Light Ears: Information Leakage via Smart Lights

ANINDYA MAITI, University of Texas at San Antonio, USA
MURTUZA JADLIWALA, University of Texas at San Antonio, USA

Modern Internet-enabled smart lights promise energy efficiency and many additional capabilities over traditional lamps. However, these connected lights also create a new attack surface, which can be maliciously used to violate users' privacy and security. In this paper, we design and evaluate novel attacks that take advantage of light emitted by modern smart bulbs, in order to infer users' private data and preferences. The first two attacks are designed to infer users' audio and video playback by a systematic observation and analysis of the multimedia-visualization functionality of smart light bulbs. The third attack utilizes the infrared capabilities of such smart light bulbs to create a covert-channel, which can be used as a gateway to exfiltrate user's private data out of their secured home or office network. A comprehensive evaluation of these attacks in various real-life settings confirms their feasibility and affirms the need for new privacy protection mechanisms.

CCS Concepts: • **Security and privacy** → *Mobile and wireless security*; • **Human-centered computing** → *Ubiquitous and mobile devices*; • **Computer systems organization** → *Embedded and cyber-physical systems*;

Additional Key Words and Phrases: Smart Light, Smart Bulb, Privacy, Multimedia, Security, Data Exfiltration

ACM Reference Format:
Anindya Maiti and Murtuza Jadliwala. 2019. Light Ears: Information Leakage via Smart Lights. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 98 (September 2019), 27 pages. https://doi.org/10.1145/3351256

1 INTRODUCTION

Smart lighting products, marketed as energy-efficient replacements of traditional incandescent and fluorescent lamps with several novel functionalities enabled by their Internet connectivity, have soared in popularity in recent years [13, 16]. A common feature among most smart lights is the ability to remotely control the lights, over a Wi-Fi, Bluetooth, or Zigbee network. Many of the current generation smart lights are also LED-based, which enables fine-grained customization of color and intensity of the light being emitted from these bulbs. Few advanced smart lights are also equipped with infrared capabilities, intended to aid surveillance cameras in low visibility environments.

Lighting products have traditionally not been an attractive target of security/privacy-related threats because conventional lamps typically do not have access to sensitive user information. However, as modern smart lights are usually connected to users' home or office network (either directly or via a communication hub) and can be controlled using users' mobile devices, they are poised to become a much more attractive target for security/privacy attacks than before. In this paper, we investigate how modern smart lights can be exploited to infer users' private information.

Authors' addresses: Anindya Maiti, Institute for Cyber Security, University of Texas at San Antonio, 1 UTSA Circle, San Antonio, Texas, 78249, USA, a.maiti@ieee.org; Murtuza Jadliwala, Department of Computer Science, University of Texas at San Antonio, 1 UTSA Circle, San Antonio, Texas, 78249, USA, murtuza.jadliwala@utsa.edu.

2474-9567/2019/9-ART98 $15.00
https://doi.org/10.1145/3351256

In this direction, we first focus on exploiting (details in Sections 5 and 7) a new feature of modern smart lights, known as *multimedia-visualization*. Multimedia-visualization is intended for use in conjunction with a song or video playing on a nearby media player, which results in a vibrant lighting effect that is synchronized with the tones present in the audio or the dominant colors in the video stream, respectively. While such immersive audio-visual or ambient lighting effects can be entertaining and relaxing, we speculate that it can also lead to loss of privacy if not properly safeguarded. Consider a scenario where a curious adversary can observe the changing light intensities/colors of a multimedia-visualizing smart light installed inside a user's residence (likely through a window). *Can the adversary determine what song/video the user is playing, only by analyzing the changing light intensities/colors of the smart light?* If yes, then these attacks can have significant privacy implications for smart light users. For instance, the US Video Privacy Protection Act (1988) was enacted to prevent abuse of users' media consumption information, which can potentially reveal fine-grained personal interests and preferences. Our first goal in this paper is to comprehensively evaluate the feasibility of such attacks that passively employ outputs from smart lights to infer users' personal media (both, audio and video) consumption.

Next, we comprehensively study and evaluate the feasibility (details in Section 9) of exploiting a smart light's infrared lighting functionality to invisibly exfiltrate a user's private data out of his/her secured personal device or network. We show that such an attack can be accomplished by carefully manipulating and controlling (possible on modern smart lights) the infrared light to create a "covert-channel" between the smart light and an adversary with infrared sensing capability. With the help of a malicious agent on the user's smartphone or computer, the adversary can encode private information residing on these devices and then later transmit it over the infrared covert-channel residing on the smart light. Moreover, as several popular brands of smart lights do not require any form of authorization for controlling lights (infrared or otherwise) on the local network, any application installed on the target user's smartphone or computer can safely act as the malicious data exfiltration agent. The overarching goal of this paper is to highlight the vulnerable state of personal information of smart light users, outline system design parameters that lead to these vulnerabilities and discuss potential protection strategies against such threats.

2 RELATED WORK

While the threats posed by smart lights demonstrated in this paper are novel, there exists several prior works on optical (visible light) side-channels and covert-channels. In this section, we discuss such prior works and contrast them with our contributions. We also discuss other modalities that can also be exploited to execute similar attacks as presented in this paper.

Private Information Leakage Through Optical Side-channels. Information leakage threats that employ some form of an optical side-channel have been exploited by many researchers to infer various types of private information. For instance, Xu et al. [88] were able to infer the video being watched on a TV using the changing light characteristics observable through the target user's window. Similarly, Schwittmann et al. [78] were able to infer a video being played on a TV, by exploiting a smartphone's ambient light sensor. An extended work by Schwittmann et al. [77] employed the ambient light sensor on a smartwatch to accomplish the same attack. Alternatively, Backes et al. [22, 24] demonstrated how reflections from objects such as, users' eyes, teapots and walls can be exploited for spying on printed or digitally displayed data. In a different kind of attack, Ferrigno et al. [35] were able to recover AES keys by observing the optical emanations released by a microcontroller.

The first inference threat that we demonstrate in this paper employs an optical side-channel to infer the source audio. While there are multiple previous research efforts [31, 41, 57, 73, 74, 84, 90] on inferring audio from different types of information side-channels, our use of optical emanations from multimedia-visualizing smart lights as an information side-channel for this task is unique and novel. Our second inference threat employs an optical side-channel to infer the video being watched by the target user, which is comparable to works by Xu et al.

[88] and Schwittmann et al. [78]. However, our video inference framework which targets optical emanations from a multimedia-visualizing smart light is significantly different from Xu et al.'s and Schwittmann et al.'s inference frameworks that targeted light emanations from a TV screen. As smart bulbs are much brighter than TV screens, they make a desirable target for distant outdoor attackers. However, the distinguishing features in a smart bulb's output is often limited to only one color (in the RGB space) at a time as outlined in Section 3. This makes it more challenging to accurately infer the source video from a smart light's output, compared to a more straightforward analysis of the various simultaneous features present in a TV screen's output.

Information Exfiltration Through Optical Covert-channels. Loughry et al. [50] were one of the first to call attention to information exfiltration attacks on air-gapped[1] systems by employing visible-light LED indicators to transmit bits of information. Zhou et al. [91] demonstrated a similar attack, but using a malicious infrared signaling hardware installed on the air-gapped system. Guri et al. [39] leveraged on the limitations of human visual perception in order to covertly exfiltrate data through a standard computer LCD display, even in the presence of a user. Guri et al. [38] extended their previous work to exploit cameras equipped with infrared lights to both exfiltrate and infiltrate data from/to an air-gapped network. Shamir [75] demonstrated how a scanner can be used to infiltrate data/malware in to an air-gapped network. In another closely related work, Ronen et al. [68] used the visible spectrum of a smart light to exfiltrate data, which makes their attack less covert than our attack using the infrared spectrum.

A common limitation of the above data exfiltration attacks is that the covert-channel's bandwidth is restricted due to the binary nature of the signaling light, and slow switching response time between individual bits. For example, Zhou et al. [91] were able to achieve a raw throughput of 2.62 bits per second, even after installing a custom infrared signaling hardware on the air-gapped system. Guri et al.'s [39] method of encoding data in the form of QR-codes can potentially achieve relatively higher throughput, but their reconstruction was successful only from a short distance of up to 8 meters. In contrast, infrared-enabled smart lights can act as a superior data exfiltration gateway because - (a) they have fine-grained control of brightness/intensity, which can be used to design communication protocols that achieve higher throughput, (b) they are brighter than LED indicators found on computers and routers, increasing the possibility of data reconstruction from a longer distance, and (c) the adversary does not have to surreptitiously place any additional malicious hardware in the target area (i.e., in addition to the smart light already installed by the user).

Other Modalities of Attack. Side-channel and covert-channel attacks have a long history ranging from Tempest attacks during the World War II [36] to latest attacks using wearables [45, 49, 52, 53, 82, 83]. The predominant modalities of such attacks are electro-magnetic (a superset of optical attacks already discussed above) [18, 47, 56, 58, 67, 79–81], acoustic [19, 20, 23, 27, 30, 32, 40], network traffic [28, 44], vibrations [25, 33, 55, 57, 61, 64, 72], and motion [29, 45, 49, 51, 52, 54, 62, 82, 83, 89]. The attacks presented in this paper may also be achievable using other modalities. For example, Nakao et al. [61] demonstrated the possibility of audio reconstruction from the vibrations of a nearby window pane, using an invisible laser beam that is reflected back from the window pane. However, such an attack requires highly specialized equipment, and a complex setup because the reflected laser beam will not be received back at the point of origin unless the window pane is perfectly perpendicular to the laser beam. Directional microphones [63] can also be used for inference from a distance, but such an attack will require partial leakage of sound from target user's room and is susceptible to wind disturbances. Network traffic analysis [44] could be used to infer videos being streamed, but requires the attacker to have access to target user's network. Air-gapped systems could be compromised using various covert-channels, but majority of such attacks [32, 56, 64] require close proximity to the target system.

Before describing the working of our attack frameworks, we next present a few technical features of smart bulbs and their optical properties, which will be helpful in fully understanding the details of the presented attacks.

[1]Systems that are physically isolated from unsecured networks, so as to prevent network-based attacks.

3 TECHNICAL BACKGROUND

LIFX [6] and Phillips Hue [12] are two of the most popular commercially-available smart light systems. These smart lighting systems support millions of colors and multiple shades of white, have fine-grained brightness and saturation controls, possess wireless communication capabilities and can be controlled (wirelessly) using software platforms and mobile apps developed by the manufacturers. Several novel applications have been developed for such smart lighting systems by leveraging on their unique wireless communication and fine-grained control capabilities. For instance, both LIFX and Phillips Hue bulbs support *multimedia-visualization* by means of the manufacturer-provided mobile app or via third-party apps such as Light DJ [8], lifxDynamic/hueDynamic [4], etc.

The multimedia-visualization feature of the smart lighting system, which can be turned on using the system's mobile app, is used to change or modulate the brightness and/or color of user-selected bulb(s) of the system in real-time. During ***audio-visualization***, the change in brightness is made to reflect either the surrounding sound levels captured using an on-device microphone (more ambient noise), or the direct output of an audio playing application (less or no ambient noise). During ***video-visualization***, the change in color and brightness is made to reflect the dominant color and brightness level in the current frame of the video being played. The mobile app remotely controls the bulb by periodically sending it specially formatted packets. A careful analysis of the local network traffic shows that audio-visualizing applications transmit approximately 10 packets to the bulb per second, whereas video-visualizing applications transmit approximately 1 packet per second. Communication in case of the LIFX bulbs happens via an 802.11 access point, whereas the Phillips Hue bulb employs 802.15.4 (Zigbee) protocol to communicate with the mobile app. Each packet in the LIFX protocol [1] consists of three headers followed by a payload:

- The headers – *frame header*, *frame address* and *protocol header* – includes details on how to process the frame and its payload, and also specifies the destination bulb.
- The *payload* describes the desired hue, brightness and saturation levels.

The Phillips Hue platform has a very similar communication protocol, but employs an additional hub between the 802.11 access point and the bulb. The hub translates TCP/IP packets in to Zigbee Light Link [17] packets comprehensible to the Hue bulbs. Upon receiving a packet, the target bulb parses the packet and changes its color, brightness and saturation based on the payload information. When multimedia-visualization (audio or video) is enabled, a stream of packets from the controlling mobile application creates a real-time visual effect with the output light.

Color Models. Next, we briefly discuss the physical properties of light emitted from multi-color LED bulbs. These properties are utilized in the attacks presented later in the paper. Most colors observed on multi-color LED bulbs (such as the LIFX and Phillips Hue), are not the same as the electromagnetic VIBGYOR spectrum. Each color in the VIBGYOR spectrum has a unique wavelength, whereas the colors 'seen' on most multi-color LED

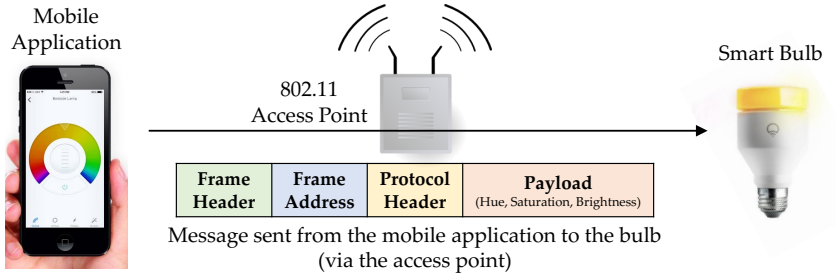


Fig. 1. The LIFX protocol, where a mobile application can control the bulb using specially formatted packets.

bulbs are made of three additive primary colors. Almost all visible colors can be constructed for the human eye by the additive color mixing of three colors that are in widely spaced regions of the visible spectrum. If the three colors of light can be mixed to produce white, they are called primary colors and the de facto additive primary colors are Red, Green and Blue (RGB). It is important to note that three-primary-color models (such as RGB) work only to 'trick' the human eye, which perceives color using three types of cone cells (S, M, and L) [69]. To illustrate with an example, the violet color seen on a multi-color LED bulb is a combination of red (λ = 620-750 *nm*), green (λ = 495-570 *nm*) and blue (λ = 450-495 *nm*) lights in $\langle 1:0:2 \rangle$ ratio, whereas violet light in a natural rainbow has a wavelength λ = 380-450 *nm*. $RGB\langle 1:0:2 \rangle$ happens to excite the S, M, and L cone cells in the same way a 380-450 *nm* wavelength light does. Therefore, we humans perceive $RGB\langle 1:0:2 \rangle$ as violet. Moreover. the RGB model enables the production of several other perceptual colors, such as the color magenta or $RGB\langle 1:0:1 \rangle$, which does not have its own place in the electromagnetic spectrum. A RGB color $\langle r, g, b \rangle$, where $r, g, b \in [0, 1]$, also self-defines the lightness (or darkness) of the color. In relation to multi-color LED bulbs, at $RGB\langle 1, 1, 0 \rangle$ the bulb outputs yellow light at maximum brightness, whereas at $RGB\langle 0.1, 0.1, 0 \rangle$ the bulb outputs yellow light at roughly 10% of maximum brightness.

An alternate representation of RGB colors is commonly done using the HSB (hue, saturation and brightness) model (Figure 2a). The *hue* represents all colors (achievable by mixing RGB) in 360 degrees, with 0°, 120° and 240° being Red, Green and Blue, respectively. At full *saturation* of compound colors, only two of the three primary colors are mixed (Figure 2b), which is represented by the outer surface of the HSB cone. Adding the third primary color reduces saturation and the saturation level moves closer to the center of the HSB cone. When all three primary colors are mixed in equal proportion, it results in white (or zero saturation). In HSB model, *brightness* is defined as the intensity of the dominant primary color(s). However, the perceived brightness (or relative luminance) depends on several biological properties of the eye and physical properties of electromagnetic radiation [43, 70].

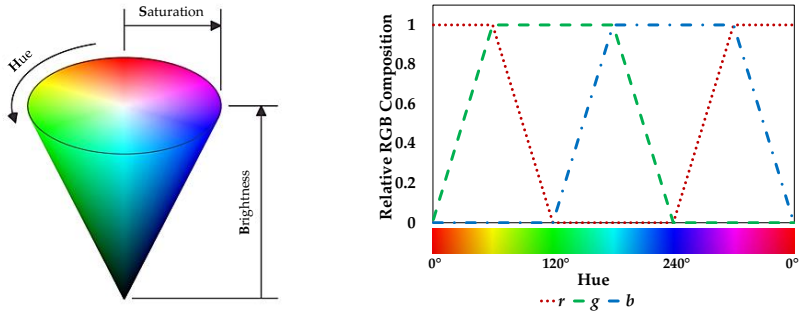
In summary, RGB can be translated to HSB using Equation 1, as follows:

$$\begin{aligned} H &= \cos^{-1}\left(\frac{\frac{1}{2}((r-g)+(r-b))}{(r-g)^2+(r-b)(g-b)}\right)^{\frac{1}{2}}; \\ S &= 1 - \frac{3}{(r+g+b)}\min(r,g,b); \\ B &= \max(r,g,b); \\ r,g,b &\in [0,1]. \end{aligned} \tag{1}$$

Infrared Lighting. In addition to RGB colors, some smart lights are also equipped with infrared capabilities. For example, the LIFX+ series of bulbs support 950 *nm* infrared lighting, which is invisible to the human eye, but useful for security cameras without built-in night-vision lighting. The brightness of the infrared light on LIFX+ bulbs can be controlled using packets with a payload describing the power level. A power level of zero indicates that the infrared LEDs will not be used, while a power level of 65535 indicates that the infrared channel should be set to the maximum possible value.

4 ADVERSARY MODEL

For the private information inference threats that take advantage of the audio-visualizing and video-visualizing functionalities of smart lights, we assume a passive adversary whose goal is to infer a target user's media consumption by visually eavesdropping on the smart bulb's output (i.e., light emitted by bulb), without actively attacking the user's wireless (Wi-Fi, Bluetooth, or Zigbee) network or appliances. The user's wireless network is assumed to be secured against eavesdropping attacks, for example using WPA2, so the adversary cannot perform direct analysis of the packets sent to the smart bulb (from the controlling mobile application). In addition to visual



(a) (b)

Fig. 2. (a) – The HSB cone; (b) – RGB composition of HSB colors, at full saturation.

eavesdropping instruments, such as light and color sensors, the adversary also needs sufficient computational and storage resources to initially create a comprehensive reference library of media items (songs and videos), and then match eavesdropped light patterns to items in the reference library.

For the data exfiltration threat using infrared-enabled smart lights, we assume an adversary whose goal is to exfiltrate data out of a target user's network or personal device, which is again assumed to be secured against wireless eavesdropping attacks. For the data exfiltration attack to work, the adversary has to additionally install a malicious software agent on the target user's device, for example, their smartphone or computer, that connects to the same network as their infrared-capable smart bulbs. This can be achieved by social engineering attacks [46], or by tricking the user in to installing a Trojan application [34]. The malicious software agent is responsible for encoding target user's private data (accessible on-device or on the network) in a format suitable for infrared communication, and transmits the encoded data using the user's infrared-enabled smart light. We also assume that the malicious software agent cannot directly communicate with an adversarial server over Internet (for example, due to a firewall), or that the user's network is air-gapped. In addition to infrared sensing hardware, for successfully executing this attack the adversary also needs sufficient computational and storage resources to record time-series of infrared light and process them to reconstruct the intended private data. In this paper, we explain and evaluate the feasibility of the above threats by means of a representative smart lighting system, namely, the LIFX system. We also assess the generalizability of our results by evaluating the proposed threats on an alternate system such as Phillips Hue.

5 AUDIO INFERENCE THREAT

We first outline the working of audio-visualizing smart lights, followed by the proposed inference attack on audio-visualizing lights. *In the audio-visualization mode, the smart bulb reacts to the high and low tones present in the input audio stream by fluctuating its output light brightness.* Smart light mobile applications typically offer two different light coloring modes during audio-visualization. In ***static hue*** mode, the bulb hue does not change during audio-visualization unless manually changed, while in the ***random hue*** mode the bulb periodically changes to a random hue at full saturation. In both modes, the input audio stream affects only the brightness level of the bulb, and the bulb hue (static or random) has no correlation to the audio.

5.1 Effect of Sound on Bulb Brightness

A simple observation with naked eyes and ears already gave us the intuition that a smart bulb's brightness fluctuates more with higher audio amplitudes. To precisely study this audio-visualization property of smart light

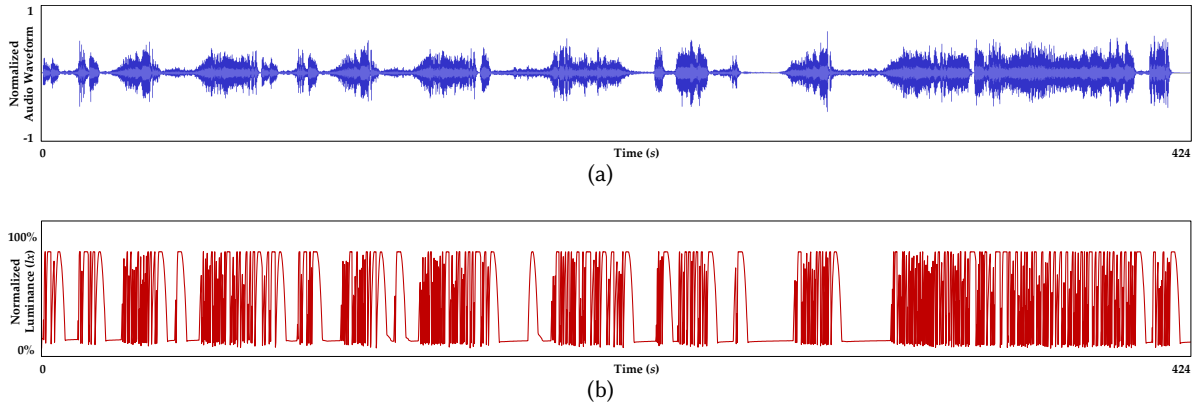
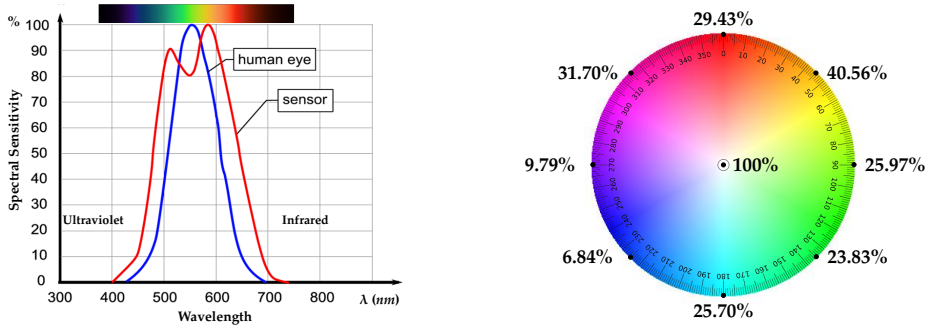


Fig. 3. (a) Audio waveform of *Beethoven's 5th Symphony* (duration of 7 minutes and 4 seconds); (b) Fluctuations in brightness of an audio-visualizing smart bulb output (in static hue mode), as captured by a BH1751FVI luminance meter.

bulbs, we conducted a systematic analysis of several raw audio tracks by precisely measuring the corresponding fluctuations in a LIFX bulb's brightness levels. For this we created an exploratory setup (details in Appendix B), where we played a few sample audio tracks multiple times in static hue mode, and observed the corresponding time-series of the bulb's brightness profile measured using a luminance meter. Our observations from this exploratory setup resulted in two significant conclusions, both of which are instrumental in the design of our attack framework. First, the observed luminance (δ_l) of the bulb fluctuates when a relatively high amplitude is detected in the audio stream. These high amplitudes may represent various types of tones, such as high vocals or drum beats, which varies on the type of song being played. The luminance fluctuation subsides when the audio amplitude is low. This phenomenon is shown with an example in Figure 3. It is evident from this example that there exists a clear correlation between the audio waveform of the song (say A_1) and its corresponding "luminance-profile" (say L_{A_1}). In other words, a luminance-profile is the time-series of brightness values of an audio-visualizing bulb when the song is playing.

Our second observation is that for a given song, the luminance-profile suffers minor distortions across multiple recordings. In other words, if the luminance-profile of a song A_1 captured during two different playbacks is denoted by L'_{A_1} and L''_{A_1}, respectively, then L'_{A_1} and L''_{A_1} will be similar but not identical. These minor distortions can be attributed to factors such as varying network latencies, packet loss due to network congestion and ambient audio noise. Also, because the brightness level is updated only every 100 milliseconds (10 Hz), an imperfect alignment of the song's starting point within the 100 milliseconds interval can cause a slightly different luminance-profile that follows. That being said, the similarity between L'_{A_1} and L''_{A_1} is generally very good. We utilize this property to design our audio inference framework which is based on *elastic* time-series matching (Section 5.2) and is immune to such minor distortions.

Let us further clarify the notion of 'brightness' in these luminance-profiles captured by the luminance sensor. The total amount of visible light coming out from a source (the bulb in our case) is described using the term *luminous flux*, which is the luminous energy released per unit time by the source and is measured in SI units of lumen or lm. *Illuminance*, on the other hand, is the amount of luminous flux incident on a surface, and is measured in lux or lx ($lx = lm/m^2$). The (normalized) brightness values present in the luminance-profile time series that we construct (Figure 3b) are nothing but the amount of luminous energy incident on the BH1751FVI photodiode (luminance meter used in our setup) divided by surface area of the sensor. In other words, the brightness values in the luminance-profile time-series are nothing but the illuminance values captured by the photodiode or luminance



(a) (b)

Fig. 4. (a) Luminance meter's response to the electromagnetic spectrum; (b) Luminance meter's response to different LIFX bulb hues (using RGB-HSB color model) at fixed brightness and full saturation.

sensor. It can be difficult to accurately measure illuminance using a photodiode that is also sensitive to infrared and/or ultraviolet radiations. However, as the BH1751FVI's spectral sensitivity is very close to the human eye (Figure 4a), the sensitivity of the photodiode to infrared and ultraviolet radiations is not significant.

5.2 Factors Affecting Song Identification

Next, let's discuss how we can use a luminance-profile in *static* and *random* hue modes to uniquely identify songs. In static hue mode the relative change in bulb's illuminance will cause proportional gain or drop on the luminance meter. This is true for all hues, including compound hues such as yellow or cyan, because the spectral sensitivity remains constant when the hue is not changing. As a result, the normalized luminance-profile of a song will be similar across different colors, as long as the hue is not changed between the start and end of a song playback. This allows us to create a reference library of luminance-profiles of songs using a single hue, and use it to match songs across any hue.

Matching luminance-profiles in random hue mode is, however, more complex. The luminance-profile in random hue mode is affected by two factors: the brightness (which is dependent on the song's amplitudes), and the luminance meter's sensitivity towards the hues that appear during the playback (which is independent of the song). From Figure 4a we already know that the luminance meter's sensitivity varies for different hues, which can create unwanted fluctuations in the luminance meter's response while recording the luminance-profile of a song (Figure 4b). This may result in incorrect song matching and poor inference performance. Moreover, characterizing the sensitivity of the BH1751FVI photodiode for each hue is not helpful because the RGB color spectrum does not have the same physical properties as the VIBGYOR spectrum (as explained in Section 3). Also, certain perceptual colors do not appear on the VIBGYOR spectrum.

A solution to this problem can be achieved by using a RGB sensor in conjunction with the luminance meter. A RGB sensor can help breakdown the composition of random hues coming out of the smart bulb, which can be used to uniquely identify the hues. Subsequently, the highest observed luminance level of each hue can be used to normalize the luminance values of all instances of corresponding hues, that appear during a song. Once normalized per hue, the resultant time-series becomes independent of changes in luminance due to changing hues, and only depends on the song amplitudes. This hue-independent luminance-profile should be similar across two different recordings, even when the order of hues is completely different. To test the feasibility of this approach, we conducted additional exploratory experiments using a Vernier GDX-LC RGB sensor. We cycled the bulb through the entire HSB hue palette at full brightness and saturation, and recorded the responses observed on

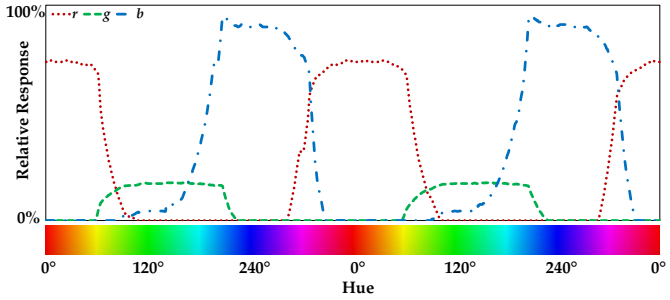
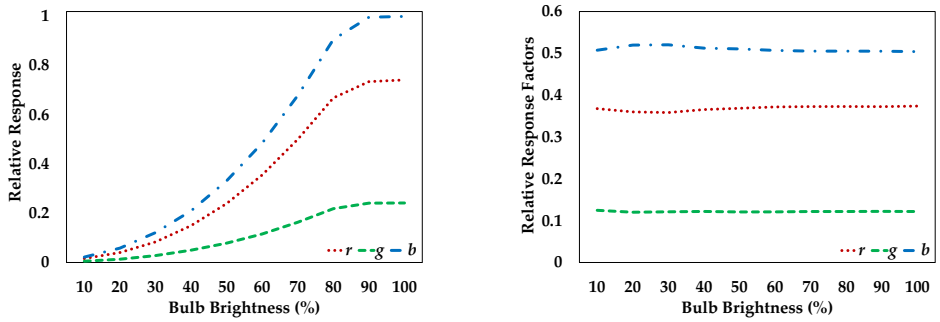


Fig. 5. Relative response of $\langle r, g, b \rangle$ components for the entire HSB hue cycle, at full brightness and saturation.

the RGB sensor. Figure 5 shows that the observed response of Red (δ_r), Green (δ_g) and Blue (δ_b) components have different characteristics. δ_b has the highest peak response, followed by δ_r and δ_g. Relative luminance calibrations made by LED bulb manufacturers (so that at fixed brightness, red, green and blue lights are perceived by the human eye to be of the same level) play a significant factor in the differing peak response. As seen in Figure 4a, our eyes are more sensitive to green than red and blue. As a result, red and blue colors are calibrated to be more luminous than green, so as avoid a greenish-white perception for $RGB\langle 1:1:1 \rangle$. We also observed that the color composition does not follow the ideal HSB model (Figure 2b), with some hues missing a primary color and some hues where all three primary colors are present (which should not happen at full saturation). This observation can be attributed to imperfect RGB sensing by the sensor [3], and/or possible design limitations of the RGB LEDs present in the bulb. These observations were consistent with another bulb of same make and model, which reduces the possibility of having manufacturing defects in our particular bulb.

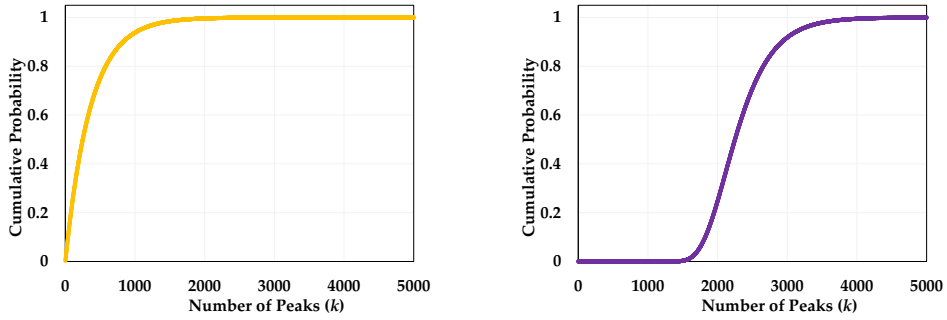
While the above observations are not ideal, we can still approximately identify individual hues based on their observed RGB response ratios $\langle \delta_r : \delta_g : \delta_b \rangle$. Certain hues neighboring to 0° and 240° cannot be uniquely identified because of only one primary color being present. However, that does not affect the proposed approach of normalization based on highest luminance of the identified hue, because hues neighboring to 0° and 240° are physically still red and blue lights, respectively. Moreover, the response of the only primary color remains roughly the same for these neighboring hues (Figure 5), at fixed brightness. In order to consistently identify a light hue at different brightness levels based on its RGB composition, the response ratio $\langle \delta_r : \delta_g : \delta_b \rangle$ should also remain constant across varying brightness levels. To test if this holds true, we analyzed $RGB\langle 1:1:1 \rangle$ at different brightness levels and recorded the corresponding δ_r, δ_g and δ_b values observed on the GDX-LC sensor. Figure 6a shows the increasing $\langle \delta_r$, δ_g and $\delta_b \rangle$ values as the brightness was increased, and Figure 6b shows the response factors. As the response factors $\left(\frac{\delta_r}{(\delta_r+\delta_g+\delta_b)}, \frac{\delta_g}{(\delta_r+\delta_g+\delta_b)} \text{ and } \frac{\delta_b}{(\delta_r+\delta_g+\delta_b)}\right)$ remain fairly constant across the entire brightness range, we can in fact use $\langle \delta_r : \delta_g : \delta_b \rangle$ to uniquely identify light hues at any brightness level.

The next challenge in normalizing a random hue luminance-profile is to ensure that we observe all hues at their highest luminance points at least once, based on which all occurrences of the corresponding hues will be normalized. The highest luminance point is generally observed at the *peaks* in fluctuations that occur when a high amplitude is detected in the input audio. The number of peaks per minute varies significantly between different types of music. With a conservative assumption of 20 peaks per minutes, it will take a minimum of 18 minutes to learn the highest luminance values of 360 hues (one per degree). However, because the hue changes randomly in random hue mode, the probability that a specific hue appears on a peak is $\frac{1}{360}$ and the probability that a hue appears within first k peaks can be represented as the cumulative probability of a geometric distribution: $1 - (1 - \frac{1}{360})^k$. The probability that all 360 hues appear at least once within first k peaks can be modeled as a classical occupancy problem in probability theory [37]. Figure 7a shows the cumulative probability distribution



(a) (b)

Fig. 6. (a) Observed response of $\langle r, g, b \rangle$ components $\langle \delta_r, \delta_g, \delta_b \rangle$ for $RGB\langle 1:1:1 \rangle$, and (b) Response factors for $RGB\langle 1:1:1 \rangle$ at different brightness levels.



(a) (b)

Fig. 7. (a) Cumulative probability distribution of a specific hue appearing within first k peaks; (b) Cumulative probability distribution of all 360 hues appearing within first k peaks.

of a particular hue appearing within first k peaks and Figure 7b shows the cumulative probability distribution of all 360 hues appearing within first k peaks. A particular hue will appear with a probability greater than 99.999% at $k \geq 5000$, whereas the probability of all 360 hues appearing at least once in the first 5000 peaks is 99.967%. In other words, at 20 peaks per minutes it will take about 250 minutes to capture the highest luminance values of all 360 hues, with a 99.967% success probability.

5.3 Audio Inference Framework

With the above understanding of audio-visualizing light properties, we can now design an inference framework for inferring the source audio from its corresponding luminance-profile. Due to the fundamental differences in the static and random hue modes, we design separate processes for each of them in the inference framework (Figure 8a), but both share a common reference library.

Capturing Luminance-profile. The inference attack starts with the adversary recording the observed luminance-profile (L'_{A_u}) of an unknown target song (A_u) using the luminance meter. The observed luminance-profile is the time-series of observed luminance values, recorded at a constant sampling rate (10 Hz). Therefore, the observed luminance-profile can be represented as $L'_{A_u} = \{\delta_l^{t=1}, \delta_l^{t=2}, \ldots, \delta_l^{t=m}\}$, where $\delta_l^{t=m}$ is the observed

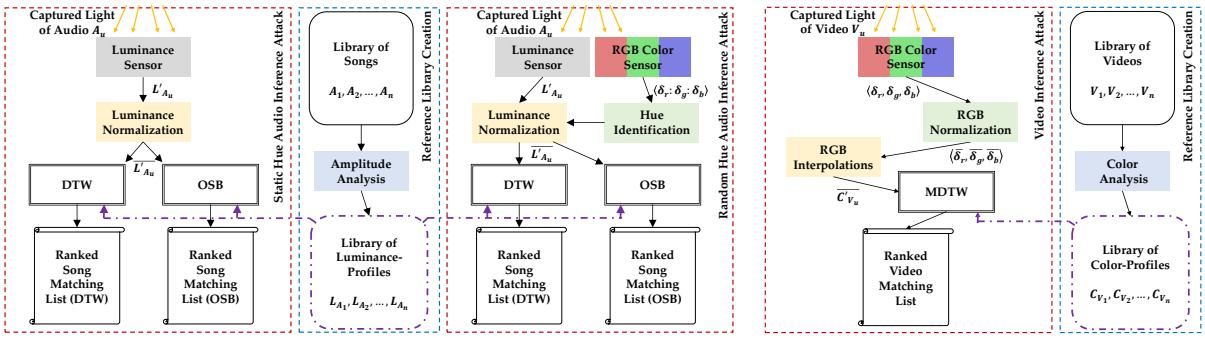


Fig. 8. (a) Audio Inference Framework; (b) Video Inference Framework.

luminance at time instance m since start of the recording. In random hue mode, the adversary additionally records the corresponding hues $\delta_h^{t=\{1,2,\ldots,m\}} = \langle \delta_r^t : \delta_g^t : \delta_b^t \rangle$ observed for all $\delta_l^{t=\{1,2,\ldots,m\}}$ in L'_{A_u}, for the purpose of normalization.

Luminance Normalization. Once the entire luminance-profile (L'_{A_u}) is recorded for a chosen observation duration, the next step for the adversary is to normalize it in order to achieve amplitude invariance during similarity search [26]. In static hue mode, L'_{A_u} is normalized with respect to maximum δ_l recorded at the point of observation. In random hue mode, the adversary normalizes each observed $\delta_l^{t=\{1,2,\ldots,m\}}$ in L'_{A_u} with respect to the maximum δ_l recorded at the point of observation for its corresponding hue. The output of this step is a normalized luminance-profile, $\overline{L'_{A_u}}$.

Creating a Reference Library. Before matching the normalized luminance-profile ($\overline{L'_{A_u}}$) against a comprehensive reference library of songs, the adversary has to create a reference library of luminance-profiles corresponding to songs in the library of songs. As seen in Figure 3, the absolute audio amplitude is directly correlated to the observed luminance. Using this observation, we can create template luminance-profiles by sampling the amplitudes in waveform audio files at 10 Hz, and converting them to absolute values. These template luminance-profiles serve as an approximate representation of how an audio-visualizing smart bulb will react.

Similarity Search. The final step for the adversary is to match the luminance-profile to songs in reference library, as follows:

- *Dynamic Time Warping.* We use a classification method based on Dynamic Time Warping (DTW) [60], an algorithm for measuring similarity between temporal sequences that are misaligned and vary in time or speed. We compute the DTW distance between the observed luminance-profile and template luminance-profiles in the reference library, selecting the song whose template yields the minimal distance, i.e., we choose the song i such that:

$$A_u = \arg\min_{A_i} DTW(\overline{L'_{A_u}}, L_{A_i}) \tag{2}$$

- *Optimal Subsequence Bijection.* We also evaluate another classification technique for measuring similarity between temporal sequences, known as the Optimal Subsequence Bijection (OSB) [48]. OSB is similar to DTW, but allows skipping of elements in the query sequence with a penalty, thus aligning noisy subsequences more efficiently. In our evaluation (Section 6) we present results obtained by using OSB and compare them to those obtained using DTW.

6 AUDIO INFERENCE EVALUATION

We comprehensively evaluate our audio inference framework (Section 5.3) across a variety of experimental parameters such as observation time duration, distance between adversary and the smart bulb, visible transmittance of a window between the smart bulb and adversary, and using different brands of smart bulbs (LIFX A19 Color Gen3 and Phillips Hue Color Gen3).

Experimental Setup. We compiled a reference library of 400 chart-topping songs released in the last two years, with equal composition of four different genres: country, dance, jazz, and rock. We chose these 400 popular songs for our dictionary because users are more likely to be playing one of these songs at the time of the attack. This is analogous to password cracking using a dictionary of most commonly used passwords. We test two different observation points, one in an indoor setting where the observation point was at 5 meters away from the bulb, another in an outdoor setting where the observation point was at 50 meters away from the bulb (Figure 9). While the outdoor setting represents a more realistic attack scenario for the audio inference attack, the indoor observation point can also be useful if the target user is listening using a headphone/earphone (and if the adversary has indoor access). For the indoor setting, a 20 *mm* 12x telephoto lens was used to focus light on the sensors. For the outdoor setting, a 80 *mm* 45-225x telescope was used.

Observation Time and Accuracy. First we evaluate the success rate of our audio inference framework without any obstruction between the smart bulb and observation point. The window was kept open for the outdoor setting. The LIFX bulb was used in this part of the evaluation, and audio-visualization was performed using the manufacturer's LIFX Android application. For 100 test songs (25 in each of the four genres), we analyzed observation durations varying from 15 to 120 seconds. The time at which the adversary starts observing each test song's audio-visualizing light output was chosen at random. In static hue mode, the hue was chosen at random before the playback of each test song. We measure the accuracy of the framework based on the rank of predicted songs matched against the entire reference library (rank of 1 being the correct prediction). From Figure 10, we can see that the mean rank of predicted songs decreases with a larger observation time window, which is intuitive. Inference accuracy in random hue mode was slightly worse than static hue, but conforms to similar improvements for longer observation time windows as static hue. Outdoor inference accuracy was lower than the indoor setting, which is expected due to the lower intensity of light reaching the sensors. Also, we did not observe a significant advantage of using OSB over DTW. OSB was slightly more accurate than DTW for shorter observation durations, however, their accuracies were very similar for the 120 *sec* window sizes. This can be attributed to the fact that, although OSB has the ability to skip outliers in the test sequence, we have very few outliers in our test sequences to begin with. This is because both audio-visualization and sensor sampling follow approximately constant sampling rates.

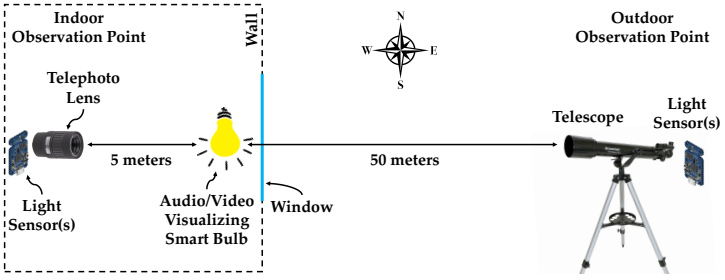


Fig. 9. Experimental setup for evaluating audio and video inference and data exfiltration attacks.

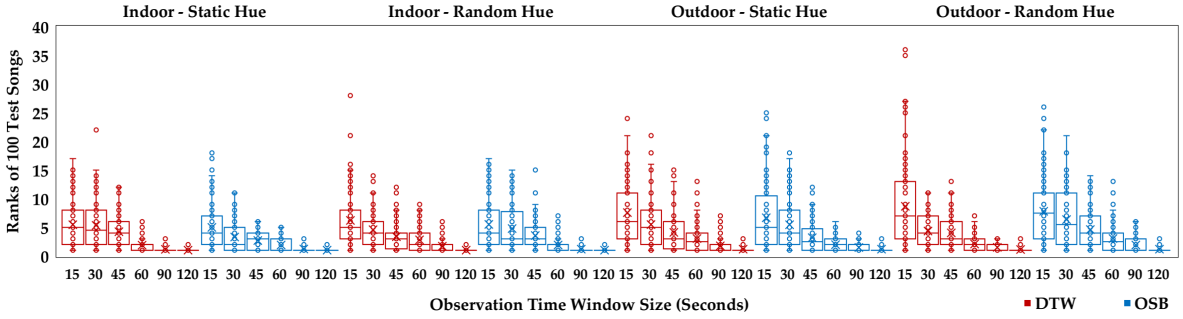


Fig. 10. Ranks of 100 test songs with respect to increasing observation time window sizes, using the full reference library.

Size of the Reference Library. A larger dictionary should intuitively increase the potential of a mismatch with another song having an audio segment with similar luminance-profile as the observed audio segment. We validate this using only half of the reference library for prediction (200 songs). In Table 1 we see that the mean rank of predicted test songs decreases with a smaller reference library. For example, the mean rank was 1.93 for random hue inference using 90 *sec* observation window (outdoor observation, OSB) and the full library, which decreased to 1.57 when only half of the library was used. However, an adversary matching against a large library of songs may be able to design more efficient matching techniques, by identifying and employing additional features that are common between songs and observed light output, which is beyond the scope of this paper.

Music Genres. Certain song genres are slower paced (such as jazz) than others (such as dance). Some other genres may have high frequency periodic beats (such as dance) while others are dominated with flat vocals (such as country). These characteristics are also reflected on the audio-visualizing smart lights, and results in a higher confusion within the same genre. Figure 11a shows the genre confusion matrix of 100 songs tested under the following settings: 60 *sec* observation duration, indoor, static hue, OSB, full library (from results in Table 1). We see that the highest confusion is more likely to happen within the genre of the test song. In this example, 51 songs were correctly predicted in the top rank, while genres of as many as 82 songs were correct. This trend was observed across all different test scenarios. This implies that even if an adversary is unable accurately match the audio-visualized data to its corresponding song, the adversary can still infer the user's media or genre preference.

Song Not in Library. As both DTW and OSB are used to calculate the lowest distance between the luminance-profile of an unknown test song and known templates in the reference library, songs absent from the library will

Table 1. Mean ranks of 100 test songs in different test settings and observation time window sizes.

	LIFX							Phillips Hue	
	No Obstruction				0.8 VT	0.65 VT	0.4 VT	No Obstruction	
	Half Library		Full Library		Full Library			Full Library	
	90 *sec*	120 *sec*	90 *sec*	120 *sec*	120 *sec*	120 *sec*	120 *sec*	90 *sec*	120 *sec*
Indoor - Static Hue - DTW	1.02	1.02	1.18	1.01	–	–	–	1.23	1.09
Indoor - Static Hue - OSB	1.03	1.02	1.22	1.02	–	–	–	1.22	1.09
Indoor - Random Hue - DTW	1.13	1.01	1.72	1.03	–	–	–	1.68	1.16
Indoor - Random Hue - OSB	1.12	1.06	1.21	1.03	–	–	–	1.24	1.17
Outdoor - Static Hue - DTW	1.36	1.07	1.78	1.20	1.63	2.95	4.47	1.79	1.35
Outdoor - Static Hue - OSB	1.29	1.05	1.50	1.19	1.65	2.42	5.10	1.62	1.56
Outdoor - Random Hue - DTW	1.51	1.09	1.53	1.15	1.84	3.64	5.85	1.77	1.27
Outdoor - Random Hue - OSB	1.57	1.10	1.93	1.21	1.81	3.67	5.79	2.01	1.25

Top Ranked Predictions

True Genre	*Country*	*Dance*	*Jazz*	*Rock*	Total
Country	19 (12 True Song)	1	2	3	25
Dance	2	22 (15 True Song)	0	1	25
Jazz	1	0	23 (16 True Song)	1	25
Rock	3	3	1	18 (8 True Song)	25
Total	25	26	26	23	Correct Songs: 51 Correct Genres: 82

(a)

Top Ranked Predictions

True Genre	*Country*	*Dance*	*Jazz*	*Rock*	Total
Country	13	2	3	7	25
Dance	3	15	1	6	25
Jazz	3	1	20	1	25
Rock	6	4	2	13	25
Total	25	22	26	27	Correct Genres: 61

(b)

Fig. 11. Genre confusion matrix of 100 test songs (a) in library, and (b) not in library.

be (incorrectly) predicted as the closest matching song in the reference library. However, based on our previous findings related to genre confusion, it may still be possible to infer the genre of a song not present in the library. To evaluate this hypothesis, we reused the same test data collected for genre classification (Figure 11a), and re-ran the genre experiment. However, for each song being tested, we removed the corresponding song from the reference library. Comparing Figures 11a and 11b, we observe that the correct prediction of song genres dropped from 82 to 61, which is still significantly higher than random guessing. This implies that an adversary can still infer target user's media or genre preference with moderate confidence, even when the reference library is not exhaustive and may not contain the song being played by the target user.

Visible Transmittance of Window. We next evaluate how our audio inference framework performs when there is partial obstruction, such as a tinted window, between the smart bulb and the adversary's light sensing hardware. We used three window glasses of varying Visible Transmittance (VT) – approximately 0.8, 0.65 and 0.4 – in the outdoor setting and evaluated the prediction accuracy. We repeated the outdoor test with 100 songs using an 120 *sec* observation time window size. In Table 1, we see that the mean ranks of the predicted song drops as the visible transmittance of the window decreases. For example, the mean rank of predicted song was 1.20 for inference using an open window (test settings: outdoor, static hue, DTW), which decreased to 1.63 using a 0.8 VT window, 2.95 using a 0.65 VT window, and 4.47 using a 0.4 VT window. This is expected because with lower VT of the window, the intensity of light reaching the sensors is diminished. Thus, highly tinted windows can act as an effective protection measure against such an inference attack.

Generalizability. We also study the generalizability of our attack framework by evaluating its inference performance on other commercial smart bulbs. We chose the Phillips Hue bulb for this evaluation because of it's popularity and it is currently one of the few smart bulbs that support audio-visualization. However, due to the lack of a native audio-visualization feature in the Phillips Hue mobile application, audio-visualization was done using hueDynamic, a third-party application. Table 1 shows the mean ranks of 100 test songs predicted after observing the audio-visualization of the Phillips Hue bulb for 90 and 120 seconds. The mean ranks in most settings, indoor or outdoor, were slightly higher than that of using the LIFX bulb. On an average across all different test settings, the mean rank of the predicted song increased by 0.06 and 0.13 for 90 *sec* and 120 *sec* observation time windows, respectively. While these results validate the generalizability of our inference attack, we investigated why the accuracy was consistently lower than when using the LIFX bulb. One of the primary contributing factors may be the difference in the maximum light output between the two bulbs. The LIFX A19 Gen3 bulb is rated at a maximum of 1100 *lm* while the Phillips Hue Gen3 is rated at a maximum of 800 *lm*. As a result, the intensity of light reaching the sensors is lower, and aligning with our previous observations, this reduces the accuracy of our inference framework.

7 VIDEO INFERENCE THREAT

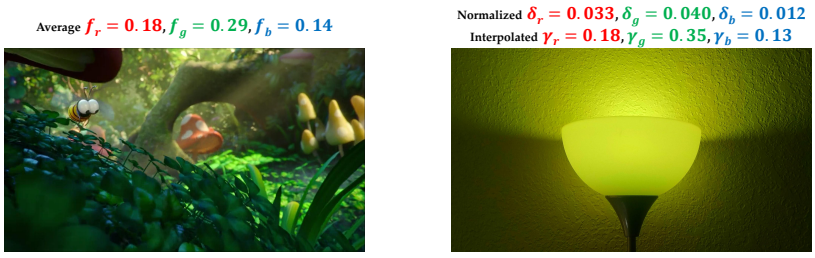
When video-visualization is turned on in the mobile app, the smart bulb reacts to the colors present in the input video stream by changing its output light color to the average RGB composition of the current frame in the video. Unlike audio-visualization which is better understood with the HSB color model, video-visualization is better represented using the RGB model. This is because in audio-visualization the saturation level never changes, whereas saturation (along with hue and brightness) in video-visualization is entirely dependent on the color composition of the current video frame. As all three variables are dynamic, using the RGB model simplifies the matching process with responses on a RGB sensor, as described next.

7.1 Effect of Video on Bulb Color

To precisely study the video-visualization properties, we created another exploratory setup (Appendix B). We played a few sample video files multiple times and observed the corresponding time-series of bulb's RGB color measured using the Vernier GDX-LC RGB sensor. The GDX-LC's RGB sensor captures responses in the 615 *nm*, 525 *nm*, and 465 *nm* peak wavelength ranges, approximately representing the observed intensities of red, green and blue colored lights, respectively. Our observations brought us to two similar conclusions as in Section 5.1. First, the observed RGB color of the bulb $\langle \delta_r, \delta_g, \delta_b \rangle$ has some correlation with the average RGB color in the current frame $\langle f_r, f_g, f_b \rangle$ of the video stream (Figure 12), which is expected. As a result, the RGB "color-profile" of a video (the time-series of $\langle \delta_r, \delta_g, \delta_b \rangle$ when the video is playing) observed on a video-visualizing bulb is unique to the video, and the probability that a completely different video also has the same color-profile is small (due to the high number of colors possible per pixel, in each frame of both videos). Second, similar to luminance-profiles in audio-visualization, color-profiles also suffers minor distortions across multiple recordings. If the color-profile of a video V_1 captured during two different playbacks is denoted by C'_{V_1} and C''_{V_1}, C'_{V_1} and C''_{V_1} will be similar but not identical. These minor distortions can be attributed to varying network latencies, packet loss due to network congestion and varying RGB calculation time. However, similarity between C'_{V_1} and C''_{V_1} is generally good, which can be utilized for designing dictionary-based elastic time-series matching attacks.

7.2 Video Matching Using Color-Profile

We already detailed some characteristics of RGB smart bulbs in random hue audio-visualization (Section 5.2). However, identifying hues based on observed response ratios $\langle \delta_r : \delta_g : \delta_b \rangle$ is relatively simpler than accurately identifying a $\langle r, g, b \rangle$ color from the observed responses. This is because all three variables (hue, saturation and brightness) of the HSB model are dynamic in video-visualization. Moreover, the non-ideal characteristics of



(a) (b)

Fig. 12. (a) – A frame from the movie *Smurfs: The Lost Village (2017)*; (b) – Corresponding light output observed on a video-visualizing LIFX bulb, $\langle \delta_r, \delta_g, \delta_b \rangle$.

the light output, and/or RGB sensor, makes it challenging to derive an accurate relation between $\langle f_r, f_g, f_b \rangle$ and $\langle \delta_r, \delta_g, \delta_b \rangle$. The two most significant problems in inferring $\langle f_r, f_g, f_b \rangle$ from $\langle \delta_r, \delta_g, \delta_b \rangle$ are: (i) the primary colors have different peak responses (Figure 5), and (ii) the responses are non-linear with respect to brightness (Figure 6a). We address both of the above problems by first normalizing all observed δ_r, δ_g and δ_b with respect to the highest observable δ_r, δ_g or δ_b. In case of both LIFX and Phillips Hue bulbs, δ_b has the highest observable response at 100% brightness (Figure 6a). This step also addresses other variable factors such distance from the light source, and optical amplifications before the sensor. After normalization, the δ_r, δ_g and δ_b are interpolated based on non-linearities learned for each of the primary colors (Appendix A). The interpolated response $\langle \gamma_r, \gamma_g, \gamma_b \rangle$ is much closer to the true composition $\langle f_r, f_g, f_b \rangle$ of the current frame (Figure 12). Even after normalization and interpolation, $\langle f_r, f_g, f_b \rangle$ and $\langle \gamma_r, \gamma_g, \gamma_b \rangle$ will not be an exact match for most colors, because the observable color responses do not adhere to ideal RGB proportions (Figure 5). However, $\langle f_r, f_g, f_b \rangle$ and $\langle \gamma_r, \gamma_g, \gamma_b \rangle$ are similar enough that we can create a reference library of color-profiles for a diverse set of videos, and use it to match the observed color-profile of a target video.

7.3 Video Inference Framework

With the above understanding of video-visualizing light properties, we can now design an video inference framework.

Capturing Color-profile. The inference attack starts with the adversary recording the observed color-profile (C'_{V_u}) of an unknown target video (V_u) using the color sensor. The observed color-profile is the time-series of observed RGB values, recorded at a constant sampling rate (1 Hz). Therefore, the observed color-profile can be represented as $C'_{V_u} = \{\langle \delta_r, \delta_g, \delta_b \rangle^{t=1}, \langle \delta_r, \delta_g, \delta_b \rangle^{t=2}, \ldots, \langle \delta_r, \delta_g, \delta_b \rangle^{t=m}\}$, where $\langle \delta_r, \delta_g, \delta_b \rangle^{t=m}$ is the observed luminance at time instance m since start of the recording.

Color Normalization and Interpolation. Once the entire color-profile (C'_{V_u}) is recorded for a chosen observation duration, the adversary next normalizes and interpolates (as described in Section 7.2) it to create the corresponding normalized and interpolated color-profile ($\overline{C'_{V_u}}$).

Creating a Reference Library. The adversary next creates a library of template color-profiles corresponding to video files in a reference library. The template color-profiles can be created by sampling the RGB composition in video files at 1 Hz. These template color-profiles serve as an approximate representation of how a video-visualizing smart bulb will react (Figure 12).

Similarity Search. The final step for the adversary is to match the color-profile ($\overline{C'_{V_u}}$) against the template color-profiles corresponding to the reference library of videos using a time-series similarity computing technique such as *Multidimensional Dynamic Time Warping (MDTW)* [85]. MDTW is a generalization of DTW for measuring similarity between temporal sequences, in two or more dimensions. We compute the 3DTW distance between the observed color-profile and template color-profiles in the reference library, selecting the video whose template yields the minimal distance. More formally, we choose the video i such that:

$$V_u = \arg\min_{V_i} MDTW(\overline{C'_{V_u}}, C_{V_i}) \tag{3}$$

8 VIDEO INFERENCE EVALUATION

We comprehensively evaluate the performance of our video inference framework (Section 7.3) for a variety of parameters. However, as video-visualization shares some of the same fundamental properties as audio-visualization, we observed similar trends in inference accuracy for factors such as window's visible transmittance and use of different bulbs. Therefore, here we analyze another critical attack condition, namely non-line-of-sight inference using reflected light.

Experimental Setup. We compiled a reference library of 500 full-length movies released on DVD and Blu-ray in the last 10 years. We test from the same two observation points as in the audio inference evaluation, indoor and outdoor (Figure 9). Also, the same telephoto lens and telescope were used in the indoor and outdoor settings, respectively. A LIFX A19 bulb was used for video visualization, and the GDX-LC RGB sensor was used to record the bulb's color output.

Observation Time, Distance and Accuracy. Similar to audio inference, we calculate the video inference accuracy based on the mean rank of 100 test videos. For each of the 100 test videos, we analyzed observation durations varying from 60 to 360 seconds. The time at which the adversary starts observing each test video's video-visualizing light output was chosen at random. In Figure 13 we see that the mean rank of predicted test videos decreases with a larger observation time window. For example, the mean rank was 6.34 for inference using a 60 *sec* observation time window (outdoor observation), which improved to 1.49 using a 360 *sec* window. Also, indoor accuracy was more accurate with mean rank of 4.24 for a 60 *sec* window to 1.11 using a 360 *sec* window.

Size of the Reference Library. As before, intuitively a larger dictionary should increase the potential of a mismatch with another video having a segment with similar color-profile as the observed segment. We validate this using only half of the reference library for prediction (250 movies). In Figure 13 we see that the mean rank of predicted test videos decreases with a smaller reference library. For example, the mean rank was 6.34 for inference using 60 *sec* observation window (outdoor observation) and the full library, which improved to 5.53 using a 60 *sec* window and half the library. However, an adversary matching against a large library of videos may be able to design more accurate matching techniques, by identifying and employing additional features that are common between video frames and observed light output, which is beyond the scope of this paper.

Video Not in Library. As MDTW calculates the distance between the color-profile of an unknown test video and known templates in the reference library, videos absent from the library will be (incorrectly) predicted as the closest matching video in the reference library. However, based on our previous findings related to audio genre confusion (in Section 6), it may still be possible to infer the genre of a video not present in the library. To evaluate this hypothesis, we analyze if the color-profiles could be used to infer video genres. Figure 14 shows the genre confusion matrix of 100 videos tested under the following settings: 240 *sec* observation duration, outdoor. For each video being tested, we removed the corresponding video from the reference library. We observe that the confusion within genres is not as high as it was in audio-visualization, with genre prediction of only 34 videos being correct. Our intuition behind the weaker genre prediction results is that time-series matching of color-profiles does not account for more complex features (compared to songs) that are shared between videos of the same genre. This implies that an adversary will have to train and utilize more complex machine learning

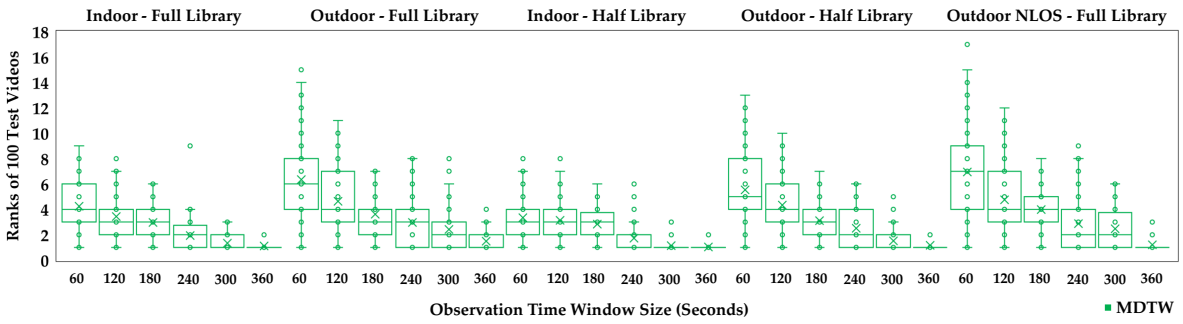


Fig. 13. Ranks of 100 test videos with respect to increasing observation time window sizes.

True Genre \ Top Ranked Predictions	*Action/ Adventure*	*Animation*	*Documentary/ Biographic*	*Romance*	*Western/ Historical*	**Total**
Action/ Adventure	14	11	2	8	4	39
Animation	3	5	0	1	3	10
Documentary/ Biographic	3	1	3	2	4	13
Romance	5	2	2	8	4	21
Western/ Historical	5	1	3	2	4	17
Total	30	20	10	21	19	**Correct Genres: 34**

Fig. 14. Genre confusion matrix of 100 test videos not in library.

algorithms, which can better utilize the RGB information present in the color-profiles, for more accurate genre prediction.

Non-line-of-sight (NLOS) Inference. So far, our experimental setup assumed a line-of-sight (LOS) observation of the smart bulb by the adversary. However, it possible that the smart bulb is not directly visible from the outside, and the adversary must infer using the light reflected off the walls of the room. To validate that our attack works in such NLOS settings, we placed the smart bulb beside the south wall of the room (Figure 9), and collected new test data in the outdoor setting based on the reflection off the west wall of the room. Then, we calculate the video inference accuracy (Figure 13) and compare the results with our previous results where the adversary had direct visual on the smart bulb. In NLOS, we observed an average difference of +0.17 in mean rank across the different observation durations, which is marginally less accurate than LOS inference. This can be attributed to the fact that the light source is omni-directional, which leads to similar level of dissemination for a distant observer in both LOS and NLOS settings. Nonetheless, these results imply that an adversary does not require LOS to the bulb for effective inference of the source video.

9 COVERT DATA EXFILTRATION THREAT

Next, we present another smart light based attack framework using which an adversary can actively and covertly exfiltrate private data from within a smart light user's personal device or network. In theory, any light source can be used to transmit data, and light based communication serves much of today's Internet backbone [2]. What makes this attack interesting is that traditional light bulbs are normally not perceived (or monitored) as an attack surface, even in high security establishments. In this attack, we not only show that a smart light can be used to transmit data, but we also describe how such an attack when carried out from within a secure air-gapped network can become a significant privacy and security threat. Moreover, unlike Internet gateways which can be protected against data exfiltration attacks using a firewall, an exfiltration gateway made out of a smart bulb has no such restrictions.

The proposed data exfiltration attack is currently possible on hub-less smart lights (e.g., LIFX) due to the lack of permission controls for controlling the light within the local network. This attack may also be possible using smart lights that connect via a hub, only if the hub does not have permission controls. However, the Phillips Hue ecosystem uses a hub with permission controls, and thus cannot be used for this attack unless the malicious software agent obtains access permissions by masquerading as an useful application (a Trojan). However, Phillips Hue currently does not offer any infrared-enabled smart lighting products, so we use the LIFX ecosystem to design and evaluate the proposed data exfiltration framework (Figure 15).

9.1 The Covert Channel

The adversary can potentially use either, or both, the visible and infrared spectrum of a smart bulb to create an one-way line-of-sight data exfiltration gateway, using transmission techniques such as amplitude and/or wavelength[2] shift keying [59, 86]. However, uninitiated changes in the visible light amplitude or color is more likely to get noticed, thus limiting the effectiveness of such a malicious gateway/channel. As human eyes are not sensitive to the infrared spectrum, it can be used to create a covert channel which can remain undetected for longer durations.

Depending on the amount of private data to be covertly exfiltrated, channel bandwidth can be a deciding factor in the success of such an attack. As a higher bandwidth channel will take less time to transmit the data, compared to a lower bandwidth channel, the likelihood of a detection and/or disruption is lessened with higher bandwidth. In our empirical tests we observed that LIFX+ bulbs can achieve a maximum infrared output (power level 65535) from an off state (power level 0) in approximately 0.45 seconds, and takes less than 0.2 seconds to completely turn off from a maximum output state. As these are the maximum switching response times, we can use the higher of the two as our clock period. After adding some padding we round up the clock period to 0.5 seconds ($ClockRate = 2\,Hz$), which results in the maximum channel bandwidth of 32 bits per second if a multilevel (M-ary) amplitude shift keying (ASK) [21] encoding technique is employed with $M = 65536$. Channel bandwidth in a M-ary ASK is $(\log_2 M) \times ClockRate$. However, distinguishing between each of the 65536 power levels (amplitudes) may be difficult, especially, if the distance of the receiver from the smart bulb is large, primarily due to signal attenuation and channel noise. As a result, the error rate at the receiver is to some extent proportional to M and the distance.

9.2 Private Data Encoding and Transmission

Once the value of M is decided by the adversary (based on the distance and acceptable level of reconstruction quality), the next step is to encode private data of interest (in binary form) using M-ary ASK. This task is undertaken by a malicious software agent installed on the target user's device or network, as outlined earlier in Section 4. The encoded data is then transmitted in blocks by controlling the infrared power level of the smart bulb connected to the same device or network. For example in 4-ary ASK, a '00' data block will be transmitted by setting the infrared power level of the bulb at 0 (off), a '01' data block will set the power level at 21845 ($65536 \div (M-1)$), a '10' data block will set the power level at 43690, and a '11' data block will set the infrared

[2] As the LIFX+ series supports only 950 nm infrared light, only amplitude shift keying is possible with the infrared spectrum.

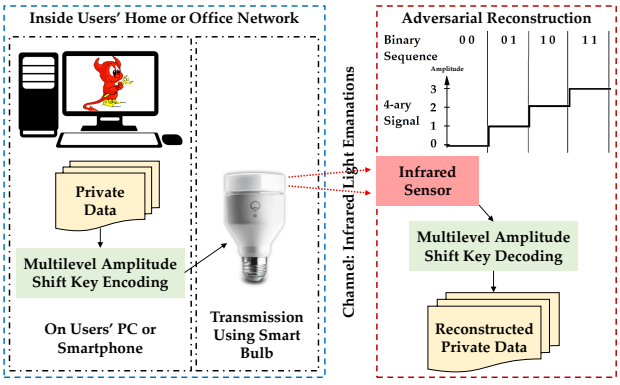


Fig. 15. Smart light-based data exfiltration framework.

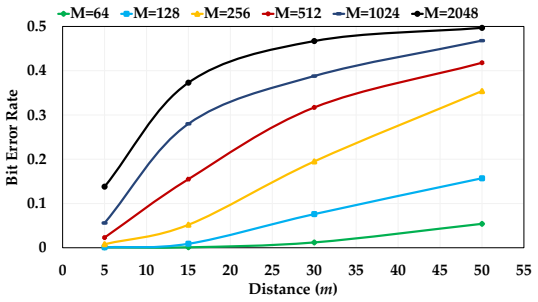


Fig. 16. Mean BER in exfiltration of both text and image.

power level at 65535 (maximum). Before data transmission, a pre-decided start symbol [65] is used to activate the receiver, synchronize the clock, and set maximum amplitude for normalization. After data transmission is complete, an end symbol is used to signal termination.

9.3 Adversarial Reconstruction

On the adversary's side, he/she observes the target user's smart bulb using an infrared sensor. As the LIFX+ bulbs supports only 950 *nm* infrared light, the adversary requires compatible infrared sensing hardware. We used a TSOP48 [15] infrared sensor and recorded data using an ATtiny85-based Arduino board. Once a start symbol is received by the infrared sensor, the adversary starts recording the observed infrared amplitudes representing the M-ary ASK encoded data, until an end symbol is received. Then the adversary normalizes the recorded data based on the maximum amplitude, and decodes it to reconstruct the private data in binary format. Depending on the amount of signal attenuation and channel noise, the reconstructed data may not be identical to the original data.

10 DATA EXFILTRATION EVALUATION

Next, we present performance evaluation results of our data exfiltration framework outlined earlier.

Experimental Setup. Standardized datasets - first 10 sets of Harvard sentences [71] and a 128×128 image of Lena (Figure 17a) - were used to evaluate the Bit Error Rate (BER) in the reconstructed data. The adversarial observation distance from the smart bulb was increased in steps (5, 15, 30, and 50 meters), and BER was calculated for different values of M (64, 128, 256, 512, 1024, and 2048) at each of these distances. A LIFX+ A19 Gen3 bulb was used for transmission and a 80 *mm* 45-255x telescope was used to focus light on the infrared sensor (TSOP48). A python script, acting as the malicious software agent, was used to encode and transmit the test data. It was executed on a Windows PC connected to the same Wi-Fi network as the bulb.

Distance and Error Rate. The infrared signal strength reduces as distance from the bulb increases. As a result, the boundaries between the M amplitude levels present in a M-ary ASK signal is also diminished, leading to higher confusion between neighboring amplitude levels and thus errors in the reconstructed data. This phenomenon was evident in our evaluation results (Figure 16), especially for higher values of M. For example, with $M = 2048$ the BER increased from 0.138 to 0.496 when the distance was increased from 5 m to 50 m. Note that when the signal is very weak or corrupted, the BER converges to 50%, provided that the binary data source used approximately follows a fair Bernoulli distribution. In Figure 17 we see how the image of Lena is degraded with higher BER for longer distances. The adversary can potentially reduce BER by employing Forward Error Correction (FEC), a

(a) Original (b) 5 m (c) 15 m (d) 30 m (e) 50 m

Fig. 17. Reconstructed image of Lena at different distances, using 128-ary ASK.

Original Text:	**A cup of sugar makes sweet fudge**
Reconstructed Text:	**A buq pf!sugbr m`kes suees hudfe**

Fig. 18. Text reconstruction using 256-ary ASK at 15 m.

digital signal processing technique used to enhance data reliability [66]. However, FEC introduces a trade-off between bandwidth overhead and BER, and the adversary must decide on optimal FEC parameters based on other variables, such as the data type, distance, transmission time and M.

***M* and Error Rate.** The boundaries between the M amplitude levels present in a M-ary ASK signal is also diminished when the value of M is increased, leading to higher confusion between neighboring amplitude levels. In Figure 16, we observe that the BER consistently increases for higher values of M. However, even with high BER the reconstructed data can still be useful for an adversary. For example, text inference from reconstructed data is easier than many other encoding schemes, because of the higher likelihood of confusion between neighboring amplitude levels. Figure 18 shows the example of a reconstructed Harvard sentence using 256-ary ASK. We see that the incorrectly reconstructed letters are neighboring to the original letters on the ASCII chart. For example, the blank-space character (ASCII value 32) was confused with '!' (ASCII value 33) and unit-separator (ASCII value 31). An adversary can perform additional semantic analysis on the reconstructed text to improve its correctness and legibility.

11 OTHER POTENTIAL ATTACKS

The three threats we investigated in this work are by no means an exhaustive set of privacy-invasive threats plausible with connected smart lights. In order to further highlight the privacy-invasive property of modern smart lights, we first outline how media consumption inferences (possible using the attacks presented in this paper) can be used to infer additional sensitive user-information and then briefly outline some other potential privacy threats from these lighting systems.

Indirect Inferences. Inference of personal information such as audio and video playback may not appear as a severe threat on the surface. However, a well informed adversary can potentially derive deeper personal information using these inference attacks. For example, Xu et al. [87] were able to correlate media preferences, personality traits, and political orientation. In that study, liberalism was correlated positively with preference for 'dark/alternative' and 'aesthetic/musical' genres, but negatively with the 'communal/popular' and 'thrilling/action' genres. Therefore, it may be possible to infer a user's personality trait and/or political orientation just by observing the user's media correlated smart light emissions. Negative public reaction to such a privacy invasive correlation was recently seen when Netflix [10], a popular media stream service, was accused of recommending TV episodes based on sexual orientation of the user [11], possibly inferred from their past media consumption habits.

Speech Inference. We analyzed the possibility of speech inference when users talk near the microphone controlling an otherwise idle audio-visualizing smart bulb (i.e., no song is playing). We found that the reaction of light to commonly spoken words is too short to derive meaningful features that can be used to uniquely identify the words (tested using LIFX and Philips Hue bulbs). However, recently we discovered a few tiled smart lighting products, such as the Nanoleaf Aurora [9] and LIFX Tiles [7]. A brief speech experiment with an audio-visualizing Nanoleaf Aurora showed that individual tiles (out of nine tiles connected in series) react differently to a spoken word, likely based on the frequency components, and the resultant pattern on the set of tiles is fairly consistent for the same word. Further analysis and attack modeling is required in order to validate if speech can indeed be reconstructed from such tiled smart lights.

Psychological Inference. Smart lights can be used to illuminate rooms with creative color combinations, often termed as 'scenes'. Based on brightness and color combination chosen by a target user, it may be possible to infer their current psychological state [76]. Moreover, Jacobs et al. [42] showed that exposure to certain primary colors can affect anxiety levels, indicating that RGB smart lights can potentially be misused to alter and/or infer target the user's anxiety levels.

Notification Inference and Correlation. Modern smart lights are heavily integrated with automation platforms (such as IFTTT [5] and Stringify [14]), which enable users to use smart bulbs as a visual notifier for events such

as calls, texts, emails, live scores, etc. A continuously observing adversary can potentially use such notifications to infer many other personal information. For example, score notifications can be used to infer the target user's favorite sports team.

12 DISCUSSION

Practical Considerations and Attack Costs. The practicality of the presented attacks rely heavily on the adversary's ability to (i) procure sensing and computational instruments, (ii) setup a secure observation point in the neighborhood of the target user, and (iii) have a stable and undisrupted view of the target user's bulb/window. In the first step towards carrying out the attacks, the adversary must be able to identify and purchase the sensing and computational instruments. Some of the instruments, such as the telescope and the PC used for running the inference algorithms, are consumer-grade and can be easily procured by an adversary at a reasonable cost ($\approx$ \$130 and $\approx$ \$500, respectively). On the other hand, some of the sensing instruments are specialized devices that are designed and sold by select manufactures. As our goal in this work was to demonstrate the feasibility of these attacks using low-cost sensors and lenses, the Yoctopuce V3 ($\approx$ \$40), Vernier GDX-LC RGB ($\approx$ \$80) and TSOP+ATtiny85 ($\approx$ \$15) are well within the budget of a moderately funded adversary. Moreover, a high-budget adversary may be able to improve the attacks by using high quality sensors (such as high speed digital cameras) and optical lenses (such as wide aperture telescopes). In addition to physical instruments, the adversary also needs digital tools for successfully carrying out the attacks. This includes creating a comprehensive dictionary for the inference attacks (high cost due to media copyrights), and developing and positioning a malicious application for the exfiltration attack (high effort).

Setting up a secure observation point in the neighborhood of the target user is the next step towards carrying out the attacks. This can be challenging for an unacquainted adversary who must first determine the target user's residence. Next, the adversary must identify and secure a covert observation point in the neighborhood, with line of sight to the target user's bulb/window. While it may be difficult for an adversary to maintain covertness in rural areas with fewer structures where he/she can hide, it may be easier to find a secretive observation point in an urban setting. Having a stable and undisrupted view of the target user's bulb/window is also essential for accurate inference and exfiltration attacks. However, several factors can disrupt or impair observation, which the adversary must account for. For example, light from passing by automobiles can introduce temporal noise, moving bodies near the target user's window can change light characteristics (especially affects inference of random hue audio-visualization), and a rainy weather can introduce a high degree of unpredictable noise in the observation channel. We did not evaluate these hindrances in this work, however they need to be considered in any practical implementation of the proposed privacy attacks.

Smartphone-based Attack. Modern smartphones (and many tablets) are being equipped with high precision light sensors for automated display power management. It may be possible to perform our inference attacks by bugging the target user's smartphone with a Trojan application that is capable of capturing and analyzing the on-device light sensor data. Unlike microphone and camera sensors, popular mobile platforms treat light sensors as a zero-permission sensor, which means that an adversary's Trojan application will not require explicit user permission to access the light sensor. This direction of attack needs further investigation, and if successful it can eliminate some of the equipment costs required for the outdoor attack scenario.

Implications and Recommendation to Manufacturers. While the idea of inferring users' media playback from their multimedia-visualizing smart light or exfiltrating data using infrared channel may not be entirely surprising to security experts, there are two major implications of our work. First, our attack frameworks showcase how the attacks are non-trivial, primarily due to the several variable parameters involved. For example, one of the most challenging aspects of inferring songs in random hue audio-visualization was not due to the physical properties of the source audio or light bulb placement, but due to the random order of hue that heavily influences

observed luminance. Second, our exfiltration attack calls attention to the lack of authentication or access control in many popular smart home systems, including smart lighting. In fact, we found that out of 10 most popular Internet-enabled smart light manufacturers, only 1 implements some form of access control (more details in Appendix C). We strongly recommend all smart home appliance manufactures to deploy secure access control mechanisms for their existing and new products. Unsecure smart home devices can become the target of data exfiltration, denial of service and other side-channel attacks, and end users will become the unfortunate victim of such attacks.

Countermeasures. Concerned smart light users can take several measures against the proposed threats. We already saw that windows with low visible transmittance causes the attacks to perform poorly. Therefore, a simple mitigation would be to cover the windows with opaque curtains and block light leakage to the outside. For the inference attacks, the maximum brightness of the bulbs can be reduced, so that the light leakage is also reduced. To prevent the exfiltration attack, strong network rules can be enforced such that unauthorized computers and smartphones cannot control smart bulbs over an IP network.

13 CONCLUSION

We designed and evaluated two attack frameworks to infer users' audio and video playback by a systematic observation and analysis of the multimedia-visualization functionality of modern smart lights. We also designed and evaluated a covert and high bandwidth data exfiltration attack by taking advantage of the infrared capabilities of these smart lighting systems. We conducted a comprehensive evaluation of these attacks in various real-life settings which confirmed the feasibility of proposed privacy threats. These threats also affirm the need for better protection mechanisms, such as mandatory access control in smart light management protocols.

REFERENCES

[1] [n. d.]. Building a LIFX Packet. https://lan.developer.lifx.com/docs/building-a-lifx-packet. Online; accessed 2018-10-30.

[2] [n. d.]. Fiber Optics Market for Telecom & Broadband, Healthcare, Defense, Private Data Networks, and Other Applications: Global Industry Perspective, Comprehensive Analysis and Forecast, 2016-2022. https://www.zionmarketresearch.com/report/fiber-optics-market. Online; accessed 2019-05-10.

[3] [n. d.]. Go Direct Light and Color Sensor User Manual - Vernier. https://www.vernier.com/files/manuals/gdx-lc/gdx-lc.pdf. Online; accessed 2018-10-30.

[4] [n. d.]. hueDynamic. https://www.huedynamic.com. Online; accessed 2018-10-30.

[5] [n. d.]. IFTTT. https://ifttt.com. Online; accessed 2018-10-30.

[6] [n. d.]. LIFX Color A19. https://www.lifx.com/products/lifx-e26. Online; accessed 2018-10-30.

[7] [n. d.]. LIFX Tile Kit. https://www.lifx.com/products/lifx-tile. Online; accessed 2018-10-30.

[8] [n. d.]. Light DJ. https://lightdjapp.com. Online; accessed 2018-10-30.

[9] [n. d.]. Nanoleaf. https://nanoleaf.me. Online; accessed 2018-10-30.

[10] [n. d.]. Netflix. https://www.netflix.com/. Online; accessed 2018-10-30.

[11] [n. d.]. Netflix is experimenting with different episode orders for 'Love, Death & Robots'. https://techcrunch.com/2019/03/19/love-death-robots-experiment. Online; accessed 2019-05-10.

[12] [n. d.]. Phillips Hue. https://www.meethue.com. Online; accessed 2018-10-30.

[13] [n. d.]. Smart home devices and controllers wholesale unit sales in the United States from 2014 to 2017. https://www.statista.com/statistics/495625/smart-home-devices-and-controllers-sales-in-the-us. Online; accessed 2018-10-30.

[14] [n. d.]. Stringify. https://www.stringify.com. Online; accessed 2018-10-30.

[15] [n. d.]. TSOP48 Product Information. http://www.vishay.com/product?docid=82459. Online; accessed 2018-10-30.

[16] [n. d.]. Worldwide intelligent lighting controls market size from 2016 to 2023. https://www.statista.com/statistics/800753/world-smart-lighting-controls-market-revenue. Online; accessed 2018-10-30.

[17] [n. d.]. Zigbee Light Link. http://www.zigbee.org/zigbee-for-developers/applicationstandards/zigbee-light-link. Online; accessed 2018-10-30.

[18] Dakshi Agrawal, Bruce Archambeault, Josyula R Rao, and Pankaj Rohatgi. 2002. The EM Side-channel(s). In *Cryptographic Hardware and Embedded Systems*.

[19] Daniel Arp, Erwin Quiring, Christian Wressnegger, and Konrad Rieck. 2017. Privacy threats through ultrasonic side channels on mobile devices. In *IEEE EuroS&P*.
[20] Dmitri Asonov and Rakesh Agrawal. 2004. Keyboard Acoustic Emanations. In *IEEE S&P*.
[21] N Avlonitis, EM Yeatman, M Jones, and A Hadjifotiou. 2006. Multilevel amplitude shift keying in dispersion uncompensated optical systems. *OptoElectronics* 153, 3 (2006).
[22] Michael Backes, Tongbo Chen, Markus Dürmuth, Hendrik PA Lensch, and Martin Welk. 2009. Tempest in a teapot: Compromising reflections revisited. In *IEEE S&P*.
[23] Michael Backes, Markus Dürmuth, Sebastian Gerling, Manfred Pinkal, and Caroline Sporleder. 2010. Acoustic Side-Channel Attacks on Printers. In *USENIX Security Symposium*.
[24] Michael Backes, Markus Dürmuth, and Dominique Unruh. 2008. Compromising reflections-or-how to read LCD monitors around the corner. In *IEEE S&P*.
[25] Andrea Barisani and Daniele Bianco. 2009. Sniffing Keystrokes with Lasers/Voltmeters. *Black Hat USA* (2009).
[26] Gustavo EAPA Batista, Xiaoyue Wang, and Eamonn J Keogh. 2011. A complexity-invariant distance measure for time series. In *SIAM International Conference on Data Mining*.
[27] Yigael Berger, Avishai Wool, and Arie Yeredor. 2006. Dictionary Attacks using Keyboard Acoustic Emanations. In *ACM CCS*.
[28] Serdar Cabuk, Carla E Brodley, and Clay Shields. 2009. IP covert channel detection. *ACM Transactions on Information and System Security* 12, 4 (2009).
[29] Liang Cai and Hao Chen. 2011. TouchLogger: Inferring Keystrokes on Touch Screen from Smartphone Motion. *HotSec* 11 (2011).
[30] Alberto Compagno, Mauro Conti, Daniele Lain, and Gene Tsudik. 2017. Don't skype & type!: Acoustic eavesdropping in voice-over-ip. In *ACM AsiaCCS*.
[31] Abe Davis, Michael Rubinstein, Neal Wadhwa, Gautham J Mysore, Fredo Durand, and William T Freeman. 2014. The visual microphone: passive recovery of sound from video. *ACM Transactions on Graphics* 33, 4 (2014).
[32] Luke Deshotels. 2014. Inaudible sound as a covert channel in mobile devices. In *USENIX Workshop on Offensive Technologies*.
[33] Irina Diaconita, Andreas Reinhardt, Delphine Christin, and Christoph Rensing. 2015. Inferring smartphone positions based on collecting the environment's response to vibration motor actuation. In *ACM Symposium on QoS and Security for Wireless and Mobile Networks*.
[34] Adrienne Porter Felt, Matthew Finifter, Erika Chin, Steve Hanna, and David Wagner. 2011. A survey of mobile malware in the wild. In *ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*.
[35] Julie Ferrigno and M Hlaváč. 2008. When AES blinks: introducing optical side channel. *IET Information Security* 2, 3 (2008).
[36] Jeffrey Friedman. 1972. Tempest: A Signal Problem. *NSA Cryptologic Spectrum* (1972).
[37] Alan M Gittelsohn. 1969. An occupancy problem. *The American Statistician* 23, 2 (1969).
[38] Mordechai Guri, Dima Bykhovsky, and Yuval Elovici. 2017. aIR-Jumper: Covert Air-Gap Exfiltration/Infiltration via Security Cameras & Infrared (IR). *arXiv preprint arXiv:1709.05742* (2017).
[39] Mordechai Guri, Ofer Hasson, Gabi Kedma, and Yuval Elovici. 2016. An optical covert-channel to leak data through an air-gap. In *IEEE Conference on Privacy, Security and Trust*.
[40] Mordechai Guri, Yosef Solewicz, Andrey Daidakulov, and Yuval Elovici. 2017. Acoustic data exfiltration from speakerless air-gapped computers via covert hard-drive noise ('DiskFiltration'). In *European Symposium on Research in Computer Security*. Springer.
[41] Jun Han, Albert Jin Chung, and Patrick Tague. 2017. Pitchln: eavesdropping via intelligible speech reconstruction using non-acoustic sensor fusion. In *ACM/IEEE International Conference on Information Processing in Sensor Networks*.
[42] Keith W Jacobs and James F Suess. 1975. Effects of four psychological primary colors on anxiety state. *Perceptual and Motor Skills* 41, 1 (1975).
[43] Dorothea Jameson and Leo M Hurvich. 1964. Theory of brightness and color contrast in human vision. *Vision Research* 4, 1 (1964).
[44] Sachin Kadloor, Xun Gong, Negar Kiyavash, Tolga Tezcan, and Nikita Borisov. 2010. Low-cost side channel remote traffic analysis attack in packet networks. In *IEEE International Conference on Communications*.
[45] Cagdas Karatas, Luyang Liu, Hongyu Li, Jian Liu, Yan Wang, Sheng Tan, Jie Yang, Yingying Chen, Marco Gruteser, and Richard Martin. 2016. Leveraging Wearables for Steering and Driver Tracking. In *IEEE INFOCOM*.
[46] Katharina Krombholz, Heidelinde Hobel, Markus Huber, and Edgar Weippl. 2015. Advanced social engineering attacks. *Journal of Information Security and applications* 22 (2015).
[47] Markus G Kuhn and Ross J Anderson. 1998. Soft Tempest: Hidden Data Transmission Using Electromagnetic Emanations. In *Information Hiding, Lecture Notes in Computer Science*.
[48] Longin Jan Latecki, Qiang Wang, Suzan Koknar-Tezel, and Vasileios Megalooikonomou. 2007. Optimal subsequence bijection. In *IEEE International Conference on Data Mining*.
[49] Xiangyu Liu, Zhe Zhou, Wenrui Diao, Zhou Li, and Kehuan Zhang. 2015. When Good Becomes Evil: Keystroke Inference with Smartwatch. In *ACM CCS*.
[50] Joe Loughry and David A Umphress. 2002. Information leakage from optical emanations. *ACM Transactions on Information and System Security* 5, 3 (2002).

[51] Anindya Maiti, Oscar Armbruster, Murtuza Jadliwala, and Jibo He. 2016. Smartwatch-Based Keystroke Inference Attacks and Context-Aware Protection Mechanisms. In *ACM AsiaCCS*.
[52] Anindya Maiti, Ryan Heard, Mohd Sabra, and Murtuza Jadliwala. 2018. Towards Inferring Mechanical Lock Combinations using Wrist-Wearables as a Side-Channel. In *ACM WiSec*.
[53] Anindya Maiti, Murtuza Jadliwala, Jibo He, and Igor Bilogrevic. 2015. (Smart)watch your taps: side-channel keystroke inference attacks using smartwatches. In *ACM International Symposium on Wearable Computers*.
[54] Anindya Maiti, Murtuza Jadliwala, Jibo He, and Igor Bilogrevic. 2018. Side-channel inference attacks on mobile keypads using smartwatches. *IEEE Transactions on Mobile Computing* 17, 9 (2018).
[55] Philip Marquardt, Arunabh Verma, Henry Carter, and Patrick Traynor. 2011. (sp)iPhone: Decoding Vibrations From Nearby Keyboards Using Mobile Phone Accelerometers. In *ACM CCS*.
[56] Nikolay Matyunin, Jakub Szefer, Sebastian Biedermann, and Stefan Katzenbeisser. 2016. Covert channels using mobile device's magnetic field sensors. In *IEEE Asia and South Pacific Design Automation Conference*.
[57] Yan Michalevsky, Dan Boneh, and Gabi Nakibly. 2014. Gyrophone: Recognizing Speech from Gyroscope Signals. In *USENIX Security Symposium*.
[58] Yan Michalevsky, Aaron Schulman, Gunaa Arumugam Veerapandian, Dan Boneh, and Gabi Nakibly. 2015. PowerSpy: Location Tracking Using Mobile Device Power Analysis. In *USENIX Security Symposium*.
[59] Eric Monteiro and Steve Hranilovic. 2014. Design and implementation of color-shift keying for visible light communications. *Journal of Lightwave Technology* 32, 10 (2014).
[60] Meinard Müller. 2007. *Information retrieval for music and motion*. Vol. 2. Springer.
[61] Masashi Nakao and Masahiko Hara. 2001. Voice monitoring system using laser beam. US Patent 6,317,237.
[62] Sashank Narain, Triet D Vo-Huu, Kenneth Block, and Guevara Noubir. 2016. Inferring user routes and locations using zero-permission mobile sensors. In *IEEE S&P*.
[63] Torsten Niederdraenk. 2007. Directional microphone. US Patent 7,245,734.
[64] Ed Novak, Yutao Tang, Zijiang Hao, Qun Li, and Yifan Zhang. 2015. Physical media covert channels on smart mobile devices. In *ACM UbiComp*.
[65] Wallace M Porter and Harry T Enmark. 1987. A system overview of the airborne visible/infrared imaging spectrometer (AVIRIS). In *Imaging Spectroscopy II*, Vol. 834. International Society for Optics and Photonics.
[66] Rohit Puri and Kannan Ramchandran. 1999. Multiple description source coding using forward error correction codes. In *Asilomar Conference on Signals, Systems, and Computers*.
[67] Jean-Jacques Quisquater and David Samyde. 2001. ElectroMagnetic Analysis (EMA): Measures and Countermeasures for Smart Cards. In *Smart Card Programming and Security, Lecture Notes in Computer Science*.
[68] Eyal Ronen and Adi Shamir. 2016. Extended functionality attacks on IoT devices: The case of smart lights. In *IEEE EuroS&P*.
[69] Austin Roorda and David R Williams. 1999. The arrangement of the three cone classes in the living human eye. *Nature* 397, 6719 (1999).
[70] Andrew F Rossi and Michael A Paradiso. 1999. Neural correlates of perceived brightness in the retina, lateral geniculate nucleus, and striate cortex. *Journal of Neuroscience* 19, 14 (1999).
[71] EH Rothauser. 1969. IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics* 17 (1969).
[72] Nirupam Roy, Mahanth Gowda, and Romit Roy Choudhury. 2015. Ripple: Communicating through physical vibration. In *USENIX NSDI*.
[73] Nirupam Roy and Romit Roy Choudhury. 2016. Listening through a Vibration Motor. In *ACM MobiSys*.
[74] Neil Savage. 2015. Visualizing sound. *Commun. ACM* 58, 2 (2015).
[75] SC Magazine UK. 2014. Light-based printer attack overcomes air-gapped computer security. (2014).
[76] Klaus Warner Schaie and Robert Heiss. 1964. Color and Personality. (1964).
[77] Lorenz Schwittmann, Christopher Boelmann, Viktor Matkovic, Matthäus Wander, and Torben Weis. 2016. Identifying TV channels & on-demand videos using ambient light sensors. *Pervasive and Mobile Computing* (2016).
[78] Lorenz Schwittmann, Viktor Matkovic, Torben Weis, et al. 2016. Video recognition using ambient light sensors. In *IEEE PerCom*.
[79] Peter Smulders. 1990. The Threat of Information Theft by Reception of Electromagnetic Radiation from RS-232 Cables. *Computers & Security* 9, 1 (1990).
[80] Wim Van Eck. 1985. Electromagnetic Radiation from Video Display Units: An Eavesdropping Risk? *Computers & Security* 4, 4 (1985).
[81] Martin Vuagnoux and Sylvain Pasini. 2009. Compromising Electromagnetic Emanations of Wired and Wireless Keyboards. In *USENIX Security*.
[82] Chen Wang, Xiaonan Guo, Yan Wang, Yingying Chen, and Bo Liu. 2016. Friend or Foe?: Your Wearable Devices Reveal Your Personal Pin. In *ACM AsiaCCS*.
[83] He Wang, Ted Tsung-Te Lai, and Romit Roy Choudhury. 2015. Mole: Motion leaks through smartwatch sensors. In *ACM MobiCom*.
[84] Teng Wei, Shu Wang, Anfu Zhou, and Xinyu Zhang. 2015. Acoustic eavesdropping through wireless vibrometry. In *ACM MobiCom*.

[85] Martin Wöllmer, Marc Al-Hames, Florian Eyben, Björn Schuller, and Gerhard Rigoll. 2009. A multidimensional dynamic time warping algorithm for efficient multimodal fusion of asynchronous data streams. *Neurocomputing* 73, 1-3 (2009).
[86] Fuqin Xiong. 2005. Amplitude shift keying. *Encyclopedia of RF and Microwave Engineering* (2005).
[87] Xiaowen Xu and Jordan B Peterson. 2017. Differences in media preference mediate the link between personality and political orientation. *Political Psychology* 38, 1 (2017).
[88] Yi Xu, Jan-Michael Frahm, and Fabian Monrose. 2014. Watching the watchers: Automatically inferring tv content from outdoor light effusions. In *ACM CCS*.
[89] Zhi Xu, Kun Bai, and Sencun Zhu. 2012. Taplogger: Inferring user inputs on smartphone touchscreens using on-board motion sensors. In *ACM WiSec*.
[90] Li Zhang, Parth H Pathak, Muchen Wu, Yixin Zhao, and Prasant Mohapatra. 2015. Accelword: Energy efficient hotword detection through accelerometer. In *ACM MobiSys*.
[91] Zheng Zhou, Weiming Zhang, and Nenghai Yu. 2018. IREXF: Data Exfiltration from Air-gapped Networks by Infrared Remote Control Signals. *arXiv preprint arXiv:1801.03218* (2018).

APPENDIX A – POLYNOMIAL INTERPOLATIONS FOR NON-LINEAR COLOR RESPONSES

The observed RGB responses (δ_r, δ_g and δ_b) have non-linear relationships with brightness (Figure 6a and Table 2). Therefore, we employ Lagrange polynomials to learn the non-linearities. We use five observed $\langle \delta_r, \delta_g, \delta_b \rangle$ points for $RGB\langle 1:1:1 \rangle$ in Table 2 to calculate Equation 5, using Lagrange basis polynomials (Equation 4). The functions in Equation 5 are then used in the attack phase (with different distance and amplification factors) to accurately interpolate all observed δ_r, δ_g and δ_b values in a color-profile. A similar set of functions were also learned and applied for attacks on the Phillips Hue bulb. γ_r, γ_g and γ_b remain approximately constant across different amalgamations of primary colors, for both LIFX and Phillips Hue.

$$\gamma_{c\in\{r,g,b\}} = F(\delta_c) = \sum_{j=1}^{n} P_j(\delta_c); \quad where \quad P_j(\delta_c) = c_j \prod_{k=1;k\neq j}^{n} \frac{\delta_c - \delta_{ck}}{\delta_{cj} - \delta_{ck}}. \tag{4}$$

Table 2. Normalized δ_r, δ_g and δ_b for different brightness levels of $RGB(1:1:1)$, using the LIFX bulb.

Composition $\langle r, g, b \rangle$	δ_r	δ_g	δ_b
$\langle 0.2, 0.2, 0.2 \rangle$	0.04052	0.013536	0.058512
$\langle 0.4, 0.4, 0.4 \rangle$	0.149747	0.050023	0.209917
$\langle 0.6, 0.6, 0.6 \rangle$	0.355869	0.115935	0.485289
$\langle 0.8, 0.8, 0.8 \rangle$	0.66887	0.218776	0.906446
$\langle 1.0, 1.0, 1.0 \rangle$	0.741883	0.242022	1

$$\begin{aligned} \gamma_r &= 9.892\delta_r^4 - 8.691\delta_r^3 - 0.203\delta_r^2 + 2.086\delta_r + 0.116; \\ \gamma_g &= 979.1\delta_g^4 - 304.9\delta_g^3 + 7.177\delta_g^2 + 5.883\delta_g + 0.119; \\ \gamma_b &= 3.535\delta_b^4 - 4.651\delta_b^3 + 0.608\delta_b^2 + 1.390\delta_b + 0.117. \end{aligned} \tag{5}$$

APPENDIX B – EXPLORATORY SETUPS

Audio-visualization. In the exploratory setup we used a Yoctopuce V3 (BH1751FVI) luminance sensor for measuring light intensity (Figure 19a). A LIFX A19 Gen3 bulb was used for audio-visualization. A 20 *mm* 12x telephoto lens was used to focus light on the sensor.

(a) (b) (c)

Fig. 19. (a) Yoctopuce V3 (BH1751FVI) luminance sensor; (b) Vernier GDX-LC RGB sensor; (c) TSOP48 infrared sensor soldered on an ATtiny85 Arduino board.

Video-visualization. In the exploratory setup we used a Vernier GDX-LC RGB sensor to measure RGB composition of observed light (Figure 19b). A LIFX A19 Gen3 bulb was used for video-visualization. A 20 *mm* 12x telephoto lens was used to focus light on the sensor.
Infrared Exfiltration. In the exploratory setup we used a TSOP48 infrared sensor soldered on an ATtiny85 Arduino board for measuring 950 *nm* infrared light (Figure 19c). A LIFX A19+ Gen3 bulb was used as a M-ary ASK transmitter, using the infrared spectrum. A 80 *mm* 45-225x telescope was used to focus light on the sensor.

APPENDIX C – STUDY OF POPULAR SMART LIGHTS

We studied the most popular Internet-enabled smart lighting products sold in the USA, and categorized them based on various attributes listed in Table 3. All of these smart lights support multimedia-visualization using official manufacturer's applications, and/or using third-party applications. Except Philips Hue, none of the smart light manufactures implement access control for controlling their bulbs, tiles or strips. While some products have well documented APIs for LAN control, others can also be controlled over LAN by straightforward reverse-engineering of their control packets.

Table 3. Popular smart light manufacturers and their product attributes.

Brand	Product(s)	LAN Control	Hub Required?	Access Control	Multimedia-Visualization
eufy	Bulbs	✓(Unofficial)	X	X	✓
Geeni	Bulbs, Strips	✓(Unofficial)	X	X	✓
LIFX	Bulbs, Tiles, Strips	✓	X	X	✓
LOHAS	Bulbs	✓(Unofficial)	X	X	✓
MagicLight	Bulbs, Strips	✓	X	X	✓
Nanoleaf	Tiles	✓	X	X	✓
Philips Hue	Bulbs, Strips	✓	✓	Button Press on Hub	✓
Sengled	Bulbs	✓(Unofficial)	✓	X	✓
TECKIN	Bulbs	✓(Unofficial)	X	X	✓
TP-Link	Bulbs	✓(Unofficial)	X	X	✓