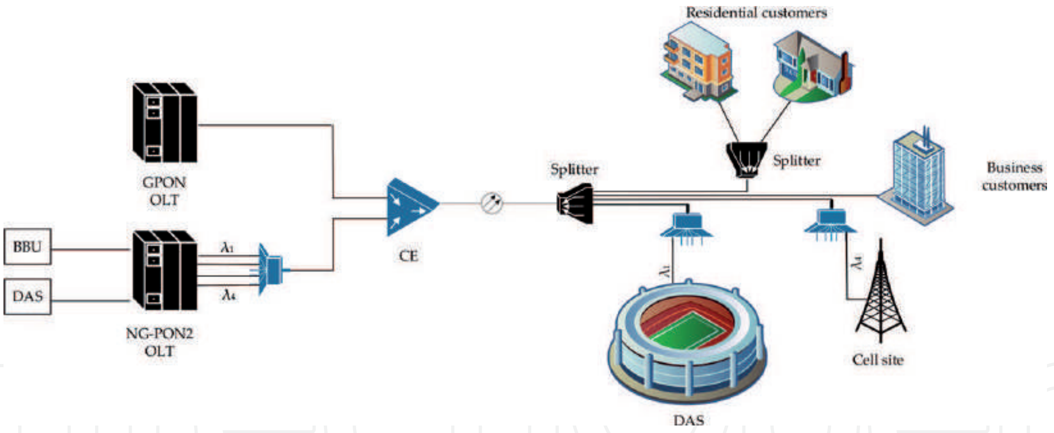

Figure 5.
Coexistence NG-PON2 and GPON scheme with dedicated lambdas for 5G networks.

and their radiations are constrained to prevent intra- and inter-cell interference. In general, the reach of NG-PON2 (up to 20 km from the OLT) is sufficient for covering an acceptable number of end users and for effective usage of its coverage (the division of covered territory into several smaller sectors/cells). The use of access technologies for data transfers or generally for triple play has already been noted out by ITU in [26]. **Figure 5** defines a possible scheme of the NG-PON2 network for its connection to the 5G network. The connection can be realized by dedicated wavelengths (λ). By using a coexistence element (CE), such a coexistence scheme for older PON standards under the ITU recommendations can be established. Regarding the aforementioned dedicated wavelengths, up to 4λ with a 10 Gbit/s transfer rate is considered. One disadvantage of this radio tower connection method is the custom lock method that is publicly available but is much more complex than in the case of the IEEE network. As a result, it will be necessary to use the conversion station to transmit the signal from the radio station toward the end customers.

5. GPON frame structure and activation process analysis

At present, GPON is one of the most promising solutions for modern access networks. Among other useful and important features, it provides us with triple play services on a single optical fiber, good scalability, DBA, simple topology management, etc. In comparison with the previous standards that only supported transmission over asynchronous transfer mode (ATM), GPON is the first standard that supports transmission over both ATM and ethernet technologies. In the ethernet mode, the ethernet frames are encapsulated using GPON encapsulation mode (GEM) and transferred inside GEM frames. As a result, some ethernet structures, such as interpacket gap, preamble, or start of frame delimiter, are not available. For more information, see **Figure 6**.

The basic GPON topology comprises the following three components: OLT, ONU, and optical distribution network (ODN). Typically, there is/are a single/more OLT/s in the network (depending on the preferences of the associated Internet service provider) performing encapsulation and de-encapsulation of downstream and upstream network traffic, respectively, for multiple end users (up to 128 end users per port). The ONU is located at the end user's premises and converts the signals from the optical to the electrical domain. Finally, an ODN is composed of the elements placed between OLTs and ONUs such as optical fibers, splitters, and connectors.

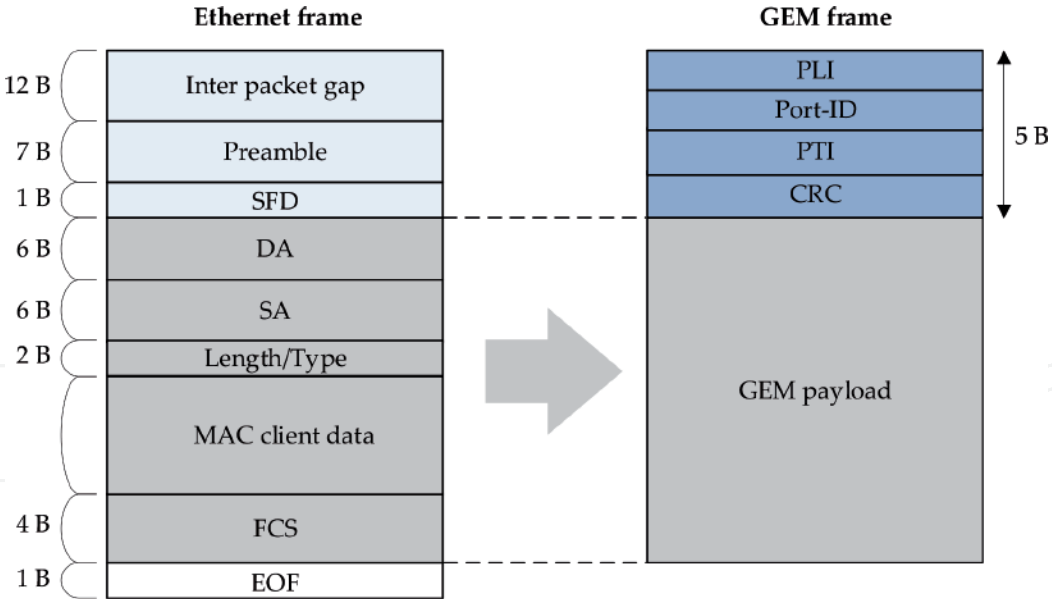

Figure 6.
Ethernet encapsulation into the GEM frame [26].

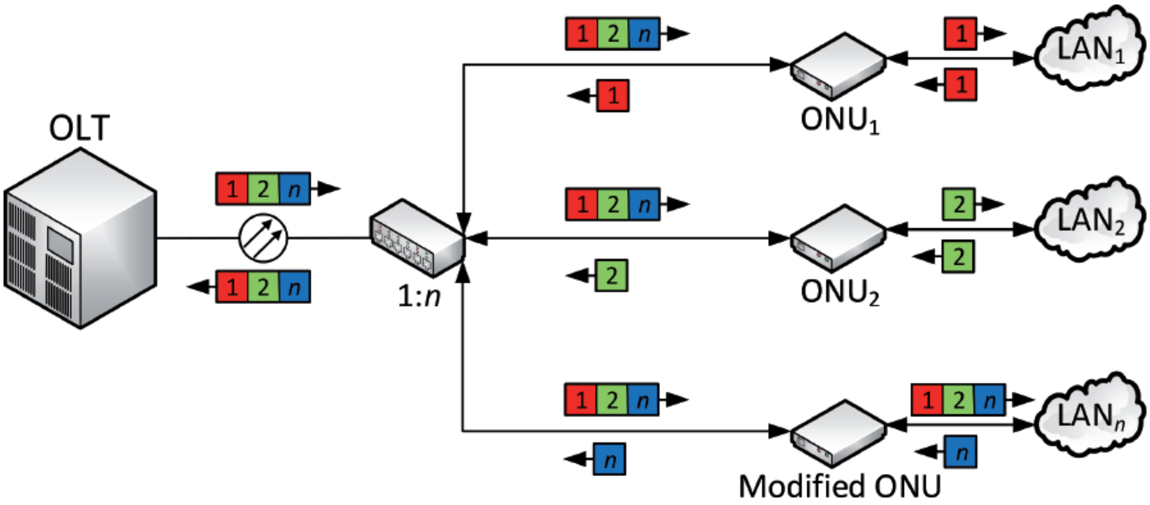

Figure 7.
Interception of downstream communications.

The risk of passive interception of communications results directly from the nature of PON communication. Downstream communication can be secured; however, the major disadvantage is that security is only optional. A potential attacker could, therefore, modify the firmware of an ONU and eavesdrop on all the communication in the downstream direction [26, 27]. The traffic in this direction can also be captured using optical radiation detectors, not necessarily an ONU detector, so encryption of data in the downstream direction had to be introduced [28]. However, the subsequent processing of the captured signal is an essential next step. The situation where the modified end unit receives all frames, including those not directly assigned to it, can be seen in **Figure** 7.

The previously mentioned passive interception could also occur in the upstream direction because no security is used for the upstream communication. This type of interception is complicated; however, it is feasible. The recommendations for use do not define any security for this direction of communication. The reason for this is based on the fact that it is not possible to capture the communication of other end users in the upstream direction via the ONU, so communication is not necessary to be encrypted. To eavesdrop on the communications in this direction, a potential attacker would have to disrupt the PON optical line. This situation would, however, affect the transmission properties of the network in question, which should be captured by the service provider's surveillance center. This way of interception is therefore very unlikely [29].

The abovementioned reason resulted in the fact that no security standard has been provided for any of the individual PON standards. In the event of encryption of the downstream transmission, e.g., using advanced encryption standard (AES) or other secret key-based technology, these keys would have to be sent in an unsecured form—plain text in the upstream direction. It was based on the assumption that upstream communication was safe; therefore, it was not necessary to provide any additional security [30].

The research described in [31] focused specifically on the possibilities of interception of the communication in the upstream direction. The authors tested whether it was possible to intercept the communication through the back reflections of the optical signal. These reflections could be caused by a variety of commonly used optical components, such as passive optical hubs and/or connectors. Moreover, the optical positive-intrinsic-negative (PIN) detectors and avalanche photodiode (APD), as well as the preamplifiers, also had an effect on capturing the communications in the upstream direction. Testing was carried out at various ODN configurations, mainly aimed at testing the back reflection of the optical signal. The success of the potential attacker depended primarily on the type of connector used and the photodetector. A polished connector (PC) was considered inappropriate in terms of network security. The angled polish connector (APC) reduced signal reflections by virtual vertical grinding. Using an APD connector, however, increased the probability of a successful interception of the communicating ONU. Nevertheless, the capability of eavesdropping in the upstream direction was not dependent on the particular bit rate; it depended mostly on the power level of the retroreflection and the type of connector in use [31].

The following demonstrates how to intercept communication in both directions with a specialized tool in hand. Real-time network analysis of the transmitted data (ONU management and control interface (OMCI) channel and GEM data units for end units) was performed. For the purpose of the demonstration, the GPONxpert tool was used. This tool has been developed specifically for passive optical networks. The tool allows for the real-time analysis of ONU-ID, performance levels, and Alloc-ID. However, a detailed analysis of the transmitted data is still necessary to be implemented in the form of postprocessing. Although the manufacturer, TraceSpan, also has other modifications to this device, for our purposes, the most popular measuring device was used. The lite versions contained support for ONU-ID analysis. The real-time analysis of levels, Alloc-IDs, and other parameters was stored using field programmable gate array (FPGA) and sent to the device manufacturer for the postprocessing. The manufacturer then sent the report from the measurement back to the customer.

This work is focused on the analysis of downstream and upstream transmission in GPON standard topology. At the start of the measurement, all ONUs search for their associated network parameters (e.g., serial number, ONU-ID, etc.) that are stored inside the previously mentioned GEM frames. Since the distance between the ONUs and the OLT are different, it was also necessary to use an equalization delay parameter that is assigned by the OLT during the activation process. For more information, see [31, 32]. Consequently, all ONUs wait for a random period prior to starting data transmission. In the frame of this work, data are broadcasted in the downstream direction. In the upstream direction, time slots assigned by the OLT are used instead. Moreover, in this work, we did not use the DBA algorithm. Consequently, all ONUs are expected to transfer data in time slots with prespecified start and stop times.

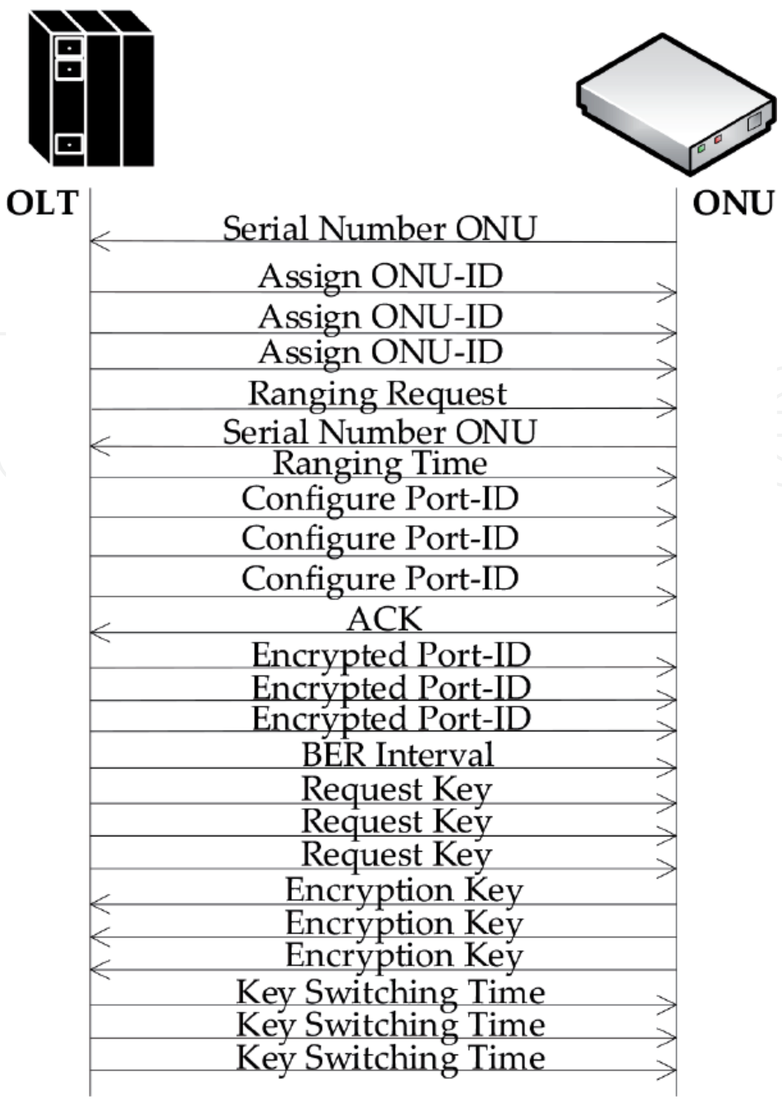

Figure 8.
GPON activation process with encryption channel establishment messages [29].

To summarize, on the one hand, this work is interested in the analysis of user data and the activation process. However, on the other hand, the description of the activation process is omitted, as has already been described in our previous work [32]. Since the user plane and control plane data are transferred using GEM frames, it is not possible to use a common packet analyzer such as Wireshark. For this purpose, we used a GPONxpert analyzer in a standalone mode in which all data are transferred and saved to a hard drive. Therefore, to perform a deeper inspection or analysis, all the data must to be postprocessed. In general, the control plane data can be divided into signaling, OMCI. First, we focused on the signaling data analysis. When the connection is established, messages such as Assign ONU-ID, Configure Port-ID, Assign Alloc-ID, Encrypted Port-ID, Encryption_key, key_request_message, and Key_switching_time are transmitted three times. This, as well as a complete GPON signalization, can be seen in **Figure 8**.

It can be seen that a physical layer operations, administrations and maintenance (PLOAM) message, specifically the "Serial number ONU," are transferred from the ONU to the OLT. This message holds information such as the vendor serial number, a list of supported data profiles, and the value of random delay of 82 μs [28]. The OLT uses these messages to extract the serial number and allocate the associated ONU-ID. Moreover, to minimize the impact of unequal distances among the ONUs and the OLT, it uses unique random delays for each of the ONUs that are based on the time between two successive "Serial number ONU" messages. As soon as the

OLT receives the ONU-ID, it sends the PLOAM message: "Assign ONU-ID." At this point, even though the OLT is aware of the assigned ONU-ID, it is not able to use unicast addressing because the ONU itself still cannot recognize the ONU-ID as its own, and therefore, broadcast addressing needs to be used (the ONU serial number is taken as the identifier) [29]. This means that every ONU receives this message; however, based on the comparison of the incoming and internal serial numbers, only the targeted ONU processes the message. In **Table 1**, it can also be seen that ZTE company is the final unit manufacturer. Based on hard-defined bytes in the MAC address, the manufacturer can be checked directly using its unique label: "0xC03B4EB4." GPON networks supported the transfer of ATM cells; however, in the last review in 2014, this support completely disappeared as these networks did not find their real application. For this reason, "ATM support Disable" can also be observed in the captured data. On the other hand, GEM support is necessary for any GPON data transfer: "GON support Enable." The captured data also have a description of the signal's power level, however, only with the following levels: low/medium/high power.

After the OLT sends the "Assign ONU-ID" message, it consequently sends the "Ranging request" message using the specific ONU-ID. Consequently, the ONU is capable of using a single grant to transmit data. The OLT unit's response to the "Serial Number ONU" message is a PLOAM message, "Assign ONU-ID." This message already carries a unique identifier for the designated end unit. From the nature of PON technology, it is clear that each end unit receives all messages. Using the unique ONU-ID, also called a serial number (if ONU-ID is not assigned), ONUs decide which messages to process. In this case, the assignment of ONU-ID = 1, i.e., the first end unit has already been replied to. The serial number of the unit equals "0x5A544547C03B4EB4", the Psync field is fixed and does not change throughout the communication. This fact is evidenced by the other messages listed in **Table 1**. "Ident Superframe Counter: 499314877" specifies the order of the transmitted frame/s. The ONU endpoint activation process in the GPON network is based on the sending of specific messages three times in a row. The second copy of the message is left for the demonstration of the Superframe counter being incremented by 1. After that, the ONU responds with the "Serial number ONU" message using the maximum priority T-CONT class (i.e., urgent data). The OLT computes a new value for the equalization delay using the "Ranging Time" message sent by the ONU. In the initial ONU report, the unit generates a random delay of 82 μs. The control unit must virtually ensure the same distance for all ONU end units. Each unit is located at a different distance, different customer stores, and/or residential units or streets. Supporting up to 20 km in the distribution part allows for the entire housing estate to be connected. The OLT sends a "Ranging request" message to specify a unique ranging time for each ONU. For this particular message, ONUs are required to respond immediately with their ONU-IDs and serial numbers. The OLT unit repeats the "Ranging request" message three times in total. It is important to note the second response, where the ONU specifies the mandatory parameters such as ONU-ID, the serial number (now omitted), and adds information about the Urgent PLOAM waiting and Traffic waiting in type 2 T-CONTs. The individual T-CONTs represent the distribution of traffic according to their classification by importance. T-CONT 1 responds to urgent data, i.e., data with the highest priority (e.g., voice over Internet protocol—VoIP) and fixed bandwidth. TCONT2 + 3 transfer Internet protocol television (IPTV) data with guaranteed bandwidth, T-CONT 4 is commonly used for best-effort data, and the last T-CONT5 is a mixed type including all types of bandwidth and services. Based on the received OLT responses, the OLT unit evaluates the assigned delay for the given ONU and sends the delay value

ID	ONU-ID	Message type		Message type
2	Unassigned ONU ID	Serial number ONU	Vendor ID: ZTEG, Vendor SN: 0xC03B4EB4, Random Delay: 82 µs, ATM Support: Disable, GEM support: Enable, ONU TX power level: high power	PLOAM
117	Broadcast message	Assign ONU-ID	ONU ID: 1; serial number: 0x5A544547C03B4EB4; Psync: 0xB6AB31E0; Ident Superframe Counter: 499314877; PLOAM CRC: 142	PLOAM
118	Broadcast message	Assign ONU-ID	Ident superframe counter: 499314878	PLOAM
120	1	Ranging request	Psync: 0xB6AB31E0; Ident FEC Indicator: 1; Ident Superframe Counter: 499315777	BWmap
1	1	Serial number ONU	ONU ID: 1; vendor ID: ZTEG; vendor SN: 0xC03B4EB4; random delay: 0	PLOAM
121	1	Ranging request	Psync: 0xB6AB31E0; Ident FEC Indicator: 1; Ident Superframe Counter: 499315777	BWmap
2	1	Serial number ONU	Delimiter: 0xAB5983; ONU ID: 1; Urgent PLOAM waiting: 1; Traffic waiting in type 2, 3, 4, 5 T-CONTs: 0	PLOAM
122	1	Ranging time	Path EqD descriptor: main path EqD; delay: 265409	BWmap
125	1	Request password	Ident Superframe Counter: 499318309	PLOAM
1	1	Password	Password (Hex): 0x47433033423445423400; password (ASCII): GC03B4EB4	PLOAM
126	1	Request key	Psync: 0xB6AB31E0	PLOAM
4	1	Encryption key	Key index: 0; fragment index: 0; key bytes: 0x681A055363E86213	PLOAM
7	1	Encryption key	Key index: 0; fragment index: 1; key bytes: 0x62677982F890BA9C	PLOAM
127	1	Key switching time	Superframe counter: 499321133	PLOAM
10	1	Acknowledge	DM_ID: key switching time	PLOAM
130	1	Configure Port-ID	Activate: enable; port-ID: 1	PLOAM
13	1	Acknowledge	DM_ID: configure port-ID	PLOAM
133	1	Encrypted Port-ID/VPI	Port-ID: 1	PLOAM
16	1	Acknowledge	DM_ID: encrypted port-ID/VPI; ONU ID: 1	PLOAM
136	1	BER interval	BER interval: 40000	PLOAM
19	1	Acknowledge	DM_ID: BER interval	PLOAM
142	1	Assign Alloc-ID	Alloc-ID: 1; Alloc-ID: Type GEM payload	PLOAM
22	1	Acknowledge	DM_ID: assign Alloc-ID	PLOAM

Table 1.
Activation process details in captured data in real GPON networks.

to the "Ranging time" message. GPON networks support so-called backup paths and link recovery systems when an alternative route is available. The message contains two fields: "Path EqD Descriptor: Main Path EqD" identifying the primary path and the backup path (the backup path was not available at the time of testing; therefore, it is not included in the message). The delay value specifies the delay for the end unit in "Delay: 265409," but this value does not match the value in μs. These steps set the basic communication parameters, the assigned ONU-ID, and the equalization delay. During the measurement, secure communication was enabled. The definition of reached states in which communication security can be performed and the prerequisites for negotiating the key are given in [33–35]. The entire process is started with the PLOAM message, "request password" containing "Ident Superframe Counter: 499318309." This message requires the end unit to respond with the same message with a password three times in a row. The captured data contain two fields: "Password (Hex): 0x47433033423445423400" and "Password (ASCII): GC03B4EB4." Next, the "Request Key" message is sent, the content of the message is not fully defined in this case; it is necessary to respond to this message with the Encryption Key message. The "Encryption Key" message consists of "Key Index: 0," "Fragment Index: 0" and "Key Bytes: 0x681A055363E86213." The sequence of these messages is followed and sent three times in a row. In our case, a single message is not enough to deliver the key, so another three messages are used to deliver the remaining part of it. This fact is illustrated by the following: "Fragment Index entry: 1," and "Key Bytes: 0x62677982F890BA9C." The next "Key Switching Time" message should define the start time when a new key is used that was not reached because the tool did not detect these fields. It only detected "Superframe Counter field: 499321133." The start time field contents must confirm the end unit using the "Acknowledge" message. The "Acknowledge" message contains the "Downstream Message Id: Key switching Time" field, confirming the previous message. Next, the OLT sends the "Configure Port-ID" message to the ONU specified by the ONU-ID. In the context of data transmission, the ONU-ID is used for the data flow allocation in a GEM frame. The ONU had to send the acknowledgement (ACK) messages three times (one for each of the received messages). As visualized in **Table 1**, the downstream message identification (DM_ID) contains a "Configure Port-ID" field that holds the confirmed message's name, and an ONU ID equaling the ONU-ID of the end unit (in our case 1). Subsequently, the OLT checks whether the Port-ID is encrypted. If it is not (i.e., the ONU remains in the registration process), the ONU sends the ACK message as a response to each correctly received message. Next, the OLT sends a "BER" (Bit Error Rate) message to specify an accumulation interval for each of the ONUs (number of downstream frames per ONU) that is used to count the number of downstream bit errors [29]. At this point, the ONU knows the Port-ID. However, to establish bidirectional data communication, the Alloc-ID is required to identify a traffic-bearing entity (e.g., T-CONT), which represents the recipient of the upstream data allocated during the BWmap procedure [29]. It is important to note that each ONU requires at least a single Alloc-ID that is equal to the ONU-ID and that is not transmitted by the OLT in the "Assign Alloc-ID" message. In this work, the following Alloc-ID was provided by the OLT: 1. The end unit must always contain at least one ONU-ID identifier, but it may contain several Alloc-IDs. Often, the initial Alloc-ID corresponds to the assigned ONU-ID, which also occurred in this case. The ONU acknowledges each of the PLOAM messages. After that, the encryption of the Port-IDs is rechecked. Nevertheless, it should be mentioned that data encryption is optional, and in reality, many ISPs do not use Port-ID encryption.

5.1 OMCI channel analysis

After the signaling phase is over, the operation, administration and maintenance (OAM) can be transferred using the OMCI channel. In our work, the OMCI procedures begin when the OLT sends a "Get/Set request" message to the ONU. When the ONU receives such a message, it responds with its own "Get/Set" message. In this work, we used a single ONU, see **Table 2**. At this point, the crucial phase of the OMCI analysis is the software image entity type inspection, as the ONU is to be authorized by its own serial number against the database of the OLT (depending on the particular ISP implementation). In the case that the OLT does not have the record of the ONU in the database, the ONU is not allowed to download the software image along with the configuration. On the other hand, if the record is present, the ONU downloads the data. It is important to stress that because the ISPs may offer different transmission speeds, functions, etc., to customers, each customer should have his or her own distinct software image. The software image message responds to the image data transfer used to set the parameters. The message parameters are reported as "inactive" as they are in the initial phase of the file download. The next analyzed message informs about the software image being valid and active. As soon as the ONU has the software image, it is capable of transferring the customer service support data as well as the metadata. To support VoIP telephony, which is a QoS-demanding service, the ONU downloads an additional configuration containing information such as the type of codec, constant bit rate allocation, and T-CONT priority. The next step is to set the parameters for VoIP service. This service is a key service used for the highest priority end units. Their setting corresponds to the priority operation, i.e., T-CONT1, in which a fixed bandwidth must be assigned. In the case of most of the service providers, this value is set to 512 kbit/s. This speed

ID	Type		Entity
367	Get	TCI priority: 1	ONU
368	Get response	TCI priority: 1; result reason: command processed successfully; vendor id: ZTEG	ONU
370	Get	TCI priority: 1	Software image
371	Get response	Result reason: command processed successfully; version: V3R016C00S917T; is committed: uncommitted; is active: inactive; is valid: valid	Software image
373	Get response	Is committed: committed; is active: active; is valid: valid	Software image
378	Get	VOIP configuration state	VOIP config data
379	Get response	VOIP configuration state inactive: configuration retrieval has not been attempted	VOIP config data
381	Get response	Profile version: 00000000	VOIP config data
400	Get all alarms		ONU DATA
401	Get all alarms response	OMCI alarms received on ME—physical path termination point ETHERNET UNI, instance—257, LAN-LOS No carrier at the Ethernet UNI.	ONU DATA

Table 2.
OMCI channel details in the analyzed GPON network.

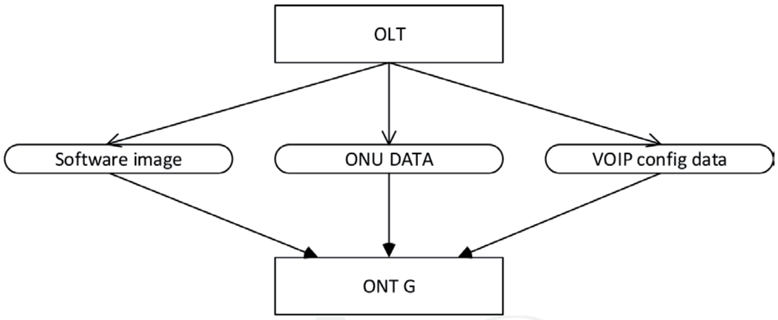

Figure 9.
OMCI channel diagram for the analyzed ONU.

must be guaranteed, even though it is considerably higher than the bandwidth of the G.711 codec (64 kbit/s). Successful reception and setting of the VoIP parameters are indicated with the message with ID 381 "Profile version: 00000000." In addition to the previously mentioned OMCI channel procedures that take place on the side of the ONU, there are also procedures on the side of the OLT: channel synchronization, verification, alarm indication, FEC monitoring, and so on, see **Table 2**. In summary, by analyzing the OMCI channel data, we performed active monitoring of the alarms of the distribution network. As seen in the "Get all alarms" message (ID: 400), the end-point ONU has reported a message signaling a failure on the Ethernet port.

In the case of OMCI channel measurement, it would be possible to summarize the transferred software image data, ONU data and the VoIP configuration file, see **Figure 9**.

A special case of the activation process is the message sequencing that can be seen in **Table 3**. This part of the activation is not mandatory for end units but is the last deactivation process aimed at the previously allocated parameters, most often the ONU-ID. This occurs when there is an immediate power outage. In the case of charged capacitors, the end unit sends a "Remote Error Indication" message. The message indicates that the ONU encountered an error. In the context of our experiments, this particular message was sent six times in total. When detecting a certain number of errors, most commonly defined by the manufacturer of the control unit, a "Dying Gasp" message follows. This message is dedicated to informing the control unit about an end unit failure, i.e., the loss of communication. The critical parameter of this message is the ONU-ID. After receiving such a message, the control unit sends the PLOAM message, "Deactivate ONU-ID," that causes this identifier to be released and consequently be reused by another end unit within the activation process. The PLOAM message is sent three times. Other parameters are discarded as internal timers have expired and communication/synchronization has not been restored in the downstream direction.

ID	ONU-ID	Message type	
1	1	Remote error indication	Sequence number: 3
6	1	Remote error indication	Sequence number: 8
7	1	Dying gasp	ONU ID: 1
9	1	Dying gasp	ONU ID: 1
59	1	Deactivate ONU-ID	Ident superframe counter: 498791449
60	1	Deactivate ONU-ID	Ident superframe counter: 498791451

Table 3.
Dying gasp PLOAM message details after the ONU lost power supply.

6. Conclusion

According to its grant policy, the European Union should contribute to building high-speed networks in the member states. This chapter introduced the state of the art in the field of Internet access technologies in the Czech Republic. The Czech Republic, as a member of the European Union, has committed to building high-speed Internet access for at least half of the households by 2020. Current market research has shown that WiFi technology is still dominant in the Czech Republic. The Czech Republic is behind the trend in FTTH/FTTB high-speed fiber optic connections by up to 10 and 5% for FTTB and FTTH, respectively.

The key part of this chapter is dedicated to the analysis of data transmitted in the GPON network. In cooperation with the Internet service provider Orange Slovakia, an active capture of transmitted data on the network was performed. As soon as the activation process of the end unit was completed successfully, data communication in both directions in GPON networks was possible. On the one hand, the sequence of the associated messages was defined by ITU-T Recommendation G.984, but on the other hand, it was only a recommendation and the specific implementation was fully within the manufacturer's competencies. Even though the end units were supposed to preserve the frame structure and the transmitted messages, as a result of the previously mentioned facts, it was often the case that the different manufacturers' end units were not compatible among themselves. Within the context of our analysis, TraceSpan's GPONxpert tool was used to capture network data. This device allowed for active listening of communication and real-time evaluation of its parameters. Detailed data analysis was a necessary form of postprocessing. To present the result of the activation process analysis, a sequence of key messages ensuring the activation of the end unit was displayed. Using these messages, it was possible to read the manufacturer and serial number of the end unit, set parameters such as ONU-ID and Alloc-ID. The OMCI channel provided end user parameters for a defined set of services, most often by downloading a profile image file corresponding to paid services and speeds. According to the reports, it was obvious that the VoIP parameters were also set.

Transmission of "Dying Gasp" messages was a special case of the activation process, or the logout and release of allocated parameters of the associated end units. These messages reflected a power outage of these units. Because the end units had unique UNU-ID/Alloc-ID parameters, the same parameters were used for other end units in the event of a power failure occurring in an already activated unit.

Acknowledgements

The presented research has been supported by a project of the Ministry of the Interior under grant no. VI20172019072, "E-infrastructure CESNET-modernization," registration no. CZ.02.1.01/0.0/0.0/16 013/0001797, and the National Sustainability Program under grant no. LO1401. For the research, the infrastructure of the SIX Center was used.

Conflict of interest

The authors declare no conflict of interest.

Notes/thanks/other declarations

Tomas Horvath would like to dedicate his part to his girlfriend (Lucie Baierova) and his family (Dagmar, Jan, and Petra). They have supported him during his University study. He also would like to give thanks to Ales Buksa for his support at the University. Ales has taught and inspired him with many things in his personal life.

Author details

Tomas Horvath[1,2]*, Petr Munster[1,2] and Josef Vojtech[2]

1 Department of Telecommunications, Brno University of Technology, Brno, Czech Republic

2 Department of Optical Networks, CESNET a.l.e., Prague, Czech Republic

*Address all correspondence to: horvath@feec.vutbr.cz

References

[1] Suzuki N, Miura H, Matsuda K, Matsumoto R, Motoshima K. 100 Gb/s to 1 Tb/s based coherent passive optical network technology. Journal of Lightwave Technology. 2018;**36**(8):1485-1491. DOI: 10.1109/JLT.2017.2785341

[2] Horvath T, Munster P, Vojtech J, Velc R, Oujezsky V. Simultaneous transmission of accurate time, stable frequency, data, and sensor system over one fiber with ITU 100 GHz grid. Optical Fiber Technology. 2018;**40**(1):139-143. DOI: 10.1016/j.yofte.2017.11.016

[3] Vojtech J, Slapak M, Skoda P, Radil J, Havlis O, Altmann M, et al. Joint accurate time and stable frequency distribution infrastructure sharing fiber footprint with research network. Optical Engineering. 2017;**56**(2):027101-027107. DOI: 10.1117/1.OE.56.2.027101

[4] G.983.1: Broadband Optical Access Systems Based on Passive Optical Networks (PON) [Internet]. International Telecommunication Union. Geneva: International Telecommunication Union; 1998. Available from: https://www.itu.int/rec/T-REC-G.983.1-199810-S/en [Accessed: 08-11-2018]

[5] Angelopoulos JD, Venieris IS, Protonotarios EN. A distributed FIFO spacer/multiplexer for access to tree APONs. In: Proceedings of ICC/SUPERCOMM'94-1994 International Conference on Communications. New Orleans, LA, USA: IEEE; 1994. pp. 70-74. Available from: http://ieeexplore.ieee.org/document/369020/

[6] 802.3ah-2004—IEEE Standard for Information technology—Local and metropolitan area networks—Part 3: CSMA/CD Access Method and Physical Layer Specifications Amendment: Media Access Control Parameters, Physical Layers, and Management Parameters for Subscriber Access Networks [Internet]. USA: IEEE; 2004. Available from: https://ieeexplore.ieee.org/document/983911 [Accessed: 08-11-2018]

[7] Kramer G, Pesavento G. Ethernet passive optical network (EPON): Building a next-generation optical access network. IEEE Communications Magazine. 2002;**40**(2):66-73. Available from: http://ieeexplore.ieee.org/document/983910/

[8] Next Generation Internet Initiative [Internet]. Next Generation Internet Initiative | Digital Single Market. EU; 2018. Available from: https://ec.europa.eu/digital-single-market/en/next-generation-internet-initiative [Accessed: 08-11-2018]

[9] G.984.1: Gigabit-Capable Passive Optical Networks (G-PON): General Characteristics [Internet]. International Telecommunication Union. Geneva: International Telecommunication Union; 2003. Available from: https://www.itu.int/rec/T-REC-G.984.1-200303-S/en [Accessed: 08-11-2018]

[10] Li Li, Xin Shouting, Duan De-Gong. Research of DBA schemes and QoS in PON system. In: 2016 2nd IEEE International Conference on Computer and Communications (ICCC). Chengdu, China: IEEE; 2016. pp. 2148-2153

[11] Arokkiam JA, Brown KN, Sreenan CJ. Optimised QoS-aware DBA mechanisms in XG-PON for upstream traffic in LTE backhaul. In: 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW). Vienna, Austria: IEEE; 2016. pp. 361-368

[12] Horvath T, Munster P, Cymorek P, Oujezsky V, Vojtech J. Implementation of NG-PON2 transmission convergence

layer into OPNET modeler. In: 2017 International Workshop on Fiber Optics in Access Network (FOAN). Munich, Germany: IEEE; 2017. pp. 1-5

[13] Horvath T, Munster P, Vojtech J, Havlis O. Modified GIANT dynamic bandwidth allocation algorithm of NG-PON. Journal of Communications Software and Systems. 2017;**13**(1):15-22. DOI: 10.24138/jcomss.v13i1.243

[14] Farmer J, Lane BW, Bourg K, Wang W. FTTx Networks: Technology Implementation and Operation. Singapore: Morgan Kaufmann; 2017

[15] Digital Czech Republic v. 2.0—The Way to the Digital Economy [Internet]. Ministry of Industry and Trade. Prague: Ministry of Industry and Trade; 2014. Available from: https://goo.gl/J6ynRj [Accessed: 21-09-2018]

[16] National Plan for the Development of Next Generation Networks [Internet]. Ministry of Industry and Trade. Prague: Ministry of Industry and Trade; 2017. Available from: https://goo.gl/V3qxUD [Accessed: 21-09-2018]

[17] Y.2001: General Overview of NGN [Internet]. 2001: General Overview of NGN. Switzerland: ITU; 2005. Available from: https://www.itu.int/rec/T-REC-Y.2001/en [Accessed: 21-09-2018]

[18] Annual Reports 2014 [Internet]. Annual Rreports 2014. Prague: Czech Telecommunications Office (CTU); 2015. Available from: https://www.ctu.eu/2014-0 [Accessed: 21-09-2018]

[19] FTTx & Gigabit Society [Internet]. EU: IDATE; 2017. Available from: https://en.idate.org/categorie-produit/fttx-gigabit-en/ [Accessed: 30-09-2018]

[20] Montagne R. FTTH/B Panorama [Internet]. Spain; 2018. Available from: http://www.valencia.ftthconference.eu

[21] Rigby P. FTTH Handbook [Internet]. FTTHCouncil. EU: FTTHCouncil; 2016. Available from: http://www.ftthcouncil.eu/documents/Publications/FTTH_Handbook_V7.pdf [Accessed: 23-09-2018]

[22] Asaka K. What will be killer devices and components for NG-PON2?. In: 2014 The European Conference on Optical Communication (ECOC). Cannes, France: IEEE; 2014. pp. 1-3

[23] Nesset D. NG-PON2 technology and standards. Journal of Lightwave Technology. 2015;**33**(5):1136-1143. DOI: 10.1109/JLT.2015.2389115

[24] Khotimsky DA. NG-PON2 transmission convergence layer: A tutorial. Journal of Lightwave Technology. 2016;**34**(5):1424-1432. DOI: 10.1109/JLT.2016.2523343

[25] Horvath T, Munster P, Vojtech J, Havlis O, Gallo M. Transmission convergence layer of NG-PON2 in VPIphotonics tool. Journal of Communications Software and Systems. 2017;**13**(3):141-147. DOI: 10.24138/jcomss.v13i3.359

[26] G.989.3: 40-Gigabit-Capable Passive Optical Networks (NG-PON2): Transmission Convergence Layer Specification [Internet]. International Telecommunication Union. Geneva: International Telecommunication Union; 2015. Available from: http://www.itu.int/rec/T-REC-G.989.3/ [Accessed: 29-09-2018]

[27] Cale I, Salihovic A, Ivekovic M. Gigabit passive optical network—GPON. In: 2007 29th International Conference on Information Technology Interfaces. Cavtat, Croatia: IEEE; 2007. pp. 679-684

[28] Hood D, Trojer E. Gigabit-Capable Passive Optical Networks. Hoboken: Wiley; 2012

[29] G.984.3: Gigabit-Capable Passive Optical networks (GPON): Transmission Convergence Layer Specification. International Telecommunication Union. Geneva: International Telecommunication Union; 2015. Available from: http://www.itu.int/rec/T-REC-G.984.3/ [Accessed: 29-09-2018]

[30] Yan Y, Yamashita S, Yen S-H, Afshar PT, Gudla V, Kazovsky LG, et al. Invited paper: Challenges in next-generation optical access networks. IET Optoelectronics. 2011;**5**(4):133-143. DOI: 10.1049/iet-opt.2011.0027

[31] Mendonca C, Lima M, Teixeira A. Security issues due to reflection in PON physical medium. In: 2012 14th International Conference on Transparent Optical Networks (ICTON). UK: IEEE; 2012. pp. 1-4

[32] Horvath T, Munster P, Jurcik M, Koci L, Filka M. Timing measurement and simulation of activation process in GPON networks. Optica Applicata. 2015;**45**(4):459-470. DOI: 10.5277/oa150403

[33] Malina L, Munster P, Hajny J, Horvath T. Towards secure gigabit passive optical networks: Signal propagation based key establishment. In: Proceedings of SECRYPT 2015. Colmar, Francie: IEEE; 2015. pp. 349-354

[34] Malina L, Horvath T, Munster P, Hajny J. Security solution with signal propagation measurement for gigabit passive optical networks. Optik—International Journal for Light and Electron Optics. 2016;**127**(16):6715-6725. DOI: 10.1016/j.ijleo.2016.04.069

[35] Horvath T, Malina L, Munster P. On security in gigabit passive optical networks. In: 2015 International Workshop on Fiber Optics in Access Network (FOAN). Brno, Czech Republic: IEEE; 2015. pp. 51-55