

# Detecting Sensitive Data Exfiltration by an Insider Attack

Yali Liu  
Electrical & Computer Engineering  
University of California, Davis  
Davis, CA 95616 USA  
yliu@ece.ucdavis.edu

Cherita Corbett  
and Ken Chiang  
Sandia\* National Laboratories  
Livermore, CA 94551, USA  
{clcorbe, kchiang}@sandia.gov

Rennie Archibald, Biswanath  
Mukherjee and Dipak Ghosal  
Department of Computer Science  
University of California, Davis  
Davis, CA 95616 USA  
{rvarchibald, mukherjee,  
ghosal}@cs.ucdavis.edu

## ABSTRACT

Detecting and mitigating insider threat is a critical element in the overall information protection strategy. By successfully implementing tactics to detect this threat, organizations avoid the loss of sensitive information and also potentially protect against future attacks. Within the broader scope of mitigating insider threat, we focus on detecting exfiltration of sensitive data through the high speed network. We propose a multilevel approach that consists of three main components: 1) network level application identification, 2) content signature generation and detection, and 3) covert communication detection. The key scientific approach used for all the above components is applying statistical and signal processing techniques on network traffic to generate signatures and/or extract features for classification purposes. We provide a summary of the approaches used in network level application identification and content signature generation and detection and briefly describe our approach in detecting covert communications. This paper touches on these issues and outlines overall directions for our research.

## Categories and Subject Descriptors

C.2.0 [Computer-Communication Networks]:  
General—Security and Protection

## General Terms

Security

## Keywords

Data exfiltration, insider attack, traffic classification, feature extraction, signatures, covert communication.

## 1. INTRODUCTION

In widely connected environments, a successful insider attack could result in serious damage to the interests of an enterprise. An insider attack can occur through inadvertent security breach by an authorized user or through a planned security breach by an authorized user or by an outsider through a compromised system. Planned insider attack can result in exfiltration or destruction of sensitive data or compromise the communication networks and various network servers and resources.

Within the broader goals of mitigating insider attacks, our work addresses the detection (not the deterrence and prevention) of deliberate and intended distribution of sensitive content outside the organization (exfiltration of sensitive content) using the organization's system and network resources by a trusted insider. The key scientific approach uses a method of generating signatures of sensitive content that are unalterable and applying wavelet to compact the signatures for purposes of scalability. These signatures are then used by detection systems placed at the network egress points to detect and prevent the leakage of sensitive information. A key technical challenge is detection despite transformations being applied to the content, e.g. encryption, modulation by the transport protocol, slight modifications for the purpose of averting detection, and hidden by steganography tools. The engineering challenge is to perform the detection in real-time on high data rate links. The ongoing collaborative research is being carried out by researchers from Sandia National Labs

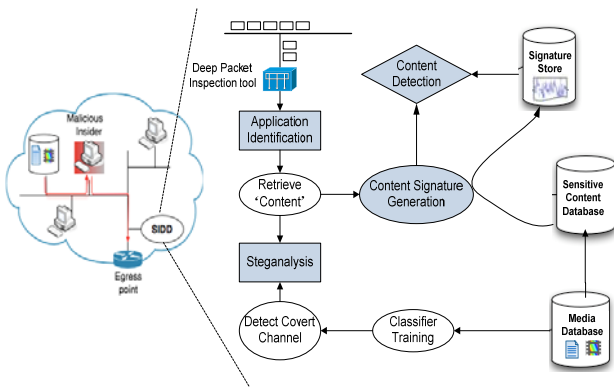
---

\*Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-04AL85000.

at Livermore and from the Department of Computer Science at the University of California, Davis.

## 2. OVERALL TECHNICAL APPROACH

Figure 1 shows a block diagram of the proposed Sensitive Information Dissemination Detection (SIDD) system and its deployment in the network. The critical data repository contains all the critical content. The content signature generator is used to generate the signatures which are stored in the signature store. The Deep Packet Inspection tool is used to inspect the network traffic to detect the exfiltration of sensitive information.



**Figure 1. Sensitive Information Dissemination Detection (SIDD) system.**

In the following subsections we outline our research directions on the key components of SIDD.

### 2.1 Application Identification

One of the key challenges facing network administrators when securing and managing the performance of a network is the anonymity of the traffic on the network. While current research has addressed the issue of identifying the application layer protocols such as SSH, HTTP, and FTP, a more fine grained identification is required for variety of applications that run over HTTP [1]. In our work, we investigate a number of applications drawn from three different classes of applications, namely, social networking (MySpace and Facebook), web-mail (Gmail and Hotmail), and streaming video applications (Youtube and Veoh) that all use HTTP and determine if there have specific features in the network traffic generated by them. We consider different signals that can be derived from the network flow, e.g. the packet inter-arrival times and packet sizes, and apply time domain, frequency spectrum,

and wavelet based analyses to identify distinguishing features of the applications. The preliminary results show that the applications do indeed have discernable characteristics for all the three analysis techniques. Furthermore, we show that these features are robust under stress tests for a range of network and host CPU loads.

### 2.2 Content Signature Generation and Detection

Our primary focus is on video content and we adopt a novel approach based on the paradigm that content itself is a signature [2]. We extract a time series (or signal) from the video content and analyze the signal to generate signatures. The key motivation is to use the temporal correlation between the video frames as the signature.

One of key issue of using the content itself as a signature is that of scalability arising due to the size of the signature. As result it is important to reduce the size of the signature by only keeping some of the key attributes of the content. In our research, we use wavelets [3] to reduce the size of the signatures. While there are many advantages of using wavelets to analyze signals, the most important ones are: a) Time-Frequency Localization that allows us not only to know what the characteristic frequencies in the signal are, but also where “approximately” in time they occur; and b) Multi-Resolution Analysis which allows the analysis to be performed at different levels of resolution or scales.

In this research project we first investigate different methods of creating unique, compact, and robust signatures for video content. We propose to use Discrete Wavelet Transforms (DWTs) and Continuous Wavelet Transforms (CWTs) to create signatures that are tolerant to noise and distortions. Treating the content as time series allows us to use signal processing techniques and tools to analyze these signals. The other benefit of this method is that any portion of the content can be used for detection. This is important because sometimes only a portion of the content starting at some arbitrary point may be available. Moreover, it is important to detect exfiltration as soon as possible so that mitigating actions can be taken. Finally, the most important benefit of our approach is the computation cost saving. Since we do not need to decode the target video sequence, the signature can be obtained by

inspecting the header of the packets carrying the video stream.

Another important focus of our work is on the signature matching technique. The specific problem is the following. Given a portion of a video stream captured from the network, the task is to determine if it is part of a protected video for which signatures have been created a priori and stored in a signature store. We apply wavelet analysis on the target video and then compare with the signatures in the signature store. For our preliminary study we have employed cross-correlation to estimate the degree to which two series are correlated. This technique is effective when the monitored video streams are identical to the video streams in the video database. Our preliminary results show that such an approach has good performance both in terms of false positive and false negative probabilities. However, the user may evade detection by trans-coding video stream and/or injecting noise which may impact the short term correlation. We are investigating different approaches that consider long term trends such as co-integration.

### 2.3 Detecting Covert Communication

Steganography is the art and science of hiding the presence of communication by embedding a secret message within an innocuous carrier medium, such as digital audio, image, and video. With the rapid proliferation of multimedia capturing tools and transmission protocol, covert channels may be relatively easy to establish [4]. By slightly altering the binary sequence of the distributed sensitive content samples with existing steganography tools, the distributed sensitive content can be encrypted, modified, and hidden to avoid detection. Therefore, one key aspect of this proposed technology will be the development of methods that can be used to detect the presence of covert communication channels.

Our preliminary focus is on audio steganalysis. The inherent redundancy in the audio signal and its transient and unpredictable characteristics imply a high hidden capacity. This is further aided by the fact that the human ear is insensitive to small distortions in the audio signal. All these properties make audio a good candidate for use as a "cover" for covert communications through hide secret messages and also for exfiltration of sensitive protected content. At this point in our preliminary work, we have studied the use of audio files as target carrier medium, but the

methods can be generalized to other types of media files.

We have proposed an audio steganalysis scheme that measures audio content distortion using Hausdorff Distance. Unlike previous work in audio steganalysis that used the traditional audio quality metrics [5], such as signal-to-noise ratio (SNR), Perceptual Audio Quality Measure (PAQM), and other such metrics, the proposed distortion measure is designed specifically to detect the existence of the covert communication (modifications to pure audio content). Given an audio object ( $x$ ) which could potential involved with a covert communication, we consider its de-noised version ( $x'$ ) as an estimate of the reference object. After appropriate segmentation, we apply wavelet decomposition to both  $x$  and  $x'$  to generate wavelet coefficients at different levels of resolution. Next, Hausdorff distances are used to test the similarities between the wavelet coefficients of sensitive audio and their reference values. The statistical moments of these Hausdorff distances are used as the features to train a classifier on the difference between known pure content and modified audio files with covert channels. Simulations with numerous audio sequences show that our algorithm has a strong discriminatory ability and the performance is significantly superior to existing methods. ~~Moreover, as the proposed scheme makes no assumptions about the embedding technique used, it may be extended to other steganography tools and algorithms.~~

### 3. REFERENCES

- [1] T. T.T. Nguyen, G. Armitage, A Survey of Techniques for Internet Traffic Classification using Machine Learning, IEEE Communications Surveys and Tutorials, 2008.
- [2] A. Hampapur and R. M. Bolle, Comparison of distance measures for video copy detection, Proceedings of IEEE International Conference on Multimedia and Expo (ICME), pp. 737–740, August 2001.
- [3] C. S. Burrus, R. A. Gopinath, H. Guo, Introduction to wavelets and wavelets transforms, a primer. Upper Saddle River, NJ (USA): Prentice Hall, 1998.
- [4] J. Dittmann, D. Hesse, R. Hillert, Steganography and steganalysis in voice-over IP scenarios: operational aspects and first experiences with a new steganalysis tool set. Proceedings of SPIE on Security, Steganography, and Watermarking of Multimedia Contents VII, Vol.5681, pages 607-618, Magdeburg, 2005.

- [5] H. Ozer, I. Avcibas, B. Sankur, N. Memon, Steganalysis of audio based on audio quality metrics. Security and Watermarking of Multimedia Contents, pages55-66. Santa Clara, 2003.