Covered Remote User Authentication in Steganographic Collaboration Framework

Gevorg Margarov
State Engineering University of Armenia (Polytechnic)
Yerevan, Armenia
e-mail: gmargarov@gmail.com

ABSTRACT
It is offered to cover a collaborative framework by the kind of social network type web site. As the covered remote user authentication mechanism is considered the idea of visual password which consists of click point's sequence that the user chooses on the visual images. The model to identify the most likely regions of the image for users to click in order to create visual passwords is developed. This model predicts probabilities of likely user click points. This enables to predict the entropy of a click point in a visual password for a given image. The model allows evaluating automatically whether a given image is well suited for the remote user authentication.

Keywords
Steganography, remote authentication, collaborative framework, visual password.

1. INTRODUCTION
Online information management and project collaboration is experiencing rapid growth across information-rich and highly fragmented sectors such as scientific researches, technical designing and especially software development. As the speed and volume of communication continues to increase, effective information storage and exchange in collaborative frameworks face problems of information security maintenance. These problems in many respects limit wide application of collaborative interaction through the Internet. Besides there are applications which demand enough high degree of confidentiality for preservation of commercial and other privacy.

The interesting decision of an information security problem in online cooperation is construction of the covered collaborative frameworks on the basis of steganographic approaches [1]. Thus one of the basic problems of information security maintenance in the Internet in general and in such environments in particular remains reliable and safe user authentication. In case of steganographic collaboration framework, obviously, it is necessary to give the basic attention to the covered forms of the access to the framework and remote user authentication.

The most common user authentication scheme in computer systems today is the alphanumeric password. Although alphanumeric passwords are used widely, they have certain well known drawbacks such as low memorability of high entropy passwords. These drawbacks are not due to the authentication system itself but arise from the interaction between the users and the system. Since users usually cannot remember high entropy passwords they tend to select short or simple passwords that can be simply enough broken [2]. Policies and mechanisms that force users to select high entropy passwords usually result in other unsafe practices, such as the passwords being written down and kept in the open. In order to improve the security of user authentication, alternatives to alphanumeric passwords have been proposed, e.g., token based authentication, biometrics, graphical passwords, or "multiple factors" based on the simultaneous use of two or more authentication mechanisms.

However for the covered authentication it is represented to the most interesting to define the user on sequence of his actions especially on the basis of visual images. As such visual image can be considered simple games starting windows, usual photos and other "natural" objects. In this case the main motivation is the hypothesis that people are better at remembering images than artificial words. Visual objects seem to offer a much larger set of usable passwords which are much easier for remembering, than high entropy alphanumeric passwords.

2. PHOTO BASED STEGANOGRAPHIC COLLABORATION FRAMEWORK
Steganographic collaboration framework can be covered by the kind of social network type web site. For example, it can be a web site intended for the publication of photos. As is shown on Figure 1 such site contains a set of usual photos.

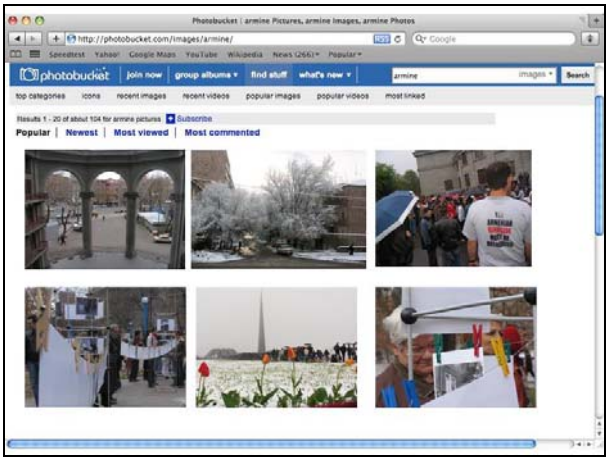

Figure 1. The web site intended for the publication of photos

For access to covered collaboration framework first of all it is necessary to click on the corresponding photo. It means that the same web site can be used as cover for a set of collaboration frameworks and the user having chosen a photo simultaneously chooses the framework for access to. After a choice of a photo the system opens it in the standard shape in a new window as it is shown on a Figure 2.

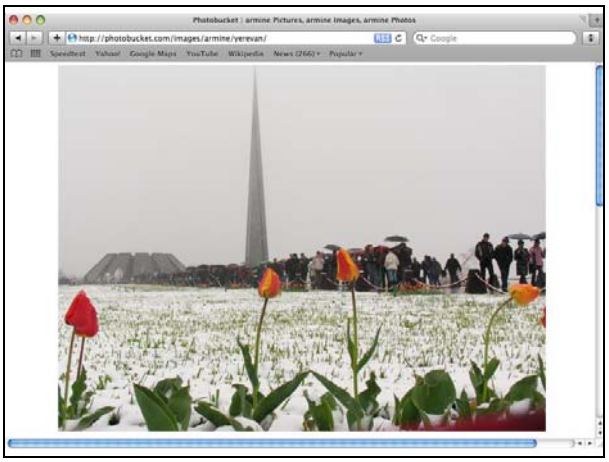

Figure 2. The photo in the new window

In this case a visual password consists of a sequence of click points (say 5 to 8) that the user chooses on the photo (Figure 3).

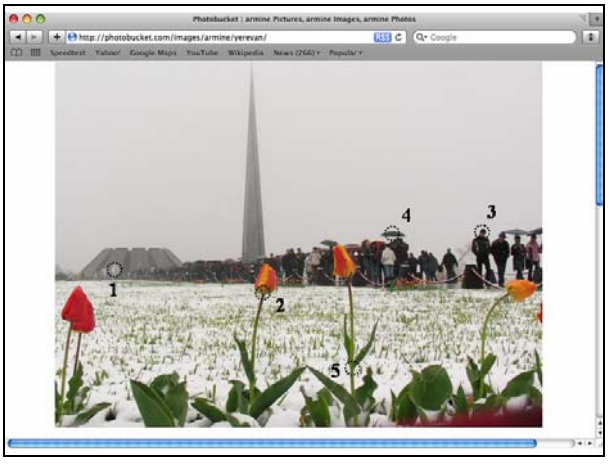

Figure 3. Five password click points (dotted circles)

It is obvious, that the photo is not secret and has no role other than helping the user remember the click points. Any pixel on the photo is a candidate for a click point.

To log in the collaboration framework, the user has to click again closely to the chosen points, in the chosen sequence. Since it is almost impossible for human users to click repeatedly on exactly the same point, the system has to allow for an error tolerance r in the click locations (e.g., a circle with radius r pixels). This can be done by discretization the click locations, using three different square grids. Each grid has width $6r$ between grid lines (horizontal or vertical). Each one of the three grids is staggered with respect to the previous grid by a distance $2r$ vertically and $2r$ horizontally (Figure 4). If there were only one discretization grid then a selected click point could be close to a grid line and small variations in the user's clicking could lead to a click in a different grid square, thus leading to the wrong password. On the other hand, it can be proved that with the three staggered grids every point in a two-dimensional photo is at distance at least r from the grid lines of at least one of the three grids.

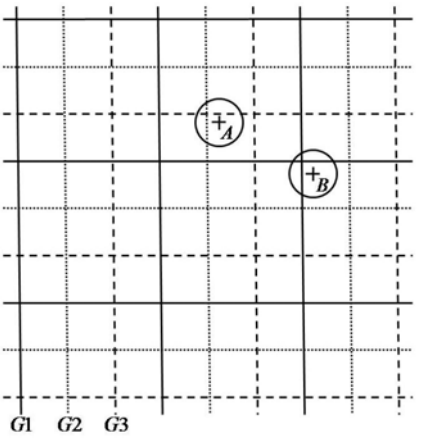

Figure 4. Three staggered grids, $G1$, $G2$, and $G3$

For definiteness it is necessary to notice, that on the Figure 3 the point A concerns to the grid $G1$, and the point B concerns grids $G2$ and $G3$.

The simultaneous use of multiple grids makes the click points "robust" against the inevitable small uncertainties in the clicking. Hence, this form of discretization is called "robust discretization". Click positions are mapped into grid squares. A sequence of click points is represented by a sequence of grids together with a sequence of grid squares. For secure storage of passwords by the system, a cryptographic hash function can be applied to the sequence of grid squares.

It is necessary to notice, that such approach does not exclude use instead of photo any other type of images, including created on the computer or scanned. Moreover on the image even it is possible to mark out in special way a set of click points. But it is possible to assume that use of natural complex real-world photos is more preferable and help users remember complex passwords better. Actually the user can not only choose one of the photos placed on a site, but also upload an own photo. This suggests that in a human context, the conditional entropy of a password will depend on the selected image. Thus there is a natural necessity to estimate conditional entropy of a click point on the chosen image, as password component. In other words, it is necessary to create a model that provides probabilities with which a user clicks on (or near) any point of the image.

3. USER CHOICE MODELLING

Classical studies on visual attention and eye movements show that most images contain a few portions that most humans focus on [3]. When asked to create a visual password a user would probably not click with the same frequency on all available pixels, but focus on some specific areas which are better remembered.

In this approach the grouping of the users' click points reduces the entropy of these click points. However, the goal is not simple observation of entropy but creation of model which enables to predict the entropy of user click points. Such model would enable to estimate productivity of automatic dictionary attacks, or to rule out certain images a priori (if they lead to low entropy). On the basis of such model it is possible to predict the most likely click locations, along with their probability values. Hence, it is possible to estimate security of the authentication system a priori (that is, before any observational user studies) and to provide a method for

selecting appropriate images that result in higher entropy of the users' click points.

In order to predict the possible click positions, a colour-based mean-shift segmentation algorithm [4] can be applied to the image. This algorithm partitions a digital image into regions, called segments, according to a given criterion in order to locate objects of interest and detects natural boundaries of visually attractive regions in an image. It produces an image partition that eliminates redundant information and highlights the important regions.

After segmentation, the centroid (centre of gravity) of each segmented region is calculated. All these centroids are weighted according to their attractiveness to humans and mapped to the grid squares of the robust discretization that was described in the previous section. The probability values of all the centroids that are mapped into the same grid square are summed, and the result is taken to be the attention probability of that square. This defines the focus of attention map which, for each of the three grids and each grid square, gives the probability that this grid square will be clicked in. At last the entropy of a click point in a given image is calculated by using the attention probabilities of the grid squares.

Some studies have shown that the user attention is influenced by both "high-level" and "low-level" factors [5]. High-level factors involve image content and memory feedback, but these factors are too complicated to be included in the model. Low-level factors are basic geometric and physical image features, such as contrast, size, shape, colour, motion, location, foreground, object category, etc. [5].

In order to compute the focus of attention map for an image, all this factors should be considered. But for modelling approach can be selected some of the above factors and combined them in a fixed way. The factors used in the model are luminosity contrast, colour contrast, and foreground of segments. Finally, studies have shown that users generally focus on people in a scene, and in particular on the eyes, mouth and hands [3].

3.1. Luminosity contrast

Contrast is the difference in visual properties that makes an object distinguishable from other objects and the background. The luminosity contrast of a segment is calculated by taking the intensity value (i.e., the grey level) of a segment and comparing it with neighbouring segments. The luminosity contrast L of a segment S_i is calculated as follows:

$$L(S_i) = \frac{1}{N_i} \sum_{j=1}^{N_i} \left| G(S_i) - G(S_{i,j}) \right|$$

where $G(S_i)$ is the grey level of the segment S_i, N_i is the number of neighbours of S_i, and $S_{i,j}$ ($j = 1, 2, ..., N_i$) are the segments that are adjacent to S_i.

3.2. Colour contrast

In addition to contrast in luminosity, contrast in hue between a segment and its surrounding is a good measure of saliency. It is computed in the HSV domain (Hue Saturation Value). Hue defines the colour value (such as blue, yellow, green) of an area, saturation measures the colourfulness of the area in proportion to its brightness. The "value" is related to the colour luminance or colour intensity. Colour contrast C is computed in the same way as luminosity contrast, but hue values $H(S_i)$ are used instead of grey levels $G(S_i)$. Before transforming RGB (Red Green Blue) into the HSV domain, RGB values are normalized to remove the brightness of the colour. Then, normalized RGB values are transformed into the HSV domain, and the hue contrast is computed as follows

$$C(S_i) = \frac{1}{N_i} \sum_{j=1}^{N_i} \left| H(S_i) - H(S_{i,j}) \right|$$

3.3. Foreground

This feature distinguishes foreground objects from background objects. The observation that background objects typically occupy very large segments, compared to foreground objects can be used. Therefore the length of the borders of segments can be used to label them as background or foreground. Next, very large regions which are likely to belong to the background of the image can be eliminated as they have lower probability of selection. The foreground feature $F(S_i)$ is calculated as follows:

$$F(S_i) = 1 - \min\left\{1, \frac{B(S_i)^{1,3}}{T}\right\}$$

where T is the total number of border pixels (for all the segments) in the image, and $B(S_i)$ is the number of border pixels of segment S_i. When $B(S_i)$ is very large then the value of the foreground feature $F(S_i)$ for segment S_i is close to zero. The exponent 1,3 in the equation is obtained empirically [5].

3.4. Combining factors

The three factors features above are combined into a final focus of attention (F_A) map for the image which takes a value between 0 and 1 and is calculated as follows:

$$F_A(S_i) = \omega_1 L(S_i) + \omega_2 C(S_i) + \omega_3 F(S_i)$$

where ω_k ($k = 1, 2, 3$) are weight factors which are fixed and can be obtained empirically or computed adaptively according to the content of the image. Contrast is the most important factor for determining the most salient regions and it is given higher weight than colour and foreground.

Once the F_A map has been computed, it is compared to a threshold which is determined empirically. Attention values under that threshold are set to zero in order to create a better focus of attention map. It can be assumed that saliency values under a certain threshold are equally likely and do not attract user attention.

For each one of the three grids, the F_A values of the points that get quantized to the same grid square are summed, which yields a focus of attention value F_G for each grid square in the grid. The F_G values of the grid squares are turned into

probabilities by dividing each F_G value by the sum of all F_G values. These probabilities are then used to predict the entropy $E(I)$ per click point in an image I :

$$E(I) = -\sum_{l=1}^{N} p_l \log_2 p_l$$

where N is the number of grid squares, p_l is the predicted probability of grid square l.

It is necessary to notice, that the given model defines the probability and entropy of a single click point. If the click points of a password with m click points were independent then the total entropy of the visual password would be $m \cdot E(I)$. However, it is not reasonable to assume independence. So, $m \cdot E(I)$ is an upper bound on the total entropy of the visual password. Obviously, dependence between click points only reduces effective value of entropy proportionally and these settlement values enough substantiate allow comparing various images. The focus of attention map and the entropy that are computed could be called a priori, as opposed to the real focus of attention map and entropy that can be obtained experimentally.

4. SUMMARY

This paper is devoted to creation of covered collaborative frameworks on the basis of steganographic approaches. In particular it is offered to cover a collaborative framework by a web site intended for the publication of photos. As the covered remote user authentication mechanism the idea of visual password is considered. The idea is to define the user on sequence of click points that he chooses on the visual images (say photo). Investigated the security of the visual password scheme and the suitability of different type of images by providing a model that predicts the users' click points and their saliency value. From this it is predicted the entropy of a click point in a visual password and the degree of suitability of the image is estimated.

The model could be improved by extending the focus of attention map so that, in addition to centroids of regions, it includes mid and end points of edges in the image, as well as corner points or tips of pointy regions. Moreover, in image segmentation, texture information may be included to get better results in natural images. In this paper it is only considered individual click points. In order to predict entire passwords the correlations between click points must considered. Finally, for the experimental test of the model would be need to collect thousands of visual password data for different types of images and different users. Even at this point we can say that when users create visual passwords they should be aware that the most salient regions can be predicted automatically with a significant probability.

REFERENCES

[1] G. Margarov, V. Markarov, A. Khachaturov, "Steganographic system with dynamically reconfigurable structure", *Proceedings of the 2009 International Conference on Security & Management, SAM'09*, Volume 1, Las Vegas, 43-45, CSREA Press, 2009.

[2] B. Ives, K. Walsh, H. Schneider, "The domino effect of password reuse", *Communications of the ACM*, Volume 47, Issue 4, 76-78, 2004.

[3] J. Senders, "Distribution of attention in static and dynamic scenes", *Proceedings of SPIE - Human Vision and Electronic Imaging II*, Volume 3016, 186-194, 1997.

[4] D. Comaniciu, P. Meer, "Mean shift: A robust approach toward feature space analysis", *IEEE Transactions on pattern analysis and machine intelligence*, 24(5), 603-619, 2002.

[5] W. Osberger, A.J. Maeder, "Automatic identification of perceptually important regions in an image", *Proceedings of 14th International Conference on Pattern Recognition*, Volume 1, 701-704, 1998.