ON PRIVACY AND SECURITY IN DISTRIBUTED VISUAL SENSOR NETWORKS

Alexandra Czarlinska, William Luh, and Deepa Kundur

Department of Electrical and Computer Engineering
Texas A&M University
College Station, TX 77843-3128 USA

ABSTRACT

There is a critical need to provide privacy assurances for distributed vision-based sensor networking in applications such as building surveillance and healthcare monitoring. To effectively address protection and reliability issues, secure networking and processing must be considered from system inception. This paper presents attacks that affect the data privacy in visual sensor networks and proposes privacy-promoting security solutions based on opponent detection via game-theoretic analysis and keyless encryption.

Index Terms— Visual sensor networks, privacy of visual data, network security, keyless privacy.

1. INTRODUCTION

In this work, we consider privacy and security mechanisms for a heterogeneous wireless visual sensor network (VSN). The network is comprised of wirelessly communicating camera nodes and scalar sensors where the sensors trigger the cameras and provide specific privacy guarantees based on event detection. The network may be deployed in one or more zones such as throughout a building and its perimeters. The scalar sensor network may contain a number of malicious nodes that have been remotely or locally re-programmed to disinform regarding the presence of an event for the purpose of enabling and disabling camera privacy settings. Upon acquisition of the privacy enabled or disabled frames, camera nodes must encrypt and wirelessly transmit them to a local or off-site sink even if secret keys have been compromised. The surveillance network must thus remain dependable and provide privacy as well as confidentiality of surveillance despite disruption and disinformation activities.

Measures to protect VSNs, to date, have focused on the problem of providing privacy by directly obfuscating the visual data. For example, Lo *et al.* [1] propose image processing conducted directly at the cameras that converts the video information into abstractions containing only the information necessary to detect salient human characteristics. Only the abstractions are communicated and processed within the network, providing privacy. Fidaleo *et al.* [2] introduce the notion of "subjective privacy" in video where only the behavior, but not the identity of an individual under surveillance is conveyed. Their approach to privacy involves processing of the raw sensor data in order to remove personally identifiable information. The resulting data, approved for public viewing, is communicated with the aid of cryptographic security measures for further processing. Wickramasuriya *et al.* [3] present a privacy preserving video surveillance system that monitors subjects in an observation region using video cameras along with localized sensors. The localized sensors include radio frequency identification (RFID) tags that subjects wear and motion detectors placed within the observation environment. 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. The information from the various sensors are fused with the video data resulting in a video stream with only authorized subjects being masked through image processing.

In this paper, we take a non-obfuscation approach to addressing privacy issues in VSNs. Given the growing trend in VSN design that promotes the interaction of both visual and scalar sensors for both performance and privacy [3], we assert in this paper that the privacy preservation problem in VSNs is intrinsically tied to certain network security issues. For example, a VSN that depends on motion detectors or RFID sensors to trigger cameras on or obfuscate sensitive information is susceptible to networking attacks that will hinder the privacy-guarantees of the system. Equally important is the fact that sensor networks are susceptible to attacks such as sensor node tampering and compromise that make secret cryptographic keys briefly unavailable, thus once again, preventing some types of privacy-enabling processing. Thus in this paper we focus on strategies that will promote privacy in the face of these attacks.

2. VISUAL SURVEILLANCE SENSOR NETWORKS

2.1. VSN Model

We consider a VSN comprised of camera nodes and scalar sensor nodes that trigger regional cameras and enable or disable local privacy rights. For ease of (re-)deployment, the camera nodes transmit their acquired visual surveillance wirelessly to a (local or remote) network sink such as a surveillance center. This heterogeneous network is deployed in a large region such as a building that is comprised of many smaller local zones z_i as shown in Figure 1 that can, for example, correspond to different rooms and can include outdoor perimeter areas.

Each camera node may rely on a decision obtained directly from one or more sensors regarding the presence of events. We note however that with some probability, each individual sensor may be compromised and re-programmed by an attacker (a sensor may also occasionally malfunction). In such cases, a sensor may neglect reporting an event or may signal for privacy settings to remain in place despite a possible intrusion. Alternatively, a captured sensor may falsely trigger events to drain the wireless resources of the network or detract attention from a zone with genuine activity. In this setting, the binary decision of an individual sensor regarding the presence of an event (under no attack) is modeled as a Bernoulli random variable $Bern(p)$ where p is the probability of the presence of event. We note that if the probability distribution function (PDF) of the phenomenon under observation is known, the probability p of an event is the area under the PDF to the right of the event-decision threshold T_h.

To mitigate the possible re-programming of a number of sensors, we consider the setup shown in Figure 1. Specifically, each zone contains an intermediate cluster head that receives decisions from sensors and performs attack detection upon these decisions. As shown in Figure 1, some of the sensors are legitimate while others may be hostile due to attack. Based on its attack detection results, the cluster head provides feedback to the camera nodes regarding privacy settings and regarding the presence of events. When the presence of an event is confirmed by the cluster head, each notified camera enciphers the acquired frames for wireless transmission to a sink. We note that the cluster head task may be performed by a dedicated *secure* node or may be rotated among several sensors to mitigate attacks and limit the energy drain of any given node.

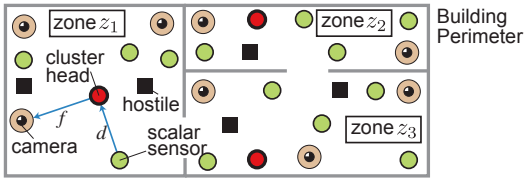

Fig. 1. Sensor sends possibly attacked decision d to cluster-head while cameras acquire frames. The frames are processed based on cluster head's feedback f.

2.2. Network Attacks on VSN Privacy

In VSN settings, an opponent may engage in active sensor attacks that disrupt event detection and privacy-preservation to perpetrate undetected intrusions on the premises [4]. The attacker may also additionally eavesdrop on the network links to intercept camera frames. Interception of camera data through eavesdropping, such as during its wireless transit to an off-site sink, provides the attacker with valuable information regarding the premises.

From the point of view of the cluster head, each sensor reporting its binary decision may be in error due to attack with some unknown probability q [5]. The effect of the active attack is to alter a decision 0 (privacy enabled) to a 1 (privacy disabled), or vice versa and must thus be mitigated. The cluster head is thus faced with the task of attack detection given a potentially pervasive attack with unknown attack probability. Sensor networks also suffer from a high likelihood of key compromise such as through node tampering that reveals keying information and jeopardizes privacy services. The ensuing network re-keying effort utilizes energy and might cause unacceptable delays. Thus, keyless methods of preserving privacy even for a short-term period are of value when pervasive privacy-protected surveillance is required. The model of wireless eavesdropping in this paper involves a distributed attacker who can capture a fraction of encrypted results from $m < n$ cameras while being transmitted to the sink motivating the study of keyless encryption.

3. PROPOSED VSN PRIVACY PROTECTION STRATEGIES

3.1. Sensor Attack Detection and Reaction

In pervasive surveillance settings, it cannot be assumed that scalar support data is always reliable and that encryption keys are available. To secure the data gathering process in the face of sensor attacks, we wish to verify whether the scalar sensor decisions d_i for $i \in \{1, n\}$ collected by the n scalar sensors are legitimate or if they have been altered. Thus, ideally we wish to distinguish between the hypothesis $\mathcal{H}_0$ where the d_i have not been tampered with and the hypothesis $\mathcal{H}_1$ where the d_i have been altered. In the case of $\mathcal{H}_0$ the random variables d_i should come from the Bernoulli distribution $Bern(p)$ where p is the probability of an event. The case of $\mathcal{H}_1$ presents a challenge since the actions of the captured nodes are unknown. If each hostile node attacks with some probability q, then d_i will come from $Bern(r)$ where $r = p + q - 2pq$ but where q (and thus r) still remain unknown.

For the attack to remain undetected at the cluster head however, the probability of attack q should be chosen such that the scalar data received by the cluster-head appears plausible. This signifies that the scalar decisions $\mathbf{d}$ where $\mathbf{d} = [d_1, \cdots, d_n]$ should have a weight (number of 1s) that is plausibly close to np (especially for large n where the actual weight approaches np on average). Since the attacker does not know the exact value of p (it depends both on the PDF of the sensed event and the threshold T_h), the optimal choice of q may be determined based on game theoretic analysis [5], the results of which we now summarize. The attacking network must select a value of q that is *small* and that *decreases* with increasing cluster size n to be stealthy. This optimal value provides an estimate for the parameter q that is missing from the $\mathcal{H}_1$ hypothesis.

We thus now consider the problem from the point of view of attack detection where $\mathcal{H}_1$ is the attack hypothesis and $\mathcal{H}_0$ is the non-attack hypothesis. The optimal Neyman-Pearson (NP) detector to distinguish between the two hypotheses is given by Eq. 1 where w is the weight of the data $\mathbf{d}$ and $\mathcal{T}$ is a threshold chosen based on a desired probability of false alarm α. Importantly as shown in Eq. 1, while $\mathcal{T}$ can be determined without knowledge of the attack parameter q, the resulting probability of detection β cannot be determined without it. Thus, the analysis of [5] provides the missing parameter required to determine the optimal detector performance.

$$w \underset{H_0}{\overset{H_1}{\gtrless}} \mathcal{T}(p, \alpha) \quad \text{where} \quad \alpha(p), \beta(q, p) \tag{1}$$

The above approach may be employed in a variety of VSN applications. For the VSN of Section 2.1, the process follows:

(1) NP test: The threshold of the detector is set based on an application-dependent probability of false alarm α which is generally chosen to be very small. Upon receiving the decisions of the n scalar sensors, the cluster-head employs the NP detector of Eq. 1. The cluster-head proceeds based on the outcome of the test ($\mathcal{H}_0$ or $\mathcal{H}_1$).

(2) $\mathcal{H}_0$ (Non-Attack Hypothesis): In this case the cluster-head trusts d_i for $i \in \{1, n\}$. Based on each $d_i \in \{0, 1\}$, the cluster-head notifies each corresponding camera if it should encipher and transmit its surveillance or if it should discard the sequence captured to-date.

(3) $\mathcal{H}_1$ (Attack Hypothesis): If the result of the detector is the $\mathcal{H}_1$ hypothesis, then an attack is assumed (since α is chosen to be small). It is known that approximately *only* $nq(n)$ nodes are expected to be in error where $nq(n)$ is small [5]. What is not known is *which* nodes are in error and appropriate intrusion detection approaches may be applied. In a VSN application, if privacy is or paramount importance, then no visual data need be communicated. If event surveillance is also to be prioritized then, cameras corresponding to a 1 can encipher and transmit their frames while the cameras that correspond to 0 employ localized image processing [5, 6] to verify whether their frames are indeed "event-empty."

3.2. Keyless Visual Encryption

Pervasive privacy-enabled VSN surveillance necessitates the need for keyless confidentiality of wirelessly communicated visual infor-

mation. To achieve confidentiality at any time, we propose a scheme intended to compliment key-based cryptographic solutions by replacing those methods during periods of key unavailability.

The proposed scheme is tailored to the case where n camera nodes in a cluster capture correlated visual surveillance but where some (or all) of them do not have encryption keys with the sink (and/or with each other). During this time of unavailability, the camera nodes should still be able to "encipher" the surveillance efficiently *without* having to transmit the frames to each other. In other words, a distributed scheme where each camera performs the enciphering separately is desirable. The scheme that we thus develop and overview in this work is *based* on the principles of distributed source coding where correlated data may be compressed separately yet optimally given that some correlation statistics about the data are known. We develop a distributed scheme for *visual* surveillance that provides *confidentiality* maintaining privacy even if $m < n$ of the "enciphered" shares are intercepted by the hostile network.

We first overview our basic scheme called S/DISCUS (secure distributed source coding using syndromes) and then present a novel algorithm for using S/DISCUS on *visual* surveillance. Suppose that a cluster contains n nodes (i.e. cameras) where each node i captures surveillance data U_i modeled as a string of k symbols; here, the nodes capture $U_1^k, U_2^k, \ldots, U_n^k$ where the symbols are from a finite field.[1] Suppose that these n strings of surveillance are not independent but correlated via a parameter t in the following sense:

$$w(U_1^k + U_2^k + \cdots + U_n^k) \leq t, \tag{2}$$

where $w(\cdot)$ is the weight (i.e. the number of symbols that do not match). The correlation model is expressing the observation that $U_1^k, U_2^k, \ldots, U_n^k$ are similar with only a few differences among the strings. We can see this more clearly if we consider a finite field of order say 2^8 (for example the number of grayscale levels in a digital grayscale image, or one RGB plane of a digital colored image). In this case the correlation in Eq. 2 restricts the number of different pixels among the n nodes to only t nodes (in practice, *visual* surveillance may not obey this correlation model thus necessitating an appropriate algorithm for its use). Given this model and representation of the surveillance data, each node using S/DISCUS inputs its U_i^k into its own simple shift register circuit (the tap coefficient design is detailed in [7]). This distributed approach using readily implementable shift-registers results in the following desirable properties: (1) The output string of each shift register circuit is shorter than the k input symbols providing compression; (2) The visual data is secure against a distributed eavesdropper (such as a hostile network) that captures $m < n$ outputs of the shift registers; and (3) The sink can reconstruct each of $U_1^k, U_2^k, \ldots, U_n^k$ perfectly from the received shares without the use of decryption keys.

Importantly the eavesdropper's ignorance is true *even* if he knows the exact coefficients of all the shift register circuits (such as through node capture). Furthermore, given infinite time and resources, the eavesdropper cannot reduce the cardinality of the message set down to 1; that is, the eavesdropper cannot solve for the message (surveillance data) [7].

We now consider the case of visual surveillance data. Normally S/DISCUS performs both lightweight encryption as well as compression when the correlation model of Eq. 2 is satisfied (and does so such that the decoder can perfectly reconstruct all the messages given all the shares). When S/DISCUS is applied to images that depict a common scene from *different* orientations and perspectives, the correlation model of Eq. 2 typically does not hold. One solution is to have the cameras locally register their images prior to using S/DISCUS. Such registration however typically requires (distributed) camera calibrations [8, 9] and may not be desirable for the VSN setting.

[1]The actual data itself may not appear in the form of a string of symbols. In practice the data collected by a node is grouped appropriately and can be mapped to symbols in a finite field at the discretion of the engineer/designer, in the same way that images and audio are often encoded.

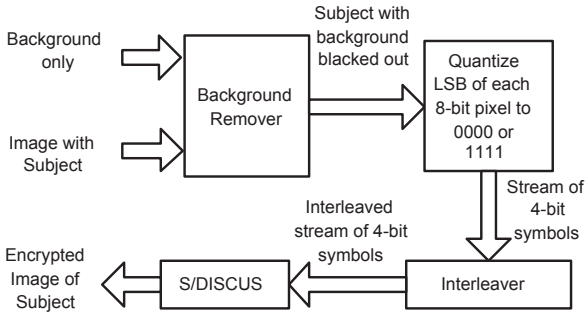

Fig. 2. Image Encoder

We propose an alternative solution with local preprocessing as outlined in Figure 2 with the goal of achieving sufficient invariance such that the correlation model of Eq. 2 may be satisfied. Importantly, the invariances and variances must be distributed uniformly since otherwise some portions of the input stream will satisfy the correlation model, while large portions (particularly important features) will not and therefore be undecodable.

As shown in Figure 2, the proposed solution requires that a background scene image be available for each camera; this may be periodically captured by the cameras (when events are not detected) and relayed to the sink without encryption. Using the background, a subtraction algorithm (such as the one used in [6]) is applied to an event image, so that the background pixels can be set to a constant (for example black) providing a basic source of invariance. To achieve a higher guarantee of invariability, the 4 least significant bits (LSB) of each 8-bit pixel are also quantized. The reasoning is that *adjacent* pixels of images will likely get quantized to the same LSB value thus providing further invariance (though the quantization process introduces some irreversible distortion, in practice the distortion is not prohibitive as shown experimentally in Section 4). Once the invariance has been obtained, it must be spread across the input which is accomplished through the use of an interleaver that permutes the pixel positions. This interleaver is deterministic in practice and may be known to the opponent without compromising the security since its only purpose is to transform the input stream into one that better satisfies the correlation model. Finally S/DISCUS is applied to the input stream as shown in Section 4.

4. PERFORMANCE AND DISCUSSION

For scalar data gathering in the stealthy sensor attack scenario, an optimal NP detector can be employed to detect hostile behavior. This step indeed also mitigates the presence of the attack due to the SANs' need to be stealthy. Conceptually, as the cluster size increases (there are more data points taken), the attacker's *optimal* attack parameter q decreases signifying that picking a larger cluster reduces the (average expected) fraction of attacked nodes. However as q becomes smaller, it is harder for the optimal NP detector to detect the attack. Thus there is an inherent trade-off in the process of cluster size selection that affects mitigation and detection simultaneously.

The relationship between detection and mitigation is depicted in Figure 3(a) for a commonly used value of probability of false alarm $\alpha = 0.05$ (other α's yield similar plots). The plot shows results for various probabilities of event p for $p \in \{0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$ (results for $1-p$ are identical due to symmetry). For each p, the cluster size n is varied over $n \in \{1, 2, 3, 5, 10, 20, 30, 40, 50\}$. For each (p, n) pair, the optimal attack probability q is determined from [5] and the corresponding probability of detection is found from Eq. 1 and [4]. The horizontal axis depicts the optimal attack parameter *and* also corresponds to the (average expected) *fraction* of attacked nodes (recalling that each q corresponds to a different n). We observe that the detection performance P_D is best for small p (which might be the case over long periods of time). Whether p is known or unknown, however, for a desired α we can examine the P_D-q curve and select a suitable trade-off point from which the required cluster size n can be determined for attack mitigation.

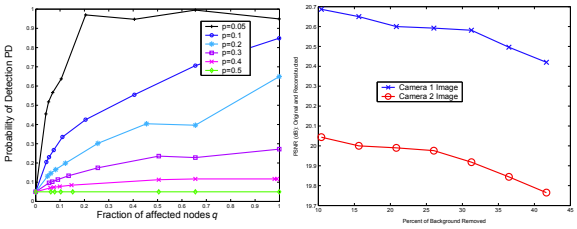

Fig. 3. (a) Probability of Detection P_D (vertical) vs. *optimal* attack q (horizontal) for various probabilities of an event p for $\alpha = 0.05$. (b) Tradeoff between Reconstruction Quality (vertical) and Background Redundancy (horizontal).

Fig. 4. Original image from (a) Camera 1 and (c) Camera 2. Reconstructed images at the sink (5×5 median filtered)(b) PSNR of 20.8 dB (d) PSNR of 20.1 dB.

Next we wish to examine the performance of the S/DISCUS paradigm for enciphering correlated images collected from cameras using *different perspectives* without image registration. The experiments were performed using the S/DISCUS scheme with *two* cameras and input blocks of 15 4-bit symbols (where the 4 LSB of each 8-bit pixel are quantized). The original images were captured in poor lighting conditions as shown in Figures 4(a) and (c) and reconstructed at the sink with PSNR (peak signal to noise ratio) of 20.8 dB and 20.1 dB as shown in Figures 4(b) and (d) respectively. Importantly, these PSNR values correspond to the *full* use of the background to achieve invariance and thus correspond to the case of *full background redundancy*. As shown in Figure 3(b) however, the PSNR and redundancy characteristics may be traded-off by selecting the percent of background material that is removed (i.e. not utilized to achieve invariance). The ability to trade-off the desired PSNR and redundancy is an important characteristics for wireless cameras. Finally we note that the encryption achieved with the low-complexity S/DISCUS for the images in Figure 4 confounds an eavesdropper by giving approximately 36 possible pixel values for *each* pixel. Based on its distributed enciphering and PSNR/redundancy flexibility, the S/DISCUS scheme may be well-suited for certain privacy-enabling VSN surveillance applications where keys are temporarily unavailable due to hostile or challenging conditions.

5. CONCLUSIONS

In this work, we present approaches for addressing attacks on privacy in emerging VSNs. Given the interaction of scalar and visual sensors within emerging VSNs, the privacy problems is intrinsically tied to aspects of network security. The proposed methodology for reliable gathering of scalar support data offers compromises between mitigation and detection that are important for network design. The proposed technique for keyless enciphering of correlated visual data is shown to perform in a distributed scenario without the need for inter-node communication thus demonstrating potential for pervasive privacy-preservation in VSNs.

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," in *Adjunct Proc. International Conference on Pervasive Computing*, Munich, Germany, May 2005, pp. 101–104.

[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," in *Proc. ACM International Workshop on Video Surveillance and Sensor Networks*, New York, NY, October 2004, pp. 46–53.

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

[4] A. Czarlinska and D. Kundur, "Attack vs. failure detection in event-driven wireless visual sensor networks," in *ACM Multimedia & Security Workshop (MM&Sec'07)*, Dallas, TX, 20-21 September 2007.

[5] A. Czarlinska and D. Kundur, "Reliable Event-Detection in Wireless Visual Sensor Networks through Scalar Collaboration and Game Theoretic Consideration," *IEEE Transactions on Multimedia, Special Issue on Multimedia Applications in Mobile/Wireless Contexts, accepted.*, Available: www.ece.tamu.edu/˜czlinska/TMM08/WVSN.pdf.

[6] M. Wu and C. W. Chen, "Collaborative image coding and transmission over wireless sensor networks," *EURASIP Journal on Advances in Signal Processing*, , no. 70481, 2007.

[7] W. Luh and D. Kundur, "Secure distributed source coding with side-information," *IEEE Communications Letters*, 2008, Accepted.

[8] D. Devarajan, R. J. Radke, and H. Chung, "Distributed metric calibration of ad hoc camera networks," *ACM Trans. on Sensor Networks*, vol. 2, no. 3, pp. 380–403, August 2006.

[9] A. Barton-Sweeney, D. Lymberopoulos, and A. Savvides, "Sensor localization and camera calibration in distributed camera sensor networks," in *Proc. of IEEE BaseNets*, San Jose, CA, October 2006.