

Remote video eavesdropping using a software-defined radio platform

Martin Marinov
St Edmund's College



*A dissertation submitted to the University of Cambridge
in partial fulfilment of the requirements for the degree of
Master of Philosophy in Advanced Computer Science*

University of Cambridge
Computer Laboratory
William Gates Building
15 JJ Thomson Avenue
Cambridge CB3 0FD
UNITED KINGDOM

Email: mtm46@cam.ac.uk

May 31, 2014

Declaration

I Martin Marinov of St Edmund's College, being a candidate for the M.Phil in Advanced Computer Science, hereby declare that this report and the work described in it are my own work, unaided except as may be specified below, and that the report does not contain material that has already been used to any substantial extent for a comparable purpose.

Total word count: ??,???

Signed:

Date:

This dissertation is copyright ©2014 Martin Marinov.
All trademarks used in this dissertation are hereby acknowledged.

Abstract

This dissertation presents a software toolkit for remotely eavesdropping video monitors using a Software Defined Radio (SDR) receiver. It exploits compromising emanations from cables carrying video signals. Analogue video is usually transmitted one line of pixels at a time encoded as a varying current. This generates a wideband electromagnetic wave that can be picked up by an SDR receiver. The presented software can map the received field strength of each pixel to a grayscale value in order to show a real-time false colour estimate of the original video signal.

The software significantly lowers the costs required for undertaking a practical attack compared to existing solutions. Furthermore, it allows for an additional digital post-processing which can aid in analysing and improving the results. It also provides mobility for a potential adversary, requiring only a commodity laptop and an USB SDR dongle. The attacker does not need to have any prior knowledge about the victim's video display. All parameters such as resolution and refresh rate can be estimated with the aid of the software.

The software comprises of a library written in C, a collection of plug-ins for various Software Define Radio (SDR) frontends and a Java based Graphical User Interface (GUI). It is designed to be a multi-platform application. All native libraries can be pre-compiled and packed into a single Java jar file which allows the toolkit to run on any supported operating system.

This report documents the digital processing techniques that have been employed in order to extract, detect and lock to a video signal. It also explains the architecture of the software system and the techniques used in order to achieve low latency and real-time interactivity. It demonstrates the usage of the system by performing a practical attack. It then gives some ideas about what could be improved further and some analysis of data that was collected during the development of the software.

Contents

1	Introduction	1
1.1	History	2
1.2	Related Work	3
1.2.1	Work of Wim van Eck	3
1.2.2	Work of Markus Kuhn	4
1.3	Achieved Goals and Motivation	5
2	Background	9
2.1	Signal Processing	9
2.2	IQ Sampling	12
3	Methodology	15
3.1	Analogue Video Signals	15
3.2	Generated Radio Wave	17
3.2.1	Signal Equation	17
3.2.2	Sampling	18
3.2.3	Repetitions	18
3.2.4	Sampling Rate	20
3.3	Reception	20
3.3.1	Theory	20
3.3.2	Practice	21
3.4	Resolution and Framerate Detection	22
3.4.1	Introduction to Autocorrelation	24
3.4.2	Aliasing	25
3.4.3	Number of Lines in a Frame	27
3.4.4	Errors	28
4	Practical attack	31
4.1	The Setup	31
4.2	Preparation	32

5	Implementation	33
5.1	Hardware	33
5.2	Architecture	35
5.2.1	Main Library	35
5.2.2	JavaGUI	35
5.3	Digital Signal Processing	35
5.3.1	Overview	35
5.3.2	Signal Reconstruction	35
5.3.3	Frame Synchronization	35
5.3.4	Resolution Detection	35
5.3.5	Extended Bandwidth	35
6	Summary and Conclusions	37

Chapter 1

Introduction

Interference is usually regarded as a simple annoyance. It is not difficult to experience TV signal or radio station getting distorted because of a device operating in the vicinity of your receiver. Manufacturers have to meet certain requirements so that their products do not cause interference with other systems. The truth is that the emanation is linked to the way the emitting device works internally. The interference is generated by alternating currents due to switches, oscillators and other electronic and mechanical components operating inside the device. Therefore it carries some information about the internal state of the device.

This means that a simple annoyance could potentially turn into a security leak, broadcasting sensitive data into the wild. A clever attacker could possibly distinguish different modes of operation or even recover raw data that is being processed. And the industry standards that manufacturers follow to minimise interference say nothing about the information such interference may carry. A very sensitive detector can pick up such signals at levels several magnitudes lower than what industry standards specify as radiation limits.

In the case of video displays the issue is particularly important. Having a secure, encrypted computer system that unintentionally broadcasts its display in clear text does not sound secure at all. The attacker could be fully

passive and the breach could be never detected. Furthermore, the repeating nature of the signal combined with the long video wires that could act as antennas, mean that such signals can travel very long distances, allowing practical attacks from vans across the street [1].

However very little research has been published into the open literature on the topic. The main reason is possibly the expensive specialised equipment required to conduct experiments with compromising emanations. One of the goals of this project is to address this issue and demonstrate that a practical attack could be undertaken using an affordable software-defined radio receiver.

1.1 History

A compass needle points north when put next to a wire that has current flowing in it. Turning off the current makes the needle twitch i.e. temporary deflecting from the north direction. The same effect could be seen if the current is turned back on. This phenomenon was noticed by the Danish physicist Hans Christian Ørsted in April 1820 [2]. It shows that a changing electric current creates a magnetic field and vice versa. Later this phenomenon was utilised to create the electrical telegraph and allow for long distance communications.

However, the undesired effects of this phenomenon received almost no attention for a couple of decades. British army noticed crosstalk between telephone wires during the Nile and Suakin expedition in 1884-85 [3]. The first reported exploitation of the phenomenon was in 1914 when earth leakage from telephone wires caused a lot of crosstalk. Listening posts were established in order to intercept enemy messages. Next year valve amplifiers were utilised which allowed for extended listening range [4].

The US National Security Agency conducted a classified research with code-name TEMPEST in 1972 [5]. The document was later partially declassified in 2007. It describes unwanted emanations coming from a Bell-telephone

mixing device that was used for encryption. The signal allowed an attacker to reconstruct the original plain text. Researchers at Bell Lab demonstrated a practical attack from about 80 feet away that allowed reconstructing about 75% of the plain text being processed by the machine.

Computer monitors started to become common in the end of the 20th century. Wim van Eck published the first unclassified technical analysis of the security risks of emanations from computer monitors in 1985 [1]. The next important publication regarding video emanations came from Markus Kuhn nearly 20 years later in 2003 [6].

1.2 Related Work

With the notable exception of van Eck and Kuhn, there is almost no research on the topic of compromising emanations from video displays in the open literature. This is worrying since the phenomenon is real and could easily be exploited by an adversary.

1.2.1 Work of Wim van Eck

Electromagnetic Radiation from Video Display Units: An Eavesdropping Risk? [1] was the first publicised research on the topic of emanations from video monitors. Wim van Eck describes the similarities between contemporary monitors and black and white TV sets. He explains that these could be exploited to build a cheap eavesdropping device.

The document describes usage of a modified black and white commercial television receiver to conduct a very successful practical attack. Furthermore, a demonstration is done outside the laboratory in real world conditions. Van Eck proves that the attack is a viable security threat. He continues discussing possible sources of the emissions and ways of defending against such attacks.

However, there are some problems with his implementation:

- The modification of the TV receiver, although inexpensive, requires advanced specialised knowledge to undertake.
- The fact that the device needs constant manual adjustments to keep the oscillators in the receiver and transmitter in sync makes it difficult to use in practice.
- Technology has changed dramatically since the publishing of the paper. The attack will no longer work on modern monitors due to the variety of video modes available which now differ significantly from broadcast television.
- No room for further automated signal processing.

We can only speculate why there was no further research on this topic for decades to come. Possibly the main reason was that a modified TV receiver will no longer work with modern day video modes and commercial off-the-shelf narrowband receivers can't be used for that purpose either. As Kuhn outlines, a researcher needed to have their hands on a very special military grade wideband receivers that are expensive and have export restrictions[6]. However, the project presented in this dissertation aims to solve this problem by using an affordable software-defined radio platform for receiving the emanations from video displays.

1.2.2 Work of Markus Kuhn

Markus Kuhn was able to improve on the results achieved by van Eck with his *Compromising emanations: eavesdropping risks of computer displays* [6]. In the technical report, Kuhn analyses in a very mathematically detailed manner the properties of the waves emitted and their possible emitting circuitry. He describes the possible equipment an attacker may need to intercept and process the signal. He conducts experiments and constructs a system for real time monitoring using an FPGA board, a specialised wideband AM radio receiver and an off-the shelf VGA video monitor.

However there are a few caveats:

- A researcher will need access to very expensive, export restricted equipment in order to repeat his experiments. This equipment also often lacks the mobility required for a practical attack.
- As with van Ecks' system, Kuhn's solution still requires manual synchronisation.
- Digital signal processing for improving reception of weak signals (like time averaging) was not attempted.
- An attacker will need to know the exact video mode the target is using.

The report also mentions possible ways of modulating hidden messages into the signal. Kuhn also discusses ways of defending against such attacks using hardware and software solutions. He demonstrate automatic character recognitions. The report also includes experiments with optical eavesdropping of CRT displays.

Overall his real-time monitoring system improves on van Eck's one by supporting a variety of modern video modes. Thus Kuhn was able to prove that the threads outlined almost two centuries ago are still valid. In a later paper [7] he focuses on digital displays and explains that the same attack vector could also be used in that context.

The presented project is inspired by Dr Kuhn's work. His report simplifies further research into the topic by giving a very detailed and mathematical description of the emanated signal and its properties. This was an invaluable starting point proving that a software-defined radio implementation could be viable.

1.3 Achieved Goals and Motivation

The project aims to raise awareness of the potential security implications of the compromising emanations from video monitors. It implements the

principles outlined in the related work section 1.2 using a completely new approach: utilising a software-defined radio receiver. It also combines them into a single portable software library. In its basic form it allows tuning to a video signal by manually controlling the vertical and horizontal synchronisation similar to the implementation that van Eck and Kuhn have used in their practical demonstrations.

The system builds on top of the existing research by having achieved the following additional goals that were never done before:

- Uses affordable unmodified off-the-shelf equipment portable enough to simplify practical attacks.
- Almost no specialised experience is necessary to operate the system.
- No prior knowledge of the target machine video parameters is required. The system can automatically estimate them remotely in real time.
- Reception of weaker signals is possible due to additional digital signal processing.
- Eliminates the need for constant manual adjustments in order to keep the picture steady on the screen.
- Open source license aiding further research on the topic.

The source code and the latest pre-compiled binaries could be obtained from github.com/martinmarinov/TempestSDR. If user's computer is running OS X, Windows or Linux, a multiplatform statically precompiled Java jar file is available for starting up the system in a couple of clicks.¹

The system supports a plug-in architecture that allows simplified development of device drivers for any software-defined radio frontend. Currently supported plug-ins:

- Pre-recorded file with raw IQ samples (Multiplatform support)

¹Some device drivers might not be available on all operating systems. At time of writing the OS X binaries need to be compiled manually. Refer to the README for more information.

- Mirics SDR (Windows support)
- UHD (Linux and OS X support)
- ExtIO (Windows support)

The system comes with a Java GUI but the underlying library could be used independently by any other system as a shared or statically linked library. Furthermore the core is written in C and has no additional dependencies making it portable and easy to compile.

This report provides some background information on the physics behind the radio wave generation. It describes the basics of a SDR radio receiver. It then explains the properties of the emanated signal and how to reconstruct it. It goes on describing the automatic remote estimation of video parameters. A practical attack is demonstrated afterwards showing that an attacker with no prior knowledge of the target video display can receive its signal from a distance using the system. It then outlines the architectural details of the library and the implementation details. Finally, a conclusion is done with some suggestions on further research.

However, the report does not discuss ways of limiting the amount of information leaked via such compromising emanations. This is beyond the scope of the project. The system, however, could be used for further research on the topic. It could possibly aid such attempts by serving as a rapid prototyping tool.

The report also does not claim that the experimental results presented are exhaustive. This is due to the fact that there is a wide range of software-defined radio front-ends that are supported, each of them having different characteristics. Therefore no attempt was done to characterise the hardware used for the experiments. The presented measurements are not provided as absolute value (i.e. voltages) but rather as relative (i.e decibels). The results will very much depend on the target video display as well. Some displays will emit strong signals that could be picked up tens of meters away, while others emit no detectable video emanations. Characterising receivers and

target video displays could be also left as a further research topic.

There are also some inherent notes about the system:

- The performance of the system depends on the speed of the CPU of the attacker. No GPU acceleration was implemented since it wasn't really necessary. Performance never dropped below 35 fps on an Intel Core i5 laptop without GPU acceleration.
- Difficult to compare its core performance to existing implementations due to the fundamental differences of the underlying architectures.
- Software-defined radio receivers have a lower sensitivity and higher noise figure due to lack of high quality analogue filters compared to specialised wideband receivers. There is also interference caused by their internal IC circuits or the USB/Ethernet/Power cables that connect them to the PC. However, Dr Kuhn showed that in practice this could be greatly improved using suitable analogue preconditioning.

Chapter 2

Background

A reader should have knowledge of undergraduate level mathematics including calculus. Some of the equations and theorems are simply stated and then used. If the reader wants to learn more about their derivations, they need to look at a relevant textbook. Furthermore, it is assumed that the reader has a computer science background and understands the basic concepts behind object oriented programming and procedural programming. Familiarity with the C programming language and some basic knowledge of Java would be helpful as well. Understanding the basics of signal processing and Fourier theory is also desirable but not strictly required.

2.1 Signal Processing

The fact that changing currents in a wire induce currents in another nearby wire comes straight from Maxwell's equations [8]. A signal is just changing current and/or voltage in the time domain. Let's imagine we have a wire that we call wire A. We want to transmit a signal over it. We connect it to a signal generator, so that current across the wire at time t could be written as $f(t) = I \cdot \sin(2\pi f_c t)$ where I is called the **amplitude** and f_c is the **frequency** of the continuous tone. Because of Maxwell's equations, the

changing current in wire A will also generate a circular changing magnetic field around itself.

Let's imagine that an adversary puts a wire B lying parallel to wire A. The changing magnetic field from wire A will generate changing current in wire B. The adversary can measure the current in wire B as a function of time $\hat{f}(t)$. If we think about the ideal case and ignore resistance and capacitance, we would get perfect coupling, i.e. the adversary will measure $\hat{f}(t) = I \cdot \sin(2\pi f_c t)$ which is precisely $f(t)$, the signal in wire A. We have unintentionally broadcasted a continuous wave and an adversary has successfully received it remotely, without physically tapping into our wire. This is the source of the compromising emanations we observe - signals in wires unintentionally inducing currents in other nearby wires. Of course in the real world we don't only broadcast sine waves. The waves could be rather complex but the same principles apply since each signal could be decomposed into a series of sine tones using Fourier analysis. In the real world we also can't ignore resistance and capacitance so we need to use antennas and amplifiers to intercept signals. We will also have interference from other sources so recovering the original signal is a challenge but it could be done using signal processing techniques.

We can apply **Fourier transform**[9] to see how the signal spectrum will look in the frequency domain. This will decompose the function into a sum of complex *sin* and *cos* terms. The Fourier transform of $f(t)$ is defined as

$$F(\omega) = \mathcal{F}\{f(t)\} = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt$$

and the inverse Fourier transform

$$f(t) = \mathcal{F}^{-1}\{F(\omega)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{j\omega t} d\omega$$

A sine wave of the form $f(t) = \sin(2\pi\Omega t)$ will result in spikes $F(\omega) = \frac{1}{2i}[\delta(\omega - \Omega) - \delta(\omega + \Omega)]$ at $F(-\Omega)$ and $F(\Omega)$ since $F(\omega) = 0$ everywhere else. Therefore we can see that the Fourier transform basically gives an indication of the frequency of the source signal. Note that the **Dirac delta** function is

defined as

$$\delta(t) = \begin{cases} +\infty, & x = 0 \\ 0, & x \neq 0 \end{cases}$$

The Dirac delta functions have the so-called “sifting” property so that if multiplied with a function $f(t)$, it will **sample** out only a single value, namely

$$\int_{-\infty}^{\infty} f(t)\delta(t-T)dt = f(T)$$

We will also encounter the “Dirac comb” defined as:

$$\sum_{k=-\infty}^{\infty} \delta(t - kT)$$

Which is basically an infinite number of Dirac delta functions repeating at regular intervals of T . We can therefore “sample” a function $f(t)$ at times T by multiplying it with a Dirac comb. This will result in discrete samples being obtained from $f(t)$ at regular intervals of T i.e. we will obtain $\dots, f(-2T), f(-T), f(0), f(T), f(2T), \dots$

We also need to mention **convolution**, denoted by the operator $*$. The convolution of two functions $f(t)$ and $g(t)$ is defined as:

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau$$

Note that for simplicity, in the current report, I have chosen to drop the (t) so if you encounter $f * g$, it should read $(f * g)(t)$.

To illustrate how convolution works, consider convolving a function $f(t)$ with a Dirac delta

$$f(t) * \delta(t - T) = \int_{-\infty}^{\infty} f(\tau)\delta(t - \tau - T)d\tau = \int_{-\infty}^{\infty} f(\tau)\delta(\tau - (t - T)) = f(t - T)$$

Where we used the fact that the Dirac delta function has an even symmetry $\delta(t) = \delta(-t)$. This results in a new function $f(t - T)$ which is a time shifted

copy of $f(t)$ with a shift of T . Convolving with a Dirac comb could be shown to produce infinite copies of $f(t)$ at regular intervals T which are basically summed together.

The term **lowpass** means a lowpass filter. This is a filter that acts on a signal, and that allows components of the signal with frequencies $f \leq f_{max}$ to pass through, attenuating the rest to 0. **Highpass** filter allows frequencies $f \geq f_{min}$ to pass through intact, attenuating the rest to 0. A **bandpass** filter is a combination of a lowpass and a highpass filter, allowing only frequencies $f_{min} \leq f \leq f_{max}$, attenuating the rest to 0.

2.2 IQ Sampling

A simple analogue to digital converter will measure the voltage at a wire by producing s samples per second (called **sampling rate**). This means that it will produce a new sample every $T_\Delta = \frac{1}{s}$ seconds. It converts an analogue signal a.k.a. continuous function $f(t)$ into a series of samples. This could be written mathematically as a multiplication with a Dirac comb:

$$\sum_{k=-\infty}^{\infty} f(t)\delta(t - kT) \quad (2.1)$$

If we want to detect signal with frequency up to f_{max} , then according to the Nyquist sampling theorem, we need to sample with a sample rate $s \geq 2f_{max}$. However, this usually gets impractical when f_{max} gets large. One way of dealing with that (assuming the signal we want to receive is band limited) is using a local oscillator that runs at frequency f_{lo} . The oscillator generates a signal $A_{lo} \cdot \cos(f_{lo}t)$. The output from the local oscillator is multiplied with the incoming **RF** (radio frequency) baseband signal (a.k.a. mixing). Say the RF signal is $f(t) = A(t)\cos(f_ct)$. It has a **frequency** of f_c and **phase** of f_ct . For simplicity, let's put $A_{lo} = 2$. After the mixing stage, our resulting signal

will be

$$\begin{aligned} f(t) \cdot A_{lo} \cos(f_{lo}t) &= A(t) \cos(f_c t) \cdot 2 \cos(f_{lo}t) = \\ &= A(t) \cos((f_c + f_{lo})t) + A(t) \cos((f_c - f_{lo})t) \quad (2.2) \end{aligned}$$

This results in a linear combination of two signals with frequencies called heterodyne frequencies - the sum and the difference between the local oscillator and the incoming frequency. A bandpass filter could be set up at an **IF** (intermediate frequency) f_{if} . If any of these heterodyne frequencies falls within the bandpass of the filter, the signal can go for further processing.

However we encounter a problem - \cos is an even function. Imagine we now have another RF baseband signal, namely $\hat{f}(t) = A(t) \cos((2f_{lo} - f_c)t)$. After mixing with the local oscillator, the resulting signal would be $A(t) \cos((f_{lo} - f_c)t)$ which is equivalent to what we obtained for $f(t)$ in Equation 2.2 for the difference $f_{lo} - f_c$ signal. So we can't distinguish $\hat{f}(t)$ from $f(t)$ if we decide to only filter out the difference part. This is called signal imaging. It could be shown that a signal $\tilde{f}(t) = A(t) \cos((-f_c - 2f_{lo})t)$ will in turn generate a sum $A(t) \cos((-f_{lo} - f_c)t)$ which again results in imaging if we decide to keep the summation term $f_{lo} + f_c$ from 2.2.

This method can preserve the amplitude of the wave and is widely used in AM radio reception (in conjunction with some image rejection techniques as well). However, it is obvious from the problem outlined above that the method is not sufficient for uniquely detecting the phase of the incoming wave. A solution to this problem is to multiply the incoming RF signal with a phase shifted version of the local oscillator at an angle of 90° in parallel with the mixing from Equation 2.2. This means multiplying by $-A_{lo} \cdot \sin(f_{lo}t)$. After this secondary mixing stage, we would get

$$\begin{aligned} -f(t) \cdot A_{lo} \sin(f_{lo}t) &= -A(t) \cos(f_c t) \cdot 2 \sin(f_{lo}t) = \\ &= -A(t) \sin((f_c + f_{lo})t) + A(t) \sin((f_c - f_{lo})t) \quad (2.3) \end{aligned}$$

Now let's put $\varphi(t) = (f_c - f_{lo})t$, and we only take the terms that contain φ

in 2.2 and 2.3, namely the difference terms using a bandpass filter of width $\frac{s}{2}$. Then we can define

$$I(t) = A(t)\cos(\varphi(t)) \quad \text{and} \quad Q(t) = A(t)\sin(\varphi(t)) \quad (2.4)$$

Sampling these continuous signals with an analogue to digital converter will result in discrete values at times T_Δ . Let's put $A_n = A(nT_\Delta)$ and $\varphi_n = \varphi(nT_\Delta)$. Then we can define our resulting discrete samples as:

$$I_n = A_n\cos(\varphi_n) \quad \text{and} \quad Q_n = A_n\sin(\varphi_n) \quad (2.5)$$

These I and Q samples form the basic principles of software-defined radio. I and Q stand for in-phase and quadrature phase[10]. These terms refer to the \cos and $-\sin$ oscillator phases used for mixing. After the IQ samples are obtained, they are received by a PC with a sample rate of s . Further analysis of the phase, frequency and amplitude could be done in software allowing for any signals that fit within the bandwidth to be demodulated.

The following identities could be used to obtain them for a sample n :

$$A_n = \sqrt{I_n^2 + Q_n^2} = \sqrt{A_n^2 \cdot (\cos^2(\varphi_n) + \sin^2(\varphi_n))} \quad (2.6a)$$

$$\varphi_n = \tan^{-1} \left(\frac{I_n}{Q_n} \right) = \tan^{-1} \left(\frac{\cos(\varphi_n)}{\sin(\varphi_n)} \right) \quad (2.6b)$$

$$\omega_n = \frac{d\varphi_n}{dt} = \varphi_n - \varphi_{n-1} \quad (2.6c)$$

Where ω_n is the frequency.

Chapter 3

Methodology

3.1 Analogue Video Signals

Almost all contemporary video monitors are raster based. The image is transferred from the video controller to the display in scanlines that occur at a specific rate. Each of these scanlines contains a number of pixels which are continuously encoded as a time varying signal. This signal is generated internally in the video controller by an oscillator which runs at the pixel clock rate. The signal is thereafter multiplied by the pixel intensities at the specific time.

Let's assume that the signal started transmitting at time $t = 0$, and each video frame contains h scanlines, each of which contains w pixels. The frequency with which frames are being generated is f_f frames per second. The duration of transmission of each individual pixel is

$$T_\Delta = \frac{1}{whf_f} \quad (3.1)$$

Note that at time t , frame number n_t started transmission where

$$n_t = \lfloor t f_f \rfloor \quad (3.2)$$

We can assume that the top left corner of a frame has coordinates $(0, 0)$, a pixel at position (x, y) in frame n_t will start to be transmitted at time

$$T_{(x,y)} = (yw + x + n_t wh)T_\Delta \quad (3.3)$$

and will finish transmission before time

$$T_{(x+1,y)} = T_{(x,y)} + T_\Delta = (yw + x + 1 + n_t wh)T_\Delta \quad (3.4)$$

at which the next pixel will start transmitting.

In practice w and h are determined by the screen resolution and f_f is simply the screen refresh rate. For a typical screen resolution of $width \times height$, it is true that $w \geq width$ and $h \geq height$. The reason is that video signals tend to have additional blanking intervals. This means more pixels are transmitted than what is in the active video region. This gives opportunity for the receiving monitor to synchronise its internal clock, calibrate its colour levels or in case of CRT, allow enough time for the electron beam to return to the beginning of the next line on the screen. The synchronisation timings for personal computers have been standardised by Video Electronics Standards Association (VESA) [11].

In order to decode an individual pixel, the receiving monitor has its own internal oscillator. It locks it to the pixel rate of the incoming signal either via an external clock source or using the blanking intervals. Once it receives the signal for an individual pixel, its amplitude (or binary content in case of digital signal) will correspond to the intensity. This allows the monitor to display the video in real time. If multiple colours are desired, they can be transmitted separately on different wires in the same fashion.

3.2 Generated Radio Wave

3.2.1 Signal Equation

Let's assume the discrete pixels in a video signal have intensities $v_i (i \in \mathbb{Z})$ and are being transmitted for a duration of $T_\Delta = \frac{1}{whf_f}$ (from 3.1). Let's also assume that the shape of the pixel is $p(t)$ where $p(t) = 0$ for $|t| > \frac{T_\Delta}{2}$. We know that pixel i starts transmitting at time $t_i = iT_\Delta$ (if we assume that that pixel 0 was transmitted at $t = 0$). Also the amplitude of the signal of the transmitted pixel is linearly dependant on its intensity v_i . Then the resulting signal between time t_i and t_{i+1} will have the form $v_i p(t - iT_\Delta)$. To generalise for all i , the resulting actual transmitted radio wave in the time domain will have the form

$$\tilde{v}(t) = \sum_{i=-\infty}^{\infty} v_i p(t - iT_\Delta) \quad (3.5)$$

Which is a continuous function. Equation 3.5 can be rewritten as a convolution with a Dirac delta function

$$\tilde{v}(t) = p(t) * \left(\sum_{i=-\infty}^{\infty} v_i \cdot \delta(t - iT_\Delta) \right) = p(t) * \hat{v}(t) \quad (3.6)$$

Where we have put

$$\hat{v}(t) = \sum_{i=-\infty}^{\infty} v_i \cdot \delta(t - iT_\Delta) \quad (3.7)$$

Note that this results in infinitesimally short (in time) spikes at values exact integer multiples of T_Δ with amplitudes weighted to the pixel intensities at that time. We can view 3.6 as the pixel shape being repeated at intervals of T_Δ modulated by v_i .

3.2.2 Sampling

We can notice that 3.7 looks like the result of a continuous signal being sampled at discrete points in time T_Δ apart. If we call this continuous signal $v(t)$, then we can rewrite the equation so that it reads

$$\hat{v}(t) = \sum_{i=-\infty}^{\infty} v(t) \cdot \delta(t - iT_\Delta) = v(t) \sum_{i=-\infty}^{\infty} \delta(t - iT_\Delta) \quad (3.8)$$

The requirement is that if $v(t)$ is sampled at discrete intervals iT_Δ , it should be equal to the corresponding pixel values i.e. $v(0) = v_0$, $v(T_\Delta) = v_1$, $v(2T_\Delta) = v_2$, etc. This would mean that we can precisely obtain the samples from the continuous function by applying the series of Dirac delta functions. However, now we would also like to be able to obtain the continuous signal by only having the sampled values v_i . This means that we need a unique perfect reconstruction i.e. to be able to obtain all v_i from $v(x)$ and $v(x)$ from all v_i . This condition is satisfied by the Whittaker–Shannon interpolation formula[12] which can give us the condition for perfect reconstruction for digitally sampled signals. It states that the continuous signal could be uniquely reconstructed using *sinc* interpolation. If we apply it to our v_i samples, it yields

$$v(t) = \sum_{k=-\infty}^{\infty} v_k \cdot \text{sinc}\left(\pi\left(\frac{t}{T_\Delta} - k\right)\right) \quad (3.9)$$

We can verify that when $t = iT_\Delta$, then $\text{sinc}(\pi(i - k))$ will always yield 0 except for when $i = k$ in which case it will be 1. Therefore we have proved that $v(iT_\Delta) = v_i$.

3.2.3 Repetitions

Now let's revise some mathematical identities. According to the convolution theorem [13]

$$\mathcal{F}\{g \cdot h\} = \mathcal{F}\{g\} * \mathcal{F}\{h\} \quad (3.10)$$

Which reads: the Fourier transform of a multiplication of two functions is equivalent to the Fourier transforms of the individual functions convolved together. And vice versa

$$\mathcal{F}\{g * h\} = \mathcal{F}\{g\} \cdot \mathcal{F}\{h\} \quad (3.11)$$

The convolution of the Fourier transform of two functions is equivalent to the individual Fourier transforms of the two functions multiplied together. It could be also shown that:

$$\mathcal{F}\left\{\sum_{n=-\infty}^{\infty} \delta(t - nk)\right\} = \frac{1}{k} \sum_{n=-\infty}^{\infty} \delta\left(t - \frac{n}{k}\right) \quad (3.12)$$

Using the last three identities, we can write the Fourier transform of 3.6 as

$$\mathcal{F}\{p(t) * \hat{v}(t)\} = \tilde{V}(f) = \left[\frac{1}{T_\Delta} \cdot P(f) \cdot V(f) \right] * \left[\sum_{i=-\infty}^{\infty} \delta\left(t - \frac{i}{T_\Delta}\right) \right] \quad (3.13)$$

Where $P(f)$ is the Fourier transform of $p(t)$, $\tilde{V}(f)$ is the Fourier transform of $\tilde{v}(t)$ and $V(f)$ is the Fourier transform of $v(t)$. If we put

$$G(f) = \frac{1}{T_\Delta} \cdot P(f) \cdot V(f) \quad (3.14)$$

Then we can arrive to our final conclusion that

$$\tilde{V}(f) = G(f) * \left[\sum_{i=-\infty}^{\infty} \delta\left(t - \frac{i}{T_\Delta}\right) \right] \quad (3.15)$$

This simply means that the signal spectrum $G(f)$ repeats at regular intervals throughout the radio spectrum with a frequency of $\frac{1}{T_\Delta} = whf_f$.

For example, for a resolution of $800 \cdot 600 @ 60fps$ (ignore blanking intervals for simplicity), we would have a frequency of $\frac{1}{T_\Delta} = 800 \cdot 600 \cdot 60 = 28.8MHz$. This means that we would expect to find such a signal centred at DC, 28.8 MHz, 57.6 MHz, 86.4 MHz, etc.

3.2.4 Sampling Rate

In order to obtain the full video signal, we need to know what is the minimum rate at which we need to sample the baseband. We saw from 3.15 that $G(f)$ repeats at the regular intervals. In fact we can see from 3.14 that it is $G(f)$ that contains the actual video information

$$G(f) = \frac{1}{T_\Delta} \cdot P(f) \cdot V(f)$$

As a consequence from the Whittaker–Shannon interpolation formula that we used to construct $v(t)$, we know that $V(f)$ is band limited so that $V(f) = 0$ for all $|f| \geq \frac{1}{2T_\Delta}$. This implies that $G(f)$ is also band limited to the same range (because of the multiplication). Therefore we need to have a receiver with a sampling rate of at least $\frac{1}{T_\Delta}$ in order to reconstruct $G(f)$. Note however that $g(t)$ contains the video signal convolved with the shape of the pixel $p(t)$. We can however assume $P(f) = 1$ for $|f| \leq \frac{1}{2T_\Delta}$ and zero otherwise. Therefore $G(f) = V(f)$ which would imply a sinc shaped $p(t)$. This in fact is not the case since different video cards have different pixel shapes[6]. However in practice, this is enough to obtain a good approximation of the signal.

3.3 Reception

3.3.1 Theory

In order to receive a copy of the video signal, we will need to apply a bandpass filter centred at a multiple of the pixel frequency. The width of the bandpass needs to be the same as the band limited video signal we want to eavesdrop. Then we need to generate an internal clock that runs at the pixel rate and use that to obtain the estimated pixel intensities. Luckily these steps are already done in a typical AM (Amplitude Modulation) receiver. Unfortunately the signal bandwidth is much wider than what a typical off-the-shelf AM receiver can handle.

However some Software Defined Radio receivers can provide the required bandwidth. The Ettus Research USRP B200 can provide more than 50 MHz of realtime bandwidth which covers what is required to eavesdrop most video signals. Software Defined Radios provide the radio samples as a quadrature vector which angle relates to the instantaneous phase in relation to the internal hardware oscillator of the hardware. The size of the vector is proportional to the received amplitude for each sample (refer to Section 2.2 for more information). For AM reception, we do not need the phase information since the video signal is being transmitted with a constant frequency. Therefore we can make a simple AM demodulator by taking the length of the vector of each sample. When our digital sampling rate matches the pixel rate, then each sample will contain an estimate of the pixel intensity at that time.

3.3.2 Practice

The VESA standard specifies an allowed frequency tolerance for f_p of 0.5%. Therefore due to hardware limitations, we might not be able to accurately tune the receiver sampling rate with enough accuracy to match the transmitted pixel rate. This means that some resampling needs to be done in software in order to create a new digital signal that matches as close as possible the transmitted pixel rate in which each sample corresponds to a pixel.

It is also true that we might be able to recover a lot of information from a signal even if our receiver is not capable of sampling the full width of the video spectrum. This means we will only receive a of $g(f)$ with a bandpass filter applied to it. In this case the digital resampling could interpolate the received samples to fit into multiple pixels. Therefore it is useful to introduce a measurement κ of how accurate we can reconstruct pixels of $g(t)$. This measurement can be the number of input samples that have gone into constructing a single output pixels. Therefore

$$\kappa = s \cdot T_\Delta = \frac{s}{whf_f} \quad (3.16)$$

Where as before, s is the receiver sampling rate. If $\kappa \geq 1$, then we can deconstruct individual pixels. If $\kappa < 1$ we are looking at a band-passed version of the signal and consecutive reconstructed pixels will have some interdependencies. Please refer to Figure 3.1 for a visual comparison of different values of κ .

Another possible source of distortions is the usage of multiple wires to transmit colour information. These signals interfere with each other and the resulting pixel intensity becomes a complicated mix of the signals generated in these wires. Therefore colour information would be very difficult to obtain. Nevertheless the outlined strategy allows an eavesdropper to detect changes in colour since this would result in different signal amplitudes arriving at the receiver for each pixel.

Dr Markus Kuhn has shown that the outlined procedure even works for decoding digital video signals[7]. If the pixel intensities are transmitted as a bit pattern rather than an analogue amplitude, the received intensity will be the average intensity in the bit pattern. The redundancy of bit patterns means that even in such case we can still receive a signal that will allow us to read text on the remote screen.

3.4 Resolution and Framerate Detection

The outlined method for decoding the video stream relies on us having the exact values for w , h and f_f . This is not very practical for a real world attack. We would require the adversary not to have any knowledge of the target system whatsoever. Therefore we would need to infer all of these parameters remotely by exploiting some typical characteristics of a video signal. Namely the fact that the signal is periodic.

As we have already seen, a video signal consists of video frames with width w and height h . The speed at which such frames are transmitted is f_f . However most of the time any two consecutive frames would be identical. This would be the case if the victim is looking at static text on the screen. Therefore we

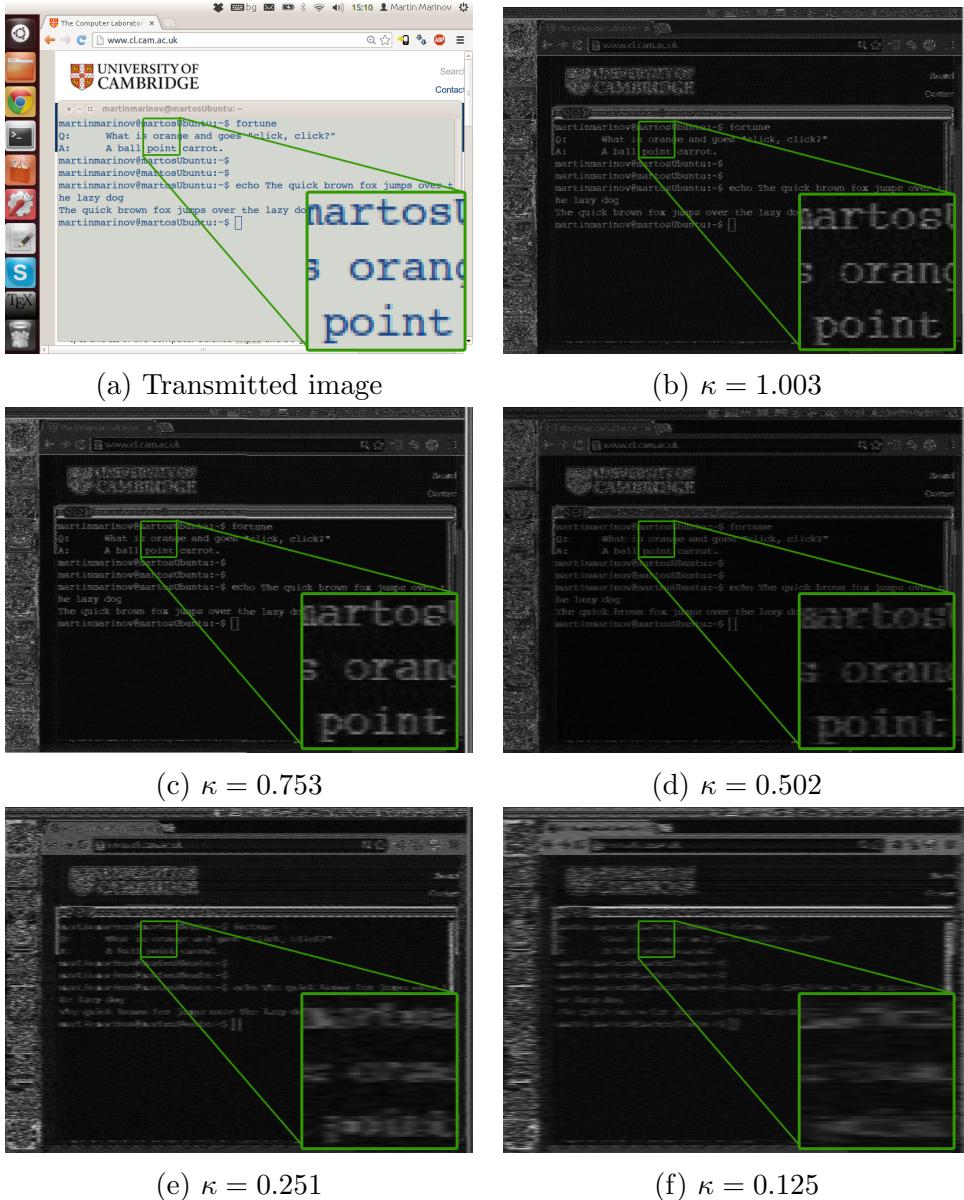


Figure 3.1: The transmitted image (a) is being sent at a pixel rate of 800x600 at 60 fps with $w = 1056$, $h = 628$ and $f_f = 60.11$. Each of the screenshots shows the reconstructed image with the receiver (USRP B2000) running at different sample rates: (b) - 40 MHz, (c) - 30 MHz, (d) - 20 MHz, (e) - 10 MHz and (f) - 5 MHz

can regard the transmitted signal as a repeating periodic signal of a single frame. Which means if we analyse the signal for repeating patterns we should be able to spot the repeating video frames in it and use that to estimate f_f .

3.4.1 Introduction to Autocorrelation

A good way of doing so would be calculating the discrete autocorrelation[14] of the incoming signal which is defined as

$$R_{vv}(j) = \sum_i v_i \bar{v}_{i-j}$$

Where \bar{v} is the complex conjugate of v and j is the **samples lag**. In our case v is the set of real samples, therefore $\bar{v} = v$. Basically the autocorrelation is a measure of the similarity of a signal with itself shifted by a lag. The higher the value of $R_{vv}(j)$ is, the more similar the function is at that lag. Therefore by analysing $R_{vv}(j)$, we would be able to spot any repeating patterns inside v_i .

Figure 3.2 shows the discrete autocorrelation of a video signal that was captured using the Mirics USB dongle at 8 MHz (i.e. 8,000,000 samples per second) sampling rate. The vertical axis is logarithmic, containing the decibels of the autocorrelation i.e. $10 \times \log_{10}(R_{vv})$. The horizontal axis contains the time lag in milliseconds. This was converted from the lag in samples j by $x = \frac{j}{s}$.

This particular example consists of 524,288 samples which is 65.536 milliseconds of recording. This allows us to estimate the autocorrelation up to half of it, namely 32.768 milliseconds. The reason is because $R_{vv}(j) = -R_{vv}(-j)$ due to symmetry. We can see a peak at 16.672 milliseconds which corresponds to a frequency of $f_f = \frac{1}{0.016672} = 59.98\text{Hz}$. This is our estimate for f_f . And it makes sense, 60Hz is a common refresh rate for contemporary video displays.

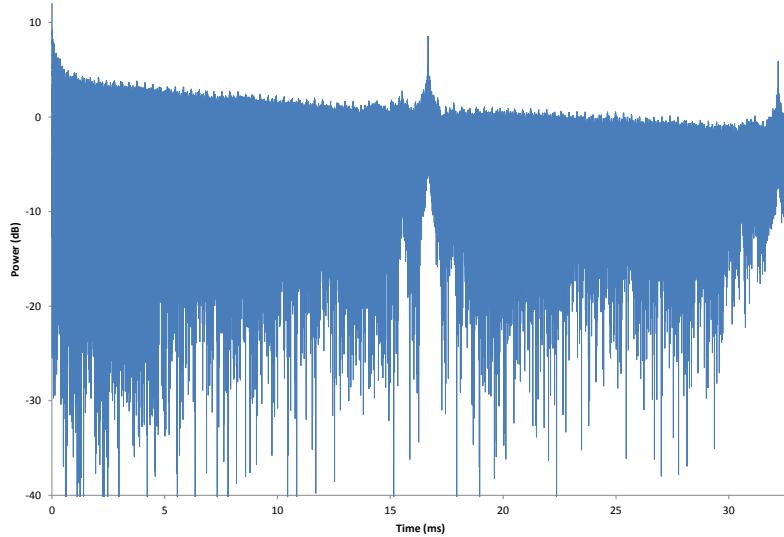


Figure 3.2: Autocorrelation of a signal

3.4.2 Aliasing

We should be able to see the peak repeating again at $2 \times 16.672 = 33.344\text{ms}$. This is analogous to comparing every two consecutive frames with every two other frames. However, as this falls just outside our window it would be seen as an alias at $65.536 - 33.344 = 32.192\text{ms}$. The same applies for the third peak at $3 \times 16.672 = 50.016\text{ms}$ which would be comparing each block of three consecutive frames with the next three. This will alias at $65.536 - 50.016 = 15.52\text{ms}$. This explains the second small peak we see on the left of the main one at 16.672ms but we see that the difference in its power is already several decibels lower. Therefore any further aliasing induced would be even smaller.

It is therefore important to choose the number of samples to be used for the autocorrelation wisely in order to avoid aliasing. We need to have enough full frames in our data so that the autocorrelation can physically work. The

minimum frequency we can get from an autocorrelation is $\frac{2s}{N}$ where N is the total number of samples that we are doing the autocorrelation on. It could be easily shown that the second alias is

$$a_2 = \frac{1}{\frac{2}{f_{min}} - \frac{2}{f}} = \frac{1}{\frac{N}{s} - \frac{2}{f}} \quad (3.17)$$

Where f is the frequency we are going to see the alias for. Therefore if we want to be able to pick up a repeating pattern of at least $f_{min} = 50\text{Hz}$, we would need $\frac{2s}{50}$ which in our case is 320,000 samples. However having such a small number of samples if we observe a signal at 85 Hz we would discover a strong alias at $\frac{1}{\frac{320,000}{8,000,000} - \frac{2}{85}} = 60.714\text{ Hz}$ which would be very misleading and will look like a legitimate signal at 60.714 Hz.

Let us assume $f_{lo} \leq f_f \leq f_{hi}$. In order to minimise aliasing, we can try to keep the second alias a_2 outside this range. Therefore we need to pick the number of samples so that

$$\frac{1}{\frac{N}{s} - \frac{2}{f_{lo}}} \leq f_{lo} \quad \text{or} \quad \frac{1}{\frac{N}{s} - \frac{2}{f_{hi}}} \geq f_{hi}$$

Which means that

$$3\frac{s}{f_{lo}} \leq N$$

or

$$3\frac{s}{f_{hi}} \geq N$$

and don't forget that in the same time we have the condition

$$2\frac{s}{f_{lo}} \leq N$$

so in the end, in order to eliminate the second alias in the desired region from f_{lo} to f_{hi} in the autocorrelation, we would need to choose the number of samples to be

$$N \geq 3\frac{s}{f_{lo}} \quad (3.18)$$

To paraphrase this result, we will need to do the autocorrelation on at least

3 frames worth of samples for the lowest frequency in our range f_{lo} to avoid second order aliasing interfering with our plot.

3.4.3 Number of Lines in a Frame

We were able to estimate f_f using an autocorrelation and eliminate aliased signals in the frequency window that we are looking to detect repetitions. However, we still have not estimated neither w , nor h . As a matter of fact we might not be able to measure w at all. The reason is that w might be a continuous signal and could potentially contain any number of pixels. However, each line of w pixels repeat h times in a frame which includes the blanking intervals. Therefore we can use their repeating nature of the blanking intervals to estimate h from the autocorrelation plot.

For this reason we need to zoom in to our plot. Refer to Figure 3.3 which is a zoomed out version of the figure we saw in the previous subsection. If we still have our allowed f_f window between f_{lo} and f_{hi} , then we can assume that h also spans between h_{lo} and h_{hi} . We would therefore expect to see a peak between $\frac{1}{f_{hi} \cdot h_{hi}}$ and $\frac{1}{f_{lo} \cdot h_{lo}}$ milliseconds. If we detect a peak at time t_{peak} , then this would correspond to $h = \left\| \frac{1}{t_{peak} \cdot f_f} \right\|$ where $\|$ represents rounding to the nearest integer. Note that we should have already estimated f_f .

Our plot shows a peak at 0.017875 milliseconds. Since we already estimated $f_f = 59.98 Hz$, then $h = \left\| \frac{1000}{0.017875 \times 59.98} \right\| = 933$ lines. If we look at a list of VESA video modes, we can see that there is a conveniently close video mode that corresponds to $h = 933$ and $f_f = 59.98$. It is the “1440x900 @ 60 Hz” video mode with $w = 1904$ pixels, $h = 932$ lines and $f_f = 60 Hz$. Therefore we could use the value of $w = 1904$ for our estimate of the number of pixels in a row. This will keep the aspect ratio correct.

We should not worry about aliasing since the number of samples in the autocorrelation is much bigger than our range. We can see on the plot the autocorrelation values repeating for each two lines, each three lines, etc. This results in a number of peaks repeating with the rate at which individual lines

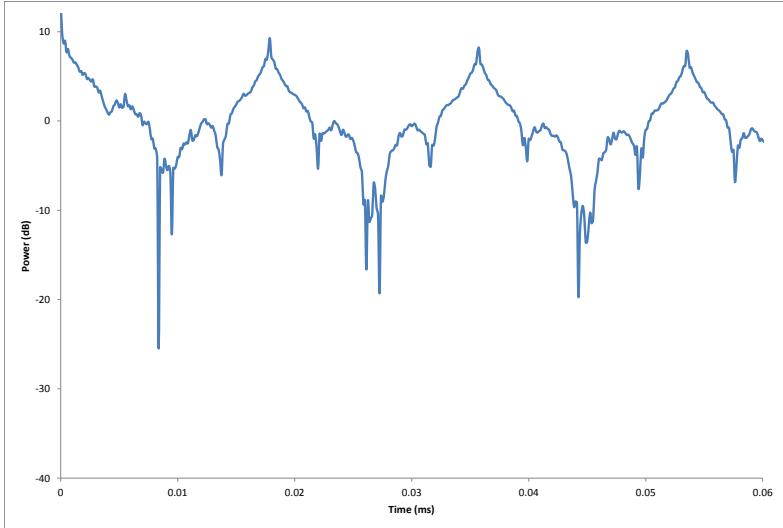


Figure 3.3: Zoomed in version of 3.2

repeat in the video signal.

3.4.4 Errors

There is a limit on how accurately one can determine f_f and h using the outlined autocorrelation method. This depends on the sampling rate of the device. The resolution of the autocorrelation is $\delta f = \frac{1}{s}$. Therefore we can only say that if there is a peak at t_x in the autocorrelation, then we have a repeating component in the signal with a frequency $f_x \pm \delta f$.

The accuracy at which we can determine h depends both on the accuracy of f_f and t_{peak} and can vary. Furthermore h should be rounded to an integer (we can only display an integer number of lines on a PC screen) which introduces further errors. Our definition for h is

$$h = \left\| \frac{1}{t_{peak} \cdot f_f} \right\| \pm (\delta h + 0.5)$$

Where 0.5 comes from the rounding. Also the uncertainties in t_{peak} and f_f are δf since both are being read from the same autocorrelation plot in milliseconds. In order to do error propagation, we need to take the partial derivatives with respect to f_f and t_{peak} .

$$\frac{\partial h}{\partial f_f} = -\frac{1}{t_{peak} \cdot f_f^2} \quad \text{and} \quad \frac{\partial h}{\partial t_{peak}} = -\frac{1}{f_f \cdot t_{peak}^2}$$

So for the uncertainty in h , we have

$$\delta h = \sqrt{\left(\frac{\partial h}{\partial f_f} \delta f_f\right)^2 + \left(\frac{\partial h}{\partial t_{peak}} \delta t_{peak}\right)^2} = \sqrt{\frac{1}{f_f^2} + \frac{1}{t_{peak}^2}} \times \frac{\delta f}{f_f \cdot t_{peak}}$$

We can also observe that the larger the sample rate gets, the smaller δf and δt get.

Chapter 4

Practical attack

Before explaining how the software works internally, let's present a demonstration of a practical video eavesdropping attack. Its main aim is to show the ease with which such an attack could happen. In the meantime this will give an opportunity to explain the characteristics of the received signal and how they are exploited.

As in the real world, we will start with no knowledge of the victim's system. We will estimate the frequency at which the emission strength has the best Signal to Noise Ratio (SNR). We will then analyse the signal to detect the resolution and refresh rate of the screen. We will afterwards lock onto the signal and try to recover the original video. We will also discuss some techniques that could be utilized to improve the quality of the image.

4.1 The Setup

The choice for a SDR front-end for this demonstration is a USRP B200¹. Depending on the particular requirements, an attacker might prefer mobility over accuracy and choose the smaller Mirics FlexiTV™MSi3101 SDR USB

¹Refer to 5.1 Hardware for discussion on currently available SDR devices.

Dongle. However, for this demonstration we will attempt to obtain the highest possible resolution. Therefore we need an SDR radio that is capable of obtaining a wide bandwidth. This makes the 32 MHz of Bandwidth that the USRP provides a much better choice than the 8 MHz available from the Mirics dongle.

[Antenna]

[Location of victim]

[Victim's fonts]

4.2 Preparation

Chapter 5

Implementation

5.1 Hardware



Figure 5.1: From left to right: MSi3101, AverMedia antenna and USRP B200

5.2 Architecture

5.2.1 Main Library

5.2.2 JavaGUI

Multiplatform

Graphical Interface

Visualisation

5.3 Digital Signal Processing

5.3.1 Overview

5.3.2 Signal Reconstruction

Multithreading

Demodulation

Re-sampling

Auto Gain

Low-pass

5.3.3 Frame Synchronization

5.3.4 Resolution Detection

5.3.5 Extended Bandwidth

Chapter 6

Summary and Conclusions

[Summarize results]

[potential future work]

Bibliography

- [1] Wim Van Eck. Electromagnetic radiation from video display units: an eavesdropping risk? *Computers & Security*, 4(4):269–286, 1985.
- [2] Karen Jelved, Andrew D Jackson, and Ole Knudsen. Selected scientific works of hans christian ørsted, 1998.
- [3] RFH Nalder. History of the royal corps of signals. *Royal Signals Institution, London*, 1958.
- [4] Ross Anderson. *Security engineering*. John Wiley & Sons, 2008.
- [5] US National Security Agency. Tempest: A signal problem. 1972.
- [6] Markus G Kuhn. Compromising emanations: eavesdropping risks of computer displays. *University of Cambridge Computer Laboratory, Technical Report, UCAM-CL-TR-577*, 2003.
- [7] Markus G Kuhn. Electromagnetic eavesdropping risks of flat-panel displays. In *Privacy Enhancing Technologies*, pages 88–107. Springer, 2005.
- [8] James Clerk Maxwell. *A treatise on electricity and magnetism*, volume 1. Clarendon press, 1881.
- [9] William L Briggs et al. *The DFT: An Owners' Manual for the Discrete Fourier Transform*. Siam, 1995.
- [10] Johan Kirkhorn. Introduction to iq-demodulation of rf-data. *EchoMAT User Manual (FA292640)*, 15:4–10, 1999.
- [11] Monitor Timing Specifications. Version 1.0. *Revision 0.8, Video Electronics Standards Association (VESA), San Jose, California*, 1998.
- [12] Claude Elwood Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

- [13] G Arfken. Convolution theorem. *Mathematical Methods for Physicists*, pages 810–814.
- [14] Ronald Bracewell. The autocorrelation function. *The Fourier transform and its applications*, pages 40–45, 1965.