On Tracing Screen Photos – A Moiré Pattern-Based Approach

Wenyuan Xu, Yushi Cheng, Xiaoyu Ji, *Member, IEEE*, and Yi-Chao Chen, *Member, IEEE*

***Abstract*—Cyber-theft of trade secrets has become a serious business threat. Digital watermarking is a popular technique to help identify the source of the file leakage, whereby a unique watermark for each insider is hidden in sensitive files. However, malicious insiders may use smartphones to photograph the secret file displayed on screens to remove the embedded hidden digital watermarks due to the optical noises introduced during photographing. To identify the leakage source despite such *screen-photo-based leakage attacks*, we leverage Moiré pattern, an optical phenomenon resulted from the optical interaction between electronic screens and cameras. As such, we present `mID`, a new watermark-like technique that can create a carefully crafted Moiré pattern on the photo when it is taken towards the screen. We design patterns that appear to be natural yet can be linked to the identity of the leaker. We implemented `mID` and evaluated it with 7 display devices and 6 smartphones from various manufacturers and models. The results demonstrate that `mID` can achieve an average bit error rate (BER) of 0.2% and can successfully identify an ID with an average accuracy of 98%, with little influence from the type of display devices, cameras, IDs, and ambient lights.**

***Index Terms*—Image forensics, Moiré patterns, screen photos.**

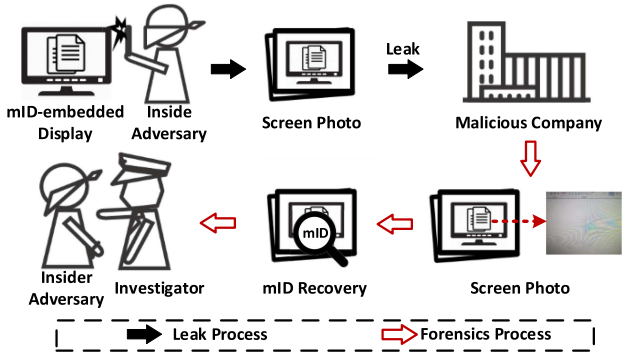


Fig. 1. Illustration of `mID` for screen photo forensics: The identity (ID) of an adversary is embedded on the screen by subtly manipulating what is being displayed and can be recovered later by analyzing the Moiré patterns on the screen photos.

I. Introduction

CYBER-THEFT of trade secrets refers to the illegal leakage of sensitive business information, e.g., digital documents, images, or cyberspaces codes. It is estimated to cause a loss of €60 billion in economic growth along with 289,000 jobs in Europe alone in 2018, and the losses will rise to one million jobs by 2025 [1]. Such cyber-thefts typically occur through insiders [2] legally accessing confidential business files and leaking them to unauthorized parties by emails or social media (e.g., WhatsApp). To identify and trace the source of the leakage, i.e., digital forensics, companies log files outbound from the network interface card or USB ports [3] and insert a digital watermark [4], [5], [6], [7], [8], [9] that is unique to an employee in each confidential file.

To avoid being tracked by digital watermarking, the adversary begins to leak secrets anonymously by photographing (usually with smartphones) the screens which display the confidential information [3]. Hereafter, we name this kind of attack as **screen-photo-based leakage attack**. Unfortunately, due to the noise caused by both the electronic screen and the camera sensors (e.g., the Gaussian and salt-and-pepper noises [10]), the digital watermark may be corrupted and unusable. Therefore, digital forensics for such attacks is in urgent need.

In this paper, we propose `mID`, a digital forensics mechanism against the aforementioned screen-photo-based leakage attack utilizing Moiré patterns [11]. Moiré patterns are optical phenomena generated during the process of photographing screens and are often observed in the photos of computer screens, TV screens, tablet screens, etc. Moiré patterns are natural optical phenomena that are virtually unnoticeable to the adversary, making them the ideal choice for screen photo forensics. As shown in Fig. 1, `mID` works as follows: once an adversary logs into a computer or an application (e.g., an email system) with her account, `mID` will modify the displayed content slightly based on her identity (ID), such that when she takes pictures of the screen, the modification will create Moiré patterns in the photos. Finally, the embedded Moiré patterns are decoded to obtain the ID.

Manuscript received 3 July 2022; revised 18 April 2023; accepted 26 July 2023. Date of publication 31 July 2023; date of current version 11 July 2024. This work was supported in part by the National Natural Science Foundation of China under Grants 62271280, 62222114, 61925109, and 62071428, and in part by Chinese Postdoctoral Science Foundation under Grant BX2021158. *(Corresponding author: Yushi Cheng.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Science and Technology Ethics Committee of Zhejiang University.

Wenyuan Xu and Xiaoyu Ji are with the Ubiquitous System Security Lab (USSLAB), Zhejiang University, Hangzhou, Zhejiang 310027, China (e-mail: wyxu@zju.edu.cn; xji@zju.edu.cn).

Yushi Cheng is with the Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing 100084, China (e-mail: yushicheng@zju.edu.cn).

Yi-Chao Chen is with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: yichao0319@gmail.com).

Digital Object Identifier 10.1109/TDSC.2023.3299983

Photo forensics via Moiré patterns is promising yet challenging since we have to encode IDs inside the Moiré patterns reliably while maintaining them like naturally generated. In this case, encoding IDs through manipulating the phases of images [11], [12], [13], [14] may not work, because it utilizes artificial stripe patterns and will deform the content of the display, like changing a straight line into a wavy one. Furthermore, where to embed Moiré patterns shall be determined by what is being displayed on the screen, and mID has to find the best display areas for encoding such that the generated Moiré patterns remain sneaky.

To overcome the aforementioned challenges, we design the encoding and decoding schemes of mID. The key to encoding is to minimize the impact on the origin display content and to locate the optimal display areas to encode so that the generated Moiré patterns can keep sneaky. Therefore, we first utilize a vertical grating scheme to simulate the natural screen-camera channel. Then, we modify the intensity levels of pixels to generate designed Moiré patterns and exploit the discretized bipolar non-return-to-zero (NRZ) encoding method. Moreover, we consider the fact that humans perceive light and color in a non-linear manner [15] and correct the luminance difference caused by the bipolar NRZ encoding to smoothen the visual effect of the raster images. In addition, mID automatically searches for appropriate display areas to embed the information, thus it maximizes its possibility of being captured in the photos. To reliably decode the ID in the presence of image distortion, we first extract the Moiré areas with image rectification and window scanning. Then, we convert the Moiré areas into the HSV (hue, saturation, value) color space [16], and perform saturation balance and enlargement for high decoding efficiency. Afterward, we utilize k-means clustering with the assistance of check codes to recover the embedded IDs. In summary, our contribution includes below:

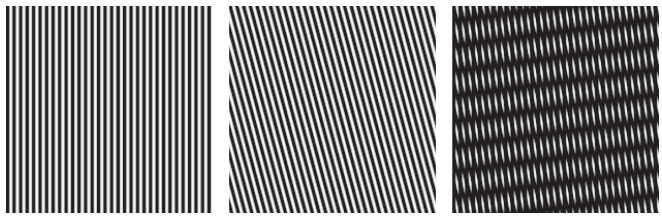
- We propose to exploit the natural Moiré phenomenon existing in the screen-camera channel for screen photo forensics. To the best of our knowledge, this is the first work that addresses screen photo forensics. We believe that mID is a promising technique and can work complementarily to several existing ones.
- We design mID, an effective digital forensics mechanism for file leakages via photos utilizing Moiré patterns.
- We evaluate mID with 7 display devices and 6 smartphones from various manufacturers and models. The results show that mID can achieve an average BER of 0.2% and an average NER (identity number error rate) of 2.0%. In addition, it can operate with little influence from display devices, cameras, IDs, and ambient lights.

II. Background

In this section, we present the background knowledge of Moiré patterns, including the principle and profiling of Moiré pattern, as well as the nonlinearity of the screen-camera channel.

A. Moiré Pattern

Moiré patterns or Moiré fringes refer to the interference patterns that can be produced when an opaque ruled pattern with transparent gaps is overlaid on another similar pattern [17]. Moiré patterns appear in many situations such as looking at two sheets of graph paper twisted 20-30 degrees. In television screen or digital photography, a pattern on an object being photographed may interfere with the shape of the light sensors to generate unwanted artifacts, i.e., Moiré patterns. In this paper, we utilize such an effect for screen photo forensics.



(a) Periodical layer l_1 (b) Periodical layer l_2 (c) Superposition of l_1 and l_2

Fig. 2. Superposition of periodical layers $l_1(x, y) = 0.5 + 0.5\cos(y)$ (a) and $l_2(x, y) = 0.5 + 0.5\cos(y\cos(15°) + x\sin(15°))$ (b) generates new frequency components (c).

B. Moiré Pattern Profiling

The appearance of Moiré patterns is attributed to the superposition of periodic layers [17] and presents a new structure that did not occur in the original layers. The periodic layers are the images or materials with optical periodic patterns, such as raster images, nylon curtains, optical filters, etc. Assume l_1 and l_2 are two periodical layers and s is the generated superposition pattern, where:

$$s(x, y) = l_1(x, y) \times l_2(x, y) \tag{1}$$

The multiplication of two periodic functions leads to nonlinearity in the frequency domain. As illustrated in Fig. 2, l_1 and l_2 are two cosine functions with the frequency of f_1 and f_2 respectively. Then, the generated structure s can be calculated as follows:

$$\begin{aligned} s &= l_1 \times l_2 \\ &= (a_1 + b_1\cos(2\pi f_1 t)) \times (a_2 + b_2\cos(2\pi f_2 t)) \\ &= a_1a_2 + a_1b_2\cos(2\pi f_2 t) + a_2b_1\cos(2\pi f_1 t) \\ &\quad + b_1b_2\cos(2\pi(f_1 + f_2)t) + b_1b_2\cos(2\pi(f_1 - f_2)t) \end{aligned} \tag{2}$$

which contains two new components $(f_1 + f_2)$ and $(f_1 - f_2)$ in the frequency domain. Moiré patterns will be captured by human eyes when the low-frequency component $(f_1 - f_2)$ is below the cutoff of the human visual system (HVS) [18]. It is because the human eyes are more sensitive to low-frequency signals and the component $(f_1 - f_2)$ has a larger amplitude as well.

C. Moiré Pattern of Screen-Camera Channel

Digital cameras often show Moiré phenomenon when photographing digital screens such as TV screens, liquid-crystal displays (LCDs), or tablet screens. The main reason for this phenomenon is that both the digital screen and the Color Filter Array (CFA) on the camera image sensors are periodic layers.

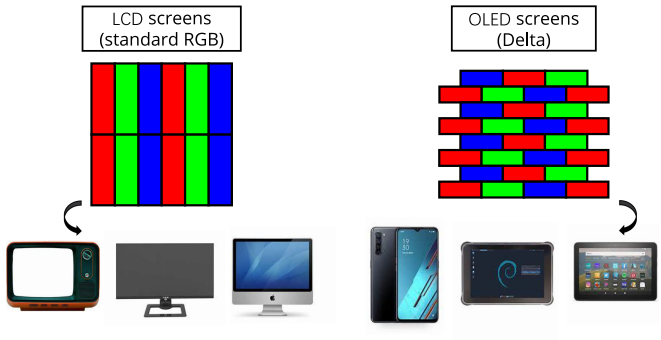


Fig. 3. Illustrations of LCDs and OLED screens.

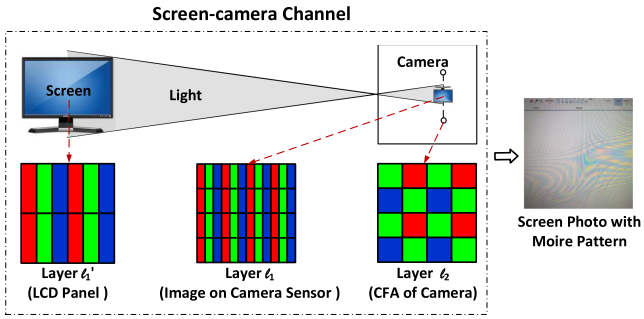


Fig. 4. Illustration of the imaging process of the screen (LCD)-camera (CFA) channel and the resulted screen photo with Moiré patterns.

The nonlinearity will arise when they are superimposed. We call this process the screen-camera channel, which is shown in Fig. 4.

Screen Image: Different devices may use different types of digital screens including LCD, OLED (Organic Light-Emitting Diode), etc. For example, computer monitors and televisions generally use LCD screens, while handheld devices like smartphones or tablets often use OLED screens. In addition, different types of screens have different pixel layouts. As shown in Fig. 3, the standard RGB horizontal layout is generally used in LCD screens, while other pixel layouts such as Delta, and Almond are more common in OLED screens. In the following, we demonstrate the screen image using an LCD screen as an instance.

The LCD screens consist of numerous small unit structures, each containing a red(R), green(G), and blue(B) filter for emitting different colors of light, which is shown on the left of Fig. 4. The unit structures of the LCD screen are organized periodically. When a camera photographs the LCD screen, it will be formed as a layer of spatial pattern, which is called the image of the LCD screen. We denote the image of the LCD screen projected on the camera sensors as layer l_1 with a frequency of f_1, which interacts with the CFA directly to generate Moiré patterns. To differentiate, we denote the layer formed by the original LCD screen as layer l_1' with a frequency of f_1'. Note that other displays such as LED screens are also applicable.

CFA: In the screen-camera channel, the image sensor is used to receive light from the screen and a CFA placed in front of the image sensor is employed to get the color information of light. Bayer filter is widely used on the built-in camera of smartphones [19]. As shown in Fig. 4, it is arrayed periodically in a 2×2 pattern and provides light information for the three color channels (RGB). As a result, the CFA forms another layer of spatial patterns, which we denote as layer l_2 with a frequency of f_2.

Nonlinear Optical Interaction: Since both l_1 (image of the screen) and l_2 (CFA of the camera) are periodic patterns, their superposition can generate new low frequency components according to the Moiré pattern profiling. When the position (distance and angle) of the camera and screen are properly adjusted to make the component $(f_1 - f_2)$ fall within the observable frequency range, ripple-like Moiré patterns can be captured in screen photos (shown in Fig. 4).

Inspired by the natural Moiré phenomenon existing in the screen-camera channel, we propose well-designed camouflaging periodical patterns presented on the screen, which can be used to nonlinear optically interact with the camera's CFA to embed a unique Moiré-pattern-based ID, i.e., `mID`, to trace the source of leakage.

III. Threat Model

For screen-photo-based leakage attacks, the goal of attackers is to leak secret information by photographing with smartphones. The photos can be transmitted to unauthorized parties via web applications (e.g., WhatsApp) or portable disks. In this attack scenario, we have the following assumptions about the company requiring file forensics and the adversary. For the companies, we assume that they have full control over confidential files. That is, they can modify the hardware and software configuration of the display device and the file system. For the adversary, we have the following assumptions:

- *Screen-capturing with Smartphones:* For the stealth of screen-photo-based leakage attacks, the attacker prefers to use her smartphone to photograph the screen which is displaying confidential information. At the same time, the attacker tries to choose the right angle and distance to obtain a high-quality photo for completely and clearly confidential information.
- *Untraceability over Internet.* The attacker's transmission path cannot be traced, i.e., the attacker can use a public network (e.g., public Wi-Fi or 5G) to anonymously deliver the screen photos to unauthorized parties.
- *Photo Processing:* We assume that the adversary may perform a variety of image processing methods to prevent the screen photo from being tracked, such as photo duplication, photo compression, image up/downscaling, format conversion, image cut, etc.

IV. Design

A. Design Requirement

To trace the source of file leakages via Moiré patterns, `mID` shall satisfy the following requirements.

Subtle Visual Difference to User: For the sake of user experience, the embedded `mID` should not interfere excessively with

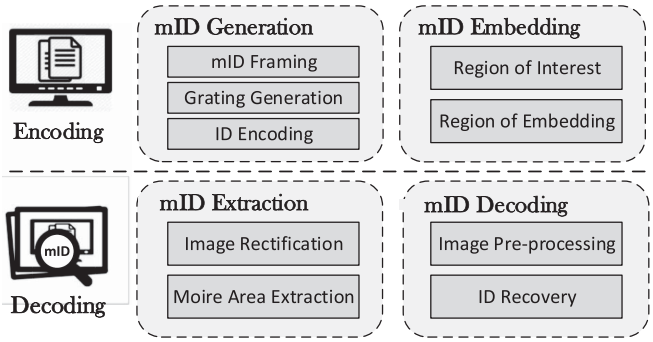


Fig. 5. System overview of mID scheme.

the visual system of the human eye. In other words, the human eye should not be able to distinguish the modifications made to the display by mID.

Vision Insensitivity to Adversary: If an attacker notices the existence of mID, she may discard the captured screen photos to prevent them from being traced. Therefore, the crafted Moiré pattern should be similar to the natural Moiré pattern produced in the screen-camera channel.

B. Overview of mID

The basic idea of mID is using different intensity levels to embed ID numbers into the superimposed Moiré pattern. The generating scheme includes mID encoding and decoding phases with four modules: (a) mID generation, (b) mID embedding, (c) mID extraction, and (d) mID decoding, as shown in Fig. 5.

Encoding Phase: In such phase, it consists two modules: the mID Generation module and the mID Embedding module. The mID Generation module is used to generate a modification pattern based on the ID number on the original display. It contains three components: (a) mID Framing that forms a proper frame, (b) Grating Generation that helps to create Moiré patterns, and (c) ID Encoding that adds the information of IDs to the Moiré patterns. In other words, designing grating is similar to finding the carrier signals and the ID encoding is similar to finding the modulation scheme in communication. After generating mID, the mID Embedding module is used to find the best application area. The aim of such a module is to achieve natural-looking Moiré patterns which can be captured by the camera but not observed by the user.

To generate the Moiré pattern, the pixels on the screen can be manipulated to form a display grating, which can create a periodical stripe pattern. Meanwhile, due to the vertical grate design on the LCD panel, generating the display grating to be vertical can make the Moiré pattern appear more natural. Second, we propose a discretized bipolar non-return-to-zero encoding method that manipulates the intensity levels of the generated Moiré pattern to represent the information. It allows mID to be encoded into the display raster more stealthily and thus unnoticeable to the users. Further, taking account of the fact that humans perceive light and color in a non-linear manner, we correct for differences in luminance caused by discretized encoding to ensure visual uniformity. Third, to embed the generated gratings into the screen and maximize their possibility of being captured in the photos, we automatically analyze the current page of the screen and search for suitable regions for embedding.

Decoding Phase: In this phase, the mID Extraction module first utilizes image rectification to remove the camera distortion, and extracts the Moiré areas of the screen photos by window scanning. Then, the mID Decoding module converts the Moiré areas into the HSV (hue, saturation, value) color space, performs saturation balance and enlargement and finally recovers mID via k-means clustering with the assistance of check codes.

C. mID Generation

1) mID Framing: To label the unique information source via Moiré-pattern-based ID, we design an N-bit digital identifier, called mID. It consists of (1) a 2-bit front check code, (2) a payload, and (3) a 2-bit end check code. The payload represents the identity of the information source and appears as a sequence of binary digits (bits), each having either the value "0" or "1". We envision it can provide photo forensics from three levels.

- *Device level:* When the device is tightly bound to the user, e.g., the devices can only be accessed by the owners, the payload can be generated based on the hardware information of the displayed device, such as the MAC (Media Access Control) address.
- *Operating system (OS) level:* When multiple users share the same device but use their own OS accounts, the payload can be generated at the OS level based on the OS user account information.
- *Application level:* For sensitive applications, e.g., the internal mail system or the database of companies, the payload can be generated based on the account information associated with the application.

We use a two-digit segment "01" as a two-end check code which is set before or after the payload of mID. Namely, a 14-bit mID with front and end check codes are presented as 01XXXXX...01. Setting such check codes can provide twofold benefits in addition to ease of decoding. First, the check code can help restore the exposure-imbalanced images and can thus improve the decoding accuracy. Second, it provides a baseline for the k-means clustering to determine which cluster maps to bit "0" or "1", as we will reveal in detail in Section IV-F.

2) Display Grating Generation: As mentioned in Section II, l_1' (the screen pixels layer) is projected onto the camera sensors to form layer l_1, and the CFA of the camera forms layer l_2, and their superposition generates mIDs. In fact, we can only manipulate l_1' among the three layers (l_1', l_1, l_2). It is because l_2 (the CFA layer) is determined by the physical structure of the smartphone's built-in camera, while l_1 (the projected screen display layer) also receives influence from the camera. Recall that a periodical grating layer can be modeled with a frequency and a phase term:

$$l(x, y) = p(\phi(x, y)) \tag{3}$$

where $l(x, y)$ represents the pixel value at the coordinate (x, y), $p(\cdot)$ is a periodic function that determines the frequency of

the grating, and $\phi(x, y)$ is a phase function that determines its geometric layout, as shown in Fig. 2. We explain how to select appropriate periodic and phase functions for layer l_1' to generate mIDs.

Periodic Function Selection: Due to the long photographing distance and the pinhole effect of cameras, layer l_1 has an increased frequency compared with that of layer l_1'. According to the Pinhole Camera Theory [20], the object size projected onto a camera sensor is inversely proportional to the distance between the object and the camera sensor:

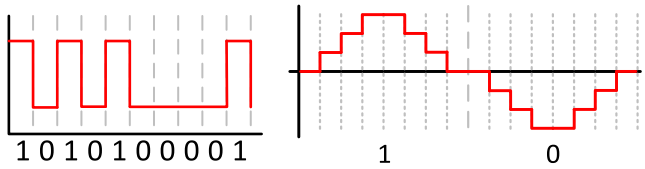
$$S_{cam} = \frac{S_{obj} \times L_f}{D} \tag{4}$$

where S_{cam} and S_{obj} are the photographed and actual sizes of the object respectively, L_f is the focal length of the camera, and D is the distance between the camera and the object. Due to that the photographing distance D is usually much larger than the focal length L_f, the size of layer l_1' shrinks to $\frac{L_f}{D}$ per unit area, which gives $f_1 = \frac{D}{L_f} \cdot f_1'$. As a result, the frequency of the generated Moiré patterns can be given as $\frac{D}{L_f} \cdot f_1' - f_2$. As the camera focal length L_f and the CFA frequency f_2 are fixed by the photographing device, for a specific device, the Moiré patterns are mainly determined by the photographing distance and the frequency of the generated grating.

As the adversaries' goal is to capture a complete and clear screen photo to extract the contents, the photographing distance D used by adversaries shall be within a range. For a 24" LCD display commonly seen on the market, the photographing distance D is usually larger than 60 cm for various smartphones while for a 12" OLED tablet, the photographing distance D is usually larger than 32 cm for various smartphones, as calculated in Section VIII. To improve the chances of the generated Moiré patterns to be captured by cameras, the frequency of the periodic function $p(\cdot)$ shall match the photographing distance. Further, the generated stripes should be as small as possible, since thinner stripes are more likely to give a uniform color than wider stripes. Thus, we set the frequency of $p(\cdot)$ to be 2 pixels, which is shown to be effective in Section VI.

Phase Function Selection: While the periodic term affects the density of the grating, the phase function determines its geometric layout and thus the Moiré patterns. Since LCD screens are most commonly used in various digital displays, we design the phase function of mID in a way that works well on LCD screens. As shown in Fig. 4, the unit structures of the same color in LCD panels are usually vertically arranged, i.e., vertical gratings of red, green, and blue are naturally present in the LCD screen. Thus, we design mID that imitates the Moiré patterns that are generated naturally by the screen-camera channel and achieve vision insensitivity to the adversary. Specifically, with the selected frequency, we generate a binary display grating for each bit of mID in the form of vertical stripes, as given below:

$$\begin{aligned} l_1'(x, y) &= p_1(\phi_1(x, y)) \\ p_1(u) &= 0.5 + 0.5\cos(\pi u) \\ \phi_1(x, y) &= y \ mod \ 2 \end{aligned} \tag{5}$$



(a) Unipolar NRZ coding for binary sequence "101010001". (b) Discretized bipolar NRZ coding for binary sequence "10".

Fig. 6. Improved discretized bipolar NRZ coding ensures a flat edge between bits "0" and "1".

As the generated mID display grating has a frequency of 2 pixels while the digital screen structure has a frequency of 1 pixel, the mID-related and the natural Moiré patterns appear at different distances and will not interfere with each other.

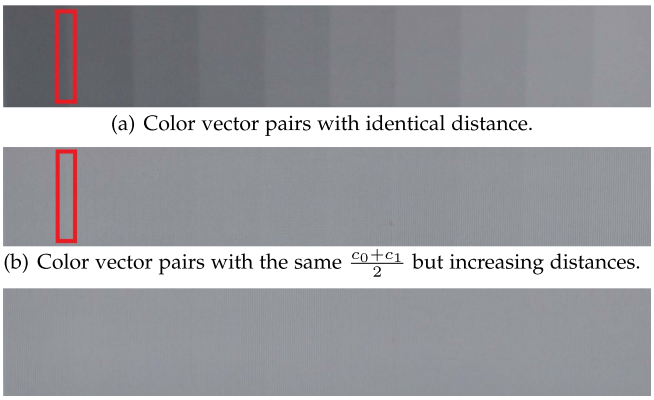
3) Intensity-Based ID Encoding: Existing work [11], [12], [13], [14] typically hide information in the Moiré patterns by manipulating the phase of one of the gratings, e.g., two secret images show no obvious patterns when observed separately, but reveal hidden information when overlapped. However, manipulating the phase may result in the bending strips which probably allows adversaries to observe them. Meanwhile, phase-based methods tend to produce significant Moiré patterns, which may alert the adversaries and is unacceptable.

Intensity of Moiré Pattern: To address it, we modify the intensity of Moiré patterns. Essentially, the mID-related moire pattern is generated by the new frequency components from overlapping the designed display grating and the camera CFA. Since the CFA layers are determined by the camera, the intensity of the mID-related Moiré patterns depends on the intensity of the display grating. Therefore, we can manipulate the intensity of the Moiré pattern by changing the pixel values of two adjacent grating stripes in RGB color space (i.e., the contrast of the stripes). Such an observation is also validated by our experiments. As the generated grating has a spatial frequency of 2, we can denote the even column in the generated grating as $c_0 = (r_0, g_0, b_0)$, and the odd one as $c_1 = (r_1, g_1, b_1)$. With Equ. 6, $c_0 = (255, 255, 255)$ and $c_1 = (0, 0, 0)$ generate the most intensive Moiré patterns. Denote the color distance between two adjacent stripes, or in other words, a pair of color vectors $\{c_0, c_1\}$, as their l^2-norm in the RGB space:

$$C_d = ||\{c_0, c_1\}||_2 = \sqrt{(r_0 - r_1)^2 + (g_0 - g_1)^2 + (b_0 - b_1)^2} \tag{6}$$

A larger C_d represents a larger contrast between two adjacent stripes and thus represents a more significant stripe grating, which results in more intensive Moiré patterns. When C_d decreases to zero, i.e., the even and odd columns are identical, the generated grating loses its periodicity and thus no Moiré patterns will be observed.

Based on it, we propose to embed identity numbers into the generated Moiré pattern by its intensity levels. Intuitively, we can utilize the high-intensity level to represent bit "1", and the low-intensity level to represent bit "0", which is also known as the unipolar non-return-to-zero (NRZ) code [21], as shown in Fig. 6(a). However, there may be a discontinuity between bit '0'



(a) Color vector pairs with identical distance.

(b) Color vector pairs with the same $\frac{c_0+c_1}{2}$ but increasing distances.

(c) Color vector pairs with luminance correction and increasing distances.

Fig. 7. Illustration of ten intensity levels for encoding using three methods, and the ones created by the proposed luminance correction scheme (c) show almost no visual difference and can embed `mID` without being noticed by adversaries.

and bit '1', i.e., if they are encoded adjacent to each other, it may make them less stealthy and more susceptible to suspicion by adversaries.

Discretized Bipolar Non-Return-to-Zero Encoding: To alleviate the problem of discontinuity, we discretize both the high and low levels to make the possible junction smooth, which we call the discretized bipolar NRZ encoding. As shown in Fig. 6(b), we discretize the high (low) intensity level into k sub-levels with each sub-level consisting of n grating columns to approximate a cosine function, for the sake of being flat at the edge of a bit. Another benefit of such an implementation is that bipolar encoding increases C_d between bit "0" and bit "1" compared with the unipolar one, which may ease the difficulty of decoding.

Nonlinearity of Color Perception: In the discretized bipolar NRZ encoding, each intensity level is represented with one pair of color vectors $\{c_0, c_1\}$ in the RGB space. Since bit "0" and bit "1" share the same baseline, the encoding requires $(2k-1)$ intensity levels in total, i.e., $(2k-1)$ pairs of color vectors with increased color distances.

Since human eyes utilize the average of contiguous objects for perception, we generate the required color vectors based on the visual average effect of the human visual system (HVS) [18], which is widely used in many images scaling methods [22]. Thus, we attempt to generate various color vectors for different intensity levels while keeping their average RGB vectors $\frac{c_0+c_1}{2}$ the same, which we assume may be possible to ensure the homogeneity of the generated grating image.

We take the mid-gray value, i.e., (128,128,128), as the background color for an instance, then generate k pairs of vectors with increased color distances such as:

$$\begin{aligned} &Level_i : \{c_0, c_1\}_i = \{(128+5i, 128+5i, 128+5i), \\ &(128-5i, 128-5i, 128-5i)\}, i \in \{1, 2, \ldots, k\} \end{aligned} \tag{7}$$

Compared with the naive color vector pairs with an identical distance shown in Fig. 7(a), the proposed ones, i.e., color vectors with the same $\frac{c_0+c_1}{2}$ but increasing distances, exhibit much fewer visual differences as shown in Fig. 7(b). However, we found that simply producing an even grating image was not enough. It is because of the Gamma Correction [23] that modern displays typically employ, thus adjusting the distance of $\{c_0, c_1\}$ may change the luminance perceived by the human eyes.

As the human visual system is non-linear, i.e., more sensitive to relative differences between darker tones than between lighter ones. We utilize gamma encoding to optimize the usage of bits when encoding the image, or bandwidth used to transmit an image [24]. Accordingly, modern display devices are gamma correction to reveal the real color. Both gamma encoding and gamma correction follow a pow-law expression [25]:

$$V_{out} = AV_{in}^{\gamma} \tag{8}$$

where the input value V_{in} is multiplied by the constant A and powered by the gamma value γ to get the output value V_{out}, with $\gamma < 1$ for encoding and $\gamma > 1$ for correction (decoding). As a result, the generated color vector (in RGB color space) is expanded before display and the luminance perceived by human eyes is not the arithmetic mean $\frac{c_0+c_1}{2}$ as supposed.

Luminance Correction: To achieve further stealthiness of the encoded pattern, we propose a luminance correction algorithm based on gamma correction and the non-uniformity color perception of HVS. Specifically, we model the average luminance Y of an RGB vector pair $\{c_0, c_1\}$ by removing the gamma compression, which transforms the image to a linear RGB color space as follows:

$$Y\{c_0, c_1\} = w_r(r_0^{\gamma} + r_1^{\gamma}) + w_g(g_0^{\gamma} + g_1^{\gamma}) + w_b(b_0^{\gamma} + b_1^{\gamma}) \tag{9}$$

where $\gamma = 2.2$ for most modern display devices [25]. w_r, w_g and w_b are the weights of the RGB channels respectively, which represent the intensity (luminance) perception of typical humans to lights of primary colors. Given that human vision is most sensitive to green and least sensitive to blue, w_g has the largest value of 0.7152 and w_b has the smallest value of 0.0722, with $w_r = 0.2126$ [26].

With luminance correction, we can generate RGB vector pairs with even luminance by optimizing the following equations:

$$\begin{aligned} &E = |Y\{c_0, c_1\} - Y_{bg}| \\ &Y_{bg} = w_r r_{bg}^{\gamma} + w_g g_{bg}^{\gamma} + w_b b_{bg}^{\gamma} \\ &\textbf{max} \quad C_d = ||\{c_0, c_1\}||_2 \\ &\quad s.t. \quad E < \varepsilon \\ &\quad s.t. \quad r_i, g_i, b_i \in \mathbb{Z} \cap [0, 255], i = 0, 1 \end{aligned} \tag{10}$$

We utilize the global search algorithm to solve the above optimization problem. However, as we can see, the solution to the formula is not unique and the number of searched vector pairs is determined by the error threshold ε. A larger ε contributes to more RGB vector pairs at the cost of less evenness of the generated grating image. Thus, ε can be determined upon the requirement of k, or in other words, the number of RGB vector pairs needed to implement the discretized bipolar NRZ encoding. After luminance correction, the generated grating is

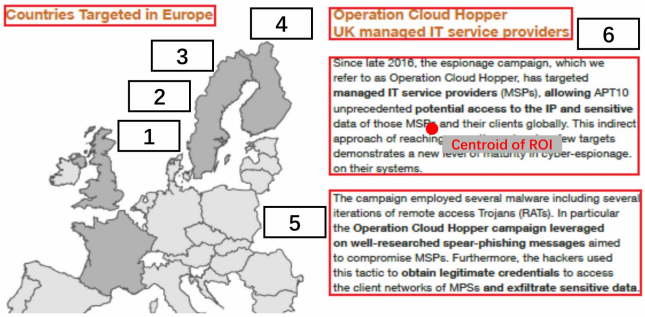


Fig. 8. Illustration of ROI (red box) and ROE (black box) of the current page [1]. The red dot is the center of ROI.

almost invisible even with increasing color distances, as shown in Fig. 7(c).

In summary, we propose a method based on the discretized bipolar NRZ encoding to embed identity numbers in the grating image so that it can be captured by the camera as a Moiré pattern and then employ luminance correction to ensure that it produces an even grating image that is guaranteed to be stealthy.

D. mID Embedding

In the mID Embedding module, we automatically analyze the current page of the screen to find the best position to embed the generated gratings and make it most possible for it to be captured in photos.

Region of Interest: Since our goal is to prevent the illegal theft of confidential information, some regions of the current page that contain confidential information such as text or images are most important to us, i.e., the region of interest (ROI). In order to find suitable regions for embedding mID, we first use computer vision (CV) techniques [27], [28], [29] to extract the position of text and images as possible ROIs, as shown in Figure Fig. 8. For each ROI, we calculate the centroid of these regions as the center of ROI for the current page. Alternatively, the defenders can manually mark the ROIs according to their demands.

Region of Embedding: To maximize the possibility that mID is captured in the screen photos, we embed the generated gratings in the vicinity of the ROI center, i.e., regions of embedding (ROEs). In general, the ROE needs to be as close to the center of the ROI as possible, and also needs to be embedded in the flat regions since (1) mID near the center of the ROI is more likely to be captured in the screen photos; (2) embedding mID in flat regions allows less detail to be lost in the page and less visual discrepancy to the user. In addition, we design to embed one bit of mID in each ROE. Since the entire mID may require a large flat region, it may be limited in the page, embedding the mID in several ROEs can help to reduce the size requirements of the ROE.

Therefore, we search for N rectangular regions close to the ROI center, where N is the number of bits of mID. Each embedding region has a size of $p \times q$, where p and q represent the height and width of a 1-bit grating, respectively. The width q can be further calculated as $q = 2k \times n$. The height p can be any value theoretically but a minimum one is required to ensure the distinguishability of Moiré patterns in the screen photos. In practice, we suggest that $p > 50$. Note that the embedding region can be any shape. We employ rectangle here for ease of encoding and decoding.

We utilize a sliding window with a size of $p \times q$ and a step of w_m to scan through the current page for ROE searching. For each image window $B(x, y)$ with (x, y) as the centroid coordinate, we evaluate its fitness $F(x, y)$ in consideration of both evenness and location:

$$D(x,y) = \frac{1}{\sum_{ch=\{r,g,b\}} \sigma(ch[x-\frac{p}{2}:x+\frac{p}{2}, y-\frac{q}{2}:y+\frac{q}{2}])}$$

$$L(x,y) = \frac{1}{abs(\frac{x}{h_B} - C_x) + abs(\frac{y}{w_B} - C_y)}$$

$$F(x,y) = w_D \cdot D(x,y) + w_L \cdot L(x,y) \tag{11}$$

where $\sigma(ch)$ refers to the standard deviation of channel $ch = \{r, g, b\}$ of the current page. h_B and w_B are the height and width of the current page, (C_x, C_y) is the centroid coordinate of ROI, and w_D and w_L are the weights of the deviation $D(x, y)$ and location $L(x, y)$ functions, respectively. In our implementation, we set $w_D = w_L = 0.5$.

We select the top N image windows in the descending order of fitness ranking as candidate ROEs. Subsequently, we sort them in ascending order by horizontal coordinates.[1] As thus, we obtain N image windows in the horizontal direction. For the obtained regions, we embed the corresponding mID bits by replacing the pixels of the original page with that of the generated gratings. Thus, we embed the generated mID gratings into the current page of the screen without causing a significant obvious visual impact on users.

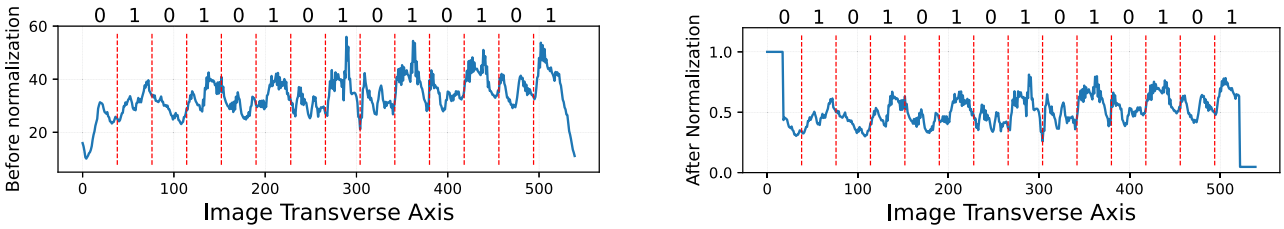
E. mID Extraction

The screen photos captured by smartphones include not only Moiré patterns but also other distracting factors. Therefore, we need to first locate the areas of the Moiré patterns in the smartphone-captured screen photos, i.e., the Moiré areas.

Image Rectification: Since the camera is not usually exactly parallel to the screen when photographing, the difference in angle between them will cause the screen photos to suffer from geometric distortion, i.e., the screen photos appear as a distorted quadrilateral rather than a rectangle. To address it, we first rectify the distorted image with the commonly-used projection transformation under the homogeneous coordinates [30], [31], and then extract the rectified rectangle that contains the screen for further Moiré area extraction.

Moiré Area Extraction: Moiré patterns tend to appear as red and green stripes, thus searching for red and green stripes is an intuitive method. However, the Moiré patterns may appear as various colors on different backgrounds and the screen-camera channel may bring noise either, so simply color searching is not sufficient enough. Therefore, we turn to the transverse coding style we employ for mID encoding, because of which the Moiré

[1] Sort by vertical coordinates if equal horizontal coordinates.



(a) Saturation distribution before normalization.

(b) Saturation distribution after normalization.

Fig. 9. After pre-processing, the saturation differences between bits "0" and "1" are enlarged.

area is likely to have larger color variations in the horizontal direction compared to the vertical one.

To extract the Moiré areas with robustness, we use a 2-dimension (2D) window W_m with a size of $h_m \times w_m$ and a step of t_m to scan through the rectified rectangle image. Specifically, we calculate the average color variation Var_h and Var_v in both the horizontal and vertical directions, and determine whether the current window belongs to the Moiré area with the following in-equation:

$$Var_v > r \cdot Var_h \tag{12}$$

where r is the ratio threshold and the window with significantly larger horizontal variation will be regarded as a part of the Moiré area. To achieve high extraction precision, the window size and step are usually supposed to be in fine granularity. In practice, we set $h_m = w_m = s_m = 10\,pixels$, and $r = 1.5$. After scanning, we obtain a number of Moiré windows in several clusters with possibly a few outliers. The number of clusters, i.e., the number of Moiré areas contained in the photo, is usually less than or equal to the number of mID bits N since two adjacent embedding regions appear as one Moiré area in the photos. To locate the Moiré areas, we first cluster those Moiré windows with mean shift clustering [32], which obtains the center of each Moiré area roughly. Then, we utilize Random Sample Consensus (RANSAC) [33] to discriminate outliers and search for the minimum rectangle that contains the rest of clustered Moiré windows for each Moiré area. We gradually iterate their boundaries until convergent, with which we extract the Moiré areas for further mID decoding.

F. mID Decoding

After extracting the Moiré areas, we perform mID decoding to recover the embedded mID. To ease burden of decoding, we arrange and connect the obtained Moiré areas together according to their horizontal coordinates. In this way, we obtain a joint Moiré area (JMA) for decoding.

1) Image Pre-Processing: The first set of decoding procedures is image pre-processing that includes (1) Color Space Transformation that makes the decoding algorithm robust across different colors, and (2) Saturation Normalization that makes saturation data that is not comparable become comparable.

Color Space Transformation: The colors of the mID-related Moiré patterns depend on the screen backgrounds. For instance, a white background will produce Moiré patterns with red and green stripes. To make the decoding algorithm robust across different RGB colors, we transform the joint Moiré area into the HSV (hue, saturation, value) color space [16]. Specifically, as we utilize the Moiré pattern intensity to encode bits and high intensity results in high color saturation, we perform mID decoding in the saturation dimension.

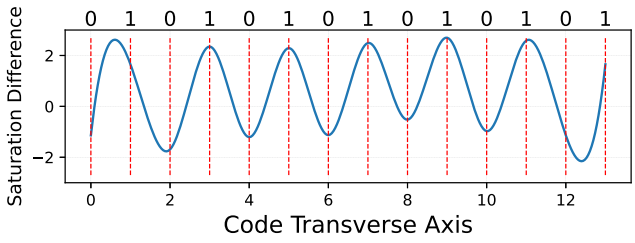


Fig. 10. Illustration of the saturation difference curve for mID.

Saturation Normalization: When taking a picture towards a screen, people tend to focus on its center to capture the whole screen. As a result, the saturation distribution in the Moiré areas may differ in the horizontal direction, as shown in Fig. 9(a). To reduce the impact of focus position, we normalize the saturation of the joint Moiré area.

Specifically, we focus on the horizontal saturation normalization since we encode mID in a transverse way and thus the horizontal saturation has a larger impact on decoding compared to that in the vertical direction. We use the z-score normalization since the maximum and minimum values of saturation are unknown:

$$x^\star = \frac{x - \bar{x}}{\sigma} \tag{13}$$

where $\bar{x}$ is the mean of the saturation, σ is the standard deviation of the saturation. After normalization, the saturation data has a mean of 0 and a standard deviation of 1. For original images, saturation normalization can enlarge the difference between bit "0" and "1" as shown in Fig. 9(b), and thus can improve the decoding accuracy.

2) ID Recovery: After image pre-processing, we recover IDs via k-means clustering with the assistance of check codes.

Saturation Difference Curving: With the normalized joint Moiré area, we first calculate the histogram of each column and obtain a $1 \times W$ saturation matrix, where W is the length

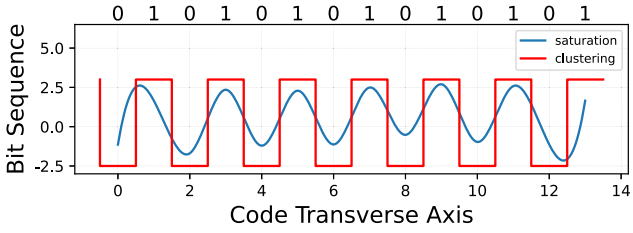


Fig. 11. Illustration of mID recovered by k-means clustering.

Algorithm 1: mID Decoding.

Input:
- $M = \{H, S, V\}$: extracted joint Moiré area
- N: number of bits of mID.
- Z: number of vertical pixels of one bit.

Output: B: decoded bit sequence

$W \leftarrow$ WIDTH(M) // get the width of the joint Moiré area
$S'(y) = normalization(\sum_x S'(x, y))$ // get the normalized saturation matrix
$S'(y) = hanning(S'(y))$ // noise suppress
for $i \in [0, N-1]$ **do**
  $S_1 = \sum S'(\frac{1}{4}\cdot\mathrm{p}: \frac{3}{4}\cdot\mathrm{p})$
  $S_2 = \sum S'(0 : \frac{1}{4}\cdot\mathrm{p}) \sum S'(\frac{3}{4}\cdot\mathrm{p}: \cdot\mathrm{p})$
  $S_d = \frac{S_1 - S_2 - S_3}{p}$
$B \leftarrow$ K-MEANS(S_d) // k-means clustering
$f \leftarrow$ CHECK_CODE_MATCHING(B)
if $f == True$: **then**
  $B = \sim B$

of the joint Moiré area. It is based on the transverse encoding we employ, which means that pixels of the same column are supposed to be identical. We then utilize a Hanning window to reduce the noise introduced during photographing and improve the SNR (signal-to-noise ratio). After that, considering that the focus position may make a negative impact on decoding accuracy, we balance the saturation with the check code. The 1st and 2nd splits correspond to the bit "0" and bit "1" of the front check code, and the $(N-1)$th and Nth splits correspond to that of the end check code. We make the saturation of the two edges equal, thus flattening the saturation distribution. Finally, we subtract the saturation values on both sides from the value in the middle to calculate the saturation difference within each bit and obtain a $1 \times N$ saturation matrix, where N is the number of encoded bits. In this way, we obtain a horizontal saturation difference curve for further decoding, as shown in Fig. 10.

Bit Clustering: For an N-bit mID, we further divide its saturation difference curve into N splits, denoted as $\{P_0, P_1, \ldots, P_{N-1}\}$. We decode the embedded ID through the intensity differences of the Moiré patterns. For bit"0", the saturation of its middle position is lower than that of its sides. For bit "1", the saturation of its middle position is higher than that of its sides. After that, we employ k-means clustering [34] to group the same bit into the same class and utilize the check codes to identify each class, i.e., bit "0" or bit "1", as shown in Fig. 11. In this way, we recover mIDs from screen photos and the whole decoding process is shown in Algorithm 1.

V. Implementation

We implement the mID scheme at both the OS and application level in Windows operating systems and it runs as a background application or script after a user logs in. For the OS level, mID employs the entire screen as the display window and creates a rendering context using the Windows API functions GetDC() and wglCreateContext. For the application level, mID employs the application window as the display window and uses its own rendering context. Then, mID captures the current page of the screen or application in real-time using the function glReadPixels() under the OpenGL (Open Graphics Library) framework [35]. After that, it searches for the ROI and ROE with the methods proposed in the mID Embedding module. With the obtained ROE, mID replaces the pixels of ROE with the gratings generated by the mID Generation module, passes the new mID-embedded screen (application) frame to the function glBufferData(), and finally renders it on the display.

VI. Evaluation

In this section, we evaluate the performance of the mID scheme. We conduct experiments under various settings and collect over 5000 photos with 7 display devices and 6 smartphones over 3 months. In particular, we evaluate the impact of (1) IDs, (2) display devices, (3) capturing devices, (4) ambient lights, (5) shooting distances, and (6) shooting angles with the metrics of bit error rate (BER) and identity number error rate (NER). In addition, we evaluate the performance of mID against several photo processing attacks. The performance of the mID scheme is summarized below:

- mID achieves an average BER of 0.2% and an average NER of 2.0%, which demonstrates promises towards screen photo forensics.
- mID performs well with little influence from the type of display devices, cameras, IDs, and ambient lights.
- mID performs well at a shooting distance of (60 cm, 80 cm) and a shooting angle of $(-20°, 20°)$, which are within the possible attack distances and angles adopted by adversaries as suggested by the theoretical calculation (in Section VIII).

A. Experiment Setup

We evaluate mID scheme in a laboratory setting with various display and capturing devices. The detailed settings are as follows.

Display Device: We use a BenQ EW Series LCD screen as the default display device. To evaluate the impact of display devices, we use 2 other LCD displays, 2 tablets and 2 laptops of different brands and models. Throughout the experiments, the display devices remain in the default settings with normal color mode and 50% screen brightness. The detailed information of each display device is shown in Table I.

TABLE I
SUMMARY OF DISPLAY DEVICES

No.	Manuf.	Model	Display Size	Aspect Ratio	Native Resolution	Panel Type	Backlight	Pixel Layout
1	BenQ	EW2440ZC	24"	16:9	1920 × 1080	MVA	LED	RGB
2	HP	24w	23.8"	16:9	1920 × 1080	IPS	LED	RGB
3	AOC	LV243XIP	23.8"	16:9	1920 × 1080	IPS	LED	RGB
4	Lenovo	IdeaPad Y700	15.6"	16:9	1920 × 1080	IPS	LED	RGB
5	ASUS	FX50J	15.6"	16:9	1920 × 1080	IPS	LED	RGB
6	HUAWEI	MatePad Pro	12.6"	16:10	2560 × 1600	IPS	OLED	Delta
7	Lenovo	Xiaoxin Pad Pro	11.5"	16:19	2560 × 1600	IPS	OLED	Delta

MVA: Multi-domain Vertical Alignment. IPS: In-Plane Switching

TABLE II
SUMMARY OF MAIN CAMERA SPECIFICATIONS OF THE CAPTURING DEVICES

No.	Manuf.	Model	Camera	Resolution	Aperture	Focal Length†	Pixel Size	Image Size	AF‡	HDR§
1	LG	Nexus 5X	Single	12.3 MP	f/2.0	5 mm, 26 mm (wide)	1.55 μm	4032 × 3024	√	√
2	HUAWEI	Mate 10	Dual	12 MP 20 MP B/W	f/1.6 f/1.6	4 mm, 27 mm (wide) 4 mm, 27 mm (wide)	1.25 μm 1.25 μm	3968 × 2976	√	√
3	HUAWEI	P9	Dual	12 MP 12 MP B/W	f/2.2 f/2.2	4.5 mm, 27 mm (wide) 4.5 mm, 27 mm (wide)	1.25 μm 1.25 μm	3968 × 2976	√	√
4	Apple	iPhone X	Dual	12 MP 12 MP	f/1.8 f/2.4	4 mm, 28 mm (wide) 6 mm, 52 mm (telephoto)	1.22 μm 1.0 μm	4032 × 3024	√	√
5	Motorola	G4 Plus	Single	16 MP	f/2.0	5 mm, 27 mm (wide)	-	4608 × 2592	√	√
6	Vivo	Xplay3S	Single	13 MP	f/1.8	4 mm, 28 mm (wide)	-	4128 × 3096	√	√

† Physical (former) and equivalent (latter) focal lengths for smartphone's built-in cameras. ‡ AF: Auto-focusing
§ HDR: High Dynamic Range Imaging

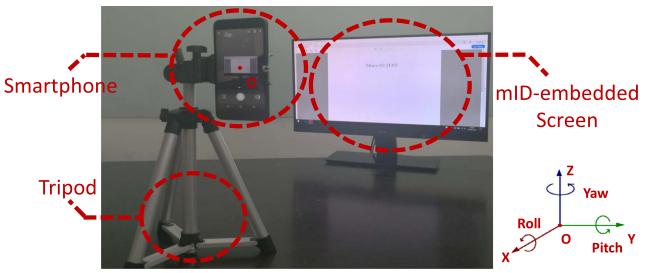


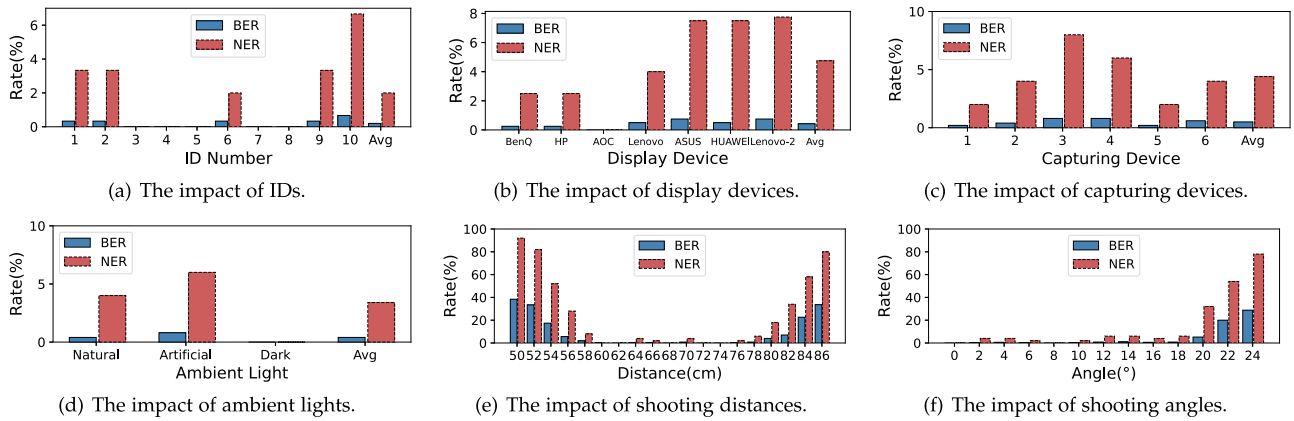
Fig. 12. Current page of the display is embedded with mID, which can be captured by the built-in cameras of smartphones.

Capturing Device: We use an LG Nexus 5X smartphone as the default capturing device. In addition, to evaluate the impact of capturing devices, we also conduct 5 other smartphones of various brands. Throughout the experiment, the capturing device is kept at a height of 30 cm from the table and aligned at the center point of the screen, as shown in Fig. 12. For the reason that the shooting distance users take photos toward the screen depends on the size of the display screens, we set different shooting distances for different display devices. For the LCD screens, the shooting distance and angle are set to 70 cm and 0° respectively. For the tablets, the shooting distance and angle are set to 35 cm and 0° respectively. We set each capturing device's main camera to the default setting mode with Auto-focusing (AF) and High Dynamic Range Imaging (HDR) turned on. At the same time, we keep other image processing techniques silent during the experiments (e.g., filters), since they are not generally set as the default mode and not all smartphones provide these techniques. The detailed parameters of the cameras are shown in Table II. Note that at the time of writing, we find no Moiré pattern filter functions available on smartphones in the current market.

Ambient Light: We conduct most experiments under the artificial lights produced by LEDs (∼200 lm), as it is the most likely attack environment in practice. In addition, we conduct experiments under the case of (1) natural lights (∼20 lm), and (2) no additional lights except for those from the display screen (<5 lm), to evaluate the impact of ambient lights.

Application Scenario: Without loss of generality, we study the PDF document as an illustration of confidential files and use Adobe Reader as the default document browser under the standard reading mode in this paper. The PDF document used in the experiments contains texts only. In addition, we conduct experiments with (1) Microsoft Word, (2) JetBrains PyCharm 2017, and (3) Google Gmail Web Client with 4 various background colors. Due to the space limitations and the similar performance across these applications, we demonstrate the results of Adobe Reader only. Note that mID scheme is applicable to both text-only and image-contained files. For instance, the Google Gmail Web Client has several images and logos in the background, and the mID scheme is able to cooperate with it as well. In addition, mID can cooperate with user operations such as zoom in and zoom out since it captures and modifies the current page in real-time.

Encoding Parameter: We choose 14-bit mID as an illustration in this paper, i.e., $N = 14$. In this way, each generated mID is



(a) The impact of IDs. (b) The impact of display devices. (c) The impact of capturing devices.

(d) The impact of ambient lights. (e) The impact of shooting distances. (f) The impact of shooting angles.

Fig. 13. Performance of mID decoding under various settings.

represented as a form of 01XXXXX...01 consists of a 2-bit front check code, a 10-bit information code, and a 2-bit end check code. For the discretized bipolar NRZ encoding, we employ 4 sub-levels with each sub-level consisting of 4 grating columns, i.e., $k = 4$ and $n = 4$. Note that all the above parameters are not mandatory and users can choose different parameters according to their requirements.

B. Performance Metrics

We use BER (bit error rate) and NER (identity number error rate) to evaluate mID from two different perspectives.

BER: BER refers to the number of bit errors divided by the total number of mID bits (excluding check codes), which evaluates the performance of mID decoding in a fine granularity.

NER: NER refers to the number of IDs that were not correctly decoded (IDs with at least one bit error) divided by the total number of mIDs. Thus, NER is a stricter criterion compared with BER and demonstrates the effectiveness of the proposed mID method.

C. Overall Performance

In this section, we first evaluate the overall performance of mID decoding with various IDs, and then evaluate the impact of the aforementioned factors including display devices, capturing devices, ambient light, etc.

1) Impact of IDs: In the first set of experiments, we evaluate the overall performance of mID with various IDs. We capture the current page of the screen in real-time, randomly generate 10 mIDs and embed them into the captured page with mID generation and mID embedding, and render the modified page on the default LCD monitor, respectively. We then capture 30 photos with the Nexus 5X smartphone for each mID. During the photographing, we use the default camera settings.

We perform mID extraction and mID decoding on the captured photos for each mID. The results in Fig. 13(a) reveals that mID scheme achieves an average BER and NER of 0.2% and 2.0%, respectively. Specifically, ID 3, 4, 5, 7, and 8 achieve the best performance of 0 BER and NER while ID 10 achieves the worst with a BER of 0.7% and a NER of 6.7%. Although the NER of each ID varies due to the limited samples as well as the randomness introduced during photographing, the BER remains relatively low and stable, demonstrating the effectiveness of the mID decoding algorithm.

2) Impact of Display Devices: In the real-world deployment, display devices may have various types and models. However, as mID does not use any explicit device attribute during design, it should be compatible with any display devices. To investigate it, we utilize two other LCD monitors, two tablets, and two laptops with different screen sizes and panel types to display the modified files. The details of each display device are summarized in Table I. With the default settings, we utilize the default capturing device and collect 40 photos for each display device.

The decoding results in Fig. 13(b) reveal that tablets show relatively higher NER and BER compared with laptop screens, while LCD monitors show the best decoding performance. Specifically, Lenovo Xiaoxin Pad Pro shows the worst performance with a BER of 0.8% and an NER of 7.8% while AOC shows the best with a BER of 0 and an NER of 0. We believe it is because laptops and tablets have smaller screen sizes compared with LCD monitors. As a result, the Moiré area occupies less area in the photos displayed on laptops or tablets, rendering it more likely to suffer from noise and more difficult to distinguish. Nevertheless, mID can still achieve an average BNR of 0.4% and an average NER of 4.8% among various display devices.

3) Impact of Capturing Devices: In practice, adversaries may use any smartphone to take pictures. To investigate whether mID works well under various smartphones, we conduct experiments with 5 other smartphones from different brands and models in addition to the default capturing device Nexus 5X. The details of each capturing device are summarized in Table II. We utilize the main camera (single or dual) of each smartphone with the auto-focusing setting to collect 50 photos respectively.

From the results shown in Fig. 13(c), we can observe that single-camera smartphones achieve better performance compared with dual-camera ones. For instance, Nexus 5X performs best in the experiments with a BER of 0.2% and an NER of 2.0%, which are both single-camera phones. By contrast, dual-camera devices suffer from relatively higher NERs, e.g., HUAWEI P9 performs the worst with a BER of 0.8% and an

NER of 8.0%. We believe it is because that dual-camera devices utilize images from both cameras to composite the final photo, which may have impact on the Moiré patterns and thus the decoding results. Overall, mID can achieve an average BER of 0.5% and an average NER of 4.4% with capturing devices various in resolution, aperture, and focal length.

4) Impact of Ambient Lights: During photographing, the ambient lights are likely to affect imaging and mID decoding. To investigate the impact of ambient lights, in addition to the artificial lights produced by LEDs (∼200 lm), we conduct experiments under two other light conditions, i.e., (1) natural lights (∼20 lm), and (2) no additional light except for the one from the display screen (i.e., dark environment (<5 lm)). For each light condition, we collect 50 photos and perform mID decoding.

The results in Fig. 13(d) demonstrate that the dark environment helps to improve the decoding performance while artificial lights have negative effects. Specifically, the dark environment achieves the best BER and NER of 0, followed by the natural environment with a BER of 0.4% and an NER of 4.0%. The artificial environment performs the worst with a BER of 0.8% and an NER of 6.0%. We believe it is because that the LEDs in the experimental room are multiple and decentralized. As a result, the light source is heterogeneous during photographing, which may cause the unevenness of exposure and thus decrease the decoding accuracy. Nevertheless, mID can still achieve an average BER of 0.4% and an average NER of 3.4% with various light conditions.

5) Impact of Photograph Distances: Theoretically, adversaries may take a photo from any distance. However, since the goal of the adversary is to record the information on the screen, the picture is likely to be taken at a reasonable distance and angle. To investigate the impact of photograph distance, we conduct experiments with the default display device. We first survey the common shooting distance adopted by normal volunteers to take a photo towards such a device, which turns out to be in the range of 50 cm–100 cm for the sake of capturing the screen well. We then conduct experiments during this range with a step of 2 cm. For each distance, we collect 50 photos and perform mID decoding.

From the results in Fig. 13(e), we can observe that mID achieves the best performance with a shooting distance around $58-80$ cm. With a photograph distance either >84 cm or <56 cm, mID decoding accuracy drops due to that the generated Moiré patterns become invisible to both human eyes and camera sensors. Overall, mID decoding can achieve an average BER of 0.3% and an average NER of 2.4% under the distance range of $(60\,\text{cm}, 80\,\text{cm})$. According to the calculation shown in Section VIII, to capture a 24" display screen completely, the photograph distance D is usually larger than 60 cm for various smartphones. Therefore, we believe that mID is basically sufficient to cover the possible attack distances adopted by adversaries.

6) Impact of Photograph Angles: In addition to the photograph distances, we also investigate the effect of photograph angle in three degrees of freedom: roll, pitch and yaw, as shown in Fig. 12. For the first degree of freedom (roll), since it rotates in the x-y plane, we can reduce its influence by image rotation. For the second and third degrees of freedom (pitch and yaw), they mainly cause the vertical and horizontal distortion of the captured image, respectively. As we use the transverse encoding in the experiments which means the vertical pixel alignment is the same, thus the vertical distortion may not affect the information representation. Besides, both the vertical and horizontal deformation can be addressed with rectification techniques [30], [31]. Therefore, we mainly evaluate the impact of the last degree-of-freedom, i.e., yaw, in this paper since it is most relevant to mID scheme.

During the experiments, we take the symmetry axis of the smartphone screen as the center axis and rotate the yaw angle with a step of 2°. We set the shooting distance to default throughout the experiments and the smartphone is tangent to the arc consisted by its motion locus. Without loss of generality, we start from the central point where the smartphone is paralleled with the display screen, i.e., 0°, and increase the angle of inclination in both clockwise (+) and anticlockwise (-) directions. For each angle, we collect 50 photos and perform mID decoding.

From the results shown in Fig. 13(f), we can observe that mID achieves a relatively low BER and NER with a photograph angle less than 20°. When the inclination angle is further increased, the distortion of Moiré stripes becomes non-negligible and difficult to be corrected, and thus may affect the performance of mID extraction and decoding. However, we argue that with an inclination angle larger than 20°, the image content is heavily distorted as well, which may also deviate from the goal of the adversaries. Overall, mID is able to achieve an average BER of 0.3% and an average NER of 3.1% with a photograph angle within $(-20°, 20°)$.

VII. User Study

We measure whether users will notice the presence of mID and how users cope with mID-related Moiré patterns in the screen photos by conducting a user study among 34 volunteers. Most of them are graduate students aged 20-30 years old. We followed the local regulations to protect the rights of human participants and have obtained the approval from the ethics committee of our institute.

To study whether users can recognize the presence of mID, we conduct the following test: on an LCD monitor in an office room, we display a PDF document using Abode Reader, which is an IELTS essay about news and we embed mID in both sides of the document body. The participants are required to sit in front of the monitor and provided 5 minutes to read the essay. After reading, we conduct a questionnaire survey for each participant, in which we ask three choice questions and three 7-point scale questions, and the detailed descriptions of each question are summarized in Table III. For comparison, we conduct another contrast test using a PDF document without mID. From Table III, we can see that the first 3 choice questions are essay-content-related, which are the superficial tasks of the test. The real aim is to learn whether the participants feel or notice any visual abnormality during the process of reading (Question 6) and we cover it up with two transitional questions (Question 4 and 5). The results shown in Fig. 14 demonstrate that the participants hardly perceive the

TABLE III
SUMMARY OF QUESTIONNAIRE SURVEY ON USERS' NOTICE OF mID

No.	Question
1-3	Essay-content-related choice questions
4	Is this test difficult? (7-point scale, where 7 indicates the most difficult)
5	Did the display device work well? (7-point scale, where 7 indicates the best functionality)
6	Do you feel abnormal during reading, e.g., display glitch/flicker or visual abnormity? (7-point scale, where 7 indicates the most abnormal)

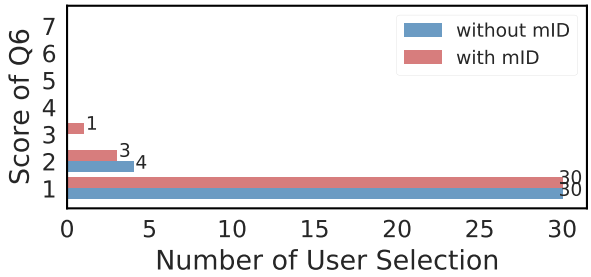


Fig. 14. Scores of Question 6 when the screen is embedded with/without mID.

existence of mID in the course of normal use (with an average of 1.147, a standard deviation of 0.429, a 95% confidence interval of [1.003,1.291] on a 7-point scale) compared with the mID-free situation (with an average of 1.118, a standard deviation of 0.322, a 95% confidence interval of [1.009,1.226] on a 7-point scale). Thus, mID should be able to satisfy the requirement of no visual impact on users.

To study how well the decoding technique works for the realistic photos taken by attackers, we conduct a real-world experiment by asking each volunteer to take 5 photos towards the mID-embedded screen after finishing the questionnaire survey, with the imagination of leaking important information to competitors and the need of capturing the information on the screen completely and clearly. The results illustrate that for the 170 photos taken by the volunteers, an average decoding accuracy of around 95% can be achieved. In addition, the results demonstrate that more than 91% (31/34) volunteers take photos with Moiré patterns as they are used to them. The other 3 users carefully adjust the shooting angle and distance to avoid Moiré patterns. However, the adjustment is not adopted by most users since it may twist the photo content. It indicates that the attackers are likely to include Moiré patterns in the screen photos.

To study whether users have the habit of taking screen photos with smartphones and their attitudes towards Moiré patterns in screen photos, we conduct another questionnaire survey for each participant as shown in Table IV. The results demonstrate that most participants have the habits of taking photos towards screens with their mobile phones to record and pass information (with an average of 4.559, a standard deviation of 1.802, a 95% confidence interval of [3.953, 5.145] on a 7-point scale). In addition, most participants have taken photos with Moiré patterns (with an average of 5.176, a standard deviation of 1.855, a 95% confidence interval of [4.552, 5.800] on a 7-point scale), and do not regard Moiré patterns in screen photos abnormal

TABLE IV
SUMMARY OF QUESTIONNAIRE SURVEY ON USERS' ATTITUDES TOWARDS MOIRÉ PATTERNS

No.	Question
1	Do you know about Moiré patterns? (7-point scale, where 7 indicates the most knowledgeable)
2	Will you take photos towards screens with smartphones to record and pass the information on screens? (7-point scale, where 7 indicates the most frequent)
3	Have you seen Moiré patterns in screen photos? (7-point scale, where 7 indicates the most frequent)
4	Will you feel odd when shooting a photo with Moiré patterns? (7-point scale, where 7 indicates the best functionality)
5	Will you take measures to remove Moiré patterns in screen photos? (7-point scale, where 7 indicates the best functionality)

(with an average of 2.029, a standard deviation of 1.150, a 95% confidence interval of [1.642, 2.416] on a 7-point scale). As a result, they will not bother to remove Moiré patterns (with an average of 1.265, a standard deviation of 0.656, a 95% confidence interval of [1.044, 1.486] on a 7-point scale). Therefore, we assume mID may not alert users and are likely to survive in screen photos.

VIII. DISCUSSION AND LIMITATIONS

In this section, we discuss several issues of mID as well as its limitations.

In-Camera Image Processing: Smartphone camera systems typically use image processing technologies such as autofocus, optical anti-vibration, HDR, and multi-camera fusion to help produce better quality images. Among them, HDR often combines multiple photos with different exposure intensities to achieve a greater dynamic range of luminosity, and we use it by default in our evaluation. To illustrate the impact of HDR, we conduct a contrast experiment with the HDR-deactivated Nexus 5 smartphone. The results demonstrate that HDR can reduce the BER from 0.4% to 0.2% for the Nexus 5 smartphone. We assume it is due to the fact that HDR helps to enhance the difference in brightness between the bit "0" and the bit "1", which can make decoding easier. In addition, image processing using multi-camera fusion combines photos from each individual camera to obtain a higher-quality image. In our evaluation, we use 3 single-camera and 3 dual-camera smartphones, and the results shown in Section VI-C2 demonstrate that dual-camera phones show a slightly higher NER but can still achieve good performance ($\sim 93\%$).

Post-Camera Image Processing: mID hides information by embedding the ID in a natural-appearance Moiré pattern. Like the same design assumptions as watermarking or steganographic techniques, we assume that the adversaries are unaware of the technique. Nevertheless, in a rare case, the adversary may process the captured screen photos to reduce the risk of being traced, as mentioned in Section III. The possible post-camera processing techniques include (1) commonly-used image processing methods such as photo duplication, photo compression, image up/downscaling, format conversion, image cut, downsampling-then-interpolating, and image filtering, and (2)

TABLE V
IMPACT OF PHOTO PROCESSING

No.	Photo Processing Technique	Defense
1	Image Duplication (copy and paste)	√
2	Image Compression (lossless)	√
3	Image Upscaling	√
4	Image Downscaling	Partial
5	Format Conversion (PNG to JPG)	√
6	Image Cut	Partial
7	Downsampling-then-interpolating	√
8	Image Filtering	Partial

specially-designed evasion algorithms for removing Moiré patterns.

For the former, we randomly select 10 photos from the screen photos collected under the default settings and conducted experiments to investigate whether mID could keep its robustness under these image operations. From the results shown in Table V, we can see that mID can successfully resist the influences of photo duplication (copy and paste), compression (lossless), upscaling (any upscaling ratio), format conversion (PNG to JPG) and downsampling-then-interpolating. It is because those photo processing methods nearly do not lose the information of the screen photo (upscaling even increases the amount of information contained in the photos) and therefore have no obvious effect on mID decoding. In contrast, image downsampling, image cut, and image filtering can affect the content of the screen photo and thus the mID decoding. For image downscaling, i.e., lossy image compression, we evaluate its impact by setting the image downscaling to 0.9, 0.8,..., 0.1, where the information loses uniformly. The results show that mID can achieve good performance (>90%) with a downscaling ratio larger than 0.6. When the ratio becomes smaller, e.g., 0.5, the performance of mID deteriorates due to excessive loss of content information (including Moiré patterns) from the screen photos. However, at the same time, the photos may become too blurry to recognize. For image cut, as we only embed mID in the vicinity of ROI, removing other photo areas does not impact the decoding of mID. If the adversary finely removes the Moiré patterns, which is possible but very difficult since they are usually surrounded by ROI, we may not obtain enough information to recover the embedded mID. For image filtering, e.g., using the median or Gaussian filters, it will blur images and thus causes information loss, impairing the mID decoding accuracy. However, image filtering cannot remove Moiré patterns but cause deteriorations in the photo quality, leading to distress in reading and comprehension. Thus, we believe that in most cases users will not bother to use it in practice. In general, mID is resistant to attacks of photo duplication, photo compression, image upscaling, format conversion, and downsampling-then-interpolating and partial attacks of image downscaling, image cut, and image filtering.

For the latter, existing Moiré pattern evasion approaches mainly have three categories: (1) adding an optical low-pass filter (OLPE) over the camera lens, (2) using an enhanced color interpolation algorithm, and (3) employing post image processing techniques. The first two approaches are both preventive measures and implemented within the cameras, thus are not capable of removing existing Moiré stripes contained in the screen photos. For the last approach, however, automatically removing Moiré patterns from a single photo is still challenging at present even with the help of deep learning [36], [37]. In most cases, it is still done manually with professional image processing software. We admit that these methods may have impacts on mID, but they can also cause a deterioration in the quality of the photo and some distress to reading and comprehension. Thus, we believe that in most cases the adversaries will not bother to remove the Moiré to ensure the completeness of the confidential information.

Display Device: In the aforementioned evaluation, we evaluate the performance of mID with display devices of various manufacturers, models, sizes, and panel types. In addition to these factors, the resolution and image rendering mode of display devices may also have impacts on the performance of mID. The dominant resolution of digital screens on the current market is 2 K and is likely to be increased to 4 K in the future. However, the resolution enhancement can enable the gratings of mID to get closer together, which can make mID smoother and more natural. For the image rendering mode, on the one hand, most users do not change the default settings of the display devices. On the other hand, if the users use a rendering mode such as low blue light, we can also obtain the current gamma value of the screen via the relevant APIs and adjust it accordingly during the brightness correction process.

Capturing Device: Due to the convenience and invisibility of smartphones, we used them as the capturing devices for our experiments. However, mID takes advantage of the interaction between the display device and the CFA of the digital camera. Other digital photography devices, such as DSLR (Digital Single Lens Reflex) cameras, can also capture Moiré patterns and can therefore be used in cooperation with mID. Further, as there are fewer image processing algorithms used in these digital photography devices, they can capture the Moiré patterns with less information loss and may contribute to higher decoding accuracy.

Transmission Over Instant-Messaging Tools: The adversary may exfiltrate the captured screen photo via instant-messaging tools, e.g., WhatsApp, Skype, and WeChat. There are two forms of image transmission via instant-messaging tools: (1) the images transferred as files, and (2) the images transferred as photos. The former generally uses a lossless approach (format and size unchanged or size unchanged with format conversion), while the latter uses a downscaled approach (compression). The adversaries are more likely to prefer the former approach to obtain a clearer photo of the confidential information. In this case, experimental results demonstrate that mID shows no performance difference in decoding 30 screen photos before/after being shared as files since mID is robust to format conversion attacks. However, in a few cases, the adversary may choose to share the screen photo directly as a photo. In this case, the screen photo is downscaled and the EXIF (exchangeable image file) information is lost. Since we encode mID in the horizontal direction and do not rely on any EXIF information, the horizontal

downscaling ratio (in the form of pixel numbers) is the main factor that may affect the decoding accuracy. Based on our experiments, the horizontal downscaling ratio depends on the photo contents and the used instant-messaging tools (different tools may use different compression algorithms) and usually ranges from 0.3 to 0.8. With the current encoding parameters shown in Section VI-A, mID can still decode screen photos with a horizontal downscaling ratio above 0.6 (i.e., a pixel loss of up to 64%) after sharing. For screen photos with smaller horizontal downscaling ratios, the decoding accuracy drops, e.g., by 63.5% for a ratio of 0.5. This can be addressed by adding more redundant pixels, i.e., increasing the value of k, for encoding. Experimental results show that with a larger $k = 8$, the decoding scheme can cope with a horizontal downscaling ratio as low as 0.3. Thus, we assume mID has the potential to survive the transmission over instant-messaging tools.

Encoding Space: The encoding space of mID mainly depends on the resolution of the display device and the composition of its current page. Specifically, a higher display resolution or a simpler page composition leads to a larger encoding space. An N-bit mID takes $q = 2k \times n \times N$ pixels in width with the capability of identifying 2^{N-4} devices. With a minimal grating height of $p = 50$, for a display device with a resolution of 1920×1080 pixels, the encoding space limit is $2^{[\frac{1920}{2k \times n} - 4] \times [\frac{1080}{p}]} = 2^{1176}$ with our default implementation. We acknowledge that the encoding space cannot reach the limit in practice since only portions of the screen can be used to embed mID. However, we believe that the encoding space of mID is still relatively large and sufficient for screen photo forensics, especially for highly-confidential scenarios.

Shooting Focus, Distance and Angle: mID applies to focus at the center of the screen and at a distance range of $(60\,\text{cm}, 80\,\text{cm})$ and $(-20^\circ, 20^\circ)$. We choose these parameters to reflect the target of the adversaries who wish to capture the content of the screen completely and clearly. Therefore, for the shooting focus, we set it in the center of the screen during the experiment, taking into account the adversaries' desire to capture confidential content completely and clearly. It is okay if the camera is not centrally focused as long as the mID-related Moiré patterns are captured in the photos. For shooting distances and angles, we agree that going over these ranges may cause the resulting Moiré pattern to be outside the visible frequency range, resulting in captured screen photos with partial or even no Moiré pattern. However, we believe that the distances and angles supported by mID can cover most of the possible photographing positions, to capture the on-screen content completely and clearly.

Minimal Photograph Distance: Considering the goal of recording the confidential information displayed on the screen with smartphones, we assume the adversary is likely to capture the screen in complete and hold the smartphone vertically to avoid the signs of secret filming. In this case, the photograph distance D adopted by the adversary shall be larger than a minimal value $D_{\min}$ to contain the entire screen in photos.

According to Equ. 4, the photograph distance D can be calculated as $D = \frac{S_{obj} \times L_f}{S_{cam}}$. For the minimal distance $D_{\min}$, S_{obj} is the physical width of the display screen, L_f is the physical focal length of the camera, and S_{cam} refers to the image width of the camera. S_{cam} can be further calculated as $S_{cam} = S_p \times N_p$, where S_p is the size of a single pixel and N_p is the number of pixels in width of the camera. As a result, $D_{\min}$ can be given as follows:

$$D_{\min} = \frac{S_{obj} \times L_f}{S_{cam}} = \frac{S_{obj} \times L_f}{S_p \times N_p} \tag{14}$$

With the device specifications in Tables I and II, we can calculate the minimal photograph distance $D_{\min}$ for various screen-camera settings. For instance, for our default setting, i.e., the BenQ EW2440ZC monitor for image display and the LG Nexus 5X smartphone for image capture, $D_{\min} = \frac{S_{obj} \times L_f}{S_p \times N_p} = \frac{53.1\,\text{cm} \times 5\,mm}{1.55\mu m \times 3024} = 56.6\,\text{cm}$. For HUAWEI Mate 10, HUAWEI P9, and Apple iPhone X, it will be 57.1 cm, 64.2 cm, and 57.6 cm, respectively. Therefore, to capture a 24" LCD display that is most commonly seen on the market with smartphones, the photograph distance shall usually be larger than 60 cm.

For the tablet screen, the photograph distance will be shorter than that of the LCD screen since it is smaller. Similarly, we can calculate the minimal photograph distance $D_{\min}$ for various tablets' screen-camera settings. For instance, for the sixth device in Table I, i.e., the HUAWEI MatePad Pro for image display and the LG Nexus 5X smartphone for image capture, $D_{\min} = \frac{S_{obj} \times L_f}{S_p \times N_p} = \frac{28.7\,\text{cm} \times 5\,mm}{1.55\mu m \times 3024} = 30.6\,\text{cm}$. For HUAWEI Mate 10, HUAWEI P9, and Apple iPhone X, it will be 30.9 cm, 30.9 cm, and 34.2 cm, respectively. Therefore, to capture a tablet completely with smartphones, the photograph distance shall be larger than 32 cm.

Comparison With Other Invisible Digital Watermark Techniques: Other invisible digital watermarking methods are embedded in the content of the confidential files, which can become unrecognizable due to noise introduced by electronic screens and camera sensors. Thus, we propose to utilize Moiré patterns for photo forensics since they are optical phenomena generated during the process of photographing screens. We compare our methods with 8 commonly-used invisible digital watermarks including 3 popular commercial tools: (1) SignMyImage [38], (2) Icemark [39], (3) OpenStego [40], and 5 open-source techniques from GitHub [41]: (4) Wavelet Transform, (5) Discrete Wavelet Transform, (6) Discrete Cosine Transform, (7) Least Significant Bit, and (8) Discrete Wavelet Transform and Singular Value Decomposition. The results show that none of the digital watermarks provided by the aforementioned methods work in the screen-photo-based leakage attacks while mID can successfully trace to the source of a screen photo with an average accuracy of 96%. Thus, we believe mID is suitable for screen photo forensics.

IX. Related Work

In this section, we present studies relevant to ours. Specifically, we discuss the aspects related to image watermarking, Moiré pattern, and optical cryptography.

Image Watermarking to Enable Digital Media Protection: Digital media requires protection when transferring through the internet or other mediums. Image watermarking techniques have been developed to fulfill this requirement [42]. Most existing image watermarking approaches are performed in the spatial [4],

[5], [6] or DWT (discrete wavelet transform) [7], [8], [9] domains and use frame synchronization methods to resist to geometric distortions. Beyond that, Riad et al. [43], [44] proposed a robust watermarking method based on Discrete Fourier Transform (DFT) for printed and scanned identity images. Gourrame et al. [45] proposed a Fourier-based watermarking method to resist print-cam attacks for real captured images and revealed that the FFT domain resists better to perspective distortions compared to the DWT domain. Thonglor and Amornraksa [46] proposed a watermarking method for posters that is robust against distortions due to printing and camera capturing processes. Different from these methods, mID is an optical watermark based on Moiré patterns and can be used for screen photo forensics.

Leveraging Moiré Patterns to Hide Invisible Messages: Moiré patterns are explored in various studies to hide messages. Lebanon et al. [47] explored ways to superimpose various patterns of gratings to create Moiré patterns of face images. Hersch et al. [48] created moving Moiré components running up and down at different speeds and orientations with the help of a revealing layer. Desmedt et al. [49] created secret sharing schemes based on Moiré patterns with shares being realistically looking images. Tsai et al. [11] enabled the creation of Moiré art and allowed visual decoding by superimposing grating images printed on separate transparencies. Walger and Hersch [13] proposed a method to embed information corresponding to up to seven level-line Moirés within a single base layer, and the information can be recovered later with a revealer printed on transparency or an array of cylindrical lenses. These studies mainly use two semi-transparent layers and overlap one on the other to reveal hidden images or information. By contrast, mID exploits the nonlinear optical interaction of the screen-camera channel to embed identity information.

Optical and Visual Cryptography to Enable Secure Information Exchange: Existing techniques [50], [51], [52] of visual cryptography (VC) usually encode a secret image into several shares with camouflaged visual patterns and stack a sufficient number of shares to reveal the original secret image. For instance, Huang and Wu [12] proposed an optical watermarking method in which a hidden binary image can be decoded by superposing a transparent key image onto a printed image. This studies [53], [54] applied VC to Quick Response (QR) codes to check the identity accessing the QR codes or control the permission to the protected data. Inspired by the aforementioned work, mID utilizes the inherent attributes of the screen-camera channel and proposes a Moiré-pattern-based optical watermarking scheme to enable screen photo forensics.

Our work is inspired by prior work and studies a digital forensics mechanism for screen photos utilizing Moiré patterns. Compared with our prior work [55], we enhance the decoding algorithm, improve the decoding accuracy, and include another type of display with different pixel layouts in this paper.

X. Conclusion

In this paper, we propose mID, a digital forensics mechanism to identify the source of the file leakages via photos utilizing Moiré patterns. We show that Moiré patterns are ideal for photo forensics because they are optical phenomena naturally generated during the process of photographing screens and are observed regularly in photos of digital screens. Leveraging it, we design an effective screen photo forensics scheme and evaluate it with 7 display devices and 6 smartphones of various manufacturers and models. The evaluation results demonstrate that mID can achieve an average BER of 0.2% and an average NER of 2.0%. In addition, the performance is barely affected by the type of display devices, cameras, IDs, and ambient lights. We believe that mID is a promising technique and can work complementarily to several existing methods to cope with illegal information leakage. Future directions that are worth studying include exploring a wider attack range and further improving the decoding accuracy.

References

[1] P. LLP, "Study on the scale and impact of industrial espionage and theft of trade secrets through cyber," 2019. [Online]. Available: https://tinyurl.com/ro5qu6o

[2] Version, "2018 data breach investigations report," 2018. [Online]. Available: https://tinyurl.com/qm3dmm2

[3] J. Morse, "Leaking anonymously is hard. Here's how to do it right, and not get caught," 2017. [Online]. Available: https://tinyurl.com/wplyk38

[4] N. Nikolaidis and I. Pitas, "Robust image watermarking in the spatial domain," *Signal Process.*, vol. 66, no. 3, pp. 385–403, 1998.

[5] W. Cheung, "Digital image watermarking in spatial and transform domains," in *Proc. IEEE Intell. Syst. Technol. New Millennium*, 2000, pp. 374–378.

[6] J. Abraham and V. Paul, "An imperceptible spatial domain color image watermarking scheme," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 31, pp. 125–133, 2016.

[7] E. Ganic and A. M. Eskicioglu, "Robust DWT-SVD domain image watermarking: Embedding data in all frequencies," in *Proc. ACM Workshop Multimedia Secur.*, 2004, pp. 166–174.

[8] Q. Li, C. Yuan, and Y.-Z. Zhong, "Adaptive DWT-SVD domain image watermarking using human visual model," in *Proc. IEEE 9th Int. Conf. Adv. Commun. Technol.*, 2007, pp. 1947–1951.

[9] C.-C. Lai and C.-C. Tsai, "Digital image watermarking using discrete wavelet transform and singular value decomposition," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 11, pp. 3060–3063, Nov. 2010.

[10] A. K. Boyat and B. K. Joshi, "A review paper: Noise models in digital image processing," 2015, *arXiv:1505.03489*.

[11] P.-H. Tsai and Y.-Y. Chuang, "Target-driven moire pattern synthesis by phase modulation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 1912–1919.

[12] S. Huang and J. K. Wu, "Optical watermarking for printed document authentication," *IEEE Trans. Inf. Forensics Secur.*, vol. 2, no. 2, pp. 164–173, Jun. 2007.

[13] T. Walger and R. D. Hersch, "Hiding information in multiple level-line moirés," in *Proc. ACM Symp. Document Eng.*, 2015, pp. 21–24.

[14] H. Pan, Y.-C. Chen, G. Xue, C.-W. B. You, and X. Ji, "Secure QR code scheme using nonlinearity of spatial frequency," in *Proc. Int. Symp. Pervasive Ubiquitous Comput. Wearable Comput.*, 2018, pp. 207–210.

[15] C. A. Poynton, "SMPTE tutorial: "Gamma" and its disguises: The nonlinear mappings of intensity in perception, CRTs, film, and video," *SMPTE J.*, vol. 102, no. 12, pp. 1099–1108, Dec. 1993.

[16] Wikipedia, "HSL and HSV," 2019. [Online]. Available: https://en.wikipedia.org/wiki/HSL_and_HSV

[17] I. Amidror, *The Theory of the Moiré Phenomenon: Volume I: Periodic Layers*, vol. 38. Berlin, Germany: Springer, 2009.

[18] Wikipedia, "Visual system," 2019. [Online]. Available: https://en.wikipedia.org/wiki/Visual_system

[19] Wikipedia, "Color filter array," 2019. [Online]. Available: https://en.wikipedia.org/wiki/Color_filter_array

[20] A. Raney, "Pinhole camera theory summary," 2017. [Online]. Available: https://tinyurl.com/s5bpf9h

[21] T. Point, "Digital communication - line codes," 2019. [Online]. Available: https://tinyurl.com/yx6gxblv

[22] C. in Colour, "Digital image interpolation," 2019. [Online]. Available: https://tinyurl.com/twxaxjk
[23] M.-J. Liaw, H.-H. Yang, and Y.-R. Shen, "Automatic gamma correction system for displays," U.S. Patent 6,593,934, 2003.
[24] C. Poynton, *Digital Video and HD: Algorithms and Interfaces, ser. Electronics & Electrical*. Amsterdam, The Netherlands: Elsevier, 2003. [Online]. Available: https://books.google.co.jp/books?id=ra1lcAwgvq4C
[25] Wikipedia, "Gamma correction," 2019. [Online]. Available: https://en.wikipedia.org/wiki/Gamma_correction
[26] Wikipedia, "Grayscale," 2019. [Online]. Available: https://en.wikipedia.org/wiki/Grayscale
[27] zlyBear, "BearOCR," 2019. [Online]. Available: https://github.com/zlyBear/BearOCR
[28] A. Gupta, A. Vedaldi, and A. Zisserman, "Synthetic data for text localisation in natural images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2315–2324.
[29] M. Liao, B. Shi, X. Bai, X. Wang, and W. Liu, "TextBoxes: A fast text detector with a single deep neural network," in *Proc. Conf. Assoc. Advance. Artif. Intell.*, 2017, pp. 4161–4167.
[30] H. S. M. Coxeter, H. S. M. Coxeter, H. S. M. Coxeter, and H. S. M. Coxeter, *Introduction to Geometry*, vol. 136. Hoboken, NJ, USA: Wiley, 1969.
[31] K. Zhang et al., "Chromacode: A fully imperceptible screen-camera communication system," in *Proc. Annu. Int. Conf. Mobile Comput. Netw.*, 2018, pp. 575–590.
[32] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.
[33] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[34] Wikipedia, "K-means clustering," 2019. [Online]. Available: https://en.wikipedia.org/wiki/K-means_clustering
[35] T. K. G. Inc, "OpenGL - The industry's foundation for high performance graphics," 2019. [Online]. Available: https://www.opengl.org/
[36] S. Yuan et al., "NTIRE 2020 challenge on image demoireing: Methods and results," 2020, *arXiv: 2005.03155*.
[37] Z. Bolun et al., "Learning frequency domain priors for image demoireing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7705–7717, Nov. 2022.
[38] A. P. Tools, "SignMyImage," 2013. [Online]. Available: http://www.adptools.com/signmyimage/
[39] P. Software, "Icemark," 2016. [Online]. Available: http://www.phibit.com/icemark/
[40] Syvaidya, "OpenStego," 2015. [Online]. Available: https://sourceforge.net/projects/openstego/
[41] Lakshitadodeja, "Image_watermarking," 2017. [Online]. Available: https://github.com/lakshitadodeja/image_watermarking
[42] L. K. Saini and V. Shrivastava, "A survey of digital watermarking techniques and its applications," 2014, *arXiv:1407.4735*.
[43] R. Riad, F. Ros, R. Harba, H. Douzi, and M. El Hajji, "Pre-processing the cover image before embedding improves the watermark detection rate," in *Proc. IEEE 2nd World Conf. Complex Syst.*, 2014, pp. 705–709.
[44] R. Riad, R. Harba, H. Douzi, M. El-hajji, and F. Ros, "Print-and-scan counterattacks for plastic card supports fourier watermarking," in *Proc. IEEE 23rd Int. Symp. Ind. Electron.*, 2014, pp. 1036–1041.
[45] K. Gourrame et al., "Robust print-cam image watermarking in fourier domain," in *Proc. Int. Conf. Image Signal Process.*, Springer, 2016, pp. 356–365.
[46] K. Thongkor and T. Amornraksa, "Robust image watermarking for camera-captured image using image registration technique," in *Proc. IEEE 14th Int. Symp. Commun. Inf. Technol.*, 2014, pp. 479–483.
[47] G. Lebanon and A. M. Bruckstein, "Variational approach to moiré pattern synthesis," *J. Opt. Soc. Amer. A*, vol. 18, no. 6, pp. 1371–1382, 2001.
[48] R. D. Hersch and S. Chosson, "Band moiré images," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 239–247, 2004.
[49] Y. Desmedt and T. Van Le, "Moiré cryptography," in *Proc. 7th ACM Conf. Comput. Commun. Secur.*, 2000, pp. 116–124.
[50] C. Blundo, A. De Santis, and M. Naor, "Visual cryptography for grey level images," *Inf. Process. Lett.*, vol. 75, no. 6, pp. 255–259, 2000.
[51] Y.-C. Hou, "Visual cryptography for color images," *Pattern Recognit.*, vol. 36, no. 7, pp. 1619–1629, 2003.
[52] P. Punithavathi and S. Geetha, "Visual cryptography: A brief survey," *Inf. Secur. J. A Glob. Perspective*, vol. 26, no. 6, pp. 305–317, 2017.
[53] X. Cao, L. Feng, P. Cao, and J. Hu, "Secure QR code scheme based on visual cryptography," in *Proc. 2nd Int. Conf. Artif. Intell. Ind. Eng.*, Atlantis Press, 2016, pp. 433–436.
[54] J. Lu, Z. Yang, L. Li, W. Yuan, L. Li, and C.-C. Chang, "Multiple schemes for mobile payment authentication using QR code and visual cryptography," *Mobile Inf. Syst.*, vol. 2017, 2017, Art. no. 4356038.
[55] Y. Cheng, X. Ji, L. Wang, Q. Pang, Y.-C. Chen, and W. Xu, "mID: Tracing screen photos via moiré patterns," in *Proc. USENIX Secur. Symp.*, 2021, pp. 2969–2986.

Wenyuan Xu received the BS degree in electrical engineering from Zhejiang University in 1998, the MS degree in computer science and engineering from Zhejiang University in 2001, and the PhD degree in electrical and computer engineering from Rutgers University in 2007. She is currently a professor with the College of Electrical Engineering, Zhejiang University. Her research interests include wireless networking, network security, and IoT security. She received the NSF Career Award in 2009, a CCS best paper award in 2017, and an ASIACCS best paper award in 2018. She was granted tenure (an associated professor) with the Department of Computer Science and Engineering, the University of South Carolina. She has served on the technical program committees for several IEEE/ACM conferences on wireless networking and security, and she is an associated editor of TOSN.

Yushi Cheng received the BS degree in electrical engineering from Zhejiang University in 2016, and the PhD degree in control science and engineering from Zhejiang University in 2021. Currently, she is a postdoctoral researcher with the Department of Automation of Tsinghua University. Her research interests include IoT security, AI security, and mobile & ubiquitous computing. She received a WST best paper runner-up award in 2017, and an ASIACCS best paper award in 2018.

Xiaoyu Ji (Member, IEEE) received the BS degree in electronic information & technology and instrumentation science from Zhejiang University, Hangzhou, China, in 2010 and the PhD degree from the Department of Computer Science from Hong Kong University of Science and Technology in 2015. From 2015 to 2016, he was a researcher with Huawei Future Networking Theory Lab in Hong Kong. He is now an associate professor with the Department of Electrical Engineering of Zhejiang University. His research interests include IoT security, including sensor, network, and AI security. He won the best paper award of ACM CCS 2017, ACM AsiaCCS 2018.

Yi-Chao Chen (Member, IEEE) received the BS and MS degrees from the Department of Computer Science and Information Engineering, National Taiwan University in 2004 and 2006, respectively and the PhD degree in computer science from the University of Texas at Austin in 2015. He joined Shanghai Jiao Tong University as a tenure-track assistant professor with the Department of Computer Science and Engineering in 2018. Prior to joining SJTU, he spent a year as a researcher in Huawei Future Network Theory Lab in Hong Kong and then worked as a co-founder in Hauoli LLC. His research interests focus on networked systems and span the areas of wireless networking, network measurement and analytics, and mobile computing.