PREVENTATIVE STEGANALYSIS IN WIRELESS VISUAL SENSOR NETWORKS: CHALLENGES AND SOLUTIONS.

Julien S. Jainsky
Texas A&M University
julienj@tamu.edu

Deepa Kundur
Texas A&M University
dkundur@tamu.edu

ABSTRACT

The goal of preventative steganalysis is to offer a proactive solution against steganography by increasing the steganalyst's knowledge of the cover-media therefore emphasizing the presence of hidden messages. This paper presents the concept of preventative steganalysis applied to wireless visual sensor networks. By means of the entropy, the uncertainty of the data captured by the network's camera is reduced, hence reducing the potential embedding capacity and discouraging the use of steganography.

Index Terms— Steganography, steganalysis, visual sensor networks.

1. INTRODUCTION

Various applications opportunities have been developed in association with wireless visual sensor networks (WVSNs). The deployment of WVSNs provides new ways to approach problems such as tactical surveillance as well as environment monitoring such as the supervision of medical patients. WVSNs via their image capturing capabilities distribute huge amounts of data which depending on the application can carry sensible information. In such cases, protecting the network against security breaches and privacy issues becomes a fundamental concern. In this paper, we introduce and study the novel yet important problem in WVSN security of preventative steganalysis. We highlight the steganographic security concerns in visual sensor networks and describe the competing goals of the steganographer (a party involved in covert communications) and the steganalyst (a party attempting to discourage covert activity) in order to derive system design principles to limit steganalytic activity in WVSNs. A steganalytic solution is developed and applied to a real case of WVSN video surveillance to illustrate the functionality of our approach.

2. BACKGROUND

Most well known measures to protect WVSNs, to date, have focused on the problem of providing privacy in vision-rich systems. Lo *et al.* [1] introduce an automated homecare monitoring system for the elderly named *UbiSense* where image processing is conducted directly at the camera to convert visual data directly into abstractions that reveal no personal information and hence protect the privacy of the monitored individuals. Fidaleo *et al.* [2] introduce the *Networked Sensor Tapestry (NeST)* architecture designed for the secure sharing, capture, and distributed processing and archiving of multimedia data. They introduce the notion of "subjective privacy" in which processing of raw sensor data is conducted to remove personally identifiable information; thus the behavior, but not the identity of an individual under surveillance is conveyed. The resulting data, approved for public viewing, is communicated in a network that employs the secure socket layer protocol and client authorization for network-level protection. Wickramasuriya *et al.* [3] present a privacy preserving video surveillance system that monitors subjects in an observation region using video cameras along with motion sensors and RFID tags. The motion detectors are used to trigger the video cameras on or off, and the RFIDs of the subjects provide authorization information in order to specify which individuals are entitled to privacy and hence have their visual information masked through image processing. Kundur *et al.*[4] present the HoLiSTiC (Heterogeneous Lightweight Sensornet for Trusted Visual Computing) framework for WVSN security that exploits secure protocols in a hierarchical directional link communication network to achieve broadband low power communications. A decentralized visual secret sharing approach is used to preserve privacy.

More recently, research has also emerged with the goal of assuring the authenticity of the data collected by sensor networks. When nodes are corrupted and provide false information, the entire network's legitimacy is compromised. The authentication of each node allows for the network to remain trusted. Several proposed solutions utilize common cryptographic concepts to provide such security. Feng et al. [5] introduce a paradigm to cryptologically embed signatures into the collected data via watermarking techniques. Their objective is to efficiently watermark the data while introducing as little distortion as possible. Zheng et al. [6] propose to offer authenticity assurance using a public key cryptographic scheme. A

978-1-61284-350-6/11/$26.00 ©2011 IEEE

derivable public key scheme is used which has the effect of simplifying the cryptography and reducing the need for key storage, therefore making it more suitable for large scale sensor networks. Because these methods still increase the workload of the WSN, Martinovic et al. [7] propose a novel paradigm that relies on the properties of wireless communications to provide authentication capabilities. They focus their study on taking advantage of frequency jamming to detect attacks and strengthen the WSN's security. Energy is always a concern when dealing with WSN where the nodes forming the network have very limited resources. As a potential solution, Blaß et al. [8] develop Extended Secure Aggregation for Wireless Sensor Networks (ESAWN). ESAWN finds a trade-off between decreasing the energy consumption of the network via data aggregation and providing authentication mechanisms although the authors agree that the security created is weaker compared to mechanisms that are not driven by energy preservation.

These existing approaches for WVSN protection all focus on protecting the *overt* data acquisition and communications systems. Fundamental questions however arise regarding the possibility of *covert* approaches for networking leading to possible breaches in both security and privacy. In this paper, we propose to study the possibility of, implications to and mitigation approaches for covert networking in the context of WVSNs.

3. SECURITY CHALLENGES

3.1. Steganography.

WVSNs are of particular interest to attackers because they convey information as frames. These images captured by every node, or camera, of the network represent a rich environment for hiding data which can be conducted via steganography. A steganographic system, as illustrated in Figure 1, usually involves two entities sharing information: the sender and the receiver. The sender can transmit data stealthily by hiding a message in an innocuous looking image or host. This embedding of information is commonly done via the use of a secret key *K* shared by both parties. The overt host information, which will be referred as *I*, is called the *cover-media*. The hidden information, denoted *W*, is also referred as a watermark. After embedding the covert information *W* within the cover-media *I*, the resulting signal is called the *stego-media, I',* and is typically perceptually identical to the cover-media hence making it difficult for another party to identify the covert communications. For instance if the media represents still or moving images then the stego-image would have to be visually identical to the cover-image. The receiver at the end of the communication channel can retrieve the hidden information *W* by means of the same key *K* used for the embedding.

From the sender's perspective, when a message *W* is to be embedded in an image *I*, it is of the utmost importance that the stego-media *I'* obtained gives up very little information about the original steganographic content. Therefore, in order for the embedded message to remain undetectable, the steganographer ideally aims at minimizing the mutual information between *I'* and *W* while doing the embedding which in this study is considered to be an additive embedding such that $I' = I + W$:

$$I(W; I') < \gamma \ \ where \ \ \gamma \to 0 \\ and \ I' = I + W \qquad (1)$$

Using any steganographic tools available to him, or her, the attacker tries to reduce the mutual information $I(W; I')$ as close to zero as possible so that the observation of *I'* gives close to no information on the watermark *W*.

The spatial redundancy of images creates an opportunity to hide large amounts of data for steganographers which is one reason why the issue of steganography is visual sensor networks requires attention.

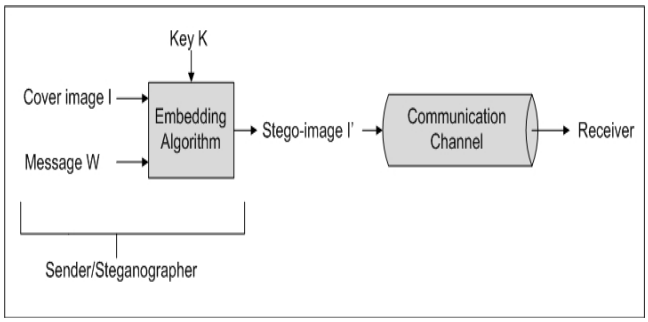

Figure 1. Steganographic system.

3.2. Steganalysis

Steganalysis describes a solution to detecting the existence of steganography. One turns to steganalysis when there is a need to detect the presence of covert communication or even eradicate potential hidden messages. Several classifications of steganalysis exist: The steganalysis can be reactive or proactive and it can involve an active or passive warden scenario. A proactive steganalysis solution is necessary when protection against a class of steganographic techniques is required. On the other hand, reactive steganalytic solutions aim at a specific steganographic embedding method. In an active warden scenario, the steganalyst acts as a middle man between the sender and the receiver of the stego-media. This means that the warden is free to apply any data processing techniques to the covert-media before

passing it to the actual receiver of the covert-communication. By doing so, the warden can deliberately transform the corrupted image and as a consequence weaken or even eradicate the presence of steganography. In a passive warden scenario however, the steganalyst can only eavesdrop on the communication between the two suspicious parties. The role of the steganalyst is therefore rather limited in the sense that it can only detect the presence of steganography but cannot prevent the covert-communication to occur. This is the reason why existing passive warden steganalytic solutions are currently not efficient enough to defend against WVSN steganography. Active warden solutions are also unsuitable for WVSNs because they usually require the use of lossy transformations in order to compromise the steganography. However, WVSNs, in the case of surveillance and monitoring applications, are required to transmit critical information which cannot be compromised by lossy data processing techniques.

The challenge that arises is for the steganalyst to create an efficient solution against covert-communications using the specifics of the network and most importantly without compromising the actual, salient data content that is carried in each image captured by the network. It is possible that certain visual processing could serve to facilitate the intended overall WVSN goal, while ensuring to a high degree that any data formerly embedded or posthumously embedded will have a chance of detection by a steganalyst. We call this concept preventative steganalysis.

The goal of preventative steganalysis is to provide security against covert communications in WVSNs by ensuring that any potential cover-media within the network has statistical characteristics such that any previously hidden data has close to no chance of being undetected by a covert receiver and any possible future data to be hidden has limited opportunity to be communicated imperceptibly within that cover-media. Stated more specifically, the presence of steganography within the cover-media has a detection probability 1-ε. We strive to design solutions that make ε approach zero so that the steganalysis can reach success rates as close to a hundred percent as possible thus discouraging any potential attacker from conducting data hiding.

For the preventative steganalysis to be successful, it becomes necessary to increase the value of the mutual information $I(W;I')$ so that when analyzing I', the presence, or absence, of steganography obviously appears. Of course, the higher the mutual information between the stego-image and the watermark is, the better the proactive steganalysis so in theory, the steganalyst needs to aim for the following maximization:

$$max[I(W;I')]\ over\ steganalytic\ tools \qquad (2)$$

This is an ideal objective for the preventative steganalysis where any tool helping the steganalyst reach this goal become maximization parameters in Equation 2.

In most cases though, the steganalyst cannot control over $I(W;I')$ because he or she only has access to the final stego-media I' without prior knowledge of the cover-media and no insight on the embedding process. This makes the steganalysis very difficult and challenging.

4. PREVENTATIVE STEGANALYSIS IN WVSNS

The sensor network environment offers elements, such as the high temporal redundancy, that could be used for steganalytic purposes so that covert-communications could be easily identified.

4.1. Theoretical considerations

Classic image steganalysis remains a difficult challenge because the steganalyst has little prior knowledge of the original cover-media. In videos however, the temporal redundancy between subsequent frames eases the job of the steganalyst who can compare close frames and decide with a low error-rate on the presence of covert communication [9].

The data collected by a node belonging to a WVSN is in a way similar to a video. In the rest of this study, we consider only one node C from the WVSN but the theory can be applied to the whole network by extending the analysis to each of the network's capturing node. The scenes recorded by C can be seen as a movie with low sampling of frames per second. This low 'fps' rate implies that, from one frame to the next, some important changes can occur. However, there is usually still a high temporal redundancy between frames which can be exploited by the steganalyst. Indeed, in such conditions, it is possible to do a frame to frame comparison in order to extract the irregularities that might occur in the presence of steganography. It is important to wisely choose a reference frame in order to make these comparisons. For instance, if comparisons are made between two corrupted frames, the steganalytic decision might be different than the decision made from the comparison against a healthy frame. This is the reason why we propose to use the first, initial captured frames which can be collected during the calibrating process for example. This reference frame can be seen as the common background for any image capture and is denoted as B.

This background frame B is an essential source of uncorrupted information for node C which can help the work of the steganalyst. Using this knowledge, the mutual information between the watermark W and the stego-frame I' can already be increased by looking at the conditional information between W and I' given the frame B:

$$I(W;I') \leq I(W;I'|B) \qquad (3)$$

This motivates the need for maximizing the conditional mutual information instead of the simplistic $I(W;I')$, which means that the steganalyst goal derived from Equation 2 is now to solve the following equation.

$$max[I(W;I'|B)]\ over\ steganalytic\ tools \quad (4)$$

As in Equation 2, Equation 4 states the ideal goal for the steganalyst which may not be achievable. However, if the maximum of $I(W;I'|B)$ is not reachable, the steganalyst should aim for a substantial increase in the mutual information $I(W;I'|B)$.
Equation 4 can be developed using entropy functions assuming the background B and the watermark W are independent:

$$\begin{aligned} I(W;I'|B) &= H(W|B) - H(W|I',B) \\ &= H(W) - H(W|I',B) \end{aligned} \quad (5)$$

To maintain initial generality of results, one often assumes as little as possible of the watermark, $H(W)$ cannot be controlled by the steganalyst to increase the result of Equation 5. Therefore increasing the conditional mutual information between W and I' given B means reducing the conditional entropy of W given I' and B:

$$\begin{aligned} &increasing\ I(W;I'|B) \\ &\Leftrightarrow reducing\ H(W|I',B) \end{aligned} \quad (6)$$

Reduction of the entropy above literally means reducing the uncertainty about the watermark when I' and B are known. This can be achieved when W is a function of I' and B. We also know from Equation 1 that $I' = I + W$ which further implies that the conditional entropy of Equation 6 can be reduced if I can be expressed as a function of B. Of course, the captured frame I usually includes more than just the background B. It conveys data and most likely some noise. This is the reason why we propose the following model:

$$I = B + D + N \quad (7)$$

where:

- B is the reference frame or common background,
- D is the data that bears interest to the network's purpose,
- N is some noise that includes the difference between the reference frame and the actual frame (excluding the data) as well as some potential additive watermark,
- B, D and N are assumed to be independent from one another.

In order for the two frames I and B to be similar, the conditional entropy of I given B must be minimized:

$$H(I|B) < \rho\ \ where\ \ \rho \to 0 \quad (8)$$

Since image I is assumed to be composed of three independent components in Equation 7, Equation 8 can be rewritten:

$$\begin{aligned} H(I|B) &\leq H(B|B) + H(N+D|B) \\ &\leq H(N|B) + H(D|B) \end{aligned} \quad (9)$$

The background and the data are important as they give information on the event happening at a specific location. On the other hand, the noise N should be removed as much as possible as they contain no important information and generally corrupt the integrity of the data. We propose to do so simply by filtering the noise in order to reduce its intensity or even erase it if possible.

4.2. Application

For the purpose of illustration, we will use two frames taken from a hall monitoring sequence [10] which can be seen in Figure 2.

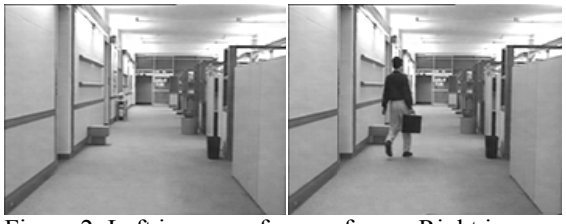

Figure 2. Left image: reference frame. Right image: Captured frame.

The left image in Figure 2 is the reference frame corresponding to the initial frame collected by the network. It represents the background B. The right image in Figure 2 corresponds to the captured frame I. By computing the difference between frames B and I, we can isolate the discrepancies between the two images which consist in a mix of noise N and important data D.

From the frame difference shown in Figure 3, it is apparent that the majority of the two images are identical. This was expected due to the high temporal correlation between B and

I. This observation helps to justify, in part, the additional assumption that the actual data conveyed by *I* is visually present in minority in I. This further implies that, when computing the difference with the reference background *B*, black pixels will be more visible and therefore the percentage of black pixels occupying the image difference between *I* and *B* is assumed to be over fifty percent:

$$P_B = \frac{number\ of black\ pixels}{total\ number\ of\ pixels} > 50\% \qquad (10)$$

Figure 3. Image difference between captured and reference frames.

Figure 3 also brings out information about the actual data which is composed of lighter pixels. In this specific case the data corresponds to a man or woman carrying a brief case. The rest of the lighter pixels constitutes the noise *N*. *N* is assumed to be a random distribution of mean equal to a neutral *128* since we are dealing with image pixel where values range from *0* to *255* on the gray scale level. The noise is expected to be rather weak compared to the main background and actual data so it is further assumed that *N* corrupt α percent of the image. This means that α percent of the black pixels and α percent of the lighter pixels are corrupted by noise.
If one were to apply an averaging filter to the image expected average pixel intensity after filtering can be computed from the equation:

$$pixel\ intensity\ of\ darker\ area$$
$$= \frac{(100-\alpha)\times 0 + \alpha \times 128}{100} = \frac{\alpha \times 128}{100} \qquad (11)$$

In the same manner, the intensity of the pixels belonging originally to the lighter area (i.e., the data) after the averaging filter has been applied will globally converge to the value:

$$pixel\ intensity\ of\ lighter\ area$$
$$= \frac{(100-\alpha)\times 255 + \alpha \times 128}{100} \qquad (12)$$

Figure 4. Filtered image on the left and mask after threshold on the right.

After filtering, a thresholding step occurs in order to eliminate or at least reduce the presence of noise. The black and white frame obtained, called mask shown on the right in Figure 4, will be used to reconstruct an estimation of the original image by *XOR* the reference background with the original image via the mask computed: the background *B* will take the place of black pixels whereas white pixels will give in to pixels from the original frame *I*. The reconstructed image is denoted I_R and is presented in Figure 5. Doing so will maximize the temporal redundancy of the data acquired since most of the image will correspond to the known background *B* and hence the conditional entropy $H(I_R|B)$ is expected to be reduced.

Figure 5. Reconstructed image $\boldsymbol{I_R}$.

4.3. Simulations

The same process as described in the previous section has been programmed with MATLAB and tested on the whole hall monitoring sequence. The conditional entropies $H(I|B)$ and $H(I_R|B)$ are estimated via the computation of the histograms of involved images and compared in order to check whether the goal we had set has been achieved.
The simulations' results of Figure 6 shows a definite improvement on the entropy due to the increase in similarities between I_R and *B* after the noise *N* has been filtered and the background in *I* replaced by the reference frame *B*.
The mutual information $I(W;I')$ and $I(W;I'|B)$ have also been computed after the images' histograms have been estimated. A 128-mean simulated watermark *W* has been added to corrupt the sequence frames. Results are shown in Figure 7.

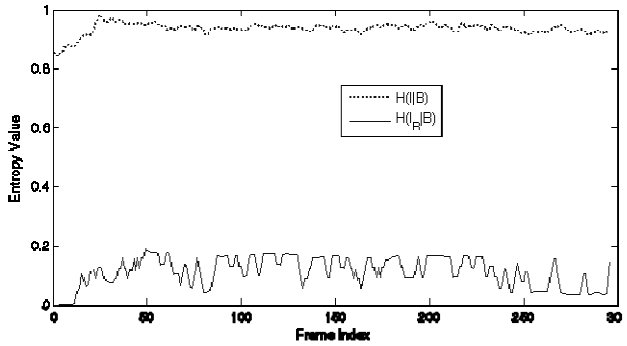

Figure 6. Comparison of conditional entropies $\boldsymbol{H(I|B)}$ and $\boldsymbol{H(I_R|B)}$.

As desired, the mutual information $I(W;I'|B)$ after applying our algorithm is increased. It is approximately twice the mutual information $I(W;I')$ when no processing is performed. This means that when inspecting a frame travelling within the network, the potential watermark is more apparent and as a consequence, the job of the steganalyst is made easier.

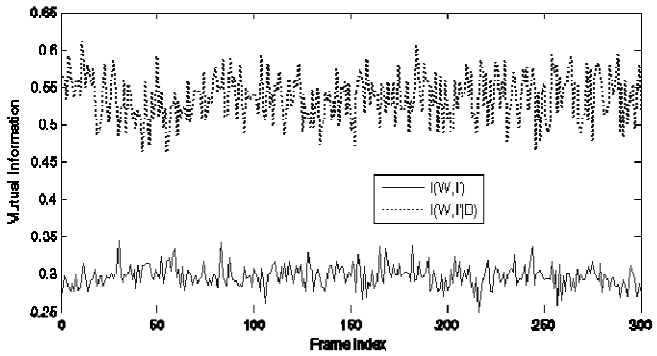

Figure 7. Comparison of Mutual Information $\boldsymbol{I(W;I')}$ and $\boldsymbol{I(W;I'|B)}$.

5. CONCLUSION

In this paper, we have introduced the novel concept of preventative steganalysis. We believe this is a promising avenue to discourage the use of steganography when the cover-media belong to the trusted party. Although the proposed solution proves to be successful in increasing the mutual information $I(W;I'|B)$ i.e. making the watermark more obvious to the steganalyst, only visual inspections have assured the data is still identifiable. While promising, preventative steganalysis obviously needs to be investigated further. It is of course necessary to find new ways to prevent steganography in a proactive manner but it is also imperative to develop customized solutions adapted to the specifics of the network such as its function, its size and its available resources.

6. REFERENCES

[1] B. P. L. Lo, J. L. Wang, and G.-Z. Yang, "From imaging networks to behavior profiling: Ubiquitous sensing for managed homecare of the elderly", *Adjunct Proceedings of the 3rd International Conference on Pervasive Computing*, Munich, Germany, pp. 101–104, May 2005.

[2] D. A. Fidaleo, H.-A. Nguyen, and M. Trivedi, "The networked sensor tapestry (NeST): A privacy enhanced software architecture for interactive analysis of data in video-sensor networks", *Proceedings of the ACM 2nd International Workshop on Video Surveillance & Sensor Networks*, New York, USA, pp. 46–53, October 2004.

[3] J. Wickramasuriya, M. Datt, S. Mehrotra and N. Venkatasubramanian, "Privacy protecting data collection in media spaces", *Proceedings of the 12th annual ACM International Conference on Multimedia*, New York, USA, pp. 48–55, October 2004.

[4] D. Kundur, W. Luh, U.N. Okorafor, and T. Zourntos, "Security and Privacy for Distributed Multimedia Sensor Networks", *Proceedings of the IEEE Special Issue on Distributed Multimedia*, vol. 96, no. 1, pp. 112-130, January 2008.

[5] J. Feng and M. Potkonjak, "Security in sensor networks: watermarking techniques", *Proceedings of the SPIE, Security and Watermarking of Multimedia Contents V,* vol. 5020, pp. 391-402, June 2003.

[6] J. Zheng, J. Li, M. J. Lee and M. Anshel, "A lightweight encryption and authentication scheme for wireless sensor networks", *International Journal of Security and Networks*, vol.1, Issue 3, pp. 138-146, December 2006.

[7] I. Martinovic, P. Pichota and J. B. Schmitt, "Jamming for Good: A Fresh Approach to Authentic Communication in WSNs", *Proceedings of the Second ACM Conference on Wireless Network Security*, Zurich, Switzerland, pp. 161-168, March 2009.

[8] E. Blaß, J. Wilke and M. Zitterbart, "Relaxed authenticity for data aggregation in wireless sensor networks", *Proceedings of the 4th International Conference on Security and Privacy in Communication Networks*, Istanbul, Turkey, pp. 1-10, September 2008.

[9] J. Jainsky, D. Kundur, and D. Halverson, "Towards digital video steganalysis using asymptotic memoryless detection", *Proceedings of the 9th Workshop on Multimedia & Security*. Dallas, TX, USA, pp. 161-168, September 2007.

[10] *Hall Monitoring Sequence*, http://trace.eas.asu.edu/yuv/index.html.