PassBYOP: Bring Your Own Picture for Securing Graphical Passwords

Andrea Bianchi, Ian Oakley, and Hyoungshick Kim

***Abstract*—PassBYOP is a new graphical password scheme for public terminals that replaces the static digital images typically used in graphical password systems with personalized physical tokens, herein in the form of digital pictures displayed on a physical user-owned device such as a mobile phone. Users present these images to a system camera and then enter their password as a sequence of selections on live video of the token. Highly distinctive optical features are extracted from these selections and used as the password. We present three feasibility studies of PassBYOP examining its reliability, usability, and security against observation. The reliability study shows that image-feature based passwords are viable and suggests appropriate system thresholds—password items should contain a minimum of seven features, 40% of which must geometrically match originals stored on an authentication server in order to be judged equivalent. The usability study measures task completion times and error rates, revealing these to be 7.5 s and 9%, broadly comparable with prior graphical password systems that use static digital images. Finally, the security study highlights PassBYOP's resistance to observation attack—three attackers are unable to compromise a password using shoulder surfing, camera-based observation, or malware. These results indicate that PassBYOP shows promise for security while maintaining the usability of current graphical password schemes.**

***Index Terms*—Graphical password, input, live video, observation, user study.**

I. Introduction

SECURE access to information underpins modern digital systems and services. We keep our communications, financial data, work documents, and personal media safe by providing identity information and then authenticating to that identity. Text passwords and personal identification numbers (PINs) are the dominant authentication method [7] as they are simple and can be deployed on systems including public terminals, the web, and mobile devices. However, passwords suffer from limitations in terms of memorability and security—passwords that are difficult to guess are also hard to remember [19]. This is a major problem as an average user possesses 25 online accounts secured with up to six different passwords [17] and representing a substantial memory burden. To deal with this problem, individuals adopt nonsecure coping strategies such as reuse of passwords across systems, noting down passwords, or simply forgetting them entirely [1]. In order to mitigate these problems, researchers have proposed *graphical password* schemes [5], [6] that rely on input such as selecting portions of an image. These systems have been shown to improve memorability without sacrificing input time or error rates [24] while also maintaining a high resistance to brute force and guessing attacks [5].

However, graphical passwords present their own problems. One issue is their susceptibility to intelligent guessing [7], [8], [32] and shoulder-surfing attacks [31]. Such attacks are effective because the sections of images that users select as password items are both easy for an attacker to observe by snooping over shoulders or setting up a camera to record input and also relatively predictable—users tend to choose *hotspots* such as the eyes in a facial portrait [11], [28], [32]. This issue is particularly problematic as the image contents for graphical password systems are typically stored on authentication servers [5] and readily presented to attackers in response to input of easily accessible user identity information [27].

To address this issue, we present a new point-click graphical password system, *PassBYOP—Bring Your Own Picture*, that increases resistance to observation attack by coupling the user's password to an image or object physically possessed. This is achieved by using live video of a physical token, such as an object, a photograph, or even an image of a body part (e.g., a palm), as the canvas for entering a graphical password. This physical object replaces easily accessible server-based images [7], and we argue that attackers will struggle to capture useful replicas of this content. We present an implementation for the scheme based on SIFT image features [20] and a demonstration of its viability through three feasibility studies covering: 1) the reliability and robustness of PassBYOP feature based input; 2) participant task performance times and error rates using PassBYOP; and 3) the security of PassBYOP against observation attack.

II. Related Work

Graphical password systems are knowledge-based authentication techniques that leverage peoples' ability to memorize and recognize visual information more readily than alphanumeric information [22]. Researchers have explored three broad types of graphical passwords: recall-based *drawmetric* schemes based on sketching shapes on screen, recognition-based *cognometric*

Manuscript received May 27, 2015; revised August 17, 2015; accepted September 26, 2015. The work of A. Bianchi was supported by the Samsung Research Fund, Sungkyunkwan University, 2014. The work of I. Oakley was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT and Future Planning (2014R1A1A1002223). This paper was recommended by Associate Editor S. Rubin.

A. Bianchi is with the Department of Industrial Design, Korea Advanced Institute of Science and Technology, Daejeon 305-338, Korea (e-mail: andrea.whites@gmail.com).

I. Oakley is with the Department of Human and Systems Engineering, Ulsan National Institute of Science and Technology, Ulsan 689-798, Korea (e-mail: ian.r.oakley@gmail.com).

H. Kim is with the Department of Computer Science, Sungkyunkwan University, Seoul 110-745, Korea (e-mail: hyoung@skku.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2015.2487511

schemes based on selecting known items from large sets of options, and cued-recall *locimetric* schemes based on selecting regions of prechosen images [5], [14]. Locimetric schemes are discussed as is multifactor authentication, as it relates to PassBYOP and its combination of a token, or something you have, on which a password, or something you know, is entered.

A. Locimetric Password Schemes

Cued-recall (locimetric) password schemes involve users selecting regions on one or more images. Blonder's [6] U.S. patent is the earliest example. A seminal example is PassPoints [30]. During login, users are shown a previously selected image, and they enter a password by clicking on a sequence of locations on the image. Authentication is successful if the *XY* coordinates of these clicks match a previously stored set of password points. A longitudinal study resulted in login times of 8.78–24.25 s and a failed authentication rate of 7–13% [30].

While simple and effective, cued-recall graphical passwords present new security issues. For instance, users typically select *hotspots* [28], locations on an image that are highly distinguishable, memorable, and also predictable to attackers. In the Microsoft Windows 8 graphical password system, the most common password involved a photo of a person and triple tapping on the face, where one of the selection points was an eye [32]. Addressing this issue, the cued-click points (CCP) [9] system presented a series of images and allowed users to select only a single point per image, reducing the need to select common hotspots. Evaluations of this technique led to authentication times in the range of 7–8 s and success rates of 90–96%.

A second key problem with locimetric systems is observation, as password click-points can be acquired by attackers after viewing a single authentication process [5]. Securing against observation attack for graphical password systems is critical. Chiasson *et al.* [11] remark: "User interface manipulations such as reducing the text size of the mouse cursor or dimming the image may offer some protection, but have not been tested." One exception is a variant of CCP that uses eye-tracking technology [16] for input. This system increased resistance to observation but negatively impacted performance: login times rose to 47.1–64.3 s and only 67% of participants successful authenticated on their first attempt. Although more secure, this technique was prohibitively slow and error prone.

B. Multifactor Authentication Schemes

Multifactor authentication [26], based on the combination of two or more independent processes, can boost security. In typical multifactor authentication schemes, physical tokens are used to generate and store secrets for user authentication. For example, Aloul *et al.* [4] used mobile phones as the hardware token for one-time password generation. Dodson *et al.* [13] proposed a challenge-response authentication system involving a user snapping a picture of a QR code with a mobile device. The data from this marker generated encrypted data that were used during login. While these tools offer increased security, they are susceptible to particular kinds of attack, such as Man-in-the-Middle schemes that snoop on, or alter, messages transmitted between a user and the system [2].

PassBYOP is a multifactor authentication system—both a physical token and a password are needed to authenticate. PassBYOP differs from prior approaches in three ways. First, it is more flexible—instead of posing restrictions on the form of tokens, any sufficiently complex image or object can be used as a PassBYOP token. Second, the two authentication factors are tightly coupled—the password factor is entered on the token factor. We suggest this close relationship will make the scheme easy to understand. Finally, the image tokens in PassBYOP are high-entropy, sufficiently so that they have been previously proposed as a single factor authentication scheme [20]. We also suggest that these physical data-rich tokens will be resistant to Man-in-the-Middle schemes as attackers will face substantial barriers in terms of capturing tokens in sufficient detail to support successful hacks.

III. PassBYOP Overview

PassBYOP seeks to make graphical passwords more secure against intelligent guessing and shoulder-surfing attacks [27], [30]. We argue these weaknesses stem from the ease with which both password contents and password canvases can be observed or, in the case of canvases, directly accessed from a server [30]. PassBYOP tackles this problem by introducing a physical token into the authentication process. This way, PassBYOP transforms a graphical password, which is traditionally a single-factor authentication mechanism, to a more secure multifactor authentication method. We argue that this makes PassBYOP *Resilient-to-Internal-Observation* [7], meaning that an attacker cannot impersonate a user simply by intercepting input on the authentication device or by eavesdropping on the communication between the authentication device and verification system.

PassBYOP authentication takes place as follows (see Fig. 1). Assuming users have previously created a password, login involves users identifying themselves at a PassBYOP terminal in a manner fitting the system and use context. For example, systems such as office door locks may assume all users are valid, while a user ID might be used on a public computer, and higher security applications, such as a bank ATM, will likely rely on a physical token such as an ATM card. PassBYOP could be integrated into any of these scenarios. Second, users place a prechosen password image or object they possess on top of a camera unit in the terminal. This is captured and displayed live on an adjacent touch screen. Third, they tap on the image locations that correspond to their password. This way, authentication requires both the physical token and the password simultaneously. We argue this raises the resistance of PassBYOP to attacks based on password observation and guessing as attackers need to possess a user's genuine token or a high fidelity copy.

IV. Implementation

The PassBYOP prototype consists of a 13.5-cm-wide $\times$ 22.5-cm-long $\times$ 12-cm-high plastic box with a transparent cover and containing an upward-facing Logitech QuickCam E3500 webcam with a resolution of 640$\times$480 pixels and a speed of 30

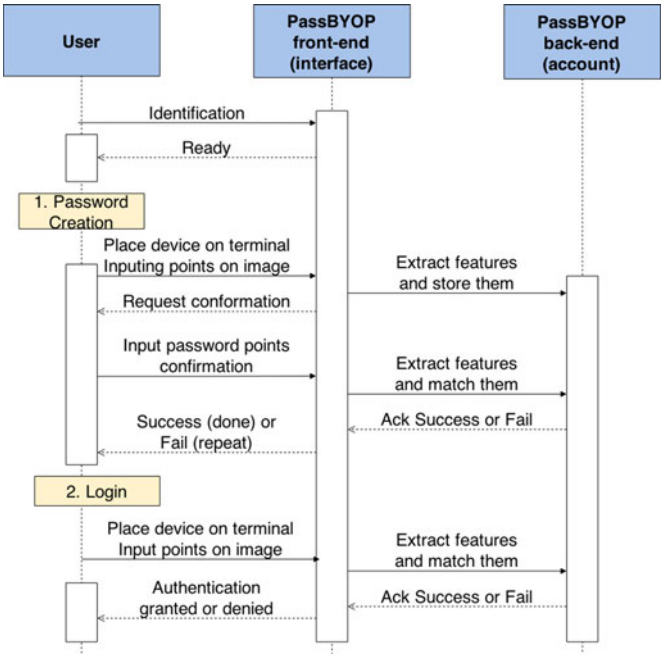

Fig. 1. Sequence diagram showing the steps involved in creating a PassBYOP password for the first time (1. Password Creation) and when attempting to login (2. Login).

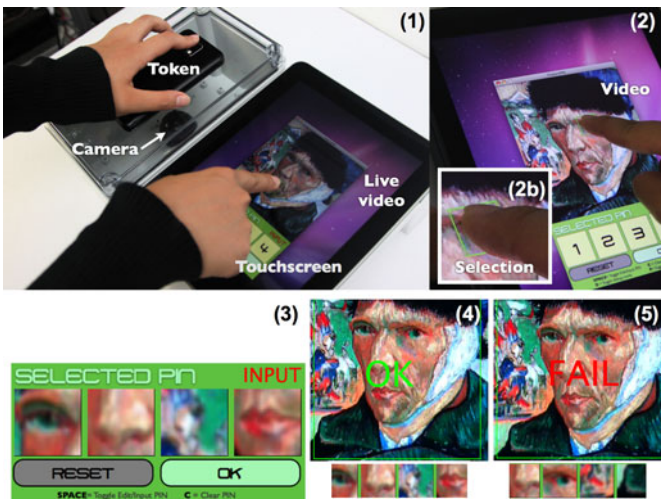

Fig. 2. (1) Overview of the PassBYOP system. (2) Input selection and closeup (2b). (3) Input selections that make up a password. (4) Successful authentication and (5) denied authentication.

frames/s. The webcam is connected to a PC running PassBYOP. The PassBYOP interface and video feed are shown on an Apple iPad that is connected wirelessly to the PC via a screen-sharing application [see (1) in Fig. 2] and fixed to the surface of a desk. The video resolution on the iPad is 450×600 pixels or approximately 8.5 cm × 14 cm. All input to the system is made on the iPad touchscreen. Specifically, as illustrated in (2) in Fig. 2, users make selections by tapping the screen to visually highlight 70 × 70 pixel (approximately 1.5 cm^2) portions of the displayed image, drag to move this region and release to select it. Once an image portion is selected, it is stored as a password item and displayed as feedback to the user at the base of the screen [see

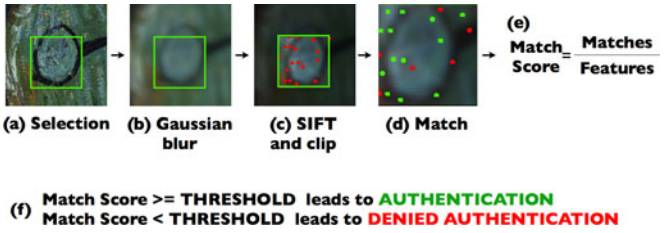

Fig. 3. PassBYOP process from image selection through feature extraction to image matching and production of a match score.

(3) in Fig. 2]. Users must input a total of four items and then press an OK button in order to enter a complete password. They can also press a reset button to clear the entered password items at any time.

In existing graphical password systems [30], the passwords are represented as the *XY* image coordinates of finger selections. This technique does not work with PassBYOP as variations in image placement on the terminal camera will lead to substantial variations in the *XY* pixel positions of image content. Instead, PassBYOP selections are stored on the authentication server as a set of *optical features* computed with the SIFT image processing algorithm [20]. This was achieved by capturing a 140 × 140 image subsection around the center point of each password item (see Fig. 3). A Gaussian blur was then applied and Lowe's [20] SIFT algorithm was computed with the peak threshold set to 2 and the edge threshold set to 10. This yields a list of image features and descriptors. Those that fell outside the central 70 × 70 selection box were discarded and the remainder used for password matching [see Fig. 3(d)].

The matching process involved minimizing the Euclidean distance between the sets of feature points in the original and entered password items (see Fig. 4). Subsequently, a threshold on the percentage of matching features was used to determine whether the entered password matched the original. Lower threshold levels result in a lenient password system, whereas higher levels are stricter. This process hinges on the fact that SIFT features are highly distinctive, robust to noise, accurate, and rotation invariant—capable of matching the features extracted from a single image against a database containing 100 000 images with an overall accuracy of 80% [20].

V. Evaluation

A. Reliability Study

This study assessed the reliability of PassBYOP in order to determine suitable thresholds for the equality of two password items in terms of the minimum number of image features they should possess and the percentage of image features that should match. As variations in token placement are inevitable with PassBYOP's camera-based setup, we also explored the robustness of the system with rotated input images. Finally, we assessed the uniqueness of feature-based password items.

1) Materials: Five source images were selected based on the image categories with highest success rate in prior work [8]. They depicted cars, a mural, toys, a statue, and a human face. These images were displayed on a Samsung Galaxy S-II

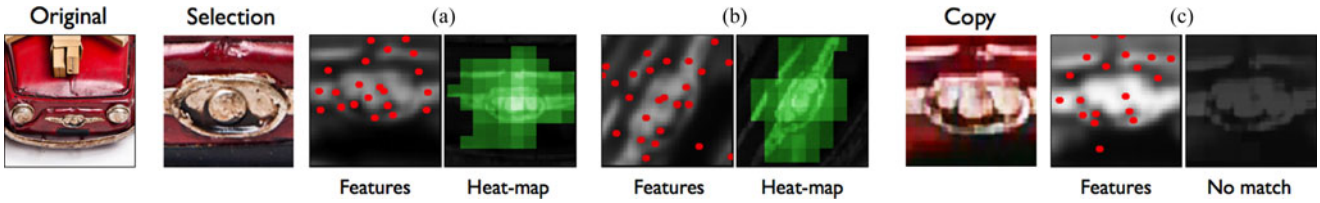

Fig. 4. Feature heatmap generated by testing the match between a selected area its transformations (rotation or translation) with the same image or a downgraded copy. Light colored zones in the heatmap indicate a match (white is 100% match). (a) Translated (b) Rotated (c) Translated.

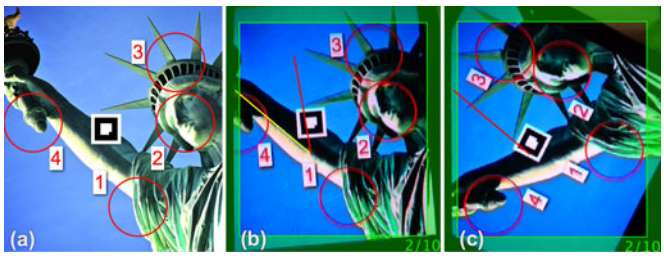

Fig. 5. (a) One of the five images used in the feasibility study. Colored lines showing the required and current angular orientation. (c) Image after the token has been rotated to match the required orientation.

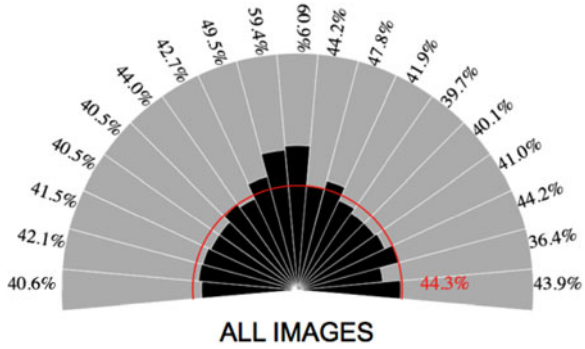

Fig. 6. Matching score break down for angles in the feasibility study. The mean value across all angles is 44.3%.

mobile phone with a screen resolution of 480 × 800 pixels and each image was preprocessed to match this screen resolution. We placed a 110-pixel square NyARToolkit fiducial marker [23] in the center of each image to enable accurate detection of its angle relative to the PassBYOP camera.

Four selection points were also marked on the image with a 110-pixel circle and labeled with numbers from 1 to 4. The selection points were chosen in a pilot study where eight users (two females, aged between 20 and 25 years) choose four passwords items on the selected images and entered them into the PassBYOP system five times. We chose prominent distinctive points from among the selections in these sessions—either those that were frequently chosen or, if there was substantial variation in the points selected by users, one of the items at random. An example of one of final images used in the study can be seen in Fig. 5. The experimental task involved users selecting these marked points in order. The use of predetermined and clearly marked selection points ensured the results were not influenced by issues such as memorability.

2) Participants: We recruited 15 volunteers (four females, two left-handed) from Sungkyunkwan University. They were a mix of students and staff, aged between 20 and 29 years (Mean: 24, SD: 2.83). None were security experts or knowledgeable in the area of security research.

3) Procedure: For each of the five preselected images, each user completed a block of 11 input trials composed of selecting the four marked points in ascending numerical order. Each user experienced the five images in a random order, and the first trial with each image was used as a reference for matching input in the subsequent ten trials. During each trial, the user also had to rotate the image to a specific angle prior to making input. For the first trial, this rotation angle always corresponded to aligning the long axis of the phone with the camera, but for all other trials, the required angle randomly varied from this vector by up to 90°, in 10° increments, in both rotational directions. The required angle was shown on screen by a short yellow line, and the angular position of the image was tracked using the AR marker and displayed as a red line. Before they were able to make selections, participants needed align these two lines. Selections made on nonaligned images were discarded and participants presented with a mild warning—an error beep. In case of inadvertent errors, participants were also able to press an on-screen reset button and start a new trial at any time. In total, this study captured 3000 valid selection events—15 participants × 5 images × 10 trials × 4 selection items. For each selection, we logged time, the number of features extracted, and the matching score.

4) Results: The mean completion time was 15.5 s (SD: 1.2), the average number of features extracted was 7.6 (SD: 2.7), and the average matching score was 44.3% (SD: 11.4). Fig. 6 shows the mean matching score for each angle studied. We examined the independent variables of image (five levels) and angle (nineteen levels) separately using one-way repeated measures ANOVA and MANOVA tests. This is because of the sparsity of the data collected—although the design was repeated measures, the large number of angles considered meant that not every participant completed a trial with every possible combination of image and angle, thus precluding the use of two-way tests. For each variable, we conducted an ANOVA on the time data and a MANOVA on the closely related measures of number of features and match score. In all cases, Mauchley's test assessed sphericity, and, if violated, Greenhouse-Geisser corrections were employed. Effect sizes are reported in the form of partial eta squared (η_p^2). Due to the exploratory nature of this investigation, and the large number of posthoc tests implied by

the 19 levels of the angle variable, we opted not to conduct a follow-up pairwise analysis for either variable.

In terms of task completion time, the image variable resulted in significant differences with a very limited effect size ($F_{(3.72,554)} = 6.219$, $p < 0.001$, $\eta_p^2 = 0.04$). Pillai's trace indicated that the image also exerted an effect on both number of features found and match score ($V = 0.973$, $F_{(8,142)} = 651.75, p < 0.001, \eta_p^2 = 0.973$). Follow-up univariate ANOVAs showed greater effect sizes: number of features found ($F_{(3.218,479.5)} = 344.3, p < 0.001, \eta_p^2 = 0.698$) and match score ($F_{(3.719,554.1)} = 219.7, p < 0.001, \eta_p^2 = 0.596$). The 19 different angles did not lead to significant variations in task completion time, but Pillai's trace indicated an effect on both number of features found and match score ($V = 0.407$, $F_{(36,864)} = 6.132$, $p < 0.001$, $\eta_p^2 = 0.204$). Follow-up univariate ANOVA showed significant variations in terms of both the number of features ($F_{(4.564,109.5)} = 4.217$, $p = 0.002$, $\eta_p^2 = 0.149$) and match score ($F_{(8.289,198.94)} = 9.147, p < 0.001, \eta_p^2 = 0.276$).

We note that times were high because participants had to perform two substantial tasks—align the image on the PassBYOP camera system and, then, select four targets. As the alignment angle and selection targets were marked in the same way throughout the study, it is not surprising that task time remains relatively stable—a significant effect was observed, but the effect size is very small. The substantial variations in the number of features and match score according to the image variable highlight the importance of the choice of source image in the PassBYOP system—some of the images used in the study were more feature rich, resulting in a more distinctive canvas on which to enter a password.

Finally, changes in the number of features and match score with the different image angles appear to clash with the notion that SIFT features are rotation invariant [20]. However, we believe these differences can be explained by the introduction of noise due to the use of a relatively low-resolution camera capturing a much higher resolution on-screen image. Furthermore, PassBYOP culls features outside a 70 × 70 pixel selection area. Consequently, with a rotated image, features situated in the corners may be lost, decreasing the overall matching score. This assertion is supported by visual inspection of the charts of the match scores—the lowest levels occur in and around 45° of rotation from the original image alignment (see Fig. 6). The rotation variations studied here represent extreme changes of up to 90°. Any realistic system implementation may constrain users to placing images in relatively consistent orientations.

To assess whether selections outside the target area would be erroneously matched using these threshold values, we created *feature heatmaps*—monochrome gradient maps where luminosity at a given location is determined by its percentage match (calculated using the SIFT algorithm) to a target area. To build these, we exhaustively executed the PassBYOP matching process for each pixel at the center of a 10 × 10 grid for each of the five images and four selection points used in the feasibility study. A blob detection algorithm was used to extract the number and size of matching regions. A total of 20 blobs were found (a one-to-one correspondence with selection points) with a mean size of 35.2 pixels (SD 11.2). Visual inspection revealed no overlaps between detected blobs—this indicates that no selections outside of the 70 × 70 selection region used in the system would lead to false positives. This evidence, in combination with the threshold results above, supports the viability and sensitivity of the overall PassBYOP approach—that SIFT feature sets are sufficiently persistent, distinctive, and unique to serve as password items. An example heatmap is shown in Fig. 7.

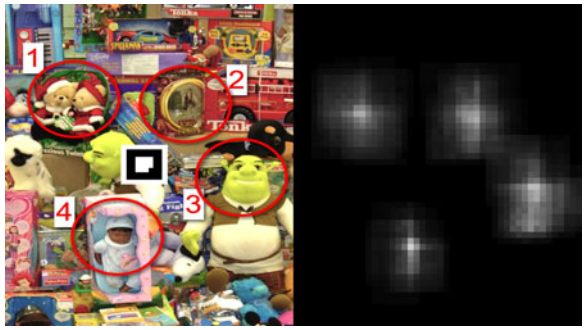

Fig. 7. Example of heatmaps extracted from one of the images used in the feasibility study.

B. Usability Evaluation

The second study in this paper explores user performance with PassBYOP in terms of entry times and error rates for comparison with prior graphical password system schemes. Users in this study authenticated in two conditions: a *private image* of their choice and a system-provided *public image*.

1) Participants: Twenty participants completed this study—six females, one left-handed and aged between 19 and 33 years (mean 23, SD: 3.8). They were students, researchers, and professionals, recruited using fliers posted in Sungkyunkwan University, emails to message boards, and word of mouth. Each participant was compensated with US$10. Participants regarded themselves as novice computer users (eight), medium users (ten), and advanced users (two). No participant was an expert in security. Participants were screened to ensure all owned a smartphone and stored personal images in its memory.

2) Materials: The *public* image depicted a parking lot populated with cars, similar to [8]. To acquire images for the *private* image condition, users were asked to select a personal authentication image in advance. They were given specific requirements: the image should be of high resolution, low granularity, and not to include large monochrome regions such as white walls. Images chosen by the participants included pictures of food (6), people (6), places (4), toys and small objects (3), and text (1). All selected images met system requirements in terms of visual richness of the contents.

3) Procedure and Measures: All participants completed both public and private image conditions in a fully balanced design—half of the participants experienced the private image condition followed by the public condition and the other half *vice versa*. All sessions took place in a quiet room using the PassBYOP terminal and a Samsung Galaxy S-II phone. PassBYOP was configured with a threshold of 40% and a minimum

number of seven features for each password item. Participants were first given an introduction to the system and its operation. They then completed a demographics form followed by the two experimental conditions.

The two conditions followed an identical structure. Each condition started with the *creation phase* in which the user placed the Galaxy S-II phone on the PassBYOP terminal and set a four-item password by selecting points on the displayed image. If any selection contained less than seven features, users were prompted to make another selection. At any point, a participant could press a reset button to clear the current selection and restart with no penalty. After the four items were entered, the participant re-entered the password. If he or she was unable to successfully do so, the system followed the typical conventions for bank passwords and required users to start afresh and create a new password. After participants successfully created a password, they moved on to the *login phase* where they authenticated two more times and then completed a visual distractor task at another computer. This task took the form of an online image-tagging game named ARTigo [3], in which users are presented with a sequence of five art images (each for 1 min) and need enter words to describe the contents. As with prior work on graphical passwords [8], the rationale for including the distractor task was to remove the graphical password from participants' working memory. After completing this distractor, participants authenticated one more time.

Performance measures used during the login phase of the study included: the mean time taken to enter the full set of four password items during successful authentication trials, the number of repeated trials required to setup the original password, the number of errors and resets that occurred during the study, and the number of features and matching score for each entered password item. Finally, at the end of both conditions, users completed a NASA TLX questionnaire [18] and answered two questions on ten-point Likert scales: "How easy was it to create a password on this image?" and "How difficult will it be to remember your password in one week?" We also conducted a poststudy interview in which we asked to participants about the usability of PassBYOP. The experiment took approximately 30 min per participant, and we captured data for a total of 80 correct authentication trials or 320 individual password item entries.

4) Results: In this analysis, significance levels were determined using $\alpha = 0.05$. Objective results from the login phase are shown in Table I. These data were first tested for normality using the Shapiro–Wilk test; error data were highly nonnormal and a subsequent Wilcoxon signed rank test showed no significant variation between conditions. Every participant successfully authenticated within three trials. With the public image, a single user contributed 42% of errors, while two users were responsible for 50% of errors in the private image condition. A paired t-test on the time data between public and private conditions also showed no significant differences. Finally, a repeated-measure MANOVA using Pillai's trace showed a significant impact of the public/private condition on the related variables of number of features and match score ($V = 0.6.54$, $F_{(2,18)} = 4.759$, $p = 0.022$, $\eta_p^2 = 0.346$). Subsequent univariate ANOVAs showed significant differences in both match score ($F_{(1,19)} = 4.851$, $p = 0.04$, $\eta_p^2 = 0.203$) and number of features ($F_{(1,19)} = 8.353$, $p = 0.009$, $\eta_p^2 = 0.305$). Although there are modest variations in these latter measures, the number of features and match scores in both conditions exceed the thresholds of seven and 40% established in the system feasibility study. Participants were capable of selecting and entering passwords with both the image selected by the experimenters and their own images.

TABLE I
RESULTS OF THE USABILITY STUDY

	System image	User image
Median creation time (s)	8.2 (5.7)	8.5 (2.9)
Median login time (s)	7.3 (2.8)	7.5 (2.1)
Password creation success rate	100%	100%
Successful login within 3 trials	100%	100%
Successful login at first trial	100%	85%
Successful login at second trial	–	100%
Total resets	6 / 87	7 / 90
Mean error items (in failed login)	1.7 / 4	2.1 / 4
Mean match score (successful)	72.9% (6.7)	77.1% (5.5)
Mean match score (fail)	27% (22.8)	13% (15.4)
Mean features (successful)	11 (2.2)	14.7 (4.5)

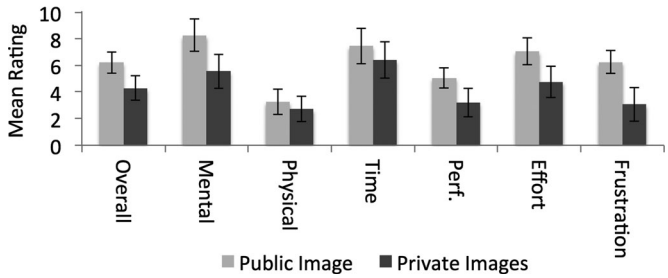

Fig. 8. TLX data showing workload in the usability study.

The TLX workload data are shown in Fig. 8. These show a general trend for reduced workload in the private image condition, an observation borne out by a significant difference in the summed measure of Overall Workload (t (19) = 2.835, $p = 0.011$, $d = 0.51$). To protect against alpha inflation, we do not report results for the component workload measures. Participants also rated ease of creating passwords with the private and public images at 8.3 (SD 5.8) and 6.95 (SD 2.68) and memorability of the private and public images at 5.8 (SD 2.19) and 4.85 (2.52), respectively. These related measures were analyzed using repeated-measure MANOVA. Pillai's trace showed a significant effect of public/private images on ease of creation and memorability ($V = 0.369$, $F_{(2,18)} = 5.263$, $p = 0.016$, $\eta_p^2 = 0.369$). However, follow-up univariate ANOVAs revealed a significant difference only in terms of ease of password creation ($F_{(2.488,18.225)} = 7.325$, $p = 0.014$, $\eta_p^2 = 0.278$). In general, these subjective data favor the private image condition over the public image condition. However, we recommend caution interpreting these effects, as they necessarily involve a number of independent tests. While opinions differ on how to handle such multiple comparisons [15], we note a lack of significance differences if corrective procedures, such as using a more conservative threshold of $\alpha = 0.01$, are applied.

In the posthoc interview, participants explained these ratings by remarking that the parking lot image contains too many similar cars, making selecting password locations challenging. Several also noted that although they felt it was easier to choose memorable locations from their private images, they still sometimes confused selection points with visually similar locations in their images. Finally, participants acknowledged that the choice of their private image was important and that their perceptions of the security and usability of the system partially reflected these choices.

C. Security Analysis

This section provides a security analysis of the PassBYOP system. We developed a threat model for PassBYOP that is based on vectors including token theft, guessing (both educated and brute-force), and observation (via shoulder-surfing, camera attacks, and via malware that takes over the PassBYOP camera). We analyze theft and guessing attacks conceptually and describe a study to assess resilience to the three different forms of observation.

1) Theft: While PassBYOP cannot prevent theft, its close coupling of a token to a password does provide benefits. Unlike many types of authentication token (e.g., door entry cards), physical possession is insufficient to crack the system—attackers must also gain access to the password. This way, PassBYOP offers advantages over purely token-based systems, including those based on secure device pairing over visual channels [21], [25]. There are also three further advantages conferred by using a token displayed on a mobile device. First, attackers must unlock the mobile device to access the token, potentially facing an additional and unrelated security scheme. Second, they must identify the precise token image, a potentially challenging process. Third, users could conceivably use software to remotely wipe a token from a stolen device. This paper argues that the relative ease with which users would be able to restrict access to obscure or remove their PassBYOP password images provides a measure of resistance to attacks based on token theft over and above that present in more traditional token-based schemes.

2) Educated Guessing or Brute Force Attacks: From a security perspective, typical cued-recall graphical passwords have practical password spaces comparable in cardinality to four- or five-digit PINs [5]. Data from the feasibility study suggest that PassBYOP has a similarly sized password space—with a matching threshold of 40%, the heatmap analysis indicates that each PassBYOP selection has a viable radius of 35 pixels (0.75 cm), leading to a valid selection area of 0.56 cm^2, a figure very close to that used in benchmark systems such as the 0.53cm^2 used in PassPoints [30]. Thus, given a total selection space of 450 $\times$ 500 pixels, the total number of discriminable selection points for each user input is approximately $\sim$220. Over a four-item PIN, according to the calculations used by Wiedenbeck *et al.* [30], this leads to a total Hartley entropy (or available password space) of $\sim \log_2(220.4^4)$, a figure greatly exceeding that of a four-digit numerical PIN [5].

We acknowledge that these entropy figures are optimistically high and represent a theoretical maximum—in reality, only a subset of the possible hotspots are actually likely to be selected [28], [29]. However, this entropy calculation appears in closely related work [30], and using this common formulation makes PassBYOP comparable with prior work. We also note that in contrast with other graphical password schemes, PassBYOPs use of a token makes guessing attacks insufficient if used alone—they must be combined with theft or observation in order to also acquire either the users token or a high fidelity copy. We argue that this increases the security of PassBYOP relative to prior approaches.

3) Observation: Cued-recall graphical passwords are vulnerable to observation attacks. A single observation can be enough to disclose a password to a bystander [11], [30]. Reflecting the importance of this vector, an observation attack was staged on the PassBYOP system to empirically assess the system's resistance to this type of threat. Three types of observation were considered: shoulder-surfing, a camera attack, and an attack based on malware that takes over the PassBYOP terminal and records the image displayed on the screen and the coordinates of the input points selected by the user. This last attack represents a worse-case scenario—a substantial and comprehensive man-in-the-middle attack akin to using the system camera to skim not only the password items entered, but also a copy of the image they are entered on. We conducted an empirical study to explore the resistance of PassBYOP to these vectors using the system configuration studied in the system feasibility study: passwords composed of four items, each with a minimum of seven features and matches recorded above a threshold of 40%.

4) Security Study: A member of our research group posed as a knowledgeable security conscious victim and repeatedly entered two PassBYOP passwords in two different attack scenarios. The first involved the use of a *public* system assigned image depicting a parking lot, as in [8], while the second involved the use of a *private* personally selected image, in this case a bowl of Japanese ramen. We argue that the public scenario mimics the case of conventional cued-recall graphical passwords, where the images used for authentication are stored on a server and disclosed at login time. On the other hand, the private scenario explores whether there is additional security value in PassBYOP's support for personally selected and maintained user-owned images.

a) Participants: Three participants (attackers) completed this study, a typical size of participant pool for this kind of experiment [12]. They were all graduate students from Sungkyunkwan University majoring in computer security. None was otherwise involved with this research, and each attacked PassBYOP in both public and private scenarios.

b) Procedure: The order of the scenarios was randomly assigned to each participant, and there was a 30-min break between attempts to crack each scenario. While attempting to crack each scenario, participants performed a series of three increasingly sophisticated attacks: 1) shoulder-surfing followed by 2) camera attack followed by 3) malware combined with camera attack. For each attack type, participants were requested to spend at least 10 min attempting to authenticate and were allowed three attempts to enter the correct password. If at any point the password was cracked, the attacker was not required to continue cracking the same scenario. If all three attempts failed, they moved on to the next attack. As an incentive, attackers who

succeeded to crack the password with shoulder-surfing were compensated with US$10, those who succeeded with camera attack received US$8, and US$5 was provided for success with the malware attack. Lunch was offered to all the attackers.

During the shoulder-surfing stage, attackers stood near the victim (within 1.5 m) during three successful logins. Note taking was encouraged. In this camera stage, attackers were provided with an HD video recording showing a closeup of the entire login process, including password item entry and a clear capture of the mobile device showing the image token. The video was shot without visual obstructions from less than 1 m away from the user with an HDR-HC3 HDV 1080i Sony camcorder. In the malware stage, attackers were provided with an additional video recording of the login phase from the point of view of the PassBYOP system camera. Attackers were able to use any tools or resources they wished during the attacks. In the public image condition, they were automatically presented with the authentication image, while in the private condition, they were able to use Internet searches and any image processing tools they wished to find, treat, create, or modify the source image and selection points observed and captured during the attack. In total, each participant spent approximately 3 h to complete the experiment.

c) Measures: We recorded the number of passwords cracked, the relative percentage matching scores, and the mean number of matching features: These last two measures indicate how well the attackers were able to reproduce the user's input images and selections—the higher the numbers, the stronger the attacks. We also distributed a questionnaire for the attackers to indicate on a ten-item Likert scale how difficult they felt the attack was and how well they self-evaluated their performance. Finally, in a poststudy interview, we asked them to describe their process.

d) Results: Table II shows the results of the attacks for authentications with both public and private images. A single observation was enough for all three attackers to crack the public image password [11]. In fact, they were able to do so quickly and confidently—in less than 10 s and with a matching score of 65%, substantially over the system threshold of 40%. In the self-reported questionnaire, the attack was declared to be easy (2.3 SD:2.3) and the attackers' performance to be good (8.3 SD:2.8). They reported that they entered the password after the shoulder surfing observation. One attacker indicated he or she had taken notes.

With private images, the shoulder-surfing attack was completely unsuccessful. Although attackers spent between 10 and 30 min trying to find a similar image using the Internet (one attacker searched on the victim's personal homepage), they were unable to authenticate within the given trials, and none of the features could be matched. Attackers reported the task to be difficult (10, SD:0) and their performance to be low (3.6, SD: 4.6). We attribute this low performance to the fact that the SIFT algorithm is capable of detecting and recognizing the features of a single image from a dataset of 100 000 keypoints [20] with an accuracy of 80%. As such, even if an attacker synthetically constructs an image where each pixel is computationally generated with a random color, the chance that any of the features required per selection will match the features of the stored password image will be 20% or lower. Based on this evidence, we argue that even with the more liberal matching threshold of 40% used in PassBYOP, the chances of a randomly generated image leading to a matching feature set is very low—certainly much lower than the one in ten chance of guessing a single numeric PIN item.

The camera attack was also unsuccessful, but two attackers were able to compromise a single password item. This attack took longer (15–45 minutes) because attackers extracted frames from the HD footage when the phone was facing the camera and used image editing tools such as Adobe Photoshop to recompose the source image used in the authentication. The attack was reported to be moderately difficult (7, SD:1) and performance to be relatively low (4, SD:2.6). One attacker explained that the difficulty was to create an image to match the original observed image. Although the footage was clear, it was challenging to reproduce an identical replica, as even small variations of size, viewing angle, or illumination led to substantially different image features.

Finally, the malware and camera attack was the most effective—it represents a worst-case scenario. Two attackers were able to compromise two of the password items—half the full password. This attack took approximately the same time as the camera attack and was not reported to be easier (7.6 SD:0.5) although it resulted in modest improvements to self-reports of performance (5.3 SD:0.5). Attackers indicated they followed an image recomposition process broadly similar to that used with the camera attack, but they encountered two unexpected difficulties. First, the low resolution of the system camera (640×480) led to downsampled image captures that could not be directly used to authenticate—features derived from low-resolution copies differ from those extracted from high-resolution originals displayed on the phone. Second, minor movements of the phone to bring the selection points into the field of view of the camera meant that attackers were not able to rely on a single frame showing the entire image and were forced to edit together multiple frames to produce their final image—a laborious task.

These results compare well with prior cued-recall password systems [8], [30], [31] that exhibit little to no resistance against shoulder-surfing. Attacks on PassBYOP took substantial time and effort and yielded a low success rate—although several items were successfully entered, no attacker managed to crack a full PassBYOP password. This result demonstrates the increased security of the PassBYOP approach against observation. It is particularly compelling as, although the attackers were partially able to crack the password, the threat model used in the malware attack was extremely generous in the type and nature of the information provided. This suggests PassBYOP would exhibit a very high resistance to observation if deployed in a real-world setting.

VI. Discussion

We presented three empirical examinations of the PassBYOP system. In the first, we established the feasibility of using image features as password items in terms of their uniqueness and the reliability with which they can be entered. In the second, we established basic user performance data while operating PassBYOP: Login took a median of 7.5 s, and although error

TABLE II
RESULTS FOR THE OBSERVATION ATTACK USING PUBLIC AND PRIVATE IMAGES SHOWING 1) NUMBER OF ITEMS CRACKED, 2) THE PERCENTAGE MEAN MATCH SCORE ATTAINED, AND 3) THE MEAN NUMBER OF MATCHING FEATURES

	Shoulder surfing			Camera			Malware & Camera		
	Items cracked	Mean score	Features	Items cracked	Mean score	Features	Items cracked	Mean score	Features
Attacker1	4/4	62.75%	6.25	–	–	–	–	–	–
Attacker2	4/4	63%	6.25	–	–	–	–	–	–
Attacker3	4/4	63.25%	6.75	–	–	–	–	–	–
Average	4/4	63%(0.2)	6.4(0.3)	–	–	–	–	–	–
	Shoulder surfing			Camera			Malware & Camera		
	Items cracked	Mean score	Mean Features	Items cracked	Mean score	Mean Features	Items cracked	Mean score	Mean Features
Attacker1	0/4	0%	0	1/4	12.25%	1.12	0/4	8%	0.5
Attacker2	0/4	0%	0	1/4	28.58%	3.41	2/4	22.12%	2.25
Attacker3	0/4	0%	0	0/4	17.87%	1.62	2/4	41.5%	4.12
Average	0/4	0%	0	0.7(0.6)	19.6%(8.3)	2.1(1.2)	1.3(1.1)	23.9%(16.8)	2.29(1.81)

data was unevenly distributed, mean rates were 9%. Finally, in the third study, we examined security and established that the use of an external token image increases the resistance to observation attack without compromising security against other vectors such as intelligent guessing or brute force. These results compare well with seminal prior work such as Passpoints [30], which yielded mean login times of 8.78–24.25 s and 1.55–2.75 failed authentication attempts prior to successfully entering a password. Similarly, Chiasson *et al.* [8] present a lab study of click-point-based graphical passwords using multiple images and report a median login time of 7 s and an error rate of 6%.

Based on these data, we argue that the use of feature extraction from captured images as the mechanism for storing and matching password items does not fundamentally change the ease with which cued-recall graphical passwords can be used. This is a highly positive conclusion as the underlying complexity of the recognition and comparison system in PassBYOP is substantial—to achieve equivalent results to prior graphical password systems is a strong endorsement of the technical viability of the approach. This result also shows that the increased resistance to observation achieved by PassBYOP does not place additional burdens on users—speed and accuracy are broadly comparable with prior systems.

Worth contextualizing is the conclusion in the light of prior work that aims to compare graphical passwords against observation attacks. For example, Forget *et al.* [16] present observation resistant graphical passwords that are entered by tracking eye movements. Login times are between 36.6 and 53.5 s, depending on the tolerance levels used in the system. Similarly, participants were only able to enter a password correctly in three attempts in 79–93% of cases. This example highlights how challenging it is to design observation resistant systems. The substantially lower task entry times, and greater accuracy of password entry with PassBYOP suggests that it is a more realistic approach to increasing the observation resistance of graphical passwords. It allows users to enter information in a comfortable and traditional way, while still introducing a hard-to-observe component—the PassBYOP tokens. Furthermore, the fact that these tokens are self-selected, rather than issued by a central certified authority, such as a bank, may also confer additional advantages. Specifically, in the usability study, participants experienced lower levels of self-reported workload and stated they preferred their own images to a standard system provided alternative.

There are a number of limitations to this study. In terms of the system, we used SIFT, a single feature extraction technique, and a more extensive investigation of alternative techniques (such as SURF) may reveal a more efficient or otherwise optimal candidate. Similarly, the feature matching algorithm we used was based on the comparison of Euclidean distance between features, as in [20]. Exploring more advanced similarity metrics could improve system performance. Furthermore, we did not perform any formal evaluation to determine the feasibility of PassBYOP across different devices and in different environmental conditions. Although we have informally tested the system with a range of mobile devices and token types and in different lighting conditions, formal study of these variables is an important next step toward demonstrating the robustness and viability of the approach. PassBYOP also used a low-resolution camera, which increased robustness against tamper-based observation attacks, but may have made it harder to recognize genuinely correct tokens and features. In the future, PassBYOP performance should be tested with a variety of cameras. Finally, the current PassBYOP system achieved multitouch input capability by wirelessly streaming video from the PassBYOP host computer to an iPad tablet. While this approach was simple and effective, greater speed and efficiency would be attained with a native application.

In summary, this paper proposed improving the security of graphical password systems by integrating live video of a physical token that a user carries with them. It first demonstrates the feasibility of the concept by building and testing a fully functional prototype. It then illustrates that user performance is equivalent to that attained in standard graphical password systems through a usability study assessing task time, error rate, and subjective workload. Finally, a security study shows that PassBYOP substantially increases resistance to shoulder-surfing

attacks compared with existing graphical password schemes [5], [16], [30]. Ultimately, we argue this paper demonstrates that PassBYOP conserves the beneficial properties of graphical passwords while increasing their security.

ACKNOWLEDGMENT

The authors would like to thank all their experimental participants for their time.

REFERENCES

[1] A. Adams and M. Sasse, "Users are not the enemy," *Commun. ACM*, vol. 42, pp. 40–46, 1999.
[2] M. Adham, A. Azodi, Y. Desmedt, and I. Karaolis, "How to attack two-factor authentication internet banking," in *Proc. 17th Int. Conf. Financial Cryptography*, 2013, pp. 322–328.
[3] ARTigo, http://www.artigo.org/.
[4] F. Aloul, S. Zahidi, and W. El-Hajj, "Two factor authentication using mobile phones," *Proc. Comput. Syst. Appl.*, 2009, pp. 641–644.
[5] R. Biddle, S. Chiasson, and P. van Oorschot, "Graphical passwords: Learning from the first twelve years," *ACM Comput. Surveys* vol. 44, no. 4, p. 19, 2012.
[6] G. E. Blonder, "Graphical passwords," U.S. Patent 5 559 961, 1996.
[7] J. Bonneau, C. Herley, P. C. van Oorschot, and F. Stajano, "The quest to replace passwords: A framework for comparative evaluation of web authentication schemes," in *Proc. IEEE Symp. Security Privacy*, 2012, pp. 553–567.
[8] S. Chiasson, R. Biddle, and P. van Oorschot, "A second look at the usability of click-based graphical passwords," in *Proc. 3rd Symp. Usable Privacy Security*, 2007, pp. 1–12.
[9] S. Chiasson, P. C. van Oorschot, and R. Biddle, "Graphical password authentication using cued click points," in *Proc. 12th Eur. Symp. Res. Comput. Security*, 2007, pp. 359–374.
[10] S. Chiasson, A. Forget, R. Biddle, and P. C. Oorschot, "User interface design affects security: Patterns in click-based graphical passwords, *Int. J. Inf. Security*, vol. 8, no. 6, pp. 387–398, 2009.
[11] S. Chiasson, E. Stobert, A. Forget, R. Biddle, and P. C. Van Oorschot, "Persuasive cued click-points: Design, implementation, and evaluation of a knowledge-based authentication mechanism," *IEEE Trans. Dependable Secure Comput.*, vol. 9, no. 2, pp. 222–235, Mar./Apr. 2012.
[12] A. De Luca, E. von Zezschwitz, N. D. H. Nguyen, M. Maurer, E. Rubegni, M. P. Scipioni, and M. Langheinrich, "Back-of-device authentication on smartphones," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2013, pp. 2389–2398.
[13] B. Dodson, D. Sengupta, D. Boneh, and M. S. Lam, "Secure, consumer-friendly web authentication and payments with a phone," in *Proc. 2nd Int. ICST Conf. Mobile Comput., Appl., Serv.*, 2010, pp. 17–38.
[14] K. M. Everitt, T. Bragin, J. Fogarty, and T. Kohno, "A comprehensive study of frequency, interference, and training of multiple graphical passwords," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2009, pp. 889–898.
[15] A. Gelman, J. Hill, and M. Yajima, "Why we (usually) don't have to worry about multiple comparisons," *J. Res. Educ. Effectiveness*, vol. 5, no. 2, pp. 189–211, 2012.
[16] A. Forget, S. Chiasson, and R. Biddle, "Shoulder-surfing resistance with eye-gaze entry in cued-recall graphical passwords," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2010 pp. 1107–1110.
[17] D. Florencio and C. Herley, "A large-scale study of web password habits," in *Proc. 16th Int. Conf. World Wide Web*, 2007, pp. 657–666.
[18] S. Hart and L. Staveland, "Development of a multi-dimensional workload rating scale," *Human Mental Workload*. New York, NY, USA: Elsevier, 1988, pp. 139–183.
[19] H. Kim and J. Huh, "Pin selection policies: Are they really effective?" *Comput. Security*, vol. 31, no. 4, pp. 484–496, 2012.
[20] G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, 91–110, 2004.
[21] J. M. McCune, A. Perrig, and M. K. Reiter, "Seeing-is-believing: Using camera phones for human verifiable authentication," *Int. J. Security Netw.*, vol. 4, no. 1/2, pp. 43–56, Feb. 2009.
[22] D. Nelson, V. Reed, and J. Walling, "Pictorial superiority effect," *J. Exp. Psychol.: Human Learning Memory*, vol. 2, no. 5, pp. 523–528, 1976.
[23] NyARToolkit. 2015. [Online]. Available: http://nyatla.jp/nyartoolkit
[24] K. Renaud and A. De Angeli, "My password is here! An investigation into visuo-spatial authentication mechanisms," *Interacting Comput.*, vol. 16, pp. 1017–1041, 2004.
[25] N. Saxena, J. E. Ekberg, K. Kostiainen, and N. Asokan, "Secure device pairing based on a visual channel (short paper)," in *Proc. IEEE Symp. Security Privacy*, 2006, pp. 306–313.
[26] B. Schneier, "Two-factor authentication: Too little, too late," *Commun. ACM*, vol. 48, no. 4, p. 136, 2005.
[27] F. Tari, A. Ozok, and S. Holden, "A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords," in *Proc. 2nd Symp. Usable Privacy Security*, 2006, pp. 56–66.
[28] J. Thorpe and P. van Oorschot, "Human-seeded attacks and exploiting hot-spots in graphical passwords," in *Proc. USENIX Security Symp.*, 2007, p. 8.
[29] P. C. van Oorschot and J. Thorpe, "Exploiting predictability in click-based graphical passwords," *J. Comput. Security* vol. 19, no. 4, pp. 669–702, 2011.
[30] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon, "Passpoints: Design and longitudinal evaluation of a graphical password system," *Int. J. Human-Comput. Stud.*, vol. 63, no. 1/2, pp. 102–127, 2005.
[31] S. Wiedenbeck, J. Waters, L. Sobrado, and J. Birget, "Design and evaluation of a shoulder-surfing resistant graphical password scheme," in *Proc. Working Conf. Adv. Visual Interfaces*, 2006, pp. 177–184.
[32] Z. Zhao and G. J. Ahn, "On the security of picture gesture authentication," in *Proc. 22nd USENIX Security Symp.*, 2013, pp. 383–398.

Andrea Bianchi received the B.S. degree in business from Università Commerciale Luigi Bocconi, Milano, Italy, in 2004, the M.S. degree in computer science from New York University, New York, NY, USA, in 2007, and the Ph.D. degree in culture technology from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2012.

He is currently an Assistant Professor with the Department of Industrial Design, KAIST. His current research interests include tangible and wearable interfaces.

Ian Oakley received the joint B.S. degree (Hons.) from the Schools of Computing Science and Psychology, University of Glasgow, Glasgow, U.K., in 1998, and the Ph.D. degree from the School of Computing Science, University of Glasgow, in 2003.

He is currently an Associate Professor with the Department of Human and System Engineering, Ulsan National Institute of Science and Technology, Ulsan, Korea. His current research interests include human–computer interaction and, specifically, multimodal, physical, tangible, and social computing.

Hyoungshick Kim received the B.S. degree from the Department of Information Engineering, Sungkyunkwan University, Seoul, Korea, the M.S. degree from the Department of Computer Science, Korea Advanced Institute of Science and Technology, Daejeon, Korea, and the Ph.D. degree from the Computer Laboratory, University of Cambridge, Cambridge, U.K., in 1999, 2001, and 2012, respectively.

He is currently an Assistant Professor with the Department of Software, Sungkyunkwan University. His current research interests include usable security and security engineering.