UNIVERSITY OF CALIFORNIA

Los Angeles

Using Inkjet Printers for Acoustic Data Exfiltration from Air-Gapped Networks

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Electrical and Computer Engineering

by

Julian de Gortari Briseno

2020

© Copyright by
Julian de Gortari Briseno
2020

ABSTRACT OF THE THESIS

Using Inkjet Printers for Acoustic Data Exfiltration
from Air-Gapped Networks

by

Julian de Gortari Briseno
Master of Science in Electrical and Computer Engineering
University of California, Los Angeles, 2020
Professor Mani B. Srivastava, Chair

The ubiquity of printers in modern office spaces provides our motivation to design a covert channel based on the emanations of one particular class of these devices which uses inkjet technology. In this thesis, we demonstrate how a covert channel can be established by leveraging the acoustic emissions of inkjet printers to exfiltrate information from an air-gapped network. In essence, malware installed on a computer with access to a printer, but no access to the Internet, can inject certain imperceptible patterns into all documents being sent to the printer, so as to control the printing process in such a way that an acoustic signal is generated which can be captured with a nearby smartphone. Throughout this work, we demonstrate how people are unable to perceive these patterns under normal light conditions. Moreover, two distinct modulation schemes are proposed for our communication system and tests are carried out considering different types of document layouts, in distinct circumstances.

The thesis of Julian de Gortari Briseno is approved.

Nader Sehatbakhsh

Omid Salehi-Abari

Mani B. Srivastava, Committee Chair

University of California, Los Angeles

2020

To my family, colleagues and friends.

TABLE OF CONTENTS

1 Introduction **1**

2 Related Work **5**

2.1 Covert Channels 5

2.2 Printers' Emissions 6

2.3 Optical Concealment 8

3 Inkjet Printers **9**

3.1 Anatomy of an Inkjet Printer 9

3.2 Printer Acoustics 12

4 Attack Model **14**

4.1 Receiver 16

4.2 Transmitter 16

4.3 Injection Patterns 19

5 Printer Subsystem in Linux **23**

5.1 Infecting the Linux Printer Subsystem 24

5.2 Injecting Patterns into PDF Files 26

6 Evaluation **28**

6.1 Obtaining the Parameters of the Modulation Schemes 29

6.2 Distance and Receiver Orientation 33

6.3 Font 35

6.4 Layout . . . 36
6.5 Color of Paper . . . 38
6.6 Background Noise . . . 38
6.7 Intercalating Modulation . . . 39
6.8 Printer Quality Settings . . . 40

7 Color Perception . . . 41
7.1 Human Perception of Yellow Dots . . . 41
7.2 Perception Survey . . . 42

8 Discussion . . . 45
8.1 Countermeasures . . . 45
8.2 Limitations and Future Work . . . 46

9 Conclusion . . . 48

A Injection Pattern Examples . . . 50

B Document Layouts . . . 55

C Source Code . . . 62

References . . . 63

LIST OF FIGURES

3.1 Picture showing the components of a Canon Pixma MG2410. **(A)** DC Motor that controls rollers' movement. **(B)** DC Motor that controls printhead movement. **(C)** Programmable Logic Controller. **(D)** Printhead. **(E)** Timing belt and linear optical encoder stretch across the width of the printer. **(F)** Excess ink depository. **(G)** Nozzle caps. **(H)** Rollers. **(I)** Loading paper tray. **(J)** Angular optical encoder attached to the end of one of the rollers. 10

3.2 Acoustical waveform and its respective spectrogram generated by an HP Photosmart D110 printer when in operation. 12

4.1 Proposed attack scenario for the establishment of the covert channel. **(1)** A trusted user on an infected computer sends a document to print. **(2)** The malware present on that computer injects particular imperceptible patterns into the document pages. **(3)** The altered document passes to the printer. **(4)** The printer produces certain deterministic acoustic emissions depending on the injected patterns. **(5)** A nearby smartphone records the noise made by the printer. 15

4.2 The acoustical waveform generated by an HP Photosmart D110 printer, which corresponds to our DPPM modulation scheme, is shown here after processing. In this case a greater difference in the time between pulses corresponds to a 0 bit, while a smaller difference to a 1 bit. 17

4.3 The acoustical waveform generated by an HP Photosmart D110 printer, which corresponds to our hybrid FPM-DPPM modulation scheme, is shown here after processing. In this case two pulses with a large time offset is interpreted as a 0 bit, while a cluster of pulses with small time offsets is considered as a 1 bit. . . . 18

4.4 The packet structure allocates 4 bits for the preamble, a variable number of bits for the payload (dependent on the printer and modulation scheme used), and a parity bit. 19

4.5 The imperceptible patterns injected into documents are arranged in two ways that correspond to the two proposed modulation schemes. **(a)** DPPM. **(b)** FPM-DPPM. 22

6.1 Five different layouts were tested as well as three different fonts for the single column case. Appendix B offers a more detailed view of these layouts. 30

6.2 The relationship between a line's length and the time it takes for the roller mechanism to activate again is shown for DPPM modulation. 31

6.3 The three printers present different relative amplitudes when using the same line offsets. The printers exhibit a positive relationship between relative amplitude and offset size. 32

6.4 The effects of distance and receiver orientation over Bit Error Ratio. Both modulation schemes were tested for each printer. 34

6.5 Bit Error Ratio according to the font type and printer used. Times and Courier are the abbreviated forms we chose for Times New Roman and Courier New, respectively. 35

6.6 Bit Error Ratio according to the layout used on different printers. 36

7.1 A magnified view of two printed letters surrounded by yellow dots. These two letters happen to be located inside one of the rectangles used for FPM-DPPM. . 42

A.1 DPPM pattern of the bit sequence **10101011100011011001010000110110** injected into a page prepared to be sent to the Epson printer (an extra line is placed at the end of the injected patterns to separate the final pulse from the page expulsion noise). The color shown here for the injected patterns was **not** the one chosen for the real attack. 51

A.2 FPM-DPPM pattern of **101010111000** injected into a page prepared to be sent to the Epson printer (an extra rectangle is placed at the end of the injected patterns to separate the final pulses from the page expulsion noise). The color shown here for the injected patterns was **not** the one chosen for the real attack. 52

A.3 FPM-DPPM pattern of **101010111000** injected into a text page prepared to be sent to the Epson printer. The color shown here for the injected patterns was **not** the one chosen for the real attack. 53

A.4 FPM-DPPM pattern of **101010111000** injected into a text page prepared to be sent to the Epson printer. The color shown here for the injected patterns was the one chosen for the real attack. 54

B.1 Single column layout with Arial font. 56

B.2 Single column layout with Times font. 57

B.3 Single column layout with Courier font. 58

B.4 Two columns layout. 59

B.5 Spreadsheet layout. 60

B.6 Complex layout. 61

LIST OF TABLES

6.1 Characteristics of the printers used to test our attack. 28

6.2 Attack design parameters for each printer. **A** corresponds to DPPM modulation and **B** to FPM-DPPM 33

6.3 Bit Error Ratio of each printer and modulation scheme when there is high environmental noise. 39

8.1 Time duration of the printing process for a normal text document vs. time duration of the printing process for an altered text document. 46

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Professor Mani B. Srivastava, for his support and guidance shown throughout the development of this work. I would also like to acknowledge my parents for helping me acquire the printers which constituted an indispensable part of this work, and letting me record their acoustic emanations in absolute silence at times. Finally, I should also recognize the effort made by all the people who voluntarily participated in the surveys designed in this work.

CHAPTER 1

Introduction

Isolating an organization's internal network from the external world, thus creating and air gap, was previously thought to be the ultimate way of securing the information maintained in a given facility from being exfiltrated by an adversary. Indeed, increasing efforts were being put into developing all kinds of security mechanisms to prevent any kind of attack through the network's normal channels, e.g., traffic monitoring, firewalls, intrusion detection systems, etc.; but unintended physical emissions from common devices were consistently ignored. As a consequence, attacks exploiting these vulnerable emanations surfaced, with the objective to break through traditional defenses and attain a foothold on the protected computer networks. One use of these unintended emissions, which has gained prominence, is in the form of covert communication channels, through which data is infiltrated and exfiltrated to and from a given network.

The concept of covert channels is in itself not new: Lampson defined covert channels for the first time in 1973 as *"those not intended for information transfer at all"* [28], giving as an example the existence of an hypothetical malicious process exploiting the paging rate in a computer, using its observable effects on performance to establish a covert channel and transmit data to a second concurrent process. Indeed, it appears the first conceived covert channels utilized shared system resources in the same host computer to transmit information between concurrent processes, either by using storage channels, in which case one process would write directly or indirectly to a storage location and another process would read its contents or acknowledge its presence; or by using timing channels, whereby processes

would modulate their usage of system resources in such a way that other processes would be capable of extracting data from the system response time [7]. Network covert channels, which involved misusing various fields in network protocols, were next developed, until finally, air-gap covert channels emerged, the focus of this thesis. This last type of intended communication channels through unintended physical emissions could be thought as an evolution of the original studies in side-channels, that started under government supervision via the TEMPEST program, which analyzed unintended communication through devices' unintended emissions [24]. Lastly, it is worth mentioning a type of covert channels which depend on embedding messages into a cover, that can be text, audio, an image, etc.; a type of covert channels associated now with the practice of steganography.

In practice, most of the covert channels relying on unintended physical emissions have considered only one way communication, either to exfiltrate sensitive information, or to infiltrate commands in order to activate latent malware, for example. The main focus on covert channels has been given to personal computers, but accessory devices connected to these computers can also actuate as transmission nodes, i.e., by using their own unintended emanations for this purpose, as demonstrated here with inkjet printers.

In this thesis, we introduce a novel way of exploiting the acoustic pulse-like noise generated by inkjet printers' mechanical components, in order to establish a covert channel through which exfiltrate information from the computer whose document is being printed. As far as we know, this is the first research work that uses printer emanations for establishing a covert channel. The attack leverages the fact that inkjet printers are sensitive to the type of graphical objects present on documents, as well as the manner in which these objects are arranged throughout the document (their layout), so that by introducing certain specific patterns through all the document, a deterministic communication channel becomes feasible. By injecting these patterns in each of the documents being sent to print, an infected device close to the printer could record the acoustic emissions and process them to obtain the sensitive data being exfiltrated. Consideration is given to the evidence left on

these documents, particularly, an effort is made to make the injected patterns imperceptible to the human eye. Two different pattern arrangements are considered that produce distinct noise sequences: one that works best on blank documents and through which we produce Differential Pulse Position Modulation (DPPM), and another that works best on text documents without images and through which we generate a hybrid form of DPPM and Frequency Pulse Modulation (FPM). Three inkjet printers are tested representing three of the biggest brands in market: HP, Epson and Canon. At a distance of 50 cm, we achieve an effective bit rate that varies between 1.76 bps and 0.12 bps, with a Bit Error Ratio (BER) lower than 30% in all cases, the exact error varying with the type of modulation employed and the type of layout used in the case of text documents. In the best case we obtain a BER lower than 1% at a distance of 50 cm. While these results might seem mediocre at best compared with conventional communication systems, for the purposes of a covert channel the resulting data bandwidth might be enough to transmit small but valuable information like passwords or malware status updates, in an environment which in theory should not let any kind of communication to the outside.

The rest of this thesis is organized as follows: related work is presented in Chapter 2, and an overview of the inkjet printer technology with an emphasis on its acoustic characteristics is given in Chapter 3. Chapter 4 details our attack model, including the design of our transmitter and receiver, and the assumptions underlying the attack. In Chapter 5 the reader will find a description of the software printer subsystem in Linux, the operating system we used throughout this work, as well as the manner in which it was exploited for the purposes of our attack. Chapter 6 details the implementation and the results of our experiments. In Chapter 7 we explain why our attack should be imperceptible to the human eye, and we present the results of a survey designed to verify this claim. We discuss the limitations, future work and countermeasures derived from this work in Chapter 8, conclusions being presented in Chapter 9. The appendices included at the end of this thesis might help the reader understand better the layouts we used for our tests and the way the injected patterns

appeared in the documents. A link to the source code of all the software artifacts used during the realization of this work is also located in the appendices.

CHAPTER 2

Related Work

In this chapter, we provide a summarized literature review of existing work on covert channels, printer emissions and techniques for optical concealment.

2.1 Covert Channels

Covert channels have been found to be feasible with almost any kind of physical emission, a few examples are introduced in here. Previous work has shown how malware can leverage the form of data transmission from a computer into a USB drive to create a communication channel based on electromagnetic signals [16]. This attack relies on the fact that sending 0 bits into a USB drive generates rapid voltage and current changes, and thus electromagnetic radiation, as a result of using Non-Return-to-Zero inverted (NRZI) line encoding. In [19], the low frequency magnetic waves emitted by computer CPUs are employed as carriers of data by controlling the CPU's workload, which by itself governs power consumption and current flows that help establish the magnetic field. Bit Whisper [17], demonstrates again how, by controlling the CPU workload, one can use the emitted heat as a signal to be captured by another nearby computer's environmental thermal sensors, and viceversa. Effectively, in this case, a full duplex channel is established which depends on the detected relative temperature offsets at certain time slots. Finally, in [20], a computer's chassis fans are manipulated in order to induce vibrations in a surface shared with a nearby smartphone, which receives the signal through its accelerometer. We will focus now on a particular class of emissions that are the basis of our attack: acoustic leakage.

Covert channels using acoustic emanations from personal computers to exfiltrate information have been the particular subject of previous research, for example: leveraging mechanical components' noise-generating capabilities, like in the case of cd/dvd drive noise [13], fan noise [22], and hard disk drive noise [23]. The hard disk drive noise exploited in [13], is the one produced in seek operations, when the actuator head moves through different tracks on the disk. Fans in a computer produce noise as a consequence of the movement of their blades, an emission exploited in [22], which can also be the source of vibrations, as explained above. Lastly, for leveraging the cd/dvd drive noise, three different mechanisms were exploited in [13]: disc tray loading, disc spinning, and the tracking device which moves the laser assembly for reading. Exploiting the noise produced by vibrations from electrical components in a switch-mode power-supply is done in [21]: basically, by modulating the power consumption of the CPU, the switch frequency of the power supply can be controlled, which in itself controls the vibration of coils and ceramic capacitors. By alternately turning speakers and headphones connected to a computer into input and output devices via jack retasking, a covert two-way communication channel can be established at the ultrasound level [18]. There has also been certain work on cyber-physical systems in this domain, in which control signals driving physical instrumentation are altered to produce specific acoustic signals without affecting significantly the closed-loop system [27].

2.2 Printers' Emissions

Regarding printers' unintended emanations, Backes et al. revealed how dot matrix printer's acoustic emissions could be leveraged to recover the text being printed [4]. Because these type of printers emit a noise whose intesity is correlated with the number of needles strinking the paper, and therefore to the shape of the letters and words being printed, it uses a combination of machine learning, audio processing and speech recognition techniques to decipher the text being printed. Laser printer emanations have also been investigated before, with results

indicating that electromagnetic emissions can be leveraged to reconstruct the document being printed, specifically, from the radiation produced by the video signal to which the image information is converted [38].

In the realm of additive technologies, also known as 3D printing, acoustic emanations have been previously utilized to recreate the G-codes used for prototype design in Fused Deposition Modeling (FDM) [3]. Basically, each of the four stepper motors used in 3D printers (three to move the printhead in one of the axes and another used for filament extrusion), emit different noise as a consequence of driving a different load. By using a series of machine learning models, all the necessary properties can be infered to reconstruct the G-codes. Capturing both acoustic emanations and magnetic emanations with a smartphone from this type of printers has also been done previously to recover the G-codes [37]. In this case, it is demonstrated that magnetic emissions also correlate with axis movement. Altering the compiler that generates these G-codes so as to modify the printing process in order to maximize the leakage information, a procedure similar in nature to the way we modify documents to establish a deterministic communication channel, has been done in [33]. Effectively the homogeneous nature of the side channel is reduced by making slight changes in fan speed, speed at the end of line segments and the power used by stepper motors, among other features. Other research work has instead focused on the printers' acoustic emissions for the purpose of defending the printing process from sabotage, producing a digital signature that can be compared with further printings of the same object and can detect changes in particular G-codes [5].

The only previous work found to exploit printers' capabilities in order to establish a covert channel is [31], where it is shown that by using a Multifunction Printer, one can establish a duplex channel by exploiting the optical characteristics of the integrated scanner, transmitting data through an optical source when the scanner is operating and receiving an acknowledgment by capturing the scanner's optical output with a camera.

2.3 Optical Concealment

As our attack is based on injecting patterns imperceptible to humans into documents, it is worth mentioning some previous research on optical concealment. The Electronic Frontier Foundation revealed in 2005 that manufacturers of laser printers had been using imperceptible tiny yellow dots arranged in grids as watermarks for each of the documents printed by that type of devices, specifically, this tracking mechanism encoded the serial number of the printer in use, as well as date and time of the printing process [35]. The injected yellow dots were less than one millimiter in diammeter, and the patterns were shown to be reapeted multiple times throughout each of the pages conforming a document. Inspired by these revelations, researchers in [6] explored the way to create a similar tracking system for inkjet printers, with yellow dots as its basis. Our attack makes use of the same yellow dots to conceal information, which means we are not presenting a new technique in this aspect, although the purpose for which we use them in this thesis may be definitely new.

Some other optical covert channels have been designed to exploit certain characteristics of our visual system that are also leveraged in this work. For example, in [15], an optical covert channel uses QR codes displayed through a computer's screen, and captured by a camera, to transmit data, without the user being able to detect it. The exfiltration method either flashes the image in a small number of the total frames shown per second, or the malware embeds the image into a uniform bright or dark surface by reducing it into a binary color image in such a way that the difference in brightness with the background surface is minimal. Our work exploits a similar effect, by which we embed lines of almost the same luminance level with respect to the page background in order for these stripes to pass unperceived. In [14], the overall brightness of the PC screen is utilized as a signal to transmit data into a video camera. Specifically, by changing rapidly the red color of each pixel within a threshold of 3%, the relative brightness is used to confer data without the user noticing, exploiting the same phenomenon mentioned previously.

CHAPTER 3

Inkjet Printers

Inkjet printing technology encompasses a series of techniques for the ejection of droplets of ink from a printhead into a substrate. Conventional inkjet technologies are divided into Continuous Inkjet (CIJ) and Drop on Demand (DoD): the first one involves the ejection of a stream of continuous liquid whose droplets are directed into a substrate by modulation of an electrostatic field, while in the latter each single ink droplet is ejected by a pressure pulse without any subsequent steering of it [36]. There are two main technologies used for DoD printers that differ on the method used to dispense the droplets: thermal and piezoelectric. Thermal inkjet printers heat a resistive element controlled by electrical current to create an ink vapor bubble that expands and generates the pressure necessary to eject a droplet from the nozzle of the printhead, piezoelectric inkjet printers contain piezoelectric actuators that change shape when applied an electric potential so that the necessary pressure to dispense the droplets is generated [36]. Further, Drop on Demand inkjet printers generally come with printheads that are part of disposable ink cartridges or are meant to last for the entire lifetime of the printer, and these printers can either have movable or stationary printheads. Our work involves Drop on Demand inkjet printers with both types of ejection technology and ink containment technology, although stationary printheads are not taken into account.

3.1 Anatomy of an Inkjet Printer

Figure 3.1 shows the components commonly found on an inkjet printer, and in this section we dedicate ourselves to describe the printing process with respect to the components found

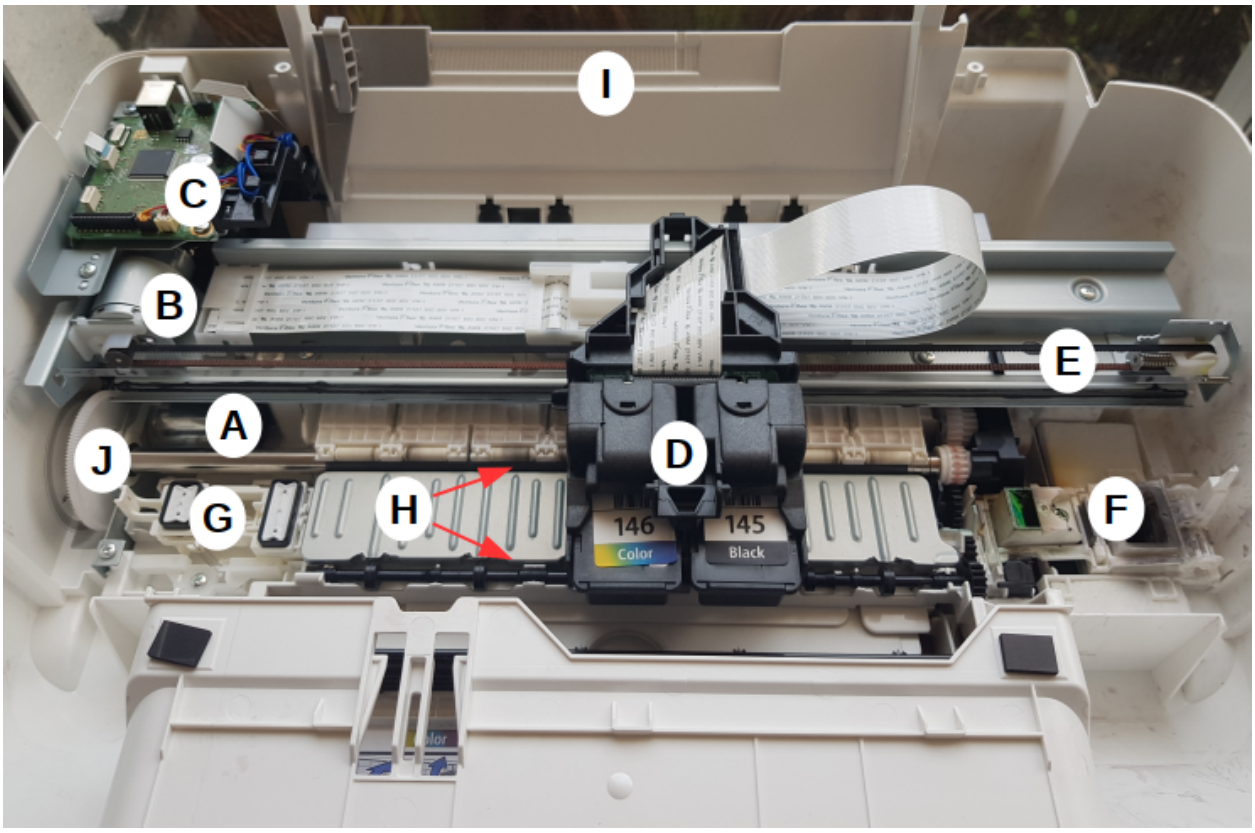


Figure 3.1: Picture showing the components of a Canon Pixma MG2410. **(A)** DC Motor that controls rollers' movement. **(B)** DC Motor that controls printhead movement. **(C)** Programmable Logic Controller. **(D)** Printhead. **(E)** Timing belt and linear optical encoder stretch across the width of the printer. **(F)** Excess ink depository. **(G)** Nozzle caps. **(H)** Rollers. **(I)** Loading paper tray. **(J)** Angular optical encoder attached to the end of one of the rollers.

on that figure. When printing a document, a paper sheet placed on the paper loading tray is first brought into the printer by the feed mechanism. This mechanism consists of a series of rollers connected through a gear train whose rotation is controlled by a DC motor, where an optical angular encoder at one end of one of the rollers is used to control the speed at which these rollers operate. The paper sheet is then advanced into the printer until this paper sheet is placed below the printhead, i.e., directly below the location in the paper were the first graphical object is to be printed. The printhead then starts sliding across a metal

rail, which spans the width of the printer, by means of a timing belt linked to another DC motor. To regulate the speed and position of the sliding printhead, a transparent plastic strip with fine black bars located parallel to the printhead's rail works as a linear optical encoder along with the sensor located in the printhead. Within the printhead, the ejection mechanism expels droplets of ink through each of the hundreds of nozzles that are part of it. When the last graphical object has been printed, the roller mechanism expels the paper sheet from the printer into another paper tray. Excess ink can be removed from the printhead at any time through a series of plastic brushes and two ink compartments located at one end of the printer, whose purpose is to collect the residues of black and color ink. At the other side of the printer, two pads cap the printhead nozzles when not used, so as to avoid ink drying. The printer's actuators are activated and controlled by a microcontroller on a PCB plaque to which all electric components connect to. The detailed disassembly of another inkjet printer is shown in [40].

Depending on the graphical objects present on the document to be printed, the printhead will pass multiple times or just one time through their location; particularly, when dealing with images or solid colors, multiple printhead passes will be required. Waasdorp et al. further explain this mechanism by stating that the printhead first moves horizontally across the width of the paper and dispenses ink droplets into it, resulting in horizontal lines with a vertical distance equal to the separation between the rows of the nozzles, so that when the printhead moves again horizontally in the other direction, the paper feed mechanism displaces the paper by a certain small distance which allows the printhead to fill the blank spots left behind [39]. Finally, the manner in which inkjet printers convert continuous tone images into dot based images is through the use of halftoning, where either by modulating the size of the dots or the frequency with which they appear, the continuous color tones can be emulated as a consequence of the human visual system low-pass filter effect, whereas the individual dots that conform the elements of the documents are blurred out [29]. As white paper is the usual printing medium for printers, bright colors may therefore contain a lower

amount of dots than darker ones.

3.2 Printer Acoustics

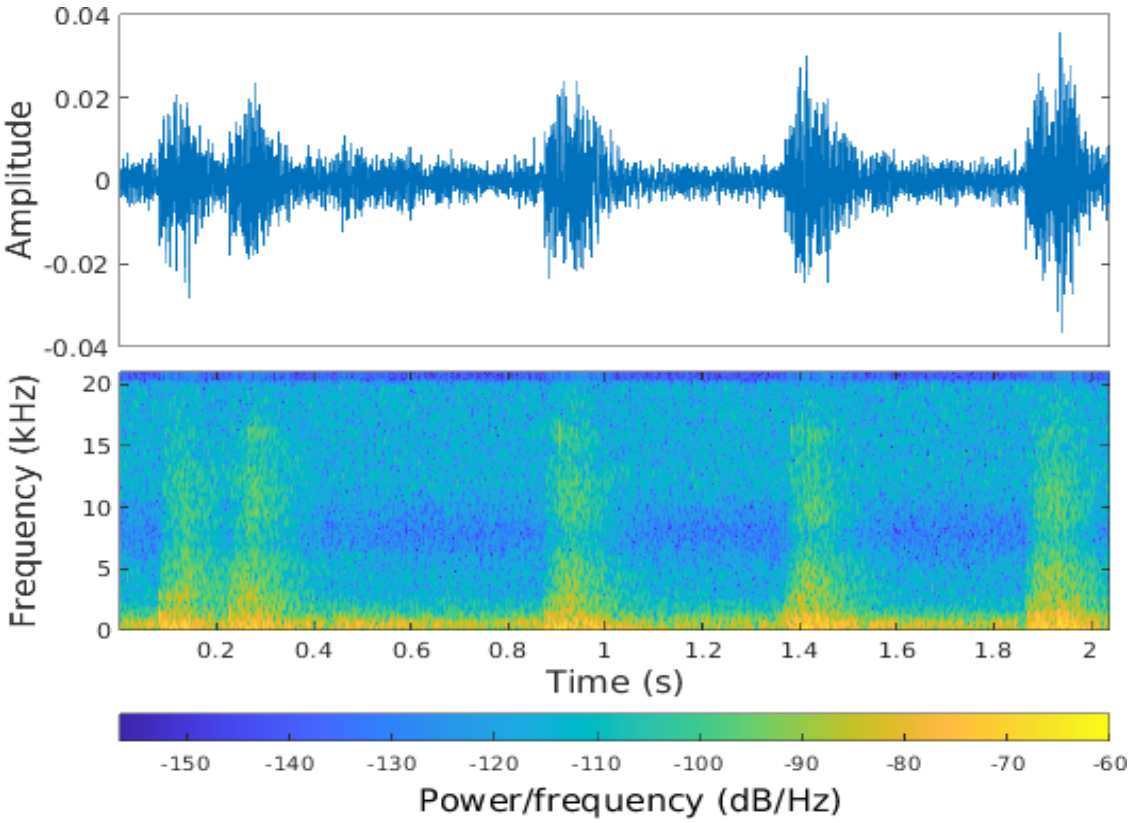


Figure 3.2: Acoustical waveform and its respective spectrogram generated by an HP Photosmart D110 printer when in operation.

Throughout our experiments, we identified two main sources of noise emitted from inkjet printers that could serve for the purposes of our covert channel: one corresponding to the sliding movement of the printhead and another which matched rollers' rotation. The noise source in which we were interested was the one generated by the roller mechanism, as it proved to produce a louder noise than the one generated by the sliding printhead, this difference being consistent across all printers. Figure 3.2 shows the waveform and spectrum of the acoustical signal in which we are interested, a signal whose pulses comprise almost the

entirety of the frequency range shown. In our attack, we use these acoustic pulses to transmit data from the infected computer, filtering the resulting signal around the empirically obtained frequency range of 3 kHz to 6 kHz, a range we use to prevent low frequency interference from the other printer noise sources and to delimit high frequency noise from the environment.

CHAPTER 4

Attack Model

In this chapter our attack model is detailed. Figure 4.1 presents the overall picture of our attack scenario: a trusted user sends a document to print from a computer infected with our malware, which intercepts the print job before it leaves the computer and injects certain imperceptible patterns into the document. The malware then sends this document to the printer, and the manner in which these previously injected patterns are laid out throughout the document will affect the operation of the printer in a way that the noise it produces serves to establish a communication channel when this noise is recorded by a nearby device. For our attack model, we then assume a situation where a computer with printer access inside the targeted air-gapped facility has already been infected with our malware.

The feasibility of infecting air-gapped networks has already been proved previously by recent cases like Brutal Kangaroo [1], ProjectSauron [26], USBFerry [9], and Ramsay [34], where malware designed to cross the air-gap by infecting employees' USB drives was used in certain sensitive networks. Perhaps the most studied and publicized instance of this type of malware has been Stuxnet [12], whose initial purpose was to sabotage an industrial control system utilized for centrifuge operation in an air-gapped facility.

We assume in our scenario that the malware has already done privilege escalation and has full permission to tinker with the printer subsystem. The network can either be totally air-gapped or its network traffic may just be heavily monitored, making it difficult to start a connection with a server outside the network without being discovered. Furthermore, we assume the attacker has the capability to record audio, either by infecting an employee's

smartphone or other mobile device with that capability. The malware in the smartphone is kept on a listening state in order to record the printer noise, processes the audio, and save the decoded data, until it detects the smartphone is connected to an outside network and finally sends the data to the attacker. The recorder should be close to the inkjet printer by a few meters when active. The site of the attack is assumed to contain at least one printer with inkjet technology and a movable printhead to which the targeted computer is connected.

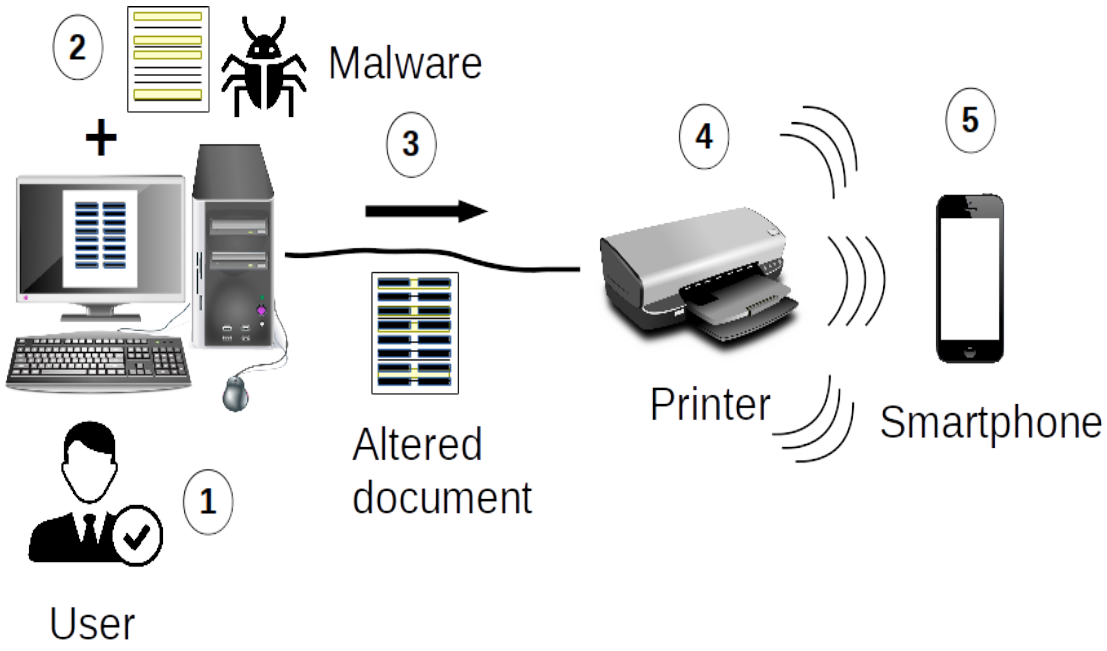


Figure 4.1: Proposed attack scenario for the establishment of the covert channel. **(1)** A trusted user on an infected computer sends a document to print. **(2)** The malware present on that computer injects particular imperceptible patterns into the document pages. **(3)** The altered document passes to the printer. **(4)** The printer produces certain deterministic acoustic emissions depending on the injected patterns. **(5)** A nearby smartphone records the noise made by the printer.

4.1 Receiver

For the design of our communication system, we exploited the acoustic pulse-like properties of the noise produced by the printers' roller mechanism. Basically, our receiver processes the acoustic signal by first breaking it into overlapping segments of 1 second. For each segment, it filters the signal in the range between 3 kHz to 6 kHz, the range proven to be informative, and it computes the Root Mean Square (RMS) envelope of the filtered signal over a sliding window varying between 20 ms to 60 ms, contingent on the printer and modulation scheme employed. The upper envelope is RMS normalized, downsampled and finally the peaks' occurrence in time are calculated based on an empirically determined minimum threshold and minimum distance between peaks. Based on the possibility of manipulating peaks' relative occurrence in time, we established two modulation methods as described in Section 4.2. Our processed signals are shown on Figure 4.2 and 4.3. Normalizing the signal before locating the signal's peaks is particularly useful for us as we can establish the same threshold value to detect peaks irrespective of their absolute values, which vary with the attenuating properties of the channel. The exact shape of the pulses not only differ from printer to printer, they may also present differences when produced by the same printer, furthermore, not only their shape, but their amplitude may change from one pulse to another as a consequence of the complex interaction between the noisy mechanical components of the printer. It is important to note that the parameters used for processing the signal require some tuning contingent on the printer used, so either previous knowledge is needed about the kind of printer being employed, or a classification model should be implemented to distinguish between printers' noise.

4.2 Transmitter

As our attack relies on the time properties of the acoustic pulses generated by the printer, two modulation schemes were found to be useful for data transmission: Differential Pulse Position

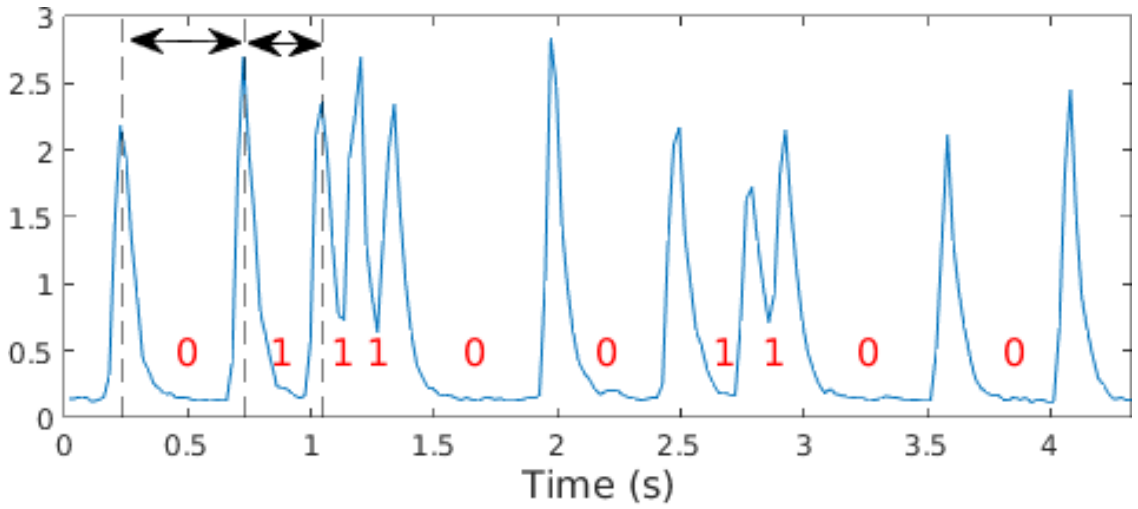


Figure 4.2: The acoustical waveform generated by an HP Photosmart D110 printer, which corresponds to our DPPM modulation scheme, is shown here after processing. In this case a greater difference in the time between pulses corresponds to a 0 bit, while a smaller difference to a 1 bit.

Modulation (DPPM) and an hybrid modulation consisting of DPPM with Frequency Pulse Modulation (FPM). DPPM encodes bits relative to the difference in position between two pulses, thus we defined two different displacements in order to establish binary transmission. An example of this type of modulation can be seen in Figure 4.2. Our hybrid method FPM-DPPM depends more on the overall pulse frequency than on particular time offsets, but the way of processing this frequency contrasts utilizes the discrete relative time differences between pulses, as in DPPM. An example of this can be seen in Figure 4.3. In this modulation scheme, 1 bits are processed as a series of small time offsets between pulses which end on a large offset that separate these type of bits, while 0 bits are composed each of them by two large time offsets. The reason 0 bits are processed in groups of two has to do with the uncertain layout of the documents being printed that can introduce extra pulses, as well as by the fact that 1 bits use one of those long offsets for separation. The reason why 1 bits require this separation, the reason why we cannot just count the resulting total number of small time differences and determine a certain quantity of them to be used as a 1 bit, has to

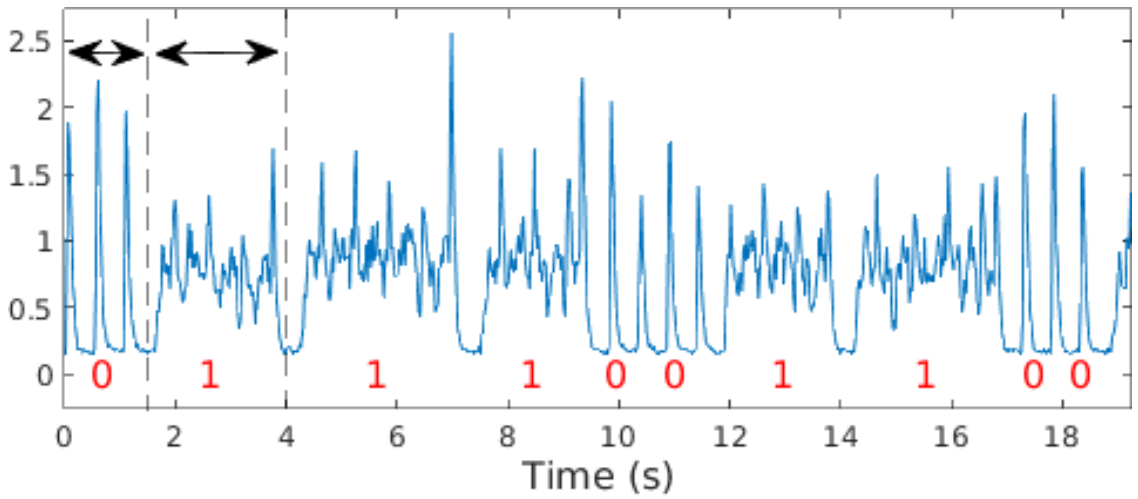


Figure 4.3: The acoustical waveform generated by an HP Photosmart D110 printer, which corresponds to our hybrid FPM-DPPM modulation scheme, is shown here after processing. In this case two pulses with a large time offset is interpreted as a 0 bit, while a cluster of pulses with small time offsets is considered as a 1 bit.

do with the fact that these clusters of pulses with small offsets do not grow linearly with the width of the patterns injected into the documents (the role of these patterns being explained in Section 4.3), so that identifying two consecutive 1 bits would be a convoluted process for some of the printers.

In our communication system data is split into packets, the payload size of these packets being printer and modulation dependent. Each page can be used to transmit a single packet, and besides the payload, each packet includes a 4-bit preamble and a parity bit, as shown in Figure 4.4. Since between each printed page a series of random noises from the printer may interrupt the flow of data and introduce extra bits, e.g., the roller mechanism is activated to expel/introduce a page, the printhead may decide to clean its nozzles, etc.; a sufficiently large preamble is needed to recognize the beginning of a packet, although even then, those extra bits may induce false positives at the packet level. Overall, a matched filter was first used to determine approximately the beginning of the packets followed by the examination of the preamble at the bit level.

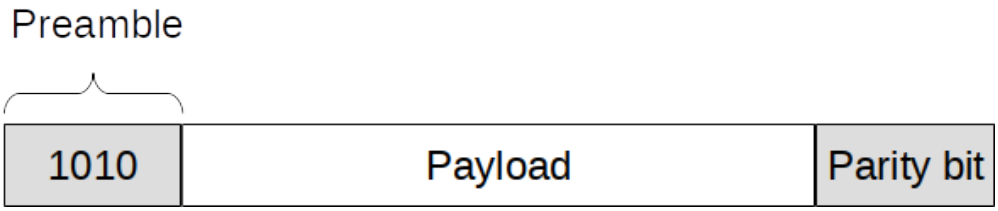


Figure 4.4: The packet structure allocates 4 bits for the preamble, a variable number of bits for the payload (dependent on the printer and modulation scheme used), and a parity bit.

4.3 Injection Patterns

Imperceptible patterns are injected into the documents being printed in order to control the printing process and establish our communication channel. While humans are unable to see these patterns directly, the printer does print them like if these where completely visible. Through the use of a series of very light yellow lines and rectangles imposed on the background of documents, as shown in Figure 4.5, our objective is to manipulate the frequency at which the roller mechanism activates, hence we define three properties in the spatial domain that will control this process: line length, rectangle width and offset between lines/rectangles. The lines and rectangles are drawn across the width of the page and an offset separates each of these elements across the length of the paper. Rectangles are used only for FPM-DPPM while lines are used for both modulation schemes. It was found that line length and rectangle width influenced the frequency of the pulses, while there was a minimum offset at which a pulse (the activation of the roller mechanism) was guaranteed. This minimum spatial offset between lines in the document translates itself into a minimum distance traveled by the paper inside the printer when in operation, thus this offset coupled with the page size limits the absolute number of lines and rectangles we can inject into a document, and as a consequence, the number of bits transmitted per document. This minimum offset is printer dependent, and it appears to be linked with the overall printing throughput, which in itself depends on the number of nozzles on the printhead, the ink droplet ejection rate and the ink

droplet size [8]; future work will be needed to ascertain this relationship. In addition to the constraints generated by the page size, the printer also establishes document margins close to the page's edges beyond which the printer does not operate, thus the number of elements we can inject is further reduced, although borderless printing can be specified in order to break free from this constraint. In the case of FPM-DPPM patterns, after a certain magnitude, the width of the rectangles triggers a frequency change in the printer's paper feed-in mechanism, explained by the fact that rectangles, as shapes with solid colors, require a more intensive ink disposal procedure, and therefore, multiple passes from the printhead. This translates in some cases as an increase in the activation frequency of the roller mechanism coupled with a reduction in the distance traversed by the paper along the printer, or in simply a decrease in that activation frequency.

For DPPM modulation, a constant offset was maintained between lines, and a change in the period of time between acoustic pulses was seen to be directly correlated to a change in the length of the lines, i.e., a longer line resulted in a greater period of time used by the printhead to draw the entire line, thus a greater period of time would elapse until the roller mechanism activated, while the opposite would be true for shorter lines. For FPM-DPPM modulation, all lines and rectangles lengths were fixed, what was important besides the width of the rectangles was the vertical space used by rectangles and clusters of lines. In this modulation, we were interested in establishing a contrast between solid color areas and more sparsely colored areas in order to trigger a change in pulse frequency, where the latter areas would produce a faster paper traversal. The clusters of lines used to define the sparsely colored areas were essential in order to force the printer to operate in a deterministic way irrespective of the elements present on the foreground, where the number of lines per cluster was insignificant as much as they did not cover all the area. To sum up, a particular vertical space with respect to the length of the paper was determined for rectangles, which corresponded to their width, and another vertical space was determined for a unit of cluster of lines. Therefore, each of the spaces occupied by a rectangle or cluster of lines corresponded

to a bit, 1 and 0 respectively, although a small separation between rectangles, to distinguish between 1 bits, is also considered, along with another small separation between rectangles and clusters of lines.

There is one caveat with the modulation schemes as presented above: DPPM modulation only works for blank documents while FPM-DPPM works mostly with text documents. The reason DPPM does not work with text documents is simply because the text occupies most of the page width, so no matter how much we change the lines' length, the printhead will take the same time to print each area, and consequently, the roller mechanism activation will not be controlled. As some documents contain blank pages, and a scenario could present itself in which the malware prints blank pages by its own initiative, DPPM was also tested. Moreover, with this last modulation scheme we can double the amount of transmitted bits per document, as FPM-DPPM rectangles and clusters of lines utilize a greater part of paper space for each of the transmitted bits than DPPM. The reason FPM-DPPM does not work well with blank documents will be discussed in Chapter 6, as it is printer-dependent. One last limitation seen in text documents with FPM-DPPM patterns is that they cannot include images which, with respect to the length of the page, are larger than the minimum width of rectangles, the same constraints being placed on font size. This is basically because they override the control we have over the printing process that depends on it being affected only by our rectangles, which have a certain width that triggers a specific behaviour used for our modulation, as discussed previously.

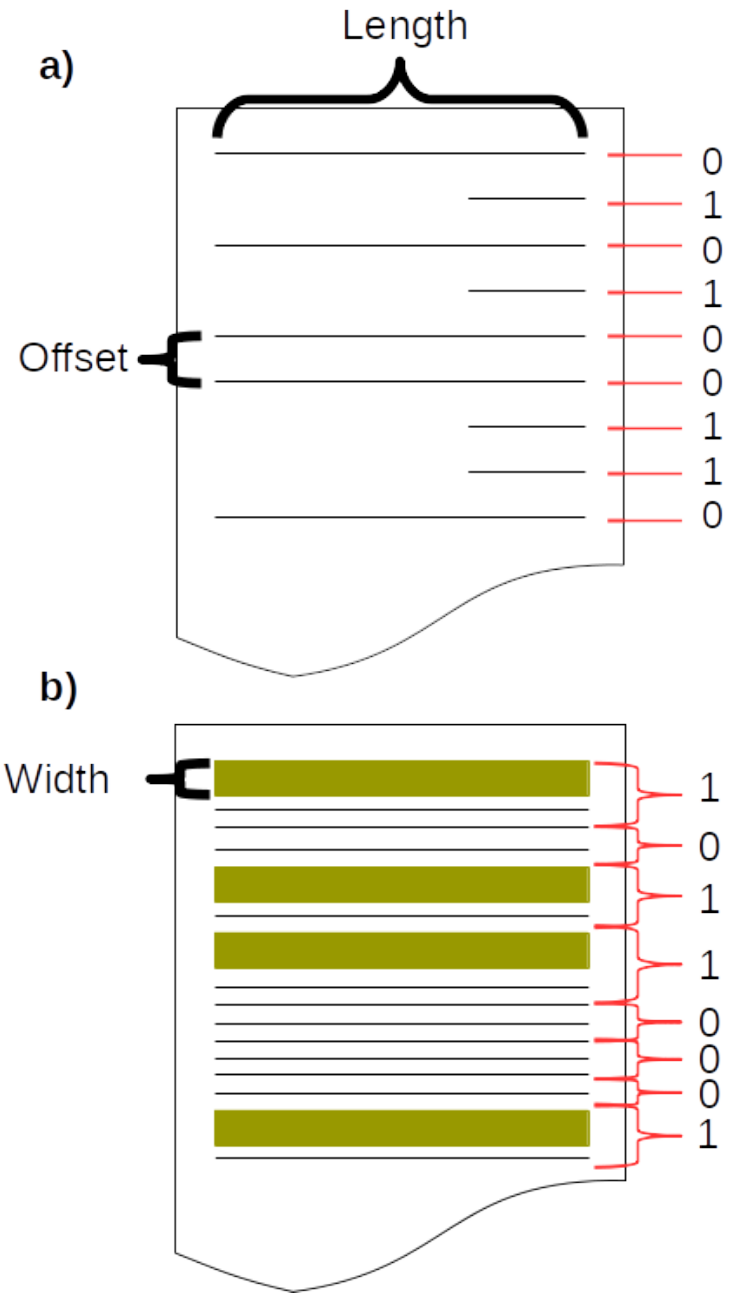


Figure 4.5: The imperceptible patterns injected into documents are arranged in two ways that correspond to the two proposed modulation schemes. **(a)** DPPM. **(b)** FPM-DPPM.

CHAPTER 5

Printer Subsystem in Linux

In Linux and macOS operating systems, the Common Unix Printing System (CUPS) is the current software tasked with managing the interface between the user's computer and its printers. In Linux based-systems, CUPS uses the PostScript Printer Definition file format (PPD) to describe printers' configuration, and a series of filters translate documents being sent to the printer into a final format understandable to it. Back-end filters are the endpoints to this chain of filters whose purpose is sending the data directly to the printers, other filters performing document format conversions and rasterization. CUPS standard print job transfer format passed from being PostScript to Portable Document Format (PDF) in 2006, so that now all important applications send print jobs in that format [30]. Consequently, our attack is designed to inject the imperceptible patterns into documents with the PDF format.

The PDF format became standardized as ISO 32000-1 with version 1.7 in 2008 [2]: it is a device and resolution independent file format, with a structured hierarchy of objects that organizes each pages' content. The former features and the possibility of addressing easily each particular page's content therefore makes the PDF file format perfect for our attack, which requires injecting specific patterns at certain pages in a consistent manner. As per the standard mentioned above, the basic PDF file structure consists of 4 sections:

- A header, which contains the version number.
- A body containing all the objects representing the contents of the document.

- A cross-reference table, which contains one-line entries specifying each object's byte offset according to the beginning of the file.
- A trailer containing the byte offset of the cross-reference table and a reference to certain special objects, such as the catalog dictionary.

In our attack we were interested in modifying the pages' content located inside the body of the PDF file, hence to address the correct structures, parsing of the trailer section was needed to obtain, first, the catalog dictionary, and from it the targeted page. This catalog dictionary serves as the root of the object hierarchy in the PDF file, and from it the page tree object can be accessed, which in itself contains a reference to each page object or leaf node. Each of these nodes contain a series of attributes along with a reference to their content stream, which is usually encoded. The content stream data consists of sequences of instructions that describe what is seen in the page, thus by including certain instructions as discussed in Section 5.2, we can inject our desired patterns and perform our attack.

5.1 Infecting the Linux Printer Subsystem

In our attack model, the malware with necessary privileges will convert the data bits to exfiltrate into a series of patterns to be injected into the documents whenever they are sent by the user's machine to the inkjet printer. By intercepting these documents before they are processed by the printer's driver or subsequent filters, we can succeed in injecting the patterns. This can be done on CUPS, by adding a malicious filter at the beginning of the filter chain used to process documents sent to the printer, or by creating a wrapper for one of the existing filters so that it first calls our malicious code and then executes the original intended code. It is worth recapitulating that each of these filters converts the file from one format into another until the document gets rasterized and processed by the manufacturer's driver, although in a driverless printer setting the PDF can be sent without passing through a special driver [32]. The filters used depend on the file format of the document sent to the

printer, but normally the file is converted at some point in the chain, or it is sent originally, as PDF. In the default filters directory (/usr/lib/cups/filter/ or /usr/libexec/cups/filter/), one finds all the filters available. The ones that end up being used are generally determined by the MIME format of the document being sent to print, a list of them defined by the configuration file *mime.types*; the particular chain of filters assigned for each file format is described on the configuration file *mime.convs*. The PPD files themselves also specify certain filters to use for a particular printer when using the **cupsFilter* attribute, and these files may be found at their default directory (/etc/cups/ppd/).

Basically, the procedure to add a malicious filter would consist of the next steps:

1. Find the MIME format definition we want to hijack in *mime.types* (*application/pdf* in our case), and substitute the label associated with the original format with a new label representing our spoofed MIME format, leaving the original label with a blank argument.

2. Define a MIME conversion in *mime.convs*, that converts our substitute MIME format into the original target format, assigning to this conversion a specific filter.

3. Design the filter previously specified in *mime.convs* and locate it in the appropriate directory. This filter will receive all the contents of the file to be printed whenever a user wants to print a document, and we can include any kind of code we want, the only condition being that at the end, the modified file should be sent to standard output so that it continues passing through the normal filter chain.

More information can be obtained from [25]. By having control of a CUPS filter, we cannot just inject patterns into any document, but include a series of checks so as to ensure the document is in the appropriate format. For example, we examine if images are present in the document by looking for certain operators, in order to ignore the pages where those are, and we check if the document is meant to be printed in grayscale by reviewing the command-line arguments passed into the filter. This last point is important for the stealthiness of

our attack, as it depends on utilizing the yellow ink color that is not employed in grayscale configurations. To deal with this, we force the printer to print in color by changing its configuration file (PPD) accordingly, and we just convert every color value, except the yellow color of the patterns we are using, into an equivalent grayscale, through the next formula:

$$gray = 0.3 \times red + 0.59 \times green + 0.11 \times blue \tag{5.1}$$

On a final note, when using FPM-DPPM for certain printers, the attack will require that the black color of the text and lines present in the document be transformed into a lighter shade of it, by just a small amount that is inappreciable. This procedure effectively reduces the influence of the foreground graphical objects and increases the influence of the imperceptible patterns.

5.2 Injecting Patterns into PDF Files

As explained above, PDF files were chosen as the preferred medium for our attack because: they are a standardized format, they are used extensively in the printing subsystem, and they present a hierarchical structure which makes it easy to locate content streams and modify them. The code we used for our attack can be reduced into the following template that presents the sequence of essential PDF operators utilized for injection:

PDF command sequence used for pattern injection

```
q
red green blue rg
x-position y-position length width re
f
Q
```

Operators q and Q save and restore the graphics state respectively, a measure necessary to avoid modifying graphics control parameters of the surrounding document environment. The *rg* operator is used with three color arguments to specify the color space for the next painting operations. The *re* operator is used to draw rectangles at some particular (x,y) coordinates of the paper with a certain length and width, although in our case we also used it to draw lines by specifying a height of 1. We use any necessary number of times this last operator to inject our patterns. The f operator fills with color the region defined by the previous rectangles. Basically, these simple commands are all what is needed to inject the patterns into the document and by writing these commands before the actual content, we can ensure the patterns are drawn in the background.

CHAPTER 6

Evaluation

In order to test our attack, we used three different printers whose characteristics are shown in Table 6.1. These printers comprise three of the biggest brands for inkjet printers and they vary in the specific technology used for ink droplet ejection (thermal vs piezoelectric) and the technology used for ink storage (cartridge vs ink tank). In further sections we will refer to a particular printer by the name of its manufacturer.

Manufacturer	**Model**	**Type**	**Ink storage technology**
HP	Photosmart D110	Thermal	Cartridge
Epson	L4150	Piezoelectric	Ink Tank
Canon	Pixma MG2410	Thermal	Cartridge

Table 6.1: Characteristics of the printers used to test our attack.

For our tests, we utilized a Samsung Galaxy S8 smartphone as a recording device with a sample rate of 44.1 kHz. Each test was executed while taking the next considerations: the designated printer was put in a small table, while at another table certain distance apart, the recording smartphone was placed. The recordings were carried out in a mostly quiet environment at night, in a room filled with furniture and with a window that faced the street. For each test, 50 pages were printed in total, although only 25 pages were printed uninterruptedly at once, as the printer trays in some models could not carry all of the 50 pages. For each modulation scheme and printer, a random bit payload was first generated

and then translated into a specific pattern to be injected into the pages of the document. The document was sent to print via USB connection to the target printer from a desktop computer with Ubuntu 18.04 and CUPS 2.2.7. Lastly, the essential parameters mentioned in Chapter 4, that define the behaviour of the printing process, were obtained experimentally as shown in Section 6.1.

For our covert channel, we measured its performance when varying the distance between the printer and receiver and the orientation of the receiver's microphone. Particularly for our FPM-DPPM modulation scheme, we conducted tests to measure the efficiency of the communication system when dealing with different types of text layouts and fonts, as summarized in Figure 6.1. Placeholder text was used in each of these documents. We further analyzed the results of our system when facing background noise, different color paper, intercalating modulation and different print quality settings. Additionally, a survey was carried out to see if humans did perceive the patterns left by the attack in the documents, presented in Chapter 7. As a final note, in Appendix A, examples of the actual injected patterns used are shown.

6.1 Obtaining the Parameters of the Modulation Schemes

As mentioned in Chapter 4, three important parameters that control the printing process in our attack are the line length, in the case of DPPM, rectangle width, in the case of FPM-DPPM, and offset between elements for both types of modulation. These parameters vary in accordance to the printer, thereby based on the experiments we performed, we obtained the results shown in Table 6.2, Figure 6.2 and Figure 6.3. The line offset sizes and rectangle widths shown in Table 6.2 correspond to the minimum possible, which does not mean those were the ones we used for the tests present on these sections, as other considerations were taken into account, e.g., in the case of the line offset for the Canon printer, we were considering not only that the printer rollers activated at that particular line

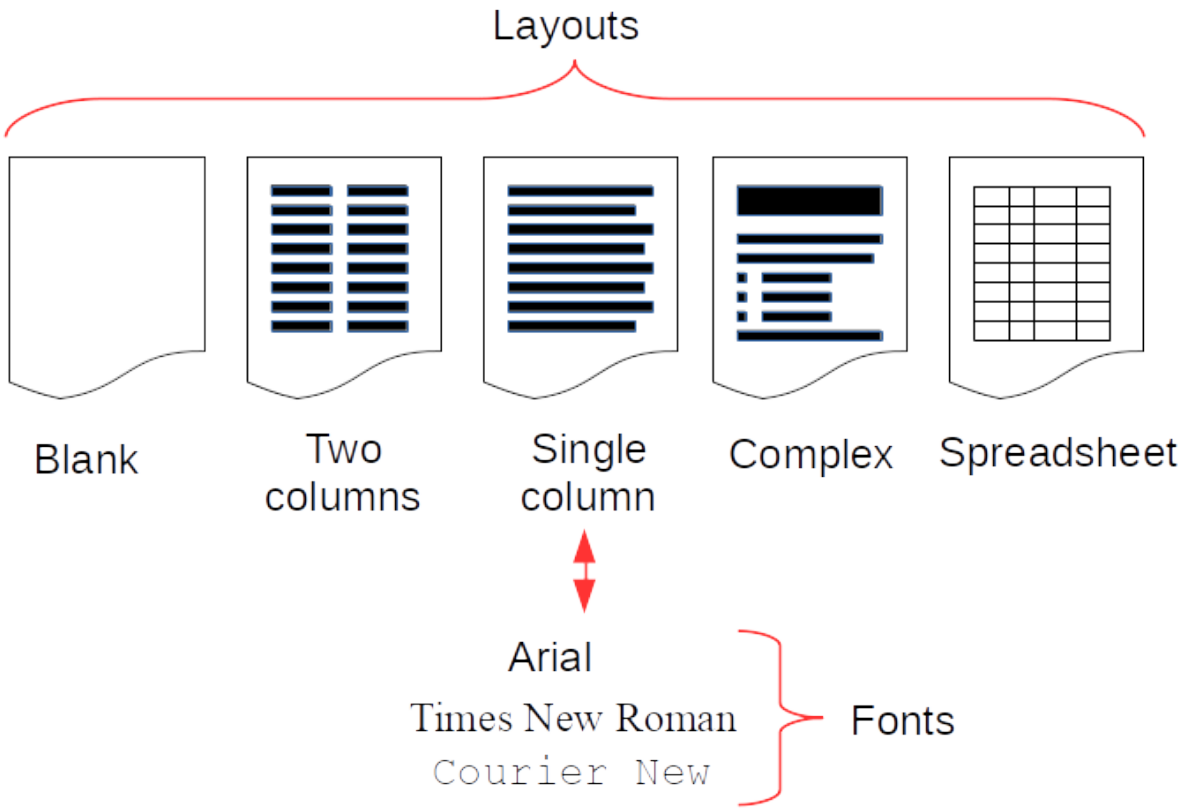


Figure 6.1: Five different layouts were tested as well as three different fonts for the single column case. Appendix B offers a more detailed view of these layouts.

separation, but that the feed-in mechanism produced a sufficient loud noise to be detected. This would only happen for a sufficiently long paper displacement along the printer, as will be revealed below with respect to Figure 6.3. The Canon printer was the only one were we could observe a difference between the minimum line offset size and rectangle width values, although all three printers varied in the size of the space required between rectangles and the space assigned to the clusters of lines when using FPM-DPPM. Indeed, it was because of this last property that the real maximum number of bits per page with the FPM-DPPM modulation scheme varied according to the bit sequence represented, i.e., small offsets of different size were added depending on the case whether a 1 or 0 bit followed another 1 bit, so that injecting a series of 1 bits would not occupy the same space as injecting a series of 0 bits. Therefore, a maximum number of bits per page was chosen for each printer that

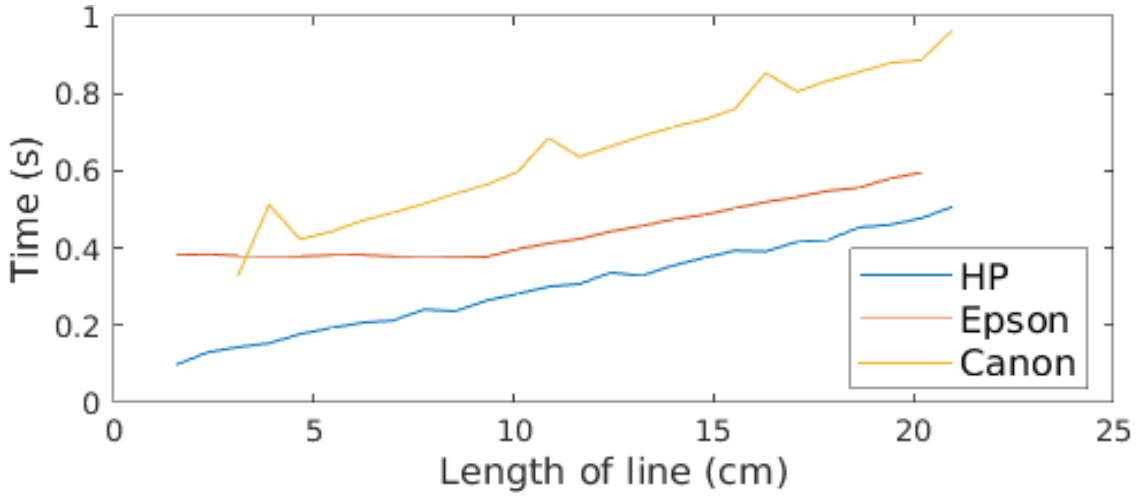


Figure 6.2: The relationship between a line's length and the time it takes for the roller mechanism to activate again is shown for DPPM modulation.

considered this constraint.

Contrary to the FPM-DPPM modulation, in DPPM modulation we had more flexibility while defining the time between pulses to use. Figure 6.2 shows that a linear relationship exists between length of line used and time between pulses, although for the Epson printer there is a point were decreasing the line length does not entail a decrease in the time period. For DPPM, in each printer, two specific line lengths, one at the greater end and another at the lower end of the possible line lengths, were used to perform binary transmission. The shorter lines were the subject of additional constraints as their position with respect to the width of the page (left, right, center), affected the printhead dynamic. The high number of line lengths available for the modulation process gives the impression that the number of bits per symbol could be increased substantially by using multiple line lengths, but in reality, the time differences between pulses exhibit high variance through most printers, making it difficult to implement anything else than binary communication.

Table 6.2 also shows some other important properties of our communication system, like the maximum number of bits per page, which all in all depends on the offset sizes and modulation scheme, the maximum bit rate, given the average printing time of documents

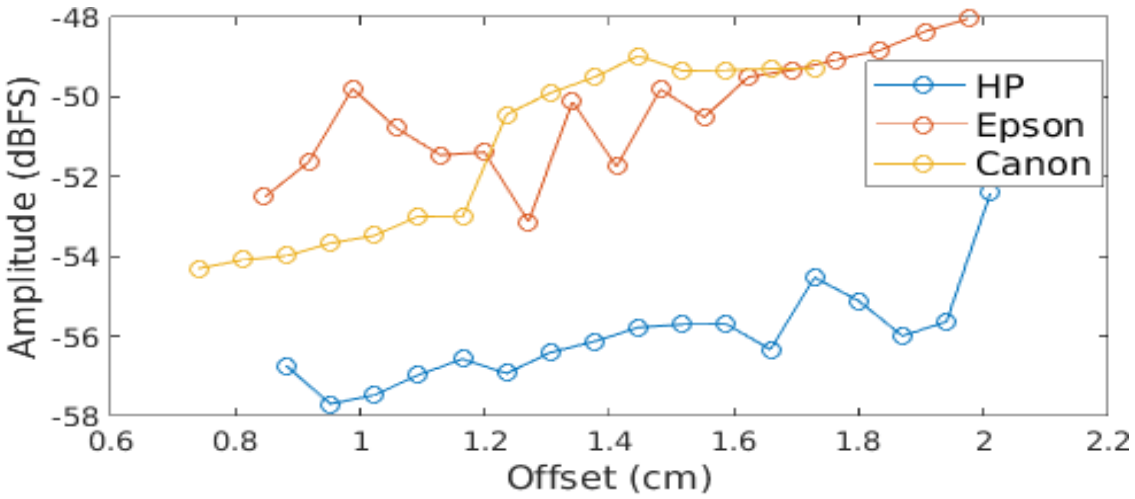


Figure 6.3: The three printers present different relative amplitudes when using the same line offsets. The printers exhibit a positive relationship between relative amplitude and offset size.

with those type of modulations, and the effective rate, which used the Bit Error Ratio results from the distance tests shown in Figure 6.4, when the smartphone is at 50 cm of distance. It is worth explaining from these results that, for example, the bit rate is higher in DPPM modulation, not just because of the greater amount of bits we can encode on the same amount of space compared to the other modulation scheme, but because overall the printer takes a lower amount of time to print a blank page with just lines than one with text and solid color rectangles. Figure 6.3 shows the relative amplitude levels per printer and its relationship with the offset size. The relative amplitudes were measured by taking the absolute values of the amplitudes of the filtered signal in question and expressing them with respect to the maximum amplitude level that can be represented digitally by the sound recorder, therefore the unit of this magnitude is decibels relative to Full Scale (dBFS). From these results we can observe that for a particular paper displacement, the roller mechanism will generate pulses of different intensity, specifically, we can observe that there appears to be a positive trend in loudness with respect to the offset size. The Canon printer's results exhibit clearly this relationship, there is in fact a certain offset value at which the roller mechanism starts

generating louder noise.

Lastly, the processes by which we obtained the relations between parameters were the following: the relation between line offset and relative amplitude was obtained by printing 50 pages and recording the noise produced at a distance of 50 cm, each page consisting of lines of the same length separated vertically by an increasing space of 0.07 cm, starting from the minimum offset possible. The relation between line length and time period was obtained similarly: we printed 50 pages, each page containing a sequence of lines separated by the same vertical offset but progressively decreasing their length by a factor of 0.77 cm.

Printer	**Minimum line offset size**	**Minimum rectangle width**	**Maximum number of bits per page**		**Maximum bit rate**		**Effective bit rate**	
			A	**B**	**A**	**B**	**A**	**B**
HP	0.88 cm		25	10	1.78 bps	0.31 bps	1.76 bps	0.29 bps
Epson	0.85 cm		27	7	1.28 bps	0.33 bps	1.24 bps	0.30 bps
Canon	0.74 cm	1.09 cm	20	6	1.09 bps	0.15 bps	1.08 bps	0.12 bps

Table 6.2: Attack design parameters for each printer. **A** corresponds to DPPM modulation and **B** to FPM-DPPM

6.2 Distance and Receiver Orientation

To test the robustness of our covert channel with respect to the distance at which our smartphone receiver was placed and its orientation relative to the printer, a series of tests were carried out to measure Bit Error Ratio, the results shown in Figure 6.4. For these tests, a 50-page document with a constant layout between pages was used: for the DPPM modulation scheme a blank document was used, while for FPM-DPPM a document with a block of justified text in 12-point Arial font repeated through all pages was used, with exception of the Epson printer, where a blank document was used instead because this

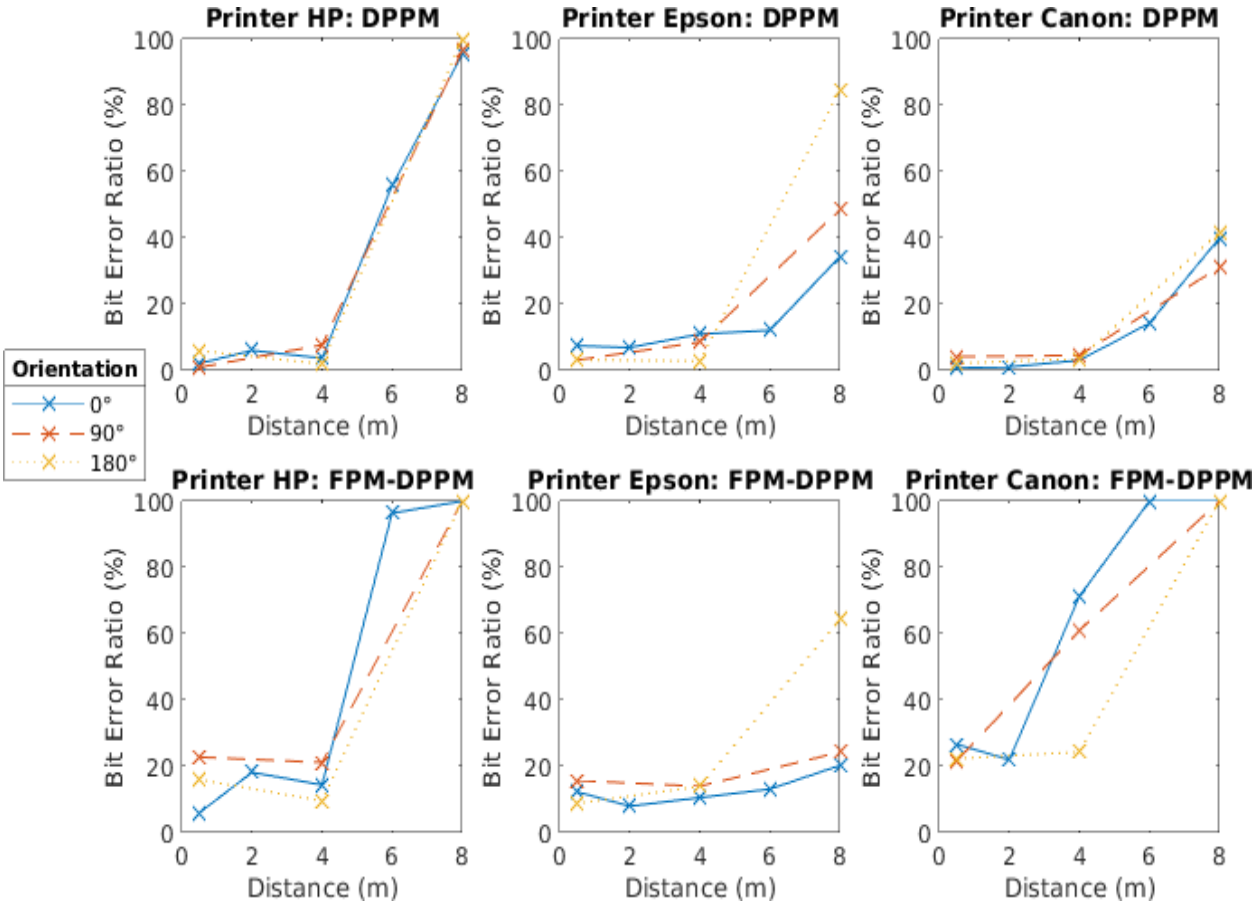


Figure 6.4: The effects of distance and receiver orientation over Bit Error Ratio. Both modulation schemes were tested for each printer.

printer did not present significant changes whether the document had text or not. Five different distances between the printer and receiver were considered: 50 cm, 2 m, 4 m, 6 m and 8 m.

Smartphones' MEMS microphones have been shown previously to be sensitive to the angles of incidence of the incoming acoustic signals, especially as the frequency of the signal increases [11], therefore three different orientations were considered: 0° (microphone facing the noise source), 90° (microphone perpendicular to noise source) and 180° (microphone opposite to noise source). These orientations were each tested at three different distances: 50 cm, 4 m, 8 m. Bit Error Ratio was calculated using the Hamming distance between the

actual received payload and the transmitted one.

From the results shown in Figure 6.4 we can note that in all cases but one, the Bit Error Ratio is lower than 20% at a close distance from the printer, and that the error increases with distance. DPPM modulation outperforms FPM-DPPM by a relatively small margin in almost all cases, except in the Canon printer, where the difference is significant due to this printer using smaller offsets in its printing process when in FPM-DPPM, therefore producing fainter noises more prone to be affected by environmental noise. Orientation does not seem to matter that much until large distances are taken into account, although there is a particular phenomenon observed with all printers at a distance of 4 m, noticeable the most with the Canon printer in FPM-DPPM, where by changing the orientation of the smartphone by 180° the Bit Error Ratio is decreased significantly, an effect which might be a consequence of the furniture arrangement and room structure modifying the signal path in such a way as to maximize the signal power at that particular orientation and position.

6.3 Font

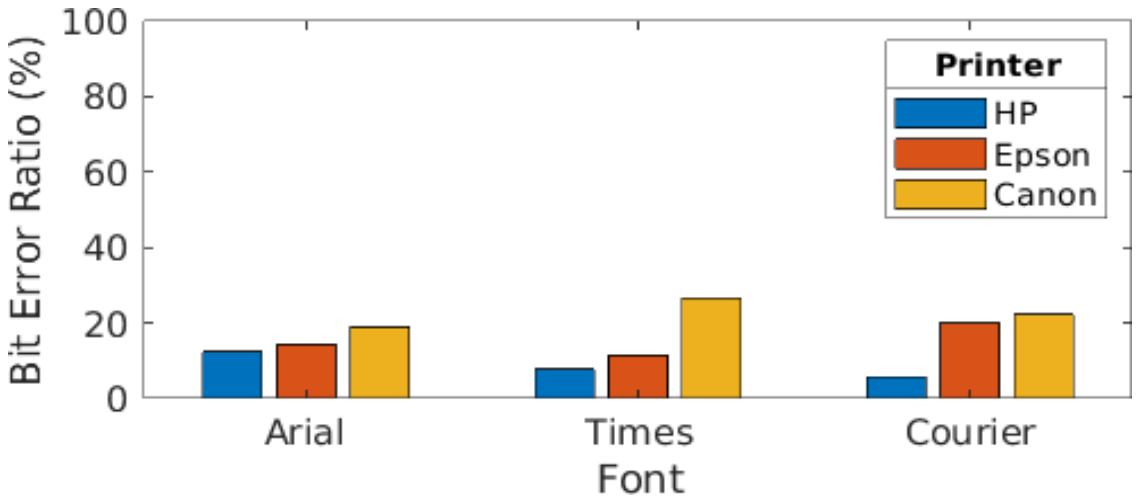


Figure 6.5: Bit Error Ratio according to the font type and printer used. Times and Courier are the abbreviated forms we chose for Times New Roman and Courier New, respectively.

Three different font types were tested on a single column document layout for FPM-DPPM, with the results shown in Figure 6.5. Each page contained left-justified paragraphs of text with 12-point font. The three different fonts used (Arial, Times New Roman and Courier New) correspond to different typeface classes: sans serif, serif and monospace. These classes represent different artistic styles for drawing letters that we thought could interfere with our printer modulation, but in reality the variation was not that significant. The best overall result was achieved with the HP printer and Courier New font, with 5.8% BER, and the worst was achieved with the Canon printer and Times New Roman font, with 26.67% BER. Each printer achieved its lowest Bit Error Ratio when using a particular font, but even then, the difference with the other results of the same printer was not that significant. That being so, we can conclude that our modulation scheme is only slightly affected by the type of font used.

6.4 Layout

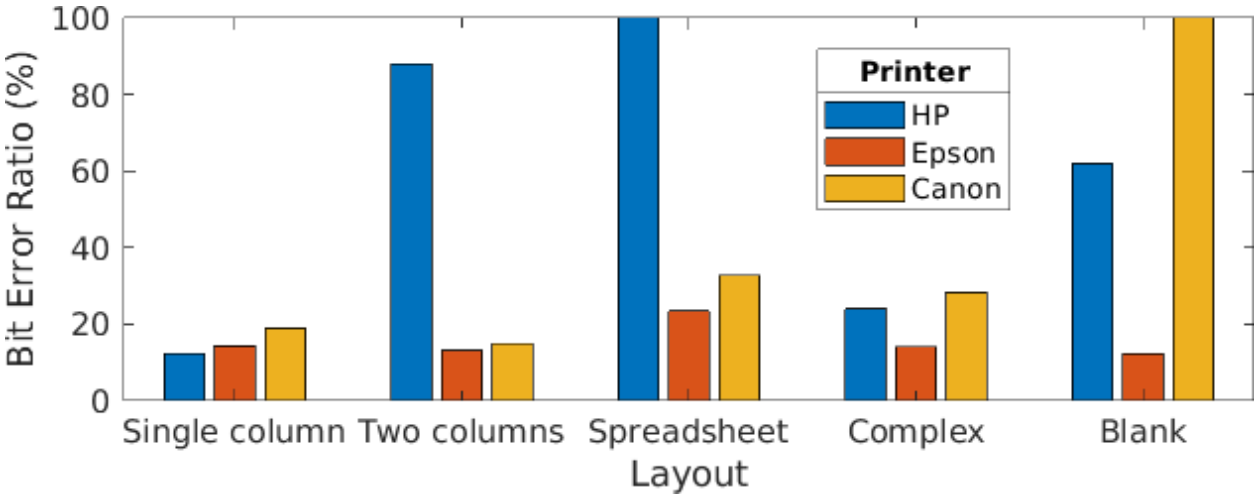


Figure 6.6: Bit Error Ratio according to the layout used on different printers.

Five different layouts were used to test the impact of this variable on the printing process. Figure 6.6 shows the results. While we are aware that the number of layouts found on

existing documents can be infinitesimally large, we chose what we thought to be the five most representative ones, consisting of one document with a single column and left-justified text, another with two columns and justified text, a spreadsheet-type document with a large table that comprised the entirety of the pages, a complex layout which combined a single column layout with subtitles of different point size and a number of lists, tables, white spaces, bold letters and italics; and lastly, a blank layout was also tested. Each of these layouts used 12-point font as the base size, except for the spreadsheet layout were 10-point font was used, and the blank layout were none was used. The single column layout worked as our base case, the two columns layout being used to test the impact of multiple columns, the spreadsheet in order to test the impact of the tables, particularly the influence of lines combined with text; the complex layout was used to test our covert channel in a more *realistic* layout, and the blank layout was used to observe if FPM-DPPM could be used when no text was present.

The Epson printer proved to be mostly insensitive to layout changes, with the highest of its Bit Error Ratio results being at 20% with the spreadsheet layout, and the lowest at 12.29% with the blank layout. The Canon printer, which achieved a 15% BER as its best result in a two column layout, did degrade its performance by a certain magnitude when using the spreadsheet and complex layouts (32.67% and 28.33% of BER respectively), although what did affect it significantly was the absence of text in the blank layout. Particularly for this printer, in the absence of text but with colored rectangles and lines of the same length filling the white space in the document, the time between each paper feed-in becomes constant, although the actual displacement between each feed-in changes according to the amount of color needed, an effect which was of no use to us. The HP printer achieved as its best result a BER of 12.4% in a single column layout, but it was overall the most sensitive to the layout variable: first of all, it was discovered that the printing process was affected by the amount of spacing between text lines. In the case of the two columns layout, where a BER of 88% was achieved, at some point the text between the two columns becomes unaligned, eliminating effectively any vertical white space between lines of text and prompting the printer to behave

like if it was printing a continuous colored object. In the case of the spreadsheet format, the fact that the tables consist of continuous vertical lines is sufficient to make ineffective the modulation process as a consequence of the same phenomenon explained previously. In the case of the blank layout, where a BER of 62% was achieved, the fact that there is no text between rectangles, only clusters of lines, allows the printer to print these patterns in a faster way, thus generating larger paper displacements between each roller mechanism activation, so that less pulses are created in total. Additionally to these constraints, the HP printer can fail to interpret correctly the spaces placed between a series of contiguous rectangles, and keep operating as if the two rectangles were joined, when a particular alignment with the text is reached inside the document.

6.5 Color of Paper

At first, the imperceptible contrast achieved by the yellow ink on white paper was thought not to be repeatable in other color substrates, e.g., blue, the complementary color of yellow. But because the ink that color printers use has some transparency and therefore is not opaque, using yellow on black or blue substrates produces instead darker colors that are even less perceptible to the human eye, effectively proving what the subtractive color model predicts when adding colors. Further tests were carried out in yellow, orange and magenta colored paper but in all of these, what dominated was the color of the paper over the sparse yellow dots.

6.6 Background Noise

To test the performance of the covert channel in a noisy setting, we played a recording of office noise through a nearby computer's speakers, at a sound level which we deemed to be similar to that found in those environments. The effect on the Bit Error Ratio was significant,

Printer	Bit Error Ratio	
	DPPM	**FPM-DPPM**
HP	52 %	53.6 %
Epson	11.33 %	10.57 %
Canon	13.3 %	52.67 %

Table 6.3: Bit Error Ratio of each printer and modulation scheme when there is high environmental noise.

as shown in Table 6.3. Only the Epson printer proved to be resilient to this test as the noise did not affect significantly its performance in any of the two modulations, while the Canon printer was resilient only for DPPM modulation. This is directly correlated to the intensity of sound produced by the activation of their roller mechanism: those cases mentioned before simply produce higher amplitude noise such that the environmental noise does not interfere with them significantly. In the case of the Canon printer, while its DPPM modulation is not really affected, its FPM-DPPM modulation is impacted because in this last modulation method, the paper displacements are smaller in size than in the other scheme, and as shown before in Figure 6.3, there is a relationship between line offset, paper displacement and intensity of noise. Overall, for the HP printer, roller noise is of a lesser magnitude than for the other printers, and the contrary situation occurs for the Epson printer, hence the results obtained.

6.7 Intercalating Modulation

The tests considered in the previous sections use only one type of modulation through all the pages of the targeted document, but in real world scenarios documents may contain both text and blank pages. To prove that the two modulation schemes could coexist in the same document, we constructed a document of 50 pages in which half were blank pages injected

with DPPM patterns, and the other half were pages with text and FPM-DPPM patterns. These two types of pages were intercalated throughout the document. For each printer, we tested one of these documents, and the final results revealed that there was no significant impact in the performance of the communication system: for the HP printer we achieved a Bit Error Ratio of 15.49%, for the Epson printer one of 5.54%, and for the Canon printer, a 10.95% BER. Only small modifications were made to the signal processing algorithm.

6.8 Printer Quality Settings

Changing the default print quality configuration from normal to any other special configuration, e.g., high quality, does render our attack ineffective as the printing process stops being modulated by these patterns. Either a constant small paper displacement is introduced, in the case of high quality, or a constant large paper displacement is used, in the case of a low quality setting, that ruins the modulation.

CHAPTER 7

Color Perception

Since our attack leaves a residue in the printed documents, i.e., yellow colored lines and rectangles (Figure 7.1), it is necessary to understand how humans perceive colors in order to make our patterns as imperceptible as possible.

7.1 Human Perception of Yellow Dots

The human visual system consists of cones and rods, the first ones used for daytime vision and the second for low light environments, thus we focus on the former. Trichromacy theory explains that cones appear in three subtypes that are sensitive to different light wavelenghts (short, middle, long) and that roughly correspond to three colors (blue, green, red). There are fewer short wavelength cones than the other types of cones, and additionally, these type of cones are less sensitive to light. Furthermore, middle and long wavelength cones overlap in their sensitivity distribution. Opponent process theory explains that cone receptors link together and form pairs of opposing colors, where activation of one of the receptors on the pair inhibits the other. These opposing color pairs are red-green, yellow-blue and black-white. The black-white opposing color pair is related to our perception of luminance and it represents mainly the combination of outputs from both the red and green cones, while the yellow-blue pair is defined as the difference between this black-white signal and the output of the blue cone. As such, yellow light stimulates both red and green cones near their peak sensitivity but inhibits the blue cone receptor, while white light stimulates all three at their peak sensitivities, but because the blue cone does not contribute that much to the luminance

perception, the difference in luminance between yellow and white is perceived to be small and hard to see [42]. As the visual system relies more on luminance contrast than on the absolute levels of luminance or than on chromatric contrast for the interpretation of information [41], it is evident that a very light tint of yellow on a white background might not be perceived by humans.



Figure 7.1: A magnified view of two printed letters surrounded by yellow dots. These two letters happen to be located inside one of the rectangles used for FPM-DPPM.

7.2 Perception Survey

The exact yellow color value used for the injected patterns varied by a small amount between printers and modulation schemes. The Canon printer used the darkest shade of yellow of the three. While all printers could in principle use the lightest tone of this color, in practice there was a point were the lines and rectangles stopped having the desired effect on the printing process and a darker shade was needed. To test whether our patterns were truly imperceptible, an experiment was carried out with 20 subjects from all genders, comprising an age range between 13 to 68. This test consisted in a total of 24 pages:

- 3 blank pages with DPPM modulation, each of them printed with a different device.

- 3 sets of 5 pages, corresponding to the layouts shown in Figure 6.1, each of them printed with a different device.
- 4 pages corresponding to the non-blank layouts mentioned above, printed without any kind of modulation.
- 2 blank pages.

Each subject would examine and write their observations regarding any *weird* feature they noticed in the page, without taking into account any possible meaning that could be given to the random text. After examining all the pages and writing their observations, the subjects were asked if they saw any yellow stripes in the documents and told to examine one of the pages in which we considered the stripes to be more visible (a blank page with FPM-DPPM patterns of the Canon printer). If they did see those stripes, we asked them if they saw them in other pages.

Our results indicate that only 2 people were able to notice the yellow patterns at first sight, and 12 other people were able to see them but only when told at the end. Indeed, for some of the FPM-DPPM Canon pages, after some initial struggles, one can notice the patterns, but only in some pages. Besides that, no one could see the patterns of the HP and Epson printers, even when told. What this survey appeared to reveal is that age is not necessarily the factor in play at first, as the 2 people who saw the lines in the first pass were in their thirties. Otherwise, when the interviewees were told of the presence of the yellow patterns, younger people were more prone to see the lines than older people (greater than 50), which is predictable, as contrast sensitivity diminishes with age [10], and is clearly aggravated by existing eye conditions. What seems to have been more in play for the younger participants was their *weirdness* threshold and the detail with which they examined the documents. Some people might think having documents with certain background patterns is not something deemed to be strange, one might think about the random defects we sometimes see in all kinds of printouts, but there are other people, which

based on their sort of professional formation, like graphical designers, are more keen to detect these type of details, as it occurred for one of the two positive cases. Obviously there is a limit to this visual acuity, which is demonstrated in the experiment, as no one could see the imperceptible patterns in the documents printed with the Epson and HP printers. Given the sample we surveyed, we can conclude that our injected patterns should pass unnoticed in normal office settings.

CHAPTER 8

Discussion

In this chapter we discuss the relevance of our attack, taking into consideration its limitations and the possible countermeasures that could be developed.

8.1 Countermeasures

The greatest caveat found on our attack concerns the physical evidence left in the printed documents. While these traces are not visible for humans in normal light conditions, there are two ways to check if a document has been injected with the attack patterns:

1. By pointing a blue LED into the document, the clusters of tiny dots become visible as black colored dots, basically as a consequence of yellow being the complementary color of blue, i.e., the blue light is absorbed completely by the dots.

2. By using a high resolution scanner to scan the document. A resolution equal or higher than 600 dpi is sufficient to capture the dots [6], as was corroborated with an Epson scanner.

More elaborate countermeasures could be based on the time it takes for the printer to print certain documents, or even one could consider the particular sounds made during the printing process, for example: text documents without images should not take that much to print in normal conditions and should produce sounds with a certain intensity. Table 8.1 shows a comparison of normal printing time of text documents against the printing time

Printer	Time Duration	
	Normal	Altered
HP	25 s	32.4 s
Epson	8.4 s	21 s
Canon	33.4 s	39.5 s

Table 8.1: Time duration of the printing process for a normal text document vs. time duration of the printing process for an altered text document.

used for documents with our injected patterns. Previous work in 3D printing has employed sound signatures to validate that the printed object is not being sabotaged [5]. From the manufacturers perspective, homogenizing the printing process in order to produce constant sounds irrespective of the type of document might prevent this type of attack, but it would entail a huge performance cost as printing would take a lot more of time. A more effortless way of dealing with this problem would be simply producing artificial noise so as to mask the noise generated by the printer, or just restricting access to the printer by moving it to a monitored soundproof room. By using instead a monochrome inkjet printer, it would be obvious when the documents get intervened, as the yellow color would be converted into grayscale, rendering visible the injected stripes. Alternatively, if using a multichrome printer and colored ink is not employed frequently, a substantial decrease on colored ink levels should sound alarms. At the level of computer software, creating a checksum for each of the CUPS filters could prevent any kind of modification of the filter chain, although control over the implementation of new filters should be also monitored.

8.2 Limitations and Future Work

Overall, one of the greatest limitations for the attack presented in this thesis is a direct consequence of the heterogeneity in the printing devices. We tested only one printer model

for each of the three brands considered, and the differences in the injected pattern's parameter values were substantial, requiring each of them some optimization. Not only that, as shown in Chapter 6, some printers with particular modulation schemes may be more susceptible to noise than others, which limits the application of this attack to settings were the environmental noise is low; with other printers, the layout of the document might be the determinant whether the attack works or not. Particularly, the fact that this attack is rendered ineffective when printing documents with images is a great limitation. Given the variety of printers in the market, an attacker would need to consider at least tuning its attack to the most sold printers in order to have any success in its endeavor, unless it manages to obtain information about the printer type used in the targeted network. Future work should try to test if the attack method we propose in this thesis is applicable to all the inkjet printers on the market, especially, it would be necessary to test if inkjet printers with stationary printheads are susceptible to this attack. Moreover, testing laser printers to see if they can be manipulated with imperceptible patterns like shown in this thesis would amplify the attack's range of operation and would therefore cover almost all printers actually used in office settings. Finally, implementing this attack on computers with operating systems other than Linux is also left for future work.

CHAPTER 9

Conclusion

In this thesis, we present a new type of acoustic covert channel attack that targets inkjet printers in order to exfiltrate information from air-gapped networks through its unintended emissions. To achieve this, first of all, a relationship was discovered that linked the layout of graphical objects in a document with the noise generated when printed. We exploited this relationship by injecting special imperceptible patterns into all documents being sent to print, thus modulating the printing process in such a way that a deterministic communication channel was established using the acoustic emanations. Two different modulation methods were presented, tuned to work better with either blank documents or text documents. Lastly, we evaluated the attack with three printers from three of the biggest brands in the market, using documents with different layouts, font types, color paper, and recording the acoustic signals at different distances, with different levels of noise, and print quality settings. The results reveal that on average, a receiver close to the printer will achieve a Bit Error Ratio lower than 30%.

The attack detailed in this work is yet another step forward in the endeavor to identify potential misuse of unintended emissions, in a world were these attack vectors have been ignored consistently. It is demonstrated again the importance of monitoring all kinds of emanations from devices kept in sensitive locations, as they have become an easy target for attackers. Defense systems relying purely on software security have become insufficient faced with the increasing number of sophisticated attacks from advance persistent threat (APT) actors, and isolation does not appears to solve the problem. Nevertheless, awareness of the

problem is the first step towards a more robust defense system.

APPENDIX A

Injection Pattern Examples

In this appendix we show some examples of pages with the proposed injection patterns.

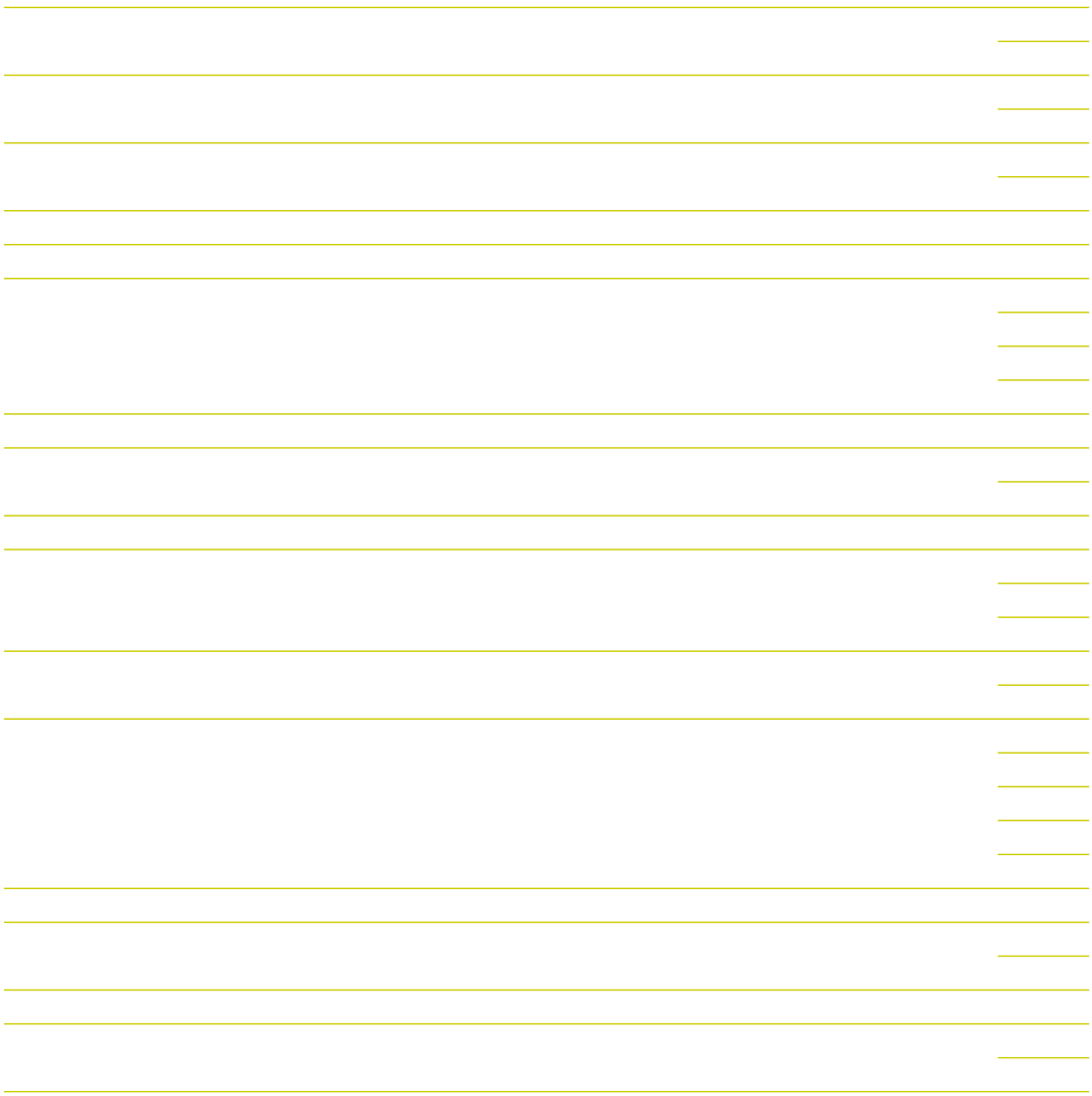


Figure A.1: DPPM pattern of the bit sequence **10101011100011011001010000110110** injected into a page prepared to be sent to the Epson printer (an extra line is placed at the end of the injected patterns to separate the final pulse from the page expulsion noise). The color shown here for the injected patterns was **not** the one chosen for the real attack.

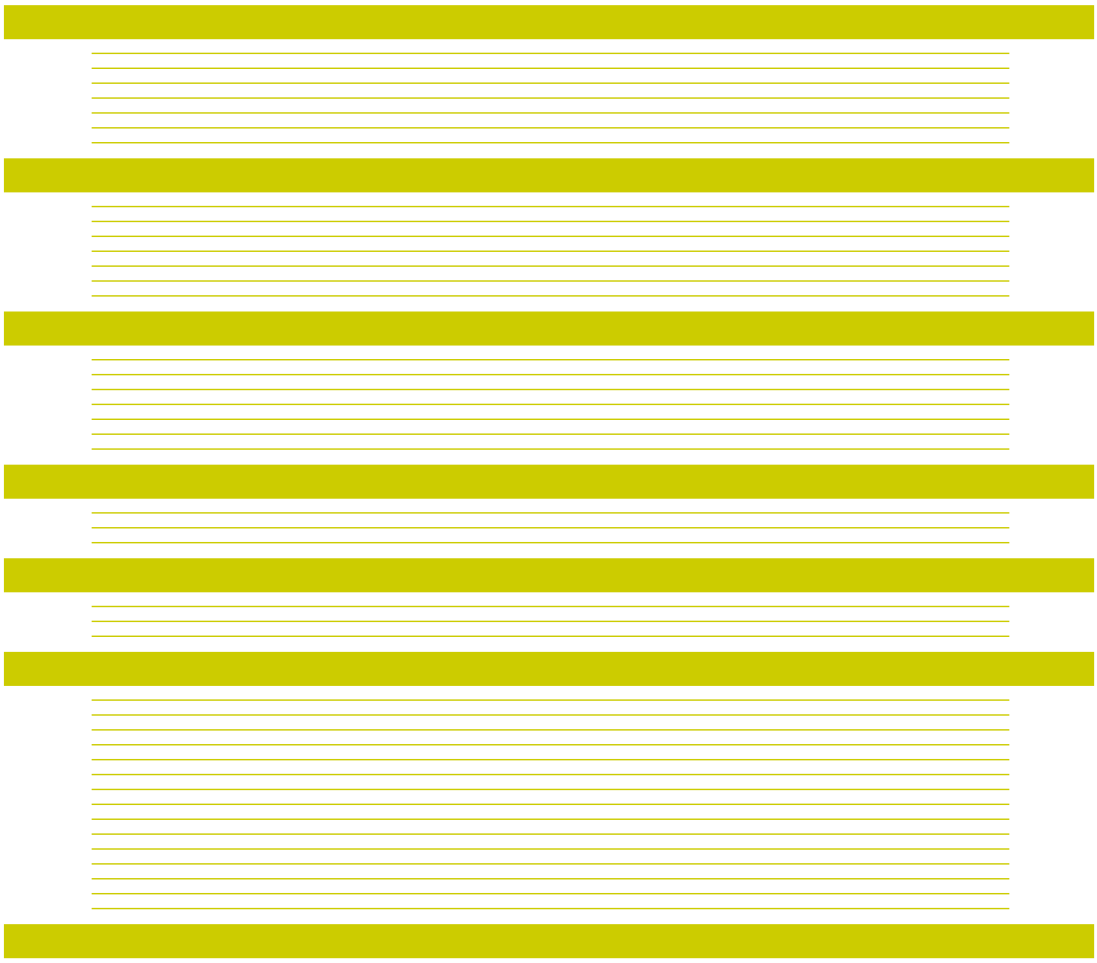


Figure A.2: FPM-DPPM pattern of **101010111000** injected into a page prepared to be sent to the Epson printer (an extra rectangle is placed at the end of the injected patterns to separate the final pulses from the page expulsion noise). The color shown here for the injected patterns was **not** the one chosen for the real attack.



Figure A.3: FPM-DPPM pattern of **101010111000** injected into a text page prepared to be sent to the Epson printer. The color shown here for the injected patterns was **not** the one chosen for the real attack.

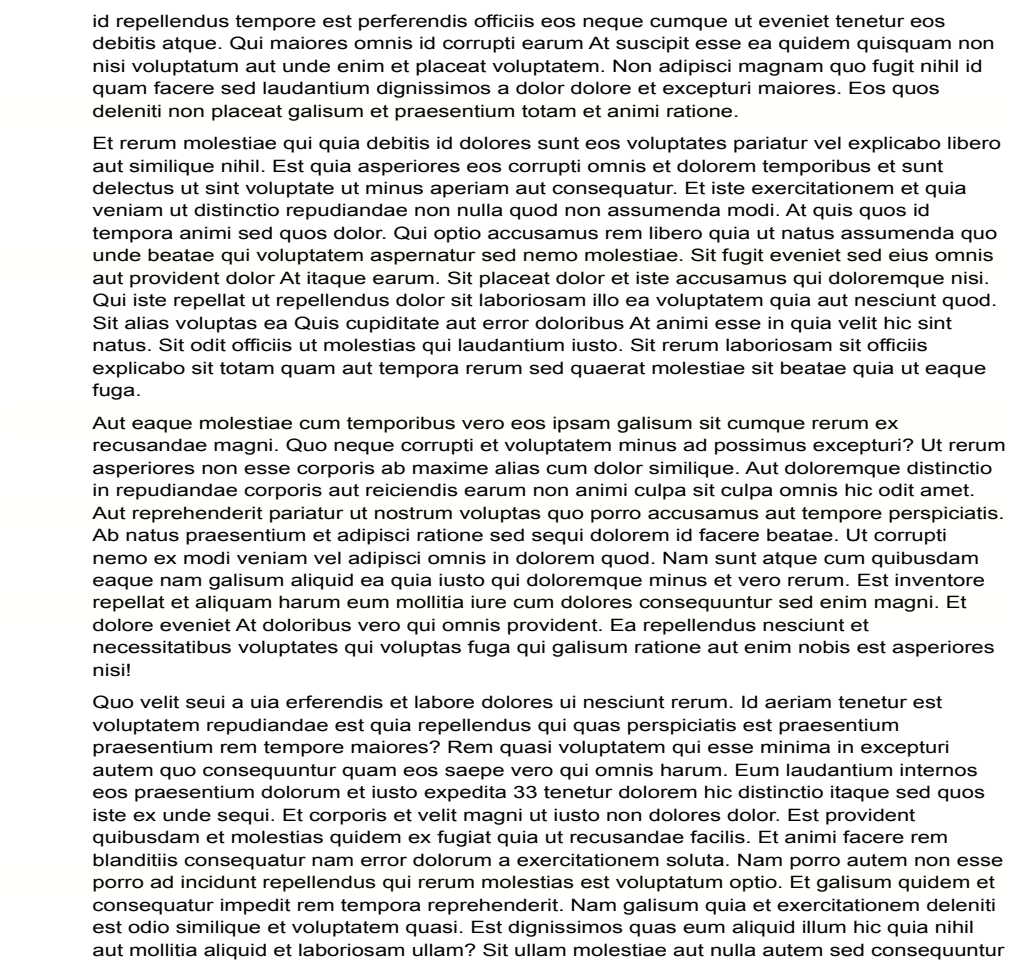


Figure A.4: FPM-DPPM pattern of **101010111000** injected into a text page prepared to be sent to the Epson printer. The color shown here for the injected patterns was the one chosen for the real attack.

APPENDIX B

Document Layouts

In this appendix we show a sample of the different layouts used for the experiments.

Lorem ipsum dolor sit amet. Eum cumque totam qui cupiditate quis ut fugit quia qui earum enim est velit consequatur ex sint distinctio qui cumque suscipit? Ut quod eius qui vitae quam non unde quasi 33 omnis possimus? Est fugiat molestiae sed doloribus expedita et galisum sapiente eum iure quasi est iste voluptates qui repellat excepturi aut nisi tenetur. Id praesentium minima est consequatur perferendis est totam molestiae non voluptatem voluptas. Ut voluptates modi et alias perspiciatis ut maiores sint id enim harum et praesentium adipisci qui accusamus quia. 33 consequatur magnam in optio enim ea omnis deleniti eos quae nemo. 33 optio amet qui illo tempore aut officiis recusandae est impedit consequatur vel ullam culpa in delectus internos? Eos temporibus quos aut dicta nihil et libero sapiente? Eum ipsum officia ad nisi laboriosam minus galisum quo numquam quidem hic explicabo voluptates aut rerum quasi qui consequatur commodi. Ut nihil ipsam ex pariatur commodi in labore quia aut dolor eveniet ut iure consequatur qui labore doloremque. Rem dolores ipsa vel placeat adipisci est molestiae necessitatibus sit aliquam earum qui fugit quasi et voluptatum amet.

Eos sunt voluptatum ut dolorum voluptas est nesciunt quae quo adipisci voluptate. Et suscipit tempora non voluptatem similique sit consequatur repellendus et voluptatibus nihil. Et excepturi consectetur aut quia laborum quo dolorem consequuntur At suscipit nulla et dolorem nemo. Qui fugit culpa quo ducimus iusto ut delectus voluptate ut eligendi repellendus et enim velit est fugit suscipit ea voluptatem quae. Non odio necessitatibus qui corporis repellat non consequatur laborum sed quasi rerum hic consequatur assumenda et saepe aspernatur? Est voluptates cupiditate ut cupiditate dolorem eos molestiae animi est autem sunt quo rerum reprehenderit ab repellat rerum! Qui ullam voluptas est atque Quis ut dicta Quis? Qui numquam dolores sit quia voluptas vel aliquam laboriosam ad labore obcaecati sit ipsa quam et deserunt neque.

Hic enim minima aut veniam minima nam vitae explicabo est internos excepturi. Et impedit tenetur ea alias dolorem et quibusdam dolorem eos galisum modi eos nulla laboriosam sit blanditiis tempore cum magni voluptatem. Id asperiores natus aut cumque earum aut adipisci autem. Et tempora magni quo architecto laboriosam et nemo dolorem. Qui ipsum odio cum odit beatae aut voluptatem dolor est dolor quae aut blanditiis vitae. Qui Quis voluptates ut voluptatem sint aut sequi galisum aut expedita minus cum quaerat laboriosam nam commodi natus ad reprehenderit voluptates. Vel maxime molestias eum numquam possimus qui cumque obcaecati eos fugit nesciunt eum itaque autem. Non vitae recusandae qui reiciendis ducimus et aliquam amet. A voluptatem accusamus eos deleniti repellat ad dignissimos aliquam. Non numquam assumenda ea ducimus quisquam et harum rerum in laudantium optio rem exercitationem voluptatem. Hic quae omnis et perferendis illum non repellat vitae sit similique officiis a autem quae sed corrupti odio.

Ad sequi possimus est reiciendis numquam et ullam vero et nobis commodi quo perferendis assumenda ut commodi nulla. Aut voluptatem libero ad quas numquam non voluptas voluptates non harum dolorem qui voluptates voluptatem sit tempora labore! Est exercitationem porro At autem sunt sed tempore magni ex tenetur galisum et aperiam minus? Ut inventore numquam qui illo quisquam eos reiciendis rerum et tempore dolores id reiciendis

Figure B.1: Single column layout with Arial font.

Lorem ipsum dolor sit amet. Eum cumque totam qui cupiditate quis ut fugit quia qui earum enim est velit consequatur ex sint distinctio qui cumque suscipit? Ut quod eius qui vitae quam non unde quasi 33 omnis possimus? Est fugiat molestiae sed doloribus expedita et galisum sapiente eum iure quasi est iste voluptates qui repellat excepturi aut nisi tenetur. Id praesentium minima est consequatur perferendis est totam molestiae non voluptatem voluptas. Ut voluptates modi et alias perspiciatis ut maiores sint id enim harum et praesentium adipisci qui accusamus quia. 33 consequatur magnam in optio enim ea omnis deleniti eos quae nemo. 33 optio amet qui illo tempore aut officiis recusandae est impedit consequatur vel ullam culpa in delectus internos? Eos temporibus quos aut dicta nihil et libero sapiente? Eum ipsum officia ad nisi laboriosam minus galisum quo numquam quidem hic explicabo voluptates aut rerum quasi qui consequatur commodi. Ut nihil ipsam ex pariatur commodi in labore quia aut dolor eveniet ut iure consequatur qui labore doloremque. Rem dolores ipsa vel placeat adipisci est molestiae necessitatibus sit aliquam earum qui fugit quasi et voluptatum amet.

Eos sunt voluptatum ut dolorum voluptas est nesciunt quae quo adipisci voluptate. Et suscipit tempora non voluptatem similique sit consequatur repellendus et voluptatibus nihil. Et excepturi consectetur aut quia laborum quo dolorem consequuntur At suscipit nulla et dolorem nemo. Qui fugit culpa quo ducimus iusto ut delectus voluptate ut eligendi repellendus et enim velit est fugit suscipit ea voluptatem quae. Non odio necessitatibus qui corporis repellat non consequatur laborum sed quasi rerum hic consequatur assumenda et saepe aspernatur? Est voluptates cupiditate ut cupiditate dolorem eos molestiae animi est autem sunt quo rerum reprehenderit ab repellat rerum! Qui ullam voluptas est atque Quis ut dicta Quis? Qui numquam dolores sit quia voluptas vel aliquam laboriosam ad labore obcaecati sit ipsa quam et deserunt neque.

Hic enim minima aut veniam minima nam vitae explicabo est internos excepturi. Et impedit tenetur ea alias dolorem et quibusdam dolorem eos galisum modi eos nulla laboriosam sit blanditiis tempore cum magni voluptatem. Id asperiores natus aut cumque earum aut adipisci autem. Et tempora magni quo architecto laboriosam et nemo dolorem. Qui ipsum odio cum odit beatae aut voluptatem dolor est dolor quae aut blanditiis vitae. Qui Quis voluptates ut voluptatem sint aut sequi galisum aut expedita minus cum quaerat laboriosam nam commodi natus ad reprehenderit voluptates. Vel maxime molestias eum numquam possimus qui cumque obcaecati eos fugit nesciunt eum itaque autem. Non vitae recusandae qui reiciendis ducimus et aliquam amet. A voluptatem accusamus eos deleniti repellat ad dignissimos aliquam. Non numquam assumenda ea ducimus quisquam et harum rerum in laudantium optio rem exercitationem voluptatem. Hic quae omnis et perferendis illum non repellat vitae sit similique officiis a autem quae sed corrupti odio.

Ad sequi possimus est reiciendis numquam et ullam vero et nobis commodi quo perferendis assumenda ut commodi nulla. Aut voluptatem libero ad quas numquam non voluptas voluptates non harum dolorem qui voluptates voluptatem sit tempora labore! Est exercitationem porro At autem sunt sed tempore magni ex tenetur galisum et aperiam minus? Ut inventore numquam qui illo quisquam eos reiciendis rerum et tempore dolores id reiciendis officiis. Sed dolore voluptatem et voluptas voluptas sed quisquam laboriosam in illo dolore et amet ipsa sed quis modi et iusto eveniet! At possimus architecto non eius praesentium qui libero harum qui deserunt deserunt in nihil repudiandae. Quo consequatur voluptas cum nostrum magnam eos consequatur tempore est alias inventore. A quos dolorem ut incidunt porro aut blanditiis quia id nesciunt quas. Est modi commodi et totam ratione non

Figure B.2: Single column layout with Times font.

Lorem ipsum dolor sit amet. Eum cumque totam qui cupiditate quis ut fugit quia qui earum enim est velit consequatur ex sint distinctio qui cumque suscipit? Ut quod eius qui vitae quam non unde quasi 33 omnis possimus? Est fugiat molestiae sed doloribus expedita et galisum sapiente eum iure quasi est iste voluptates qui repellat excepturi aut nisi tenetur. Id praesentium minima est consequatur perferendis est totam molestiae non voluptatem voluptas. Ut voluptates modi et alias perspiciatis ut maiores sint id enim harum et praesentium adipisci qui accusamus quia. 33 consequatur magnam in optio enim ea omnis deleniti eos quae nemo. 33 optio amet qui illo tempore aut officiis recusandae est impedit consequatur vel ullam culpa in delectus internos? Eos temporibus quos aut dicta nihil et libero sapiente? Eum ipsum officia ad nisi laboriosam minus galisum quo numquam quidem hic explicabo voluptates aut rerum quasi qui consequatur commodi. Ut nihil ipsam ex pariatur commodi in labore quia aut dolor eveniet ut iure consequatur qui labore doloremque. Rem dolores ipsa vel placeat adipisci est molestiae necessitatibus sit aliquam earum qui fugit quasi et voluptatum amet.

Eos sunt voluptatum ut dolorum voluptas est nesciunt quae quo adipisci voluptate. Et suscipit tempora non voluptatem similique sit consequatur repellendus et voluptatibus nihil. Et excepturi consectetur aut quia laborum quo dolorem consequuntur At suscipit nulla et dolorem nemo. Qui fugit culpa quo ducimus iusto ut delectus voluptate ut eligendi repellendus et enim velit est fugit suscipit ea voluptatem quae. Non odio necessitatibus qui corporis repellat non consequatur laborum sed quasi rerum hic consequatur assumenda et saepe aspernatur? Est voluptates cupiditate ut cupiditate dolorem eos molestiae animi est autem sunt quo rerum reprehenderit ab repellat rerum! Qui ullam voluptas est atque Quis ut dicta Quis? Qui numquam dolores sit quia voluptas vel aliquam laboriosam ad labore obcaecati sit ipsa quam et deserunt neque.

Hic enim minima aut veniam minima nam vitae explicabo est internos excepturi. Et impedit tenetur ea alias dolorem et quibusdam dolorem eos galisum modi eos nulla laboriosam sit blanditiis tempore cum magni voluptatem. Id asperiores natus aut cumque earum aut adipisci autem. Et tempora magni quo architecto laboriosam et nemo dolorem. Qui ipsum odio cum odit beatae aut voluptatem dolor est dolor quae aut blanditiis vitae. Qui Quis voluptates ut voluptatem sint aut sequi galisum aut expedita minus cum quaerat laboriosam nam commodi natus ad reprehenderit voluptates. Vel maxime molestias eum numquam possimus qui cumque obcaecati eos fugit nesciunt eum itaque autem. Non vitae recusandae qui reiciendis ducimus et aliquam amet. A

Figure B.3: Single column layout with Courier font.

Lorem ipsum dolor sit amet. Eum cumque totam qui cupiditate quis ut fugit quia qui earum enim est velit consequatur ex sint distinctio qui cumque suscipit? Ut quod eius qui vitae quam non unde quasi 33 omnis possimus? Est fugiat molestiae sed doloribus expedita et galisum sapiente eum iure quasi est iste voluptates qui repellat excepturi aut nisi tenetur. Id praesentium minima est consequatur perferendis est totam molestiae non voluptatem voluptas. Ut voluptates modi et alias perspiciatis ut maiores sint id enim harum et praesentium adipisci qui accusamus quia. 33 consequatur magnam in optio enim ea omnis deleniti eos quae nemo. 33 optio amet qui illo tempore aut officiis recusandae est impedit consequatur vel ullam culpa in delectus internos? Eos temporibus quos aut dicta nihil et libero sapiente? Eum ipsum officia ad nisi laboriosam minus galisum quo numquam quidem hic explicabo voluptates aut rerum quasi qui consequatur commodi. Ut nihil ipsam ex pariatur commodi in labore quia aut dolor eveniet ut iure consequatur qui labore doloremque. Rem dolores ipsa vel placeat adipisci est molestiae necessitatibus sit aliquam earum qui fugit quasi et voluptatum amet.

Eos sunt voluptatum ut dolorum voluptas est nesciunt quae quo adipisci voluptate. Et suscipit tempora non voluptatem similique sit consequatur repellendus et voluptatibus nihil. Et excepturi consectetur aut quia laborum quo dolorem consequuntur At suscipit nulla et dolorem nemo. Qui fugit culpa quo ducimus iusto ut delectus voluptate ut eligendi repellendus et enim velit est fugit suscipit ea voluptatem quae. Non odio necessitatibus qui corporis repellat non consequatur laborum sed quasi rerum hic consequatur assumenda et saepe aspernatur? Est voluptates cupiditate ut cupiditate dolorem eos molestiae animi est autem sunt quo rerum reprehenderit ab repellat rerum! Qui ullam voluptas est atque Quis ut dicta Quis? Qui numquam dolores sit quia voluptas vel aliquam laboriosam ad labore obcaecati sit ipsa quam et deserunt neque.

Hic enim minima aut veniam minima nam vitae explicabo est internos excepturi. Et impedit tenetur ea alias dolorem et quibusdam dolorem eos galisum modi eos nulla laboriosam sit blanditiis tempore cum magni voluptatem. Id asperiores natus aut cumque earum aut adipisci autem. Et tempora magni quo architecto laboriosam et nemo dolorem. Qui ipsum odio cum odit beatae aut voluptatem dolor est dolor quae aut blanditiis vitae. Qui Quis voluptates ut voluptatem sint aut sequi galisum aut expedita minus cum quaerat laboriosam nam commodi natus ad reprehenderit voluptates. Vel maxime molestias eum numquam possimus qui cumque obcaecati eos fugit nesciunt eum itaque autem. Non vitae recusandae qui reiciendis ducimus et aliquam amet. A voluptatem accusamus eos deleniti repellat ad dignissimos aliquam. Non numquam assumenda ea ducimus quisquam et harum rerum in laudantium optio rem exercitationem voluptatem. Hic quae omnis et perferendis illum non repellat vitae sit similique officiis a autem quae sed corrupti odio.

Ad sequi possimus est reiciendis numquam et ullam vero et nobis commodi quo perferendis assumenda ut commodi nulla. Aut voluptatem libero ad quas numquam non voluptas voluptates non harum dolorem qui voluptates voluptatem sit tempora labore! Est exercitationem porro At autem sunt sed tempore magni ex tenetur galisum et aperiam

Figure B.4: Two columns layout.

loc	mun	est	est_d	ageb	t_loc	cd_a	ent	con	upm	d_sem	n_pro_viv	v_sel	n_ent
1	1	30	2	0	1	14	1	40001	101725	101	92	1	5
1	1	30	2	0	1	14	1	40001	101725	101	119	2	5
1	1	30	2	0	1	14	1	40001	101725	101	50	3	5
1	1	30	2	0	1	14	1	40001	101725	101	77	4	5
1	1	30	2	0	1	14	1	40001	101725	101	19	5	5
1	1	30	2	0	1	14	1	40002	100812	101	60	1	5
1	1	30	2	0	1	14	1	40002	100812	101	14	2	5
1	1	30	2	0	1	14	1	40002	100812	101	34	3	5
1	1	30	2	0	1	14	1	40002	100812	101	84	4	5
1	1	30	2	0	1	14	1	40002	100812	101	104	5	5
1	1	20	1	0	1	14	1	40003	101251	101	1	1	5
1	1	20	1	0	1	14	1	40003	101251	101	23	2	5
1	1	20	1	0	1	14	1	40003	101251	101	87	3	5
1	1	20	1	0	1	14	1	40003	101251	101	108	4	5
1	1	20	1	0	1	14	1	40003	101251	101	48	5	5
1	1	40	3	0	1	14	1	40004	101471	101	45	1	5
1	1	40	3	0	1	14	1	40004	101471	101	74	2	5
1	1	40	3	0	1	14	1	40004	101471	101	107	3	5
1	1	40	3	0	1	14	1	40004	101471	101	144	4	5
1	1	40	3	0	1	14	1	40004	101471	101	14	5	5
1	1	30	2	0	1	14	1	40005	101206	101	25	1	5
1	1	30	2	0	1	14	1	40005	101206	101	57	2	5
1	1	30	2	0	1	14	1	40005	101206	101	128	3	5
1	1	30	2	0	1	14	1	40005	101206	101	117	4	5
1	1	30	2	0	1	14	1	40005	101206	101	85	5	5
1	5	20	4	0	2	14	1	40006	102058	101	36	1	5
1	5	20	4	0	2	14	1	40006	102058	101	54	2	5
1	5	20	4	0	2	14	1	40006	102058	101	74	3	5
1	5	20	4	0	2	14	1	40006	102058	101	94	4	5
1	5	20	4	0	2	14	1	40006	102058	101	22	5	5
1	1	30	2	0	1	14	1	40007	101273	201	31	1	4
1	1	30	2	0	1	14	1	40007	101273	201	4	2	4
1	1	30	2	0	1	14	1	40007	101273	201	54	3	4
1	1	30	2	0	1	14	1	40007	101273	201	76	4	4
1	1	30	2	0	1	14	1	40007	101273	201	98	5	4
1	1	20	1	0	1	14	1	40008	100737	201	69	1	4
1	1	20	1	0	1	14	1	40008	100737	201	35	2	4
1	1	20	1	0	1	14	1	40008	100737	201	104	3	4
1	1	20	1	0	1	14	1	40008	100737	201	7	4	4
1	1	20	1	0	1	14	1	40008	100737	201	89	5	4
1	1	40	3	0	1	14	1	40009	100663	201	42	1	4
1	1	40	3	0	1	14	1	40009	100663	201	61	2	4
1	1	40	3	0	1	14	1	40009	100663	201	83	3	4
1	1	40	3	0	1	14	1	40009	100663	201	19	4	4
1	1	40	3	0	1	14	1	40009	100663	201	39	5	4
1	1	30	2	0	1	14	1	40010	101014	201	96	1	4
1	1	30	2	0	1	14	1	40010	101014	201	62	2	4
1	1	30	2	0	1	14	1	40010	101014	201	1	3	4
1	1	30	2	0	1	14	1	40010	101014	201	24	4	4
1	1	30	2	0	1	14	1	40010	101014	201	45	5	4

Figure B.5: Spreadsheet layout.

6. Vel mollitia ipsam non accusantium accusantium eum fuga cupiditate?

Sed velit officiis ut exercitationem minus aut adipisci.

Lorem ipsum dolor sit amet. Eos soluta velit sit dolorem consectetur ut sequi mollitia? Ut galisum dolorem aut impedit eaque ut tempore aliquam et quasi culpa quo quia ducimus aut dolores eligendi et debitis dolore? Et enim ipsum *Qui blanditiis* et rerum corrupti ut quia ipsa!

At suscipit culpa ut dolore aperiam ea perspiciatis repellat.

Sit totam amet **Est expedita** cum voluptas voluptas vel quae delectus. Ut excepturi maiores sed molestiae vero At aliquam debitis qui sapiente quae eos eaque aut eaque odio ut officia rerum. Vel odio galisum *Vel natus et tempora esse rem dolores repellat* est facilis exercitationem. Aut fugit maiores ea quasi dolor aut voluptatem labore non accusamus delectus et placeat facere qui sint consequuntur?

Qui velit dolores et nostrum numquam.
: Id quae odio quo facere natus non eius minima.

Aut nihil sint et laboriosam deleniti.
: Qui porro nemo eum iusto facilis sit expedita earum aut vitae odit.

Et dignissimos quia qui ipsum.

Cum alias distinctio in ipsam adipisci **Eum consequatur qui alias accusantium est atque iure est nobis atque**. Ut quia consequatur vel molestiae enim *Rem cumque et minima minima eos necessitatibus quisquam ut itaque accusantium*. Ut ipsam doloribus in labore dolorem id accusamus facilis! Qui rerum consequatur ut facilis esse sit ratione sint ut obcaecati nesciunt aut repudiandae sint?

1. Aut totam veritatis a molestiae quae sed molestias iusto.
2. Qui nemo animi ut corrupti fugit cum veniam corrupti.

Eum dolores Quis non odio voluptatem.

Id incidunt laboriosam *Ab doloribus eum iste dignissimos qui ullam quam* quia assumenda qui fuga rerum et culpa recusandae? Aut dolor laudantium hic quae consectetur **Qui architecto**. Ut odio temporibus sed illum placeat ut quas nemo est corrupti quaerat.

Est ipsam dignissimos et corrupti sapiente!

Excepturi vitae *Non nobis et quisquam iusto qui minima similique qui minus optio* id dignissimos nemo ut esse consequuntur. Aut quibusdam consequatur **Non laborum eos**

Figure B.6: Complex layout.

APPENDIX C

Source Code

The source code of all the software artifacts used during the realization of this thesis may be found online at: https://github.com/nesl/InkFiltration

REFERENCES

[1] *Brutal Kangaroo – Drifting Deadline v1.2 - User Guide*, February 2016 (accessed October 11, 2020). `https://wikileaks.org/vault7/document/Brutal_Kangaroo-DriftingDeadline-V1_2-User_Guide/`.

[2] Adobe Systems Incorporated. *Document management — Portable document format — Part 1: PDF 1.7*, first edition, 07 2008 (accessed September 15, 2020). `https://www.adobe.com/content/dam/acom/en/devnet/pdf/PDF32000_2008.pdf`.

[3] M. A. Al Faruque, S. R. Chhetri, A. Canedo, and J. Wan. Acoustic side-channel attacks on additive manufacturing systems. In *2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS)*, pages 1–10, 2016.

[4] Michael Backes, Markus Dürmuth, Sebastian Gerling, Manfred Pinkal, and Caroline Sporleder. Acoustic side-channel attacks on printers. In *Proceedings of the 19th USENIX Conference on Security*, USENIX Security'10, page 20, USA, 2010. USENIX Association.

[5] S. Belikovetsky, Y. A. Solewicz, M. Yampolskiy, J. Toh, and Y. Elovici. Digital audio signature for 3d printing integrity. *IEEE Transactions on Information Forensics and Security*, 14(5):1127–1141, 2019.

[6] J. A. Briffa, C. Culnane, and H. Treharne. Imperceptible printer dot watermarking for binary documents. In Peter Schelkens, Touradj Ebrahimi, Gabriel Cristóbal, Frédéric Truchetet, and Pasi Saarikko, editors, *Optics, Photonics, and Digital Technologies for Multimedia Applications*, volume 7723, pages 166 – 174. International Society for Optics and Photonics, SPIE, 2010.

[7] Brent Carrara and Carlisle Adams. A survey and taxonomy aimed at the detection and measurement of covert channels. IH&MMSec '16, page 115–126, New York, NY, USA, 2016. Association for Computing Machinery.

[8] Rafael Castrejon-Pita, W. Baxter, J. Morgan, S. Temple, G. Martin, and I. Hutchings. Future, opportunities and challenges of inkjet technologies. *Atomization and Sprays*, 23:541–565, 08 2013.

[9] Joey Chen. Tropic trooper's usbferry targets air-gapped networks, 2020 (accessed October 2, 2020). `https://www.trendmicro.com/en_us/research/20/e/tropic-troopers-back-usbferry-attack-targets-air-gapped-environments.html`.

[10] GUNILLA DEREFELDT, GUNNAR LENNERSTRAND, and BJÖRN LUNDH. Age variations in normal human contrast sensitivity. *Acta Ophthalmologica*, 57(4):679–690, 1979.

[11] I. Djurek, T. Grubeša, and N. Orlić. Measurements of analog mems microphones. In *2019 2nd International Colloquium on Smart Grid Metrology (SMAGRIMET)*, pages 1–4, 2019.

[12] Nicolas Falliere, Liam O Murchu, and Eric Chien. W32.stuxnet dossier. Technical report, Symantec, February 2011 (accessed October 11, 2020). `https://www.wired.com/images_blogs/threatlevel/2011/02/Symantec-Stuxnet-Update-Feb-2011.pdf`.

[13] M. Guri. Cd-leak: Leaking secrets from audioless air-gapped computers using covert acoustic signals from cd/dvd drives. In *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 808–816, 2020.

[14] M. Guri, D. Bykhovsky, and Y. Elovici. Brightness: Leaking sensitive data from air-gapped workstations via screen brightness. In *2019 12th CMI Conference on Cybersecurity and Privacy (CMI)*, pages 1–6, 2019.

[15] M. Guri, O. Hasson, G. Kedma, and Y. Elovici. An optical covert-channel to leak data through an air-gap. In *2016 14th Annual Conference on Privacy, Security and Trust (PST)*, pages 642–649, 2016.

[16] M. Guri, M. Monitz, and Y. Elovici. Usbee: Air-gap covert-channel via electromagnetic emission from usb. In *2016 14th Annual Conference on Privacy, Security and Trust (PST)*, pages 264–268, 2016.

[17] M. Guri, M. Monitz, Y. Mirski, and Y. Elovici. Bitwhisper: Covert signaling channel between air-gapped computers using thermal manipulations. In *2015 IEEE 28th Computer Security Foundations Symposium*, pages 276–289, 2015.

[18] M. Guri, Y. Solewicz, and Y. Elovici. Mosquito: Covert ultrasonic transmissions between two air-gapped computers using speaker-to-speaker communication. In *2018 IEEE Conference on Dependable and Secure Computing (DSC)*, pages 1–8, 2018.

[19] M. Guri, B. Zadov, and Y. Elovici. Odini: Escaping sensitive data from faraday-caged, air-gapped computers via magnetic fields. *IEEE Transactions on Information Forensics and Security*, 15:1190–1203, 2020.

[20] Mordechai Guri. Air-viber: Exfiltrating data from air-gapped computers via covert surface vibrations, 2020.

[21] Mordechai Guri. Power-supplay: Leaking data from air-gapped systems by turning the power-supplies into speakers, 2020.

[22] Mordechai Guri, Yosef Solewicz, Andrey Daidakulov, and Yuval Elovici. Fansmitter: Acoustic data exfiltration from (speakerless) air-gapped computers, 2016.

[23] Mordechai Guri, Yosef Solewicz, Andrey Daidakulov, and Yuval Elovici. Acoustic data exfiltration from speakerless air-gapped computers via covert hard-drive noise ('diskfiltration'). In Simon N. Foley, Dieter Gollmann, and Einar Snekkenes, editors, *Computer Security – ESORICS 2017*, pages 98–115, Cham, 2017. Springer International Publishing.

[24] Harold Joseph Highland. The tempest over leaking computers. *Abacus*, 5(2):10–18, January 1988.

[25] Jsmeix. *SDB:Using Your Own Filters to Print with CUPS*, 2010. `https://en.opensuse.org/SDB:Using_Your_Own_Filters_to_Print_with_CUPS`.

[26] Kaspersky Lab. *THE PROJECTSAURON APT*, August 2016 (accessed October 11, 2020). `https://media.kasperskycontenthub.com/wp-content/uploads/sites/43/2018/03/07190154/The-ProjectSauron-APT_research_KL.pdf`.

[27] P. Krishnamurthy, F. Khorrami, R. Karri, D. Paul-Pena, and H. Salehghaffari. Process-aware covert channels using physical instrumentation in cyber-physical systems. *IEEE Transactions on Information Forensics and Security*, 13(11):2761–2771, 2018.

[28] Butler W. Lampson. A note on the confinement problem. *Commun. ACM*, 16(10):613–615, October 1973.

[29] Daniel L. Lau and Gonzalo R. Arce. *Modern Digital Halftoning*, pages 1–19. 02 edition, 2008.

[30] The Linux Foundation. *PDF as Standard Print Job Format*, 2016 (accessed October 3, 2020). `https://wiki.linuxfoundation.org/openprinting/pdf_as_standard_print_job_format`.

[31] B. Nassi, A. Shamir, and Y. Elovici. Xerox day vulnerability. *IEEE Transactions on Information Forensics and Security*, 14(2):415–430, 2019.

[32] Brian Potkin. Cupsdriverlessprinting, 2017 (accessed October 3, 2020). `https://wiki.debian.org/CUPSDriverlessPrinting`.

[33] S. Rokka Chhetri, A. Barua, S. Faezi, F. Regazzoni, A. Canedo, and M. A. Al Faruque. Tool of spies: Leaking your ip by altering the 3d printer compiler. *IEEE Transactions on Dependable and Secure Computing*, pages 1–1, 2019.

[34] Ignacio Sanmillan. Ramsay: A cyber-espionage toolkit tailored for air-gapped networks, May 2020 (accessed October 12, 2020). `https://www.welivesecurity.com/2020/05/13/ramsay-cyberespionage-toolkit-airgapped-networks/`.

[35] Seth Schoen. Secret code in color printers lets government track you. *Electronic Frontier Foundation*, 2005 (accessed October 10, 2020). `https://www.eff.org/press/archives/2005/10/16`.

[36] Atasheh Soleimani-Gorgani. 14 - inkjet printing. In Joanna Izdebska and Sabu Thomas, editors, *Printing on Polymers*, pages 231 – 246. William Andrew Publishing, 2016.

[37] Chen Song, Feng Lin, Zhongjie Ba, Kui Ren, Chi Zhou, and Wenyao Xu. My smartphone knows what you print: Exploring smartphone-based side-channel attacks against 3d printers. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, page 895–907, New York, NY, USA, 2016. Association for Computing Machinery.

[38] Cihan Ulaş, Ulaş Aşık, and Cantürk Karadeniz. Analysis and reconstruction of laser printer information leakages in the media of electromagnetic radiation, power, and signal lines. *Computers & Security*, 58:250 – 267, 2016.

[39] Rick Waasdorp, Oscar van den Heuvel, Floyd Versluis, Bram Hajee, and Murali Ghatkesar. Accessing individual 75-micron diameter nozzles of a desktop inkjet printer to dispense picoliter droplets on demand. *RSC Advances*, 8:14765–14774, 04 2018.

[40] Matthias Wandel. Un-building an ink jet printer, (accessed October 8, 2020). `https://woodgears.ca/tech/printer.html`.

[41] Colin Ware. *Information Visualization: Perception for Design: Second Edition*, pages 95–141. 04 2004.

[42] Colin Ware. *Visual Thinking: For Design*, volume 53, pages 65–85. 04 2008.